FORMAS FINITAS DO
TEOREMA DE DE FINETTI:
A VISÃO PREDITIVISTA DA
INFERÊNCIA ESTATÍSTICA
EM POPULAÇÕES FINITAS

Pilar Loreto Iglesias Zuazola

TESE DEFENDIDA
NO
INSTITUTO DE MATEMÁTICA E ESTATÍSTICA
DA
UNIVERSIDADE DE SÃO PAULO
PARA OBTENÇÃO DO GRAU DE DOUTOR
EM
ESTATÍSTICA

Área de concentração: Estatística
Orientador: Prof. Dr. Carlos Alberto de Bragança Pereira

– São Paulo, janeiro de 1993 –

Dedico este trabalho ao meu marido

Juan

*cujo apoio, carinho e compreensão em inúmeros
momentos antes e durante o trabalho foram
fundamentais.*

À minha

Mãe

*por tudo aquilo que me ensinou, em especial por ter
me ensinado a levar a vida com alegria e otimismo.*

Aos meus irmãos e sobrinho

Manuel, Irma, José Luis e Derek

e à minha grande amiga

Raquel

pelo carinho e a força que sempre me deram.

AGRADECIMENTOS

Ao meu orientador, Carlos Alberto de Bragança Pereira, pelo conjunto de idéias que possibilitaram a realização deste trabalho e pela confiança em mim depositada.

Ao professor Sérgio Wechsler, pela sua amizade e ajuda indispensável, pelas idéias e importantes sugestões e pelas variadas e estimulantes discussões em torno das idéias de de Finetti.

Ao professor Nelson Tanaka, pelas inúmeras sugestões neste trabalho, pelo seu apoio e incansável ajuda. Especialmente pela sua contribuição no desenvolvimento do Capítulo 2 deste trabalho.

Ao grupo de Bayesianas do IME-USP, pelo seu constante apoio e entusiasmo. Em especial à Olga que apesar de seu pouco tempo, sempre encontrou um espaço para ler cada uma das páginas deste trabalho, fazendo sugestões e ajudando a consertar os erros de português, pela sua amizade antes e durante este trabalho.

Ao meu amigo Reinaldo, que me apoiou, animou em todos os momentos que precisei, especialmente pela sua amizade sincera e transparente.

À turma 88, Eufrásio, Fábio, João, Márcia, Mônica e Victor que tornaram mais leve minha adaptação no meu primeiro ano no Brasil. Especialmente ao João cujo apoio foi fundamental.

À Regina, pelo suporte computacional, que não se importou nem com domingos e feriados para dar-me uma mãozinha.

Aos meus amigos Arnoldo e Ivan, pela sua amizade, estímulo e apoio.

A todos os meus colegas de Valparaíso e Santiago que contribuíram com a minha formação, pelo estímulo e incentivo.

A todos os meus amigos que me apoiaram em inúmeros momentos e facilitaram o desenvolvimento deste trabalho.

Ao Walter, pela paciência ao agüentar múltiplas alterações na fase de digitação deste trabalho.

Ao Brasil, pela sua contribuição ao desenvolvimento científico de diversos países latino-americanos.

À CAPES e à FAPESP, pelo apoio financeiro.

RESUMO

Este trabalho apresenta formas finitas do teorema de de Finetti para distribuições uniformes sobre a reta. Soluções inferenciais em populações finitas são apresentadas sob a perspectiva preditivista seguindo de Finetti. Os resultados obtidos são comparados com a abordagem bayesiana de superpopulação. A equivalência é obtida quando extensões para populações infinitas são possíveis. Neste contexto são estabelecidas proposições que garantem estendibilidade de seqüências finitas especiais de variáveis aleatórias permutáveis. Finalmente um procedimento não bayesiano para inferências em populações finitas é proposto.

ABSTRACT

The present work introduces finite forms of de Finetti's Theorem for uniform models on the real line. Inference solutions in finite populations are presented under the predictivist approach, following de Finetti. Results are compared to those obtained under a Bayesian superpopulation perspective. The equivalence is obtained whenever extendibility to infinite populations is possible. Propositions are stated in this set up which imply extendibility of some special finite sequences of exchangeable random quantities. Finally a non-Bayesian procedure for inferences in finite populations is presented.

ÍNDICE

Introdução	1
Capítulo 1. Fortalecimentos do teorema de de Finetti	8
1.1 Introdução	8
1.2 Permutabilidade e o teorema da representação de de Finetti	9
1.3 Formas infinitas do teorema de de Finetti	25
1.4 Formas finitas do teorema de de Finetti	40
1.5 Permutabilidade em populações finitas	64
Capítulo 2. Formas finitas do teorema de de Finetti para distribuições	
uniformes univariadas	68
2.1 Introdução	68
2.2 Distribuição Uniforme sobre $\{0, 1, \dots, \theta\}$, $\theta \in \mathbb{N}$	70
2.3 Distribuição Uniforme sobre $(0, \theta)$, $\theta > 0$	79
2.4 Distribuição Uniforme sobre $(-\theta, \theta)$, $\theta > 0$	93
2.5 Distribuição Uniforme sobre (θ_1, θ_2) , $\theta_1, \theta_2 \in \mathbb{R}$	102
Capítulo 3. Abordagem preditivista para inferência em populações finitas	127
3.1 Introdução	127
3.2 Preditivismo em populações finitas	130
3.3 O modelo de predição	133
3.4 Modelos discretos localmente uniformes	136
3.4.1 Conexões com estatísticas suficientes e de resumo	141
3.4.2 Conexões com a abordagem bayesiana usual de superpopulação:	
modelos Bernoulli, Geométrico e Uniforme	144
3.4.3 Aplicações	166

3.5 Modelos não discretos localmente uniformes	184
3.5.1 Predição do valor máximo da característica estudada na população ..	186
3.5.2 Predição de outras quantidades populacionais	193
3.6 Comentários finais	203
<hr/>	
Capítulo 4. Verossimilhança preditiva em populações finitas.	
Uma solução não bayesiana	202
4.1 Introdução	202
4.2 Predição	205
4.3 Predição do total populacional	212
4.4 Comentários	219
Conclusões e comentários	222
Referências bibliográficas	228

INTRODUÇÃO

Na visão bayesiana da estatística, ênfase é dada à especificação de uma distribuição a priori para os parâmetros envolvidos no estudo. Por outro lado, o modelo estatístico adotado (a verossimilhança) é considerado como determinado por situações físicas tendo assim um caráter objetivo em oposição ao caráter subjetivo da distribuição a priori.

Na visão bayesiana mais “pura” devida a de Finetti (1937), tanto a priori quanto a verossimilhança são consideradas elementos subjetivos. Na verdade, a opinião se expressa como uma distribuição conjunta para todas as quantidades envolvidas no estudo. Visão preditivista é o termo usado para o enfoque descrito por de Finetti. Alguns trabalhos publicados recentemente (veja, por exemplo, Dawid (1982), Lindley (1981), Cifarelli e Regazzini (1982), Barlow e Mendel (1992), Barlow e Spizzichino (1992), Hayakawa (1992) e Wechsler (1993)) se utilizam deste enfoque. A especificação do modelo probabilístico é a principal ênfase destes trabalhos e o uso de seqüências permutáveis e suas variações produz a base da teoria. No caso de variáveis 0-1 a verossimilhança é obtida explicitamente apenas pelo julgamento de permutabilidade. O teorema de de Finetti (1937) estabelece

exatamente esta caracterização:

Teorema de de Finetti: *Consideremos uma seqüência de variáveis permutáveis (todas as subseqüências finitas de mesmo comprimento possuem distribuições iguais) tomando valores em $\{0, 1\}$. Então a distribuição de qualquer subseqüência de comprimento finito, digamos $\{x_1, \dots, x_n\}$ pode ser sempre expressa da forma*

$$f(x_1, \dots, x_n) = \int_0^1 \theta^t (1 - \theta)^{n-t} d\mu(\theta) , \quad (1)$$

para alguma medida de probabilidade μ concentrada no intervalo $[0, 1]$.

A representação em (1) ilustra a representação preditivista da avaliação subjetiva de permutabilidade em um processo de variáveis 0-1.

Quando consideramos seqüências finitas, a representação (1) não é necessariamente verdadeira, como é mostrado a seguir.

Sejam X_1, X_2 variáveis aleatórias assumindo valores em $\{0, 1\}$ tal que

$$\begin{aligned} P(X_1 = 1, X_2 = 0) &= P(X_1 = 0, X_2 = 1) = \frac{1}{2} , \\ P(X_1 = 0, X_2 = 0) &= P(X_1 = 1, X_2 = 1) = 0 . \end{aligned}$$

O par (X_1, X_2) é permutável, mas se existe uma medida de probabilidade μ tal que

$$0 = P(X_1 = 1, X_2 = 1) = \int_0^1 \theta^2 d\mu(\theta) ,$$

então necessariamente μ concentra sua massa no ponto 0, e portanto não podemos ter

$$0 = P(X_1 = 0, X_2 = 0) = \int_0^1 (1 - \theta)^2 d\mu(\theta) .$$

Porém, se uma seqüência finita de variáveis aleatórias permutáveis 0-1 de comprimento n pode ser estendida (Diaconis (1977), Spizzichino (1982)) para uma seqüência de comprimento N ($N > n$), então a lei da seqüência de comprimento n pode ser expressa como

quase mistura de leis de variáveis aleatórias Bernoulli independentes e identicamente distribuídas, com erro tendendo para zero conforme N cresce.

Várias questões podem ser levantadas neste ponto:

- (i) Qual a justificativa preditivista para a adoção de outros modelos paramétricos padrão tais como Poisson, Geométrica, Normal etc?
- (ii) Que tipo de representações surge quando tratamos com seqüências finitas de variáveis aleatórias permutáveis e qual é o erro ao aproximar estas pela correspondente representação infinita?
- (iii) Como deve ser encarada a visão preditivista no contexto de inferência em populações finitas?

Soluções para a primeira questão já fazem parte da literatura. Teoremas de representação para variados modelos e em diferentes graus de generalidade têm sido fornecidos por diversos autores. Uma ampla discussão sobre estes resultados é fornecida por Aldous (1985) e Diaconis (1988). Recentemente Diaconis e Freedman (1990) derivaram resultados gerais que envolvem a família exponencial, para os casos univariado e multivariado. Estes tipos de resultado são denominados na literatura atual de “Formas Infinitas do Teorema de de Finetti”.

A segunda questão é relativamente mais nova. Alguns resultados em torno desta foram fornecidos inicialmente por Diaconis (1977, 1980), Diaconis e Freedman (1987). Os resultados que respondem a esta questão são usualmente denominados de “Formas Finitas do Teorema de de Finetti”. Dentro desta questão, estavam abertas (pois não temos conhecimento de publicações sobre este tema) formas finitas do teorema de de Finetti envolvendo distribuições uniformes univariadas.

Com relação à terceira questão, Ericson (1969) forneceu uma abordagem subjetivista para resolver problemas de inferência em populações finitas, utilizando-se de prioris per-

mutáveis. Porém a suposição de estendibilidade de uma seqüência finita para uma seqüência infinita de variáveis permutáveis é fundamental naquele trabalho.

Os objetivos deste trabalho são:

- (i) Estabelecer formas finitas do teorema de de Finetti que envolvam distribuições uniformes univariadas, dependendo de um e dois parâmetros.
- (ii) Fornecer uma abordagem preditivista para fazer inferências em populações finitas, levando em consideração características de população finita.

No Capítulo 1, fazemos uma revisão sobre os conceitos básicos associados com a noção de permutabilidade e as formas infinita e finita do teorema de de Finetti. A discussão não se centra apenas nos aspectos técnicos destes resultados, mas também nas questões fundamentais que surgem a partir destes. Os resultados são provados em alguns casos não seguindo necessariamente o padrão dos trabalhos em que foram publicados. Em especial, na seção 3, com o intuito de facilitar a compreensão, padronizamos os resultados seguindo o estilo de Heath e Sudderth (1976).

No Capítulo 2, estabelecemos formas finitas do teorema de de Finetti para distribuições uniformes dependendo de um e dois parâmetros. A seção 2 trata o modelo discreto e as seções 3 a 5 os modelos uniformes contínuos. Os resultados são derivados a partir da construção de distribuições uniformes sobre os seguintes espaços

- (i) $\mathcal{X}_1^N = \{(x_1, \dots, x_N) \in \mathbb{N}_0^N : \max\{x_1, \dots, x_N\} = m\}, m \in \mathbb{N}_0$
- (ii) $\mathcal{X}_2^N = \{(x_1, \dots, x_N) \in \mathbb{R}_+^N : \max\{x_1, \dots, x_N\} = m\}, m > 0$
- (iii) $H_m^N = \{(x_1, \dots, x_N) \in \mathbb{R}^N : \max\{|x_1|, \dots, |x_N|\} = m\}, m > 0$
- (iv) $\mathcal{X}_{m_1, m_2}^N = \{(x_1, \dots, x_N) \in \mathbb{R}^N : \min\{x_1, \dots, x_N\} = m_1, \max\{x_1, \dots, x_N\} = m_2\},$
 $m_1 < m_2, m_1, m_2 \in \mathbb{R}.$

É mostrado, por exemplo, que a lei das primeiras n ($n < N$) coordenadas de um ponto uniformemente distribuído sobre $\mathcal{X}_m^N$ (no caso (i)) é próxima em distância de variação total da lei de n variáveis aleatórias independentes com distribuição comum uniforme sobre o conjunto $\{0, 1, \dots, m\}$. Resultados similares são estabelecidos nos outros casos. Seguem a partir destes resultados, formas finitas do teorema de de Finetti que caracterizam classes de variáveis aleatórias permutáveis, como quase mistura de modelos associados a n variáveis aleatórias independentes com distribuição comum uniforme. A forma infinita do teorema de de Finetti associada a estas distribuições é derivado utilizando a forma finita aqui estabelecida.

O Capítulo 3 é endereçado essencialmente a discutir uma postura completamente preditivista no tratamento de populações finitas. Dentro deste contexto, a especificação de uma distribuição a priori para a coleção finita de quantidades aleatórias $X_1, X_2, \dots, X_N$ (representando características desconhecidas associadas a N indivíduos de uma população) é feita em dois estágios:

- (i) Especificação de uma distribuição a priori para a quantidade populacional de interesse, digamos $T(X_1, \dots, X_N)$.
- (ii) Especificação (para cada possível valor t de T) de uma distribuição a priori para a distribuição condicional de $X_1, X_2, \dots, X_N | T = t$.

As seções 1 e 2 são dedicadas a esta discussão. Na seção 3 estabelecemos formalmente estas idéias. Na seção 4 são discutidos alguns modelos discretos particulares. Conexões com a usual abordagem bayesiana em populações finitas são estabelecidas através do conceito de estendibilidade. Dentro deste contexto, fornecemos resultados que estabelecem condições para estendibilidade de classes particulares de seqüências finitas de variáveis aleatórias permutáveis.

Finalmente, na seção 5, são dadas soluções preditivistas para predição do máximo,

do total e da variância na população, sob esta perspectiva, utilizando-nos dos modelos estudados nos Capítulos 1 e 2. A distribuição de Pareto, enquanto distribuição a priori para as quantidades populacionais, resulta possuir propriedades interessantes em termos de tratabilidade matemática.

No Capítulo 4 é proposta uma metodologia dentro da perspectiva clássica para o problema de predição em populações finitas. A idéia é transformar modelos de superpopulação (Cassel et al (1977)), em modelos preditivos que permitam o uso de técnicas de inferência clássica. As idéias são ilustradas com vários exemplos, resgatando a partir disto alguns resultados utilizados na teoria de amostragem clássica.

Notações

As seguintes notações serão utilizadas ao longo deste trabalho.

$\mathcal{B}_N :=$ sigma álgebra de Borel do $\mathbb{R}^N$

Se $(\Omega, \mathcal{A})$ espaço mensurável e $B \in \mathcal{A}$, então

$$\mathcal{A} \cap B = \{A \cap B : A \in \mathcal{A}\}$$

$\mathcal{B}_N^+ := \mathcal{B}_N \cap \mathbb{R}_+^N$

$\#A :=$ cardinalidade do conjunto A

Se X uma variável aleatória e $x \in \mathbb{R}$, então

$$I_{[X \leq x]} = \begin{cases} 1 & \text{se } X \leq x \\ 0 & \text{se } X > x \end{cases}$$

$$I_A(x) = \begin{cases} 1 & , \text{ se } x \in A \\ 0 & , \text{ se } x \notin A \end{cases} \quad (\text{função Indicadora})$$

$$\Gamma(a) = \int_0^\infty x^{a-1} e^{-x} dx, \quad a > 0 \quad (\text{função Gama})$$

$$B(a, b) = \int_0^1 x^{a-1} (1-x)^{b-1} dx, \quad a > 0, b > 0 \quad (\text{função Beta})$$

Se $\mathbf{X}_N$ e $\mathbf{Y}_N$ são dois vetores aleatórios N -dimensionais e $A \subset \mathbb{R}^N$ ($N \in \mathbb{N}$), então

$\mathbf{X}_N \sim U(A) := \mathbf{X}_N$ é uniformemente distribuído sobre o conjunto A

$\mathbf{X}_N \amalg \mathbf{Y}_N := \mathbf{X}_N$ e $\mathbf{Y}_N$ são vetores aleatórios independentes

$\mathbf{X}_N \stackrel{\mathcal{D}}{=} \mathbf{Y}_N := \mathbf{X}_N$ e $\mathbf{Y}_N$ são identicamente distribuídos

Se $\{a_n\}_{n \geq 1}$ e $\{b_n\}_{n \geq 1}$ são duas seqüências de números reais, então

$$a_n = o(b_n) \Leftrightarrow \lim_{n \rightarrow \infty} \frac{a_n}{b_n} = 0 .$$

CAPÍTULO 1

Fortalecimentos do Teorema de de Finetti

1.1 Introdução

O famoso teorema da representação de de Finetti (1937) é um resultado dos mais importante e profundo dentro da literatura estatística. Neste teorema, resultante da noção de permutabilidade, percebemos a conexão entre os conceitos de probabilidade e frequência relativa. O objetivo deste capítulo é discutir em detalhe este aspecto do teorema e alguns resultados subsequentes.

A seção 2 é dedicada aos conceitos básicos relacionados com a noção de permutabilidade e estendibilidade. O teorema de de Finetti e resultados associados serão apresentados nessa seção.

Formas infinitas do teorema de de Finetti serão discutidas na seção 3. A menos do

caso 0-1, os resultados apresentados nessa seção serão desenvolvidos utilizando um estilo diferente do comumente usado na literatura contendo os casos que serão abordados nessa seção.

Outros desenvolvimentos em torno do teorema de de Finetti - formas finitas do teorema de de Finetti - serão discutidas na seção 4, com ênfase nos casos abordados no Capítulo 3 deste trabalho.

Finalmente, uma breve discussão do papel da noção de permutabilidade em populações finitas é feita na seção 5.

1.2 Permutabilidade e o teorema de representação de De Finetti

Nesta seção os conceitos básicos associados à noção de permutabilidade serão revisados. O famoso teorema de de Finetti e resultados relacionados serão apresentados, não apenas enfatizando os aspectos técnicos, como também suas conseqüências nos fundamentos da teoria estatística. A questão de estendibilidade que relaciona seqüências finitas com seqüências infinitas de variáveis aleatórias permutáveis será discutida ao final da seção.

Definição 1: As variáveis aleatórias $X_1, X_2, \dots, X_N$ são ditas permutáveis quando sua distribuição conjunta F é invariante sob permutações de seus argumentos, isto é, quando

$$F(x_1, \dots, x_N) = F(x_{\pi(1)}, \dots, x_{\pi(N)}) ,$$

para cada permutação π do conjunto $U = \{1, \dots, N\}$ e cada $(x_1, \dots, x_N) \in \mathbb{R}^N$.

É imediato a partir desta definição que se $X_1, X_2, \dots, X_N$ são variáveis aleatórias permutáveis com função de densidade conjunta f , então f é invariante sob permutações de seus argumentos e vice-versa.

Propriedades

1. Todas as distribuições condicionais e marginais de subcoleções de variáveis aleatórias permutáveis, são permutáveis.

Prova: Seja $I = \{i_1, \dots, i_r\} \subset U$ e π uma permutação sobre U , tal que $\pi(k) = k$ se $k \notin I$. Deixando fixos $x_{i_1}, \dots, x_{i_r}$ na Definição (1) e fazendo os termos restantes irem para infinito, conclui-se que $X_{i_1}, \dots, X_{i_r}$ são variáveis aleatórias permutáveis. De maneira análoga mostra-se a permutabilidade da distribuição condicional de $X_{i_1}, \dots, X_{i_r}$ dado $X_{j_1} = x_{j_1}, \dots, X_{j_s} = x_{j_s}$, onde $\{i_1, \dots, i_r, j_1, \dots, j_s\} \subset U$. $\square$

2. As variáveis aleatórias $X_1, X_2, \dots, X_N$ são permutáveis se e somente se

$$(X_{i_1}, \dots, X_{i_r}) \stackrel{D}{=} (X_{j_1}, \dots, X_{j_r}) ,$$

para cada $r = 1, \dots, N$ e $\{i_1, \dots, i_r\}, \{j_1, \dots, j_r\}$ subconjuntos de $U = \{1, \dots, N\}$.

Prova: Análoga à anterior. $\square$

3. Se $X_1, X_2, \dots, X_N$ são variáveis aleatórias permutáveis com $E(X_1^2) < \infty$, então $\text{Corr}(X_1, X_2) \geq -\frac{1}{N-1}$.

Prova: Seja $\mu = E(X_1)$, $\sigma^2 = V(X_1)$ e $\lambda = \text{Corr}(X_1, X_2)$. Da propriedade anterior temos que todas as distribuições marginais univariadas e bivariadas coincidem, portanto o vetor de médias e a matriz de covariância são dados por

$$\underline{\mu} = \mu \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix} = \mu \mathbf{1}_N \quad \text{e} \quad \Sigma = \sigma^2 \begin{bmatrix} 1 & \lambda & \dots & \dots & \lambda \\ \lambda & 1 & \lambda & \dots & \lambda \\ \vdots & & & & \\ \lambda & \lambda & \dots & \dots & 1 \end{bmatrix} = (1 - \lambda)\sigma^2 I_N + \lambda\sigma^2 J_N$$

onde $J_N = \mathbf{1}_N \mathbf{1}_N^t$ e I_N a matriz identidade de ordem N . Agora, dado que Σ é semidefinida positiva temos que $v^t \Sigma v \geq 0$ para cada $v \in \mathbb{R}^N$. Escolhendo $v = \mathbf{1}_N$, temos que

$$N(1 - \lambda) + \lambda N^2 \geq 0 ,$$

concluindo-se a partir desta desigualdade que $\lambda \geq -\frac{1}{N-1}$. $\square$

4. O conjunto de funções de distribuição permutáveis sobre $\mathbb{R}^N$ é um conjunto convexo.

Prova: Sejam F_1, F_2 duas funções de distribuição permutáveis sobre $\mathbb{R}^N$ e $F = \alpha F_1 + (1 - \alpha) F_2$ com $0 \leq \alpha \leq 1$. Então para $(x_1, \dots, x_N) \in \mathbb{R}^N$ e π permutação de $\{1, \dots, N\}$ temos que

$$\begin{aligned} F(x_1, \dots, x_N) &= \alpha F_1(x_1, \dots, x_N) + (1 - \alpha) F_2(x_1, \dots, x_N) \\ &= \alpha F_1(x_{\pi(1)}, \dots, x_{\pi(N)}) + (1 - \alpha) F_2(x_{\pi(1)}, \dots, x_{\pi(N)}) \text{ (Definição 1)} \\ &= F(x_{\pi(1)}, \dots, x_{\pi(N)}) . \end{aligned}$$

$\square$

Um conceito importante relacionado a esta última propriedade é o de mistura de funções de distribuição, definido a seguir.

Definição 2: Seja $\mathcal{F} = \{F(\cdot/\theta) : \theta \in \Theta\}$, família de funções de distribuição sobre $\mathbb{R}^N$, e G uma função de distribuição sobre Θ . A função de distribuição

$$F(\cdot) = \int_{\Theta} F(\cdot/\theta) dG(\theta)$$

é denominada mistura sobre $\mathcal{F}$, definida pela função de distribuição G .

Não é difícil ver a partir desta definição, que se $F(\cdot/\theta)$ é uma função de distribuição permutável, então a função de distribuição mistura F é permutável. Por exemplo, se $F(\cdot/\theta)$ é a função de distribuição de N variáveis aleatórias independentes e identicamente distribuídas, então F é uma função de distribuição permutável.

A proposição a seguir mostra que toda função de distribuição permutável sobre $\mathbb{R}^N$ pode ser expressa de maneira única como mistura de funções de distribuição de permutação aleatória. De fato, se $\mathbf{Y} = (Y_1, \dots, Y_N)$ são as correspondentes estatísticas de

ordem associadas a $X_1, X_2, \dots, X_N$ e

$$\Omega_y = \{(z_1, \dots, z_N) \in \mathbb{R}^N : (z_{\pi(1)}, \dots, z_{\pi(N)}) = (y_1, \dots, y_N) \},$$

para alguma permutação π de $\{1, \dots, N\}$,

onde $y_1 \leq y_2 \leq \dots \leq y_N$, então temos a partir da definição que se $X_1, X_2, \dots, X_N$ são variáveis aleatórias permutáveis

$$P(X_1 = x_1, \dots, X_N = x_N | \mathbf{Y} = \mathbf{y}) = \frac{\prod_{j=1}^m N_j!}{N!} I_{\Omega_y}(x_1, \dots, x_N)$$

onde m é o número de componentes distintos em $\mathbf{y}$ com multiplicidades $N_1, N_2, \dots, N_m$.

Assim, se $G(\cdot, \mathbf{y})$ denota a função de distribuição condicional de $X_1, \dots, X_N \mid \mathbf{Y} = \mathbf{y}$, então

$$G(x_1, \dots, x_N | \mathbf{y}) = \frac{\prod_{j=1}^m N_j!}{N!} N(\mathbf{x}, \mathbf{y}) ,$$

onde $N(\mathbf{x}, \mathbf{y})$ é o número de $\mathbf{z} \in \mathbb{R}^N$ tal que $(z_{\pi(1)}, \dots, z_{\pi(N)}) = (y_1, \dots, y_N)$ para alguma permutação π de $\{1, \dots, N\}$ e $z_1 \leq x_1, \dots, z_N \leq x_N$.

Proposição 1: Se $X_1, X_2, \dots, X_N$ são variáveis aleatórias permutáveis com função de distribuição conjunta F , então

$$F(x_1, \dots, x_N) = \int_{\Omega_N} G(x_1, \dots, x_N | \mathbf{y}) d\mu(\mathbf{y}) ,$$

onde $\Omega_N = \{\mathbf{y} \in \mathbb{R}^N : y_1 \leq y_2 \leq \dots \leq y_N\}$ e μ a medida de probabilidade induzida por $\mathbf{Y}$, o vetor de estatísticas de ordem.

Prova: Segue dos resultados acima e da definição de probabilidade condicional. $\square$

Não é difícil ver a partir disto que a família de distribuições de permutação aleatória é o conjunto de pontos extremos de distribuições permutáveis sobre $\mathbb{R}^N$. A Proposição 1 é denominada versão finita exata do teorema de de Finetti (de Finetti B, (1969)).

O conceito de permutabilidade é definido de maneira análoga para vetores aleatórios.

Definição 3: Os vetores aleatórios $\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_N$ assumindo valores em $\mathbb{R}^p$ ($p \in \mathbb{N}$) são ditos permutáveis se sua função de distribuição conjunta satisfaz

$$F(\mathbf{x}_1, \dots, \mathbf{x}_N) = F(\mathbf{x}_{\pi(1)}, \dots, \mathbf{x}_{\pi(N)}) ,$$

para toda permutação π de $U = \{1, \dots, N\}$ e $\mathbf{x}_1, \dots, \mathbf{x}_N \in \mathbb{R}^p$.

Notemos que se $\mathbf{X}_1, \dots, \mathbf{X}_N$ são vetores aleatórios permutáveis, então para qualquer função ψ mensurável, os vetores aleatórios $\psi(\mathbf{X}_1), \dots, \psi(\mathbf{X}_N)$ são permutáveis.

Definição 4: Uma seqüência infinita de variáveis aleatórias $\{X_n : n \in \mathbb{N}\}$ é dita permutável se e somente se

$$(X_1, \dots, X_N) \stackrel{D}{=} (X_{\pi(1)}, \dots, X_{\pi(N)}) ,$$

para toda permutação π de $\{1, \dots, N\}$ e todo $N \in \mathbb{N}$.

O teorema de de Finetti (1937) estabelece que qualquer processo estocástico do tipo acima definido pode ser representado como mistura de processos estocásticos cujas componentes são independentes e identicamente distribuídas. Demonstraremos este teorema seguindo essencialmente Bühlman, H. (1987). As definições e lemas a seguir serão necessários para a demonstração do teorema.

Seja S o grupo de permutações finitas e $S^N = \{\pi_j^N : j = 1, \dots, N!\}$ o subgrupo de permutações operando sobre os N primeiros elementos. Claramente $S^N \subset S^{N+1}$. A ação de π sobre $x \in \mathbb{R}^\infty$ é definida por $\pi x = (x_{\pi(1)}, x_{\pi(2)}, \dots)$ e se $B \in \mathcal{B}^\infty$, então $\pi B = \{\pi x : x \in B\}$. A classe de eventos S^N invariantes (S invariantes) será denotada por $\mathcal{I}^N$ ($\mathcal{I}$). Note que $\mathcal{I}^N \downarrow \mathcal{I}$. Seja $\mathbf{X} = (X_1, X_2, \dots)$ processo estocástico assumindo valores em $\mathbb{R}^\infty$ e P a lei de $\mathbf{X}$.

Lema 1: *Seja $\mathbf{X} = (X_1, X_2, \dots)$ seqüência infinita de variáveis aleatórias permutáveis e f função real definida sobre $\mathbb{R}^\infty$ tal que $f \in L^1$, então*

$$E(f) = E(f \circ \pi) , \quad \forall \pi \in S .$$

Prova: Seja $f = I_B$, $B \in \mathcal{B}^\infty$, então pela hipótese de permutabilidade

$$P(B) = P(\pi B) \text{ para cada } \pi \in S .$$

Mas $I_B(\pi x) = I_{\pi B}(x)$. Logo

$$\int_{\mathbb{R}^\infty} I_B(x) dP(x) = \int_{\mathbb{R}^\infty} I_B(\pi x) dP(x) ,$$

provando assim o lema para $f = I_B$.

Seja agora $f(x) = \sum_{i=1}^n \alpha_i I_{B_i}(x)$, com $B_i \in \mathcal{B}^\infty$ para cada i , então

$$\begin{aligned} E(f) &= \int_{\mathbb{R}^\infty} f(x) dP(x) = \sum_{i=1}^n \alpha_i \int_{\mathbb{R}^\infty} I_{B_i}(x) dP(x) \\ &= \sum_{i=1}^n \alpha_i \int_{\mathbb{R}^\infty} I_{B_i}(\pi x) dP(x) , \end{aligned}$$

esta última igualdade é consequência do caso anterior. Logo

$$E(f) = \int_{\mathbb{R}^\infty} \sum_{i=1}^n \alpha_i I_{B_i}(\pi x) dP(x) = E(f \circ \pi) .$$

A prova se conclui seguindo o procedimento usual para f não negativa e a partir disto para f , P -integrável. □

Lema 2: *Sob as condições do Lema 1, tem-se que*

$$E(f|\mathcal{I}^N) = E(f \circ \pi|\mathcal{I}^N) , \quad q.c. P ,$$

para cada $\pi \in S^N$.

Prova: Seja $C \in \mathcal{I}^N$ e $\pi \in S^N$, então

$$\begin{aligned} \int_C E(f|\mathcal{I}^N)(x) dP(x) &= \int_C f(x) dP(x) \\ &= \int_{\mathbb{R}^\infty} f(x) I_C(x) dP(x) = \int_{\mathbb{R}^\infty} f(\pi x) I_C(\pi x) dP(x) . \end{aligned}$$

Esta última igualdade vem do Lema 1. Logo

$$\begin{aligned} \int_C E(f|\mathcal{I}^N)(x) dP(x) &= \int_{\mathbb{R}^\infty} f(\pi x) I_C(\pi x) dP(x) \\ &= \int_{\mathbb{R}^\infty} f(\pi x) I_{\pi C}(\pi x) dP(x) = \int_{\mathbb{R}^\infty} f(\pi x) I_C(x) dP(x) , \end{aligned}$$

já que $C \in \mathcal{I}^N$ e $\pi \in S^N$. Consequentemente,

$$\int_C E(f|\mathcal{I}^N)(x) dP(x) = \int_C f(\pi x) dP(x) = \int_C E(f \circ \pi | \mathcal{I}^N) dP$$

Como $C \in \mathcal{I}^N$ é arbitrário, o resultado segue. $\square$

Lema 3: *Sob as condições do Lema 1, tem-se que*

$$E(f|\mathcal{I}^N) = \frac{1}{N!} \sum_{j=1}^{N!} f \circ \pi_j^N , \quad \text{q.c.} P .$$

Prova: A partir do Lema 2, temos que

$$E(f|\mathcal{I}^N) = E(f \circ \pi_j^N | \mathcal{I}^N) , \quad \text{q.c.} P \quad \text{e} \quad j = 1, \dots, N! .$$

Logo,

$$N! E(f|\mathcal{I}^N) = \sum_{j=1}^{N!} E(f \circ \pi_j^N | \mathcal{I}^N) , \quad \text{q.c.} P .$$

Consequentemente,

$$E(f|\mathcal{I}^N) = \frac{1}{N!} E \left(\sum_{j=1}^{N!} f \circ \pi_j^N | \mathcal{I}^N \right) , \quad \text{q.c.} P .$$

Mas, $g = \sum_{j=1}^{N!} f \circ \pi_j^N$ é S^N invariante, logo $g \in \mathcal{I}^N$. Portanto

$$E(f|\mathcal{I}^N) = \frac{1}{N!} \sum_{j=1}^{N!} f \circ \pi_j^N, \quad \text{q.c } P. \quad \square$$

Lema 4: *Sob as condições do Lema 1, temos que*

$$E(f|\mathcal{I}^N) \xrightarrow[L^1]{\text{q.c}} E(f|\mathcal{I})$$

Prova: Primeiro notemos que se $Y_N = E(f|\mathcal{I}^N)$, então $\{Y_N : N \geq 1\}$ é uma martingala reversa com relação a $\mathcal{I}^N$. De fato, $\mathcal{I}^N \supset \mathcal{I}^{N+1}$ para cada $N \in \mathbb{N}$ e

$$E(Y_N|\mathcal{I}^{N+1}) = E(E(f|\mathcal{I}^N)|\mathcal{I}^{N+1}) = E(f|\mathcal{I}^{N+1}) = Y_{N+1}.$$

Também, $E(f|\mathcal{I}^N) = E(Y_1|\mathcal{I}^N)$. Logo, $E(|Y_N|) < \infty$ para cada $N \in \mathbb{N}$. Consequentemente, pelo teorema de convergência de martingais reversas (Loeve, p. 394) temos que $Y_N \xrightarrow[L^1]{\text{q.c}} Y_\infty$, onde $Y_\infty = E(f|\mathcal{I})$ q.c P . $\square$

Uma importante consequência dos lemas anteriores, que será de utilidade mais para frente, é a Lei Forte dos Grandes Números. Antes de estabelecer o teorema de representação mostraremos esse resultado.

Teorema 1: *Seja $\mathbf{X} = (X_1, X_2, \dots)$ seqüência infinita de variáveis aleatórias permutáveis tal que $E(|X_1|) < \infty$, então*

$$\frac{X_1 + X_2 + \dots + X_N}{N} \xrightarrow[L^1]{\text{q.c}} E(X_1|\mathcal{I})$$

Prova: Escolhendo f definida por $f(x) = x_1$, temos a partir do Lema 3 que

$E(f|\mathcal{I}^N)(x) = \frac{1}{N!} \sum_{j=1}^{N!} x_{\pi_j(1)}$. Mas $\sum_{j=1}^{N!} x_{\pi_j(1)} = (N-1)! \sum_{i=1}^N x_i$, pois existem $(N-1)!$ permutações $\pi \in S^N$ tal que $\pi(1) = i$, para qualquer $i = 1, \dots, N$. Utilizando agora o lema 4 conclui-se a prova. $\square$

Observação 1: Utilizando os mesmos argumentos no teorema acima pode ser mostrado, por exemplo, que se $X = (X_1, X_2, \dots)$ é uma seqüência de variáveis aleatórias permutáveis, então

- i) $\frac{1}{N} \sum_{i=1}^N I_{[X_i \leq t]} \xrightarrow[\text{q.c.}]{L^1} P(X_1 \leq t/\mathcal{I})$
- ii) $\frac{1}{N(N-1)} \sum_{\substack{i,j \in \{1, \dots, N\} \\ i \neq j}} I_{[X_i \leq s, X_j \leq t]} \xrightarrow[\text{q.c.}]{L^1} P(X_1 \leq s, X_2 \leq t/\mathcal{I})$
- iii) Se $E(X_1^2) < \infty$, então $\frac{1}{N} \sum_{i=1}^N X_i^2 \xrightarrow[\text{q.c.}]{L^1} E(X_1^2|\mathcal{I})$.

Teorema 2 (de Finetti): *Uma seqüência infinita de variáveis aleatórias*

$X = (X_1, X_2, \dots)$ *é permutável se e somente se existe uma σ -álgebra $\mathcal{C}$ de eventos tal que para cada $m \geq 1$*

$$P(X_1 \leq x_1, \dots, X_m \leq x_m) = \int \prod_{j=1}^m P(X_1 \leq x_j | \mathcal{C}) dP.$$

Prova: $\Rightarrow$ Seja $N \in \mathbb{N}$ e fixe $m < N$. Definamos

$$\begin{aligned} M &= \{ \alpha : \alpha = (\alpha_1, \dots, \alpha_m) \text{ com } \alpha_i \in \{1, \dots, N\} \text{ se } i = 1, \dots, m \}, \\ H &= \{ \alpha : \alpha \in M, \alpha_i \neq \alpha_j, \text{ para cada } i \neq j \text{ e } 1 \leq i, j \leq m \} \quad \text{e} \\ I(\alpha) &= I_{[X_{\alpha_1} \leq x_1, \dots, X_{\alpha_m} \leq x_m]}. \end{aligned}$$

A partir do Lema 3 e com as notações acima, temos que

$$P(X_1 \leq x_1, \dots, X_m \leq x_m | \mathcal{I}^N) = \frac{1}{N(N-1) \dots (N-m+1)} \sum_{\alpha \in H} I(\alpha)$$

e

$$P(X_1 \leq x_1 | \mathcal{I}^N) \dots P(X_m \leq x_m | \mathcal{I}^N) = \frac{1}{N^m} \left(\sum_{\alpha \in M-H} I(\alpha) + \sum_{\alpha \in H} I(\alpha) \right).$$

Seja

$$\Delta_N = |P(X_1 \leq x_1, \dots, X_m \leq x_m | \mathcal{I}^N) - P(X_1 \leq x_1 | \mathcal{I}^N) \dots P(X_m \leq x_m | \mathcal{I}^N)|$$

$$= \left| \left(\frac{1}{N(N-1)\cdots(N-m+1)} - \frac{1}{N^m} \right) \sum_{\alpha \in H} I(\alpha) - \frac{1}{N^m} \sum_{\alpha \in M-H} I(\alpha) \right|$$

Seja $A = N(N-1)\cdots(N-m+1)$ e $B = N^m$, então

$$\begin{aligned} \Delta_N &= \frac{1}{AB} \left| (B-A) \sum_{\alpha \in H} I(\alpha) - A \sum_{\alpha \in M-H} I(\alpha) \right| \\ &\leq \frac{1}{AB} \left\{ (B-A) \sum_{\alpha \in H} I(\alpha) + A \sum_{\alpha \in M-H} I(\alpha) \right\} \end{aligned}$$

Agora

$$\sum_{\alpha \in H} I(\alpha) \leq \#H = A \quad \text{e} \quad \sum_{\alpha \in M-H} I(\alpha) \leq \#(M-H) = B-A,$$

logo

$$\Delta_N \leq \frac{2A(B-A)}{AB} = 2 \left\{ 1 - \frac{N(N-1)\cdots(N-m+1)}{N^m} \right\} \xrightarrow{N \rightarrow \infty} 0.$$

A partir disto concluímos que

$$\lim_{N \rightarrow \infty} P(X_1 \leq x_1, \dots, X_m \leq x_m | \mathcal{I}^N) = \lim_{N \rightarrow \infty} P(X_1 \leq x_1 | \mathcal{I}^N) \cdots P(X_m \leq x_m | \mathcal{I}^N) \text{ q.c.}$$

Utilizando o Lema 4, temos a partir desta igualdade que

$$P(X_1 \leq x_1, \dots, X_m \leq x_m | \mathcal{I}) = \prod_{i=1}^m P(X_i \leq x_i | \mathcal{I}) \text{ q.c.}$$

Por outro lado,

$$P(X_1 \leq x_1 | \mathcal{I}^N) = \frac{1}{N} \sum_{i=1}^N I_{[X_i \leq x_1]} = P(X_i \leq x_1 | \mathcal{I}^N) \text{ q.c.},$$

para $i = 1, \dots, m$. Tomando limite a ambos os lados nesta última igualdade, obtemos

$$P(X_1 \leq x_1 | \mathcal{I}) = P(X_i \leq x_1 | \mathcal{I}) \text{ q.c.}, \quad i = 1, \dots, m.$$

Logo, escolhendo $\mathcal{C} = \mathcal{I}$, temos que

$$P(X_1 \leq x_1, \dots, X_m \leq x_m | \mathcal{I}) = \prod_{i=1}^m P(X_1 \leq x_i | \mathcal{I}) \text{ q.c.}$$

concluindo-se assim a primeira parte da prova.

⇐ Basta notar a partir da hipótese que

$$P(X_1 \leq x_{\pi(1)}, \dots, X_m \leq x_{\pi(m)}) = \int \prod_{j=1}^m P(X_1 < x_j | \mathcal{C}) dP . \quad \square$$

Notemos que na prova do Teorema 2, $\mathcal{C}$ pode ser escolhido como $\overline{\mathcal{I}}$, o complemento de $\mathcal{I}$. Mais ainda, $\mathcal{C}$ pode ser escolhida como a σ -álgebra caudal $\mathcal{F} = \bigcap_{n=1}^{\infty} \mathcal{F}_n$, onde $\mathcal{F}_n = \sigma(X_n, X_{n+1}, \dots)$ (Chow e Teicher (1978), p. 220).

Alguns autores (por exemplo, Kingman (1978) e Cifarelli e Regazzini (1982)) preferem a seguinte formulação do teorema: a seqüência infinita de variáveis aleatórias

$\{X_n : n \in \mathbb{N}\}$ é permutável se, e somente se, para cada $m \in \mathbb{N}$, $(x_1, \dots, x_m) \in \mathbb{R}^m$

$$P(X_1 \leq x_1, \dots, X_m \leq x_m) = \int_{\mathcal{F}} \prod_{j=1}^m F(x_j) d\mu(F) ,$$

onde $\mathcal{F}$ é a família de funções de distribuição de probabilidade sobre a reta. E mais ainda, se $F_N(x) = \frac{1}{N} \sum_{j=1}^N I_{[X_j \leq x]}$ e $F(x)_{q.c.} = \lim_{N \rightarrow \infty} F_N(x)$ (o qual existe em virtude do Lema 4), então tal formulação resulta da escolha de $\mathcal{C} = \sigma\{F(x) : -\infty < x < \infty\}$ na prova do Teorema 2. De fato, $\mathcal{C}$ coincide essencialmente com $\mathcal{I}$ (Loève (1963) e Sudgen (1978)).

Notemos que a medida μ na mistura possui significado como a opinião inicial a respeito do limite do processo $\{F_N : N \in \mathbb{N}\}$. Por exemplo, se em nossa opinião μ é uma medida de probabilidade concentrada sobre a classe de distribuições normais com média zero, isto é, se $\mu(\mathcal{F}_{0,N}) = 1$ com

$$\mathcal{F}_{0,N} = \{F | F(x) = \Phi\left(\frac{x}{\sigma}\right), \quad \forall x \in \mathbb{R}, \text{ algum } \sigma^2 > 0\} ,$$

onde Φ é a função de distribuição $N(0, 1)$, então a lei do processo é descrita por

$$P(X_1 \leq x_1, \dots, X_m \leq x_m) = \int_{(0, \infty)} \prod_{j=1}^m \Phi\left(\frac{x_j}{\sigma}\right) d\nu(\sigma^2) , \quad (1)$$

para todo $m = 1, 2, \dots$ e todo $(x_1, \dots, x_m) \in \mathbb{R}^m$, com ν a medida de probabilidade induzida pela função $\psi : \mathcal{F}_{0,N} \mapsto (0, \infty)$ tal que $\psi(F) = \int x^2 dF$.

A forma funcional do modelo na representação (1) está explícita somente pela particular opinião a respeito da lei do limite do processo $\{F_N : N \in \mathbb{N}\}$. Em geral o teorema de de Finetti é essencialmente um resultado de existência, não fornecendo necessariamente a forma funcional do modelo na representação, a menos que especificações como a mencionada acima, sejam feitas (ou então que a estrutura do espaço amostral seja muito simples, isto é, finita). Se no exemplo acima tivéssemos julgado que o processo $\{X_n : n \in \mathbb{N}\}$ é ortogonalmente invariante (uma avaliação mais forte que permutabilidade), a representação em (1) seria uma consequência de tal avaliação (Kingman (1972)), deixando como único elemento inicial de incerteza a medida ν da mistura, a qual, como veremos na próxima seção, representa a opinião inicial a respeito da variável aleatória $\lim_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N X_i^2$ q.c. Desta forma, é possível reduzir a questão de expressar incertezas a respeito de quantidades de dimensão infinita, para quantidades finito-dimensionais. Mais ainda, podemos obter (veja, por exemplo, Diaconis et al (1985)) a representação completa, isto é, tanto a forma funcional do modelo na mistura como a medida misturadora através de sucessivos julgamentos de invariância. Note que não existe em qualquer destes processos parametrizações arbitrárias. A atitude de subordinar (Wechsler (1993)) a consideração de parâmetros para julgamentos de invariância é denominada preditivismo. Neste sentido, o teorema de de Finetti descreve como um subjetivista inclui parâmetros no modelo. Assim, por exemplo, o modelo normal na identidade (1) é uma consequência de julgar se o processo invariante sob transformações ortogonais. Justificativas preditivistas de outros modelos paramétricos têm sido fornecidos por diversos autores. Por exemplo, Freedman (1962a, 1962b), Kingman (1971), Smith (1981), Spizzichino (1988), Ressel (1985), Diaconis e Freedman (1990), entre outros, forneceram resultados para classes de variáveis aleatórias permutáveis que podem ser expressas como mistura de membros da família exponencial. Diaconis e Freedman (1981), denominam isto de permutabilidade parcial. Algumas destas situações serão revisadas na próxima seção.

Como menciona Barlow (1991), o teorema de de Finetti possui um caráter multidisciplinar, tendo impacto não apenas para os estatísticos, mas também para matemáticos e filósofos. No âmbito matemático, o teorema tem sido provado sob variadas formas (veja, por exemplo, Loève (1963), Chow e Teicher (1978), Aldous (1985) Dynkin (1978)) e em variados graus de generalidade. Por exemplo, Hewitt e Savage (1955) provaram o teorema em espaços métricos completos e separáveis, utilizando resultados de análise convexa. Olshen (1974) prova o teorema neste mesmo tipo de espaço utilizando o teorema ergódico de Birkoff. Outros resultados que se utilizam do teorema de representação - teoremas centrais de limite, caracterização de distribuição de extremos associadas seqüências infinitas de variáveis permutáveis - foram fornecidos por Chernoff e Teicher (1958) e Berman (1962). Até a existência de espaços (métrico e separável), onde o teorema da representação não vale, foi mostrada (Dubins e Freedman (1986)). Um outro aspecto interessante associado ao teorema de de Finetti é a validade da representação (como mistura de seqüências de variáveis aleatórias independentes e identicamente distribuídas), quando permutabilidade está restrita a seqüências finitas apenas. Esta questão é de especial interesse quando consideramos populações finitas, (Spizzichino (1982)) e está conceitualmente associada à possibilidade de prolongar infinitamente uma coleção finita de variáveis aleatórias permutáveis (De Finetti (1969)). A definição a seguir estabelece formalmente estas idéias.

Definição 5: *Uma coleção finita de variáveis aleatórias permutáveis $X_1, X_2, \dots, X_N$ é dita “estendível” se existe um processo $Z = (Z_1, Z_2, \dots)$ permutável tal que*

$$(X_1, X_2, \dots, X_N) \stackrel{D}{=} (Z_1, Z_2, \dots, Z_N)$$

Proposição 2: Se $X = (X_1, X_2, \dots)$ é um processo estocástico permutável, então

$$\text{Cov}(X_1, X_2) \geq 0 .$$

Prova: A partir do teorema da representação temos que $X_1, X_2, \dots$ são condicionalmente independentes e identicamente distribuídas, dado $\mathcal{I}$, a σ -álgebra de eventos permutáveis.

Logo,

$$\begin{aligned} \text{Cov}(X_1, X_2) &= E(E(X_1 X_2 | \mathcal{I})) - [E(E(X_1 | \mathcal{I}))]^2 \\ &= E(E^2(X_1 | \mathcal{I})) - [E(E(X_1 | \mathcal{I}))]^2 = V(E(X_1 | \mathcal{I})) \geq 0 . \end{aligned}$$

□

Notemos que a partir da Definição 2 e desta última proposição, que correlação não negativa é uma condição necessária para estendibilidade. O exemplo a seguir exhibe uma coleção finita de variáveis aleatórias não estendível.

Exemplo 1: Seja $\mathbf{y} \in \mathbb{R}^N$ fixado (com pelo menos um par de componentes distintas) e $\mathcal{X}_{\mathbf{y}} = \{\pi \mathbf{y} : \pi \in S^N\}$. Sejam $X_1, X_2, \dots, X_N$ variáveis aleatórias discretas com distribuição de probabilidade definida por

$$P(X_1 = x_1, \dots, X_N = x_N) = \frac{1}{N!} I_{\mathcal{X}_{\mathbf{y}}}(x_1, \dots, x_N) ,$$

então $X_1, X_2, \dots, X_N$ são variáveis aleatórias permutáveis e

$$\Pr\{X_1 = y_i, X_2 = y_j\} = \frac{1}{N(N-1)} \quad \text{se } i, j \in \{1, \dots, N\}, i \neq j .$$

Logo,

$$E(X_1) = \frac{1}{N} \sum_{i=1}^N y_i = \bar{y} \quad \text{e} \quad E(X_1 X_2) = \frac{1}{N(N-1)} \sum_{\substack{i, j \in \{1, \dots, N\} \\ i \neq j}} y_i y_j ,$$

portanto,

$$\text{Cov}(X_1, X_2) = \frac{1}{N(N-1)} \sum_{\substack{i,j \in \{1, \dots, N\} \\ i \neq j}} y_i y_j - (\bar{y})^2.$$

Agora $V(X_1) = \frac{1}{N} \sum_{i=1}^N (y_i - \bar{y})^2$, portanto, $\text{Cov}(X_1, X_2) = -\frac{V(X_1)}{(N-1)}$, concluindo-se a partir disto que $\text{Corr}(X_1, X_2) = -\frac{1}{N-1}$, que é exatamente o limite estabelecido na propriedade 3, no início deste capítulo.

Correlação não negativa, não é uma condição suficiente para estendibilidade, como mostra o exemplo a seguir.

Exemplo 2: Sejam X_1, X_2, X_3 variáveis aleatórias permutáveis tomando valores em $\{0, 1\}$ tal que

$$\begin{aligned} P(X_1 = x) &= \frac{1}{2} \quad \text{se } x = 0, 1, \\ P(X_1 = x_1, X_2 = x_2) &= \frac{1}{4} \quad \text{se } (x_1, x_2) \in \{0, 1\}^2, \\ P(X_1 = x_1, X_2 = x_2, X_3 = x_3) &= \begin{cases} \frac{1}{4} & \text{se } \sum_{i=1}^3 x_i = 1 \\ \frac{1}{4} & \text{se } x_1 = x_2 = x_3 = 1 \\ 0 & \text{em caso contrário} \end{cases} \end{aligned}$$

Se a sequência for estendível, então

$$\begin{aligned} &P(X_1 = 0, X_2 = 0, X_3 = 0, X_4 = 0) + P(X_1 = 1, X_2 = 0, X_3 = 0, X_4 = 0) \\ &= P(X_2 = 0, X_3 = 0, X_4 = 0) = 0 \end{aligned}$$

e

$$\begin{aligned} &P(X_1 = 1, X_2 = 0, X_3 = 0, X_4 = 0) + P(X_1 = 1, X_2 = 1, X_3 = 0, X_4 = 0) \\ &= P(X_1 = 1, X_3 = 0, X_4 = 0) = \frac{1}{4} \end{aligned}$$

Logo,

$$P(X_1 = 0, X_2 = 0, X_3 = 0, X_4 = 0) = P(X_1 = 1, X_2 = 0, X_3 = 0, X_4 = 0) = 0$$

e a partir disto

$$P(X_1 = 1, X_2 = 1, X_3 = 0, X_4 = 0) = \frac{1}{4} .$$

Por outro lado,

$$\begin{aligned} & P(X_1 = 1, X_2 = 1, X_3 = 0, X_4 = 0) + P(X_1 = 1, X_2 = 1, X_3 = 1, X_4 = 0) \\ &= P(X_1 = 1, X_2 = 1, X_4 = 0) = 0 . \end{aligned}$$

Portanto,

$$P(X_1 = 1, X_2 = 1, X_3 = 0, X_4 = 0) = 0 ,$$

em contradição com a solução das duas primeiras equações. Consequentemente, a seqüência não é estendível, mas $\text{Cov}(X_1, X_2) = 0$.

A seguir estabelecemos uma condição suficiente para estendibilidade.

Proposição 3: *Sejam $X_1, X_2, \dots, X_N$ variáveis aleatórias definidas sobre o mesmo espaço de probabilidade $(\Omega, \mathcal{A}, P)$. Seja Y uma quantidade aleatória definida sobre o mesmo espaço com valores em $\mathcal{Y}$. Se $X_1, X_2, \dots, X_N$ são condicionalmente independentes, dado Y , então a seqüência finita $X_1, X_2, \dots, X_N$ é estendível.*

Prova: Denotemos por P_y uma versão de probabilidade condicional regular (a qual existe pois os espaços são bem comportados (Ash, pp. 262-266)) de $X_1|Y = y$ e $\mu = PY^{-1}$ a medida de probabilidade induzida por Y . Seja F_y a função de distribuição associada a P_y . Consideremos o seguinte conjunto de distribuições finito dimensionais $\{F_n : n \in \mathbb{N}\}$, onde

$$F_n(z_1, \dots, z_n) = \int_{\mathcal{Y}} \prod_{i=1}^n F_y(z_i) d\mu(y) , \quad (z_1, \dots, z_n) \in \mathbb{R}^n .$$

É claro a partir da construção que o sistema de distribuições finito dimensionais é consistente, logo existe um processo $Z = (Z_1, Z_2, \dots)$ com este sistema de distribuições

finito dimensionais (teorema de extensão de Kolmogorov, Ash, pp. 189-193). O processo Z é permutável e

$$F_N(z_1, \dots, z_N) = P(X_1 \leq z_1, \dots, X_N \leq z_N)$$

para cada $(z_1, \dots, z_N) \in \mathbb{R}^N$, concluindo-se assim a prova. $\square$

No Capítulo 3 algumas classes especiais de coleções finitas de variáveis aleatórias permutáveis serão estudadas sobre este ponto de vista. Relações com os problemas clássicos de momentos (Shohat and Tamarkin (1943), de Finetti (1937)) nos permitirão estabelecer uma condição necessária e suficiente para alguns dos casos considerados.

1.3 Formas infinitas do Teorema de de Finetti

Como foi dito na seção anterior, o teorema de de Finetti é essencialmente um resultado de existência, não fornecendo necessariamente a forma funcional do modelo na representação. Somente quando a estrutura do espaço de estados do processo é muito simples (isto é, finita), a forma funcional do modelo na representação é obtida apenas pelo julgamento de permutabilidade. O teorema de de Finetti para processos permutáveis 0-1 é um exemplo desta situação. Quando o espaço de estados de um processo permutável não é finito, são necessários julgamentos adicionais de invariância para obter a forma funcional do modelo na representação.

O propósito desta seção é discutir alguns teoremas de representação que caracterizam classes de processos permutáveis com espaço de estados enumerável, começando com uma discussão detalhada do caso 0-1.

Em cada caso abordado, são discutidas tanto a representação para seqüências finitas como para seqüências infinitas de variáveis aleatórias permutáveis. Os resultados são apresentados mantendo a linha de demonstração de Heath e Sudderth (1977). A discussão

baseia-se essencialmente nos trabalhos de Diaconis (1977), Diaconis e Freedman (1987), Barlow (1991) e Wechsler (1993).

O Modelo Bernoulli

A seguir, são caracterizadas seqüências de variáveis aleatórias permutáveis tomando valores em $\{0, 1\}$. A importância desta classe de variáveis aleatórias é discutida à luz do teorema de de Finetti. Começaremos expondo algumas propriedades desta classe de processos.

Proposição 4: *A coleção de variáveis aleatórias $X_1, X_2, \dots, X_N$ tomando valores em $\{0, 1\}$ é permutável se, e somente se,*

$$\Pr\{X_1 = x_1, \dots, X_N = x_N \mid \sum_{i=1}^N X_i = t\} = \binom{N}{t}^{-1} I_{\{t\}} \left(\sum_{i=1}^N x_i \right)$$

para cada $t \in \{0, 1, \dots, N\}$ e $(x_1, \dots, x_N) \in \{0, 1\}^N$.

Prova: $\Rightarrow$ A partir da hipótese de permutabilidade temos que

$$P(X_1 = x_1, \dots, X_N = x_N) = P(X_1 = x_{\pi(1)}, \dots, X_N = x_{\pi(N)}) ,$$

para cada $(x_1, \dots, x_N) \in \{0, 1\}^N$ e $\pi \in S^N$.

Seja $\mathcal{X}_t = \{(x_1, \dots, x_N) \in \{0, 1\}^N : \sum_{i=1}^N x_i = t\}$, $t \in \{0, 1, \dots, N\}$. Claramente se $y \in \mathcal{X}_t$ então, $\{\pi y : \pi \in S^N\} = \mathcal{X}_t$. Denotemos por p_t a probabilidade (constante) de $(X_1, \dots, X_N)$ assumir um valor de $\mathcal{X}_t$. Então

$$\Pr\left\{\sum_{i=1}^N X_i = t\right\} = \sum_{(y_1, \dots, y_N) \in \mathcal{X}_t} P(X_1 = y_1, \dots, X_N = y_N) = (\#\mathcal{X}_t)p_t = \binom{N}{t} p_t .$$

Portanto,

$$P(X_1 = x_1, \dots, X_N = x_N \mid \sum_{i=1}^N X_i = t) = I_{\{t\}} \left(\sum_{i=1}^N x_i \right) \binom{N}{t}^{-1}$$

⇐ Basta notar que

$$P(X_1 = x_1, \dots, X_N = x_N) = \sum_{t=0}^N \binom{N}{t}^{-1} I_{\{t\}}(\sum_{i=1}^N x_i) P\left(\sum_{i=1}^N X_i = t\right)$$

e que cada termo na somatória é invariante sob permutações. $\square$

Denotemos por Q_{Nt} a distribuição de um vetor aleatório uniformemente distribuído sobre $\mathcal{X}_t = \{(x_1, \dots, x_N) \in \{0, 1\}^N : \sum_{i=1}^N x_i = t\}$, $t \in \{0, 1, \dots, N\}$. A proposição acima estabelece que cada medida de probabilidade permutável, P sobre $\{0, 1\}^N$, pode ser obtida misturando-se os elementos da família $\{Q_{Nt} : t = 0, 1, \dots, N\}$. Na realidade tal família corresponde aos pontos extremos desta classe de medidas de probabilidade. Mostramos isto a seguir

Proposição 5: *Seja $\mathcal{P}^N$ a classe de medidas de probabilidade permutável definidas sobre $\{0, 1\}^N$. Então $\mathcal{P}^N$ é um conjunto convexo com pontos extremos*

$$\varepsilon_N = \{Q_{Nt} : t \in \{0, 1, \dots, N\}\}.$$

Além disto cada $P \in \mathcal{P}^N$ admite representação única como combinação linear convexa de elementos de ε_N . Em outras palavras $\mathcal{P}^N$ é um simplex.

Prova: A convexidade é imediata a partir da definição de permutabilidade. Também a partir da Proposição 4, cada $P \in \mathcal{P}^N$ pode ser representado como mistura de elementos de ε_N . Para mostrar que ε_N é o conjunto de pontos extremos de $\mathcal{P}^N$, suponhamos que existe $P_1, P_2 \in \mathcal{P}^N$, ($P_1 \neq P_2$) tal que

$$Q_{Nt} = \alpha P_1 + (1 - \alpha) P_2, \quad 0 < \alpha < 1.$$

Então, notando que $Q_{Nt}(x_1, \dots, x_N) = 0$ para cada $(x_1, \dots, x_N) \notin \mathcal{X}_t$, temos a partir desta igualdade que

$$0 = P_1(x_1, \dots, x_N) = P_2(x_1, \dots, x_N)$$

para cada $(x_1, \dots, x_N) \notin \mathcal{X}_t$.

Agora, P_1, P_2 são medidas de probabilidade, portanto $P_i(x_1, \dots, x_N) > 0$ para algum $(x_1, \dots, x_N) \in \mathcal{X}_t$, $i = 1, 2$. Mas

$$P_i(x_1, \dots, x_N) = P_i(x_{\pi(1)}, \dots, x_{\pi(N)}) , \quad i = 1, 2$$

para cada permutação π de $\{1, \dots, N\}$. Logo,

$$P_i(x_1, \dots, x_N) = \frac{1}{\binom{N}{t}} I_{\mathcal{X}_t}(x_1, \dots, x_N) , \quad i = 1, 2$$

em contradição com a suposição inicial.

Para unicidade, suponhamos que μ e ν são duas medidas de probabilidade distintas definidas sobre $\{0, 1, \dots, N\}$ tal que

$$P(\cdot) = \sum_{t=0}^N Q_{Nt}(\cdot) d\mu(t) = \sum_{t=0}^N Q_{Nt}(\cdot) d\nu(t) ,$$

para algum $P \in \mathcal{P}^N$.

Agora $Q_{Nt}(x_1, \dots, x_N) = \binom{N}{t}^{-1} I_{\{t\}}\left(\sum_{i=1}^N x_i\right)$, $(x_1, \dots, x_N) \in \{0, 1\}^N$. Consequentemente, para $t' \in \{0, 1, \dots, N\}$ e $(x_1, \dots, x_N) \in \{0, 1\}^N$ tal que $\sum_{i=1}^N x_i = t'$ a igualdade acima implica em

$$Q_{Nt'}(x_1, \dots, x_N) \mu(t') = Q_{Nt'}(x_1, \dots, x_N) \nu(t') .$$

Portanto, $\mu = \nu$, em contradição com a suposição de μ e ν serem distintas. $\square$

A partir das proposições anteriores, podemos concluir que se uma medida de probabilidade P definida sobre $\{0, 1\}^N$ é permutável, então existe uma única medida de probabilidade μ sobre $\{0, 1, \dots, N\}$ tal que

$$P(\cdot) = \int_{\{0, 1, \dots, N\}} Q_{Nt}(\cdot) d\mu(t) .$$

A representação acima pode ser interpretada como segue.

Considere uma urna contendo bolas rotuladas com os números 0 e 1. O número N de bolas na urna é conhecido. As bolas são selecionadas sucessivamente e sem reposição. Seja $\mathbf{x}_N = (x_1, \dots, x_N) \in \{0, 1\}^N$, com x_i sendo o rótulo da bola na i -ésima extração. A distribuição de probabilidade Q_{Nt} corresponde à probabilidade de seleção das bolas numa dada ordem, de uma urna contendo t bolas rotuladas com o número 1. A medida μ na mistura, reflete a incerteza a respeito do número t de bolas rotuladas com o número 1 na urna.

Suponhamos agora que n ($n < N$) bolas são selecionadas sucessivamente da urna contendo t bolas rotuladas com o número 1. Então, a probabilidade de obtermos t_s bolas com o número 1 ($t_s < t$) numa dada ordem é

$$\frac{\frac{t!}{(t-t_s)!} \frac{(N-t)!}{(N-t-(n-t_s))!}}{\frac{N!}{(N-n)!}} = \frac{\binom{N-n}{t-t_s}}{\binom{N}{t}}.$$

Assim, se Q_{Ntn} denota a distribuição n -dimensional de Q_{Nt} , então

$$Q_{Ntn}(x_1, \dots, x_n) = \frac{\binom{N-n}{t-t_s}}{\binom{N}{t}} I_{\{0,1,\dots,t\}}(t_s),$$

com $t_s = \sum_{i=1}^n x_i$.

Consequentemente, a distribuição do número de bolas rotuladas com o número 1 ao selecionar n bolas sem reposição da urna é hipergeométrica com parâmetro t , tamanho da população N e tamanho da amostra n . A relação entre a distribuição hipergeométrica e a distribuição binomial, junto com a lei Forte dos Grandes Números, são a chave da versão do teorema de de Finetti para processos permutáveis tomando valores em $\{0, 1\}$.

Teorema 3 (de Finetti): *Se $X_1, X_2, \dots$ é uma seqüência infinita de variáveis aleatórias permutáveis tomando valores em $\{0, 1\}$, então existe uma única medida de probabilidade μ sobre $[0, 1]$, tal que para cada $n \in \mathbb{N}$*

$$P(X_1 = x_1, \dots, X_n = x_n) = \int_{[0,1]} \theta^{\sum_{i=1}^n x_i} (1 - \theta)^{n - \sum_{i=1}^n x_i} d\mu(\theta), \quad (x_1, \dots, x_n) \in \{0, 1\}^n.$$

Prova: Seja $N \in \mathbb{N}$ e fixe $n < N$, então

$$P(X_1 = x_1, \dots, X_n = x_n) = \sum_{t=0}^N \frac{\binom{N-n}{t-t_s}}{\binom{N}{t}} P\left(\sum_{i=1}^N X_i = t\right) = \int_{\{0,1,\dots,N\}} \frac{\binom{N-n}{t-t_s}}{\binom{N}{t}} d\nu_N(t),$$

onde ν_N é a P -lei de $\sum_{i=1}^N X_i$ e $t_s = \sum_{i=1}^n x_i$.

Agora,

$$\frac{\binom{N-n}{t-t_s}}{\binom{N}{t}} = \prod_{i=0}^{t_s-1} \frac{\left(\frac{t}{N} - \frac{i}{N}\right) \prod_{i=0}^{n-(t_s+1)} \left(\frac{N-i}{N} - \frac{t}{N}\right)}{\prod_{i=0}^{n-1} \left(1 - \frac{i}{N}\right)}.$$

Portanto, fazendo $\theta = \frac{t}{N}$ na identidade acima, obtemos

$$P(X_1 = x_1, \dots, X_n = x_n) = \int_{[0,1]} f_N(\theta) d\mu_N(\theta), \quad (2)$$

onde μ_N é a P -lei de $\frac{\sum_{i=1}^N X_i}{N}$ e

$$f_N(\theta) = \frac{\prod_{i=0}^{t_s-1} \left(\theta - \frac{i}{N}\right) \prod_{i=0}^{n-(t_s+1)} \left(\frac{N-i}{N} - \theta\right)}{\prod_{i=0}^{n-1} \left(1 - \frac{i}{N}\right)}.$$

Agora, pela lei Forte dos Grandes Números, $\mu_N \xrightarrow{N \rightarrow \infty} \mu$, onde μ é a lei da variável aleatória $\lim_{N \rightarrow \infty} \frac{\sum_{i=1}^N X_i}{N}$ e f_N converge uniformemente para f onde $f(\theta) = \theta^{t_s} (1-\theta)^{n-t_s} I_{[0,1]}(\theta)$.

Consequentemente, fazendo N tender para infinito na identidade (2), obtemos

$$P(X_1 = x_1, \dots, X_n = x_n) = \int_{[0,1]} \theta^{t_s} (1-\theta)^{n-t_s} d\mu(\theta). \quad \square$$

O teorema estabelece a conexão entre probabilidade - a qual de Finetti entende como subjetiva - e frequência relativa. Um valor particular da variável aleatória $\lim_{N \rightarrow \infty} \frac{\sum_{i=1}^N X_i}{N}$ representa a propensão (Dawid, (1982)) da ocorrência de um dado evento, numa série infinita de ensaios. A medida de probabilidade μ na mistura expressa a nossa opinião inicial a respeito destas propensões. Tal opinião será modificada através da fórmula de Bayes, conforme a ocorrência dos eventos.

O processo Bernoulli na mistura é uma consequência da avaliação de permutabilidade. Dawid (1982) usa o termo inter-subjetivo para tal modelo. O sentido deste termo pode ser obtido da seguinte situação: um grupo de pesquisadores concorda com o julgamento de permutabilidade de uma sequência de variáveis dicotômicas, contudo, descrevem suas incertezas por diferentes distribuições de probabilidades. Pelo teorema de de Finetti, podemos concluir que os pesquisadores discordam apenas na distribuição μ da variável aleatória $\lim_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N X_i$, mas não na distribuição condicional (I -modelo) da sequência dado $\lim_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N X_i = \theta$ (I -parâmetro).

No processo acima descrito, o julgamento de permutabilidade junto com a distribuição a priori μ determinam completamente a distribuição inicial do processo. Contudo, a lei do processo não precisa ser especificada nesta ordem. Avaliações além de permutabilidade podem nos levar a tal especificação, sem ter que proceder em dois estágios como é comumente feito na prática. Por exemplo, se julgamos que o processo se comporta como uma sequência de indicadores em um esquema de urna de Pólya (Feller (1966), Vol II, p. 226-227), então a medida μ na mistura pertence necessariamente à família Beta. Na realidade, não existem muitas escolhas; Hill et al (1987) mostraram que um processo de urna permutável somente pode ser Pólya, Bernoulli ou determinístico.

O teorema anterior pode ser estendido de maneira natural a processos permutáveis com espaço de estados finito. Os modelos hipergeométrico multivariado e multinomial surgem como consequência da avaliação de permutabilidade, nestes casos. Tratamos os tais casos a seguir.

O Modelo Multinomial

Sejam $X_1, X_2, \dots, X_N$ variáveis aleatórias permutáveis tomando valores no conjunto $A = \{a_1, \dots, a_M\}$, $a_i \in \mathbb{R}$, $i = 1, \dots, M$. Seja $\mathbf{Y}_i = (I_{\{a_1\}}(X_i), \dots, I_{\{a_M\}}(X_i))$,

$i = 1, \dots, N$. Claramente, $\mathbf{Y}_1, \mathbf{Y}_2, \dots, \mathbf{Y}_N$ são vetores aleatórios permutáveis tomando valores em $\{\mathbf{e}_1, \dots, \mathbf{e}_M\}$, base padrão de $\mathbb{R}^M$. Além disto, para qualquer subconjunto $\{x_1, \dots, x_N\}$ de A ,

$$P(X_1 = x_1, \dots, X_N = x_N) = P(\mathbf{Y}_1 = \mathbf{y}_1, \dots, \mathbf{Y}_N = \mathbf{y}_N) ,$$

onde $\mathbf{y}_i = (I_{\{a_1\}}(x_i), \dots, I_{\{a_m\}}(x_i))$, $i = 1, \dots, N$. Consequentemente, a caracterização do processo permutável $\mathbf{X} = (X_1, X_2, \dots)$ é equivalente à caracterização do processo permutável $\mathbf{Y} = (\mathbf{Y}_1, \mathbf{Y}_2, \dots)$.

Seja $\mathbf{T}_N = \sum_{i=1}^N \mathbf{Y}_i$ e $\mathbf{t} = (t_1, t_2, \dots, t_M)$ um valor particular de $\mathbf{T}_N$.

Proposição 6: *A coleção de vetores aleatórios $\mathbf{Y}_1, \mathbf{Y}_2, \dots, \mathbf{Y}_N$ tomando valores em $\{\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_M\}$ é permutável se, e somente se,*

$$P(\mathbf{Y}_1 = \mathbf{y}_1, \dots, \mathbf{Y}_N = \mathbf{y}_N | \mathbf{T}_N = \mathbf{t}) = \binom{N}{t_1, \dots, t_M}^{-1} I_{\{\mathbf{t}\}} \left(\sum_{i=1}^N \mathbf{y}_i \right) ,$$

$$\text{onde } \binom{N}{t_1, \dots, t_M} = \binom{N}{t_1} \cdots \binom{N - \sum_{i=1}^{M-2} t_i}{t_{M-1}} \text{ e } \mathbf{t} \in \mathcal{Y} = \{(z_1, \dots, z_M) \in \{0, 1, \dots, N\}^M : \sum_{i=1}^M z_i = N\}.$$

Prova: Será omitida, pois é análoga à da Proposição 5. □

Denotemos por $Q_{N\mathbf{t}}$ a lei de um vetor aleatório uniformemente distribuído sobre $\mathcal{X}_{\mathbf{t}} = \{(\mathbf{y}_1, \dots, \mathbf{y}_N) \in \{\mathbf{e}_1, \dots, \mathbf{e}_M\}^N : \sum_{i=1}^N \mathbf{y}_i = \mathbf{t}\}$. A proposição anterior estabelece que toda medida de probabilidade P permutável definida sobre $\{a_1, \dots, a_M\}^N$, pode ser expressa como mistura de elementos da família $\{Q_{N\mathbf{t}} : \mathbf{t} \in \mathcal{Y}\}$. De fato, procedendo de maneira análoga à da Proposição 5 (onde $M = 2$), pode-se mostrar que tal família é o conjunto de pontos extremos desta classe de medidas de probabilidade permutáveis. Além disto, a representação é única.

A lei Q_{Nt} é a distribuição de probabilidade associada ao seguinte esquema de urna: Uma urna contém N bolas rotuladas com os números $a_1, a_2, \dots, a_M$ diferentes. O vetor $t = (t_1, t_2, \dots, t_M)$ é o vetor de totais de bolas de cada tipo na urna. As bolas são selecionadas sucessivamente e sem reposição. A probabilidade de obtermos as bolas numa dada ordem é $\left(\frac{N}{t_1 t_2 \dots t_M}\right)^{-1}$. A medida que mistura os elementos da família $\{Q_{Nt}, t \in \mathcal{Y}\}$ expressa incertezas a respeito do número de bolas de cada tipo na urna.

Assim, se Q_{Ntn} denota a distribuição n -dimensional ($n < N$) de Q_{Nt} (isto é, a distribuição associada à seleção de n bolas sem reposição da urna), então

$$Q_{Ntn}(y_1, \dots, y_n) = \frac{\binom{N-n}{t_1-t_{1s}, \dots, t_M-t_{Ms}}}{\binom{N}{t_1 \dots t_M}} I_{\prod_{i=1}^M \{0, 1, \dots, t_i\}}(t_s), \quad (3)$$

onde $t_s = \sum_{i=1}^n y_i = (t_{1s}, \dots, t_{Ms})$ e $\sum_{i=1}^M t_{is} = n$.

A versão do teorema de de Finetti para processos permutáveis com espaço de estados finito é dada a seguir.

Teorema 4: Se $Y_1, Y_2, \dots, Y_N, \dots$ é um processo permutável tomando valores em $\{e_1, \dots, e_M\}$, então existe uma única medida de probabilidade μ sobre

$$S = \{(z_1, \dots, z_M) \in [0, 1]^M : \sum_{i=1}^M z_i = 1\}$$

tal que para cada $(y_1, \dots, y_n) \in \{e_1, \dots, e_M\}^n$ e $n \in \mathbb{N}$

$$\Pr \{Y_1 = y_1, \dots, Y_n = y_n\} = \int_S \prod_{i=1}^M \theta_i^{t_{si}} d\mu(\theta),$$

onde $\sum_{i=1}^n y_i = t_s = (t_{s1}, t_{s2}, \dots, t_{sM})$.

Prova: Seja $N \in \mathbb{N}$ e fixe $n < N$ ($n \in \mathbb{N}$), então

$$P\{Y_1 = y_1, \dots, Y_n = y_n\} = \sum_{t \in \mathcal{Y}} \frac{\binom{N-n}{t_1-t_{1s}, \dots, t_M-t_{Ms}}}{\binom{N}{t_1 t_2 \dots t_M}} P(T_N = t),$$

onde $\mathcal{Y} = \{(z_1, \dots, z_M) \in \{0, 1, \dots, N\}^M : \sum_{i=1}^M z_i = N\}$ e $t_s = \sum_{i=1}^n y_i$.

Notando que

$$\frac{\binom{N-n}{t_1-t_1, \dots, t_M-t_M}}{\binom{N}{t_1, \dots, t_M}} = \frac{1}{\prod_{i=0}^{n-1} \left(1 - \frac{i}{N}\right)} \prod_{i=1}^M \prod_{l=0}^{t_i-1} \left(\frac{t_i}{N} - \frac{l}{N}\right)$$

e fazendo $\theta_i = \frac{t_i}{N}$, $i = 1, \dots, M$, $\boldsymbol{\theta} = (\theta_1, \dots, \theta_M)$, obtemos

$$\Pr \{Y_1 = y_1, \dots, Y_n = y_n\} = \int_S \frac{\prod_{i=1}^M \prod_{l=0}^{t_{is}-1} \left(\theta_i - \frac{l}{N}\right)}{\prod_{i=0}^{n-1} \left(1 - \frac{i}{N}\right)} d\mu_N(\boldsymbol{\theta}),$$

onde μ_N é a P -lei de $\frac{1}{N} \sum_{i=1}^N Y_i$ e $S = \{(z_1, \dots, z_M) \in [0, 1]^M : \sum_{i=1}^M z_i = 1\}$. Seja $f_N(\boldsymbol{\theta}) = \frac{\prod_{i=1}^M \prod_{l=0}^{t_{is}-1} \left(\theta_i - \frac{l}{N}\right)}{\prod_{i=0}^{n-1} \left(1 - \frac{i}{N}\right)} I_S(\boldsymbol{\theta})$. Então, fazendo N tender para infinito, temos que f_N converge uniformemente para f , onde $f(\boldsymbol{\theta}) = \prod_{i=1}^M \theta_i^{t_{is}} I_S(\boldsymbol{\theta})$. Por outro lado, pelo Lema 4, $\frac{1}{N} \sum_{i=1}^N Y_i$ converge quase certamente. Denotando por μ a lei da variável aleatória $\lim \frac{1}{N} \sum_{i=1}^N Y_i$ e fazendo N tender para infinito na identidade (3), conclui-se a prova. $\square$

O modelo Multinomial $(1, \theta_1, \dots, \theta_M)$ (na sua versão singular) é uma consequência da permutabilidade do processo $\{Y_n : n \in \mathbb{N}\}$. Um valor particular do vetor $\boldsymbol{\theta} = (\theta_1, \dots, \theta_M)$ descreve a distribuição de frequências dos valores em A associada a uma realização do processo $\{Y_n : n \in \mathbb{N}\}$ (ou equivalentemente do processo $\{X_n : n \in \mathbb{N}\}$). Com efeito, de acordo com o Lema 4, a seqüência $\left\{ \frac{1}{N} \left(\sum_{i=1}^N I_{\{a_1\}}(X_i), \dots, \sum_{i=1}^N I_{\{a_m\}}(X_i) \right) : N \in \mathbb{N} \right\}$ converge quase certamente e $\boldsymbol{\theta}$ simboliza este limite. A medida μ na mistura expressa então a nossa opinião em relação às propensões de ocorrência dos valores em A associados a uma seqüência infinita de ensaios. Um exemplo de processos permutáveis deste tipo é o esquema de urna de Polya com M cores (Hill et al (1987)). Neste caso, a medida μ na mistura necessariamente pertence à família Dirichlet.

A consideração de processos permutáveis com espaço de estados finito, permitiu obter explicitamente a forma funcional do modelo na mistura e a partir disto uma justificativa preditivista do modelo multinomial. Esta situação é na realidade excepcional, gerada pela estrutura simples do espaço de estados. Quando o espaço de estados não é finito,

são necessários julgamentos além de permutabilidade para obtermos a forma funcional do modelo na representação. Algumas destas situações são ilustradas a seguir.

O Modelo Geométrico

Sejam $X_1, X_2, \dots, X_N$ variáveis aleatórias tomando valores em $\mathbb{N}_0$. Suponha que a distribuição condicional de $X_1, \dots, X_N | \sum_{i=1}^N X_i = t$ é uniforme sobre

$$\mathcal{X}_t = \{(x_1, \dots, x_N) \in \mathbb{N}_0^N : \sum_{i=1}^N x_i = t\}, \quad t \in \mathbb{N}_0.$$

Se Q_{Nt} denota tal distribuição condicional, então

$$Q_{Nt}(x_1, \dots, x_N) = \binom{N+t-1}{t} I_{\{t\}} \left(\sum_{i=1}^N x_i \right).$$

De fato, o número de elementos em $\mathcal{X}_t$ corresponde ao número de maneiras em que t defeitos (não distinguíveis) podem ser distribuídos em N itens (Feller, Vol. 1 (1968)), cada um dos quais pode conter de 0 até t defeitos.

É evidente que os elementos da família de distribuições de probabilidade $\{Q_{Nt} : t \in \mathbb{N}_0\}$ são medidas de probabilidade permutáveis. Consequentemente, se P denota a lei de $X_1, \dots, X_N$, então

$$P((x_1, \dots, x_N)) = \int Q_{Nt}(x_1, \dots, x_N) d\mu(t), \quad (4)$$

com μ a P -lei de $T_N = \sum_{i=1}^N X_i$. A medida μ na mistura expressa a nossa opinião a respeito do número total de defeitos que podem ser encontrados no lote de N itens.

Seja $\mathcal{P}^N$ a classe de medidas de probabilidade que podem ser expressas como mistura de elementos da família $\{Q_{Nt}, t \in \mathbb{N}_0\}$. Uma caracterização da classe $\mathcal{P}^N$ é dada a seguir (Diaconis e Freedman (1987)).

Seja $\mathcal{N} = \{(a_1, \dots, a_N) \in \mathbb{Z}^N : \sum_{i=1}^N a_i = 0\}$ e para cada $A \subset \mathbb{R}^N$ definamos $A + \mathbf{a} = \{\mathbf{x} + \mathbf{a} : \mathbf{x} \in A\}$, $\mathbf{a} \in \mathbb{R}^N$.

Proposição 7: *Sejam $X_1, \dots, X_N$ variáveis aleatórias assumindo valores em $\mathbb{N}_0$ e P a lei de $(X_1, \dots, X_N)$. Então $P \in \mathcal{P}^N$ se, e somente se, para cada $A \subset \mathbb{N}_0^N$ tal que $A + \mathbf{a} \subset \mathbb{N}_0^N$, $\mathbf{a} \in \mathcal{N}$ temos $P(A) = P(A + \mathbf{a})$.*

Prova: $\Rightarrow$ Se $P \in \mathcal{P}^N$, então

$$P(A) = \sum_{t=0}^{\infty} Q_{Nt}(A) P\left(\sum_{i=1}^N X_i = t\right).$$

Mas,

$$Q_{Nt}(A) = \sum_{\{(y_1, \dots, y_N) \in A : \sum_{i=1}^N y_i = t\}} Q_{Nt}(y_1, \dots, y_N) = \#(A \cap \mathcal{X}_t) \binom{N+t-1}{t}^{-1}.$$

Agora, se $\mathbf{a} \in \mathcal{N}$ é tal que $A + \mathbf{a} \subset \mathbb{N}_0^N$, temos que $\#(A \cap \mathcal{X}_t) = \#((A + \mathbf{a}) \cap \mathcal{X}_t)$.

Para ver isto, consideremos a função $\varphi : A \cap \mathcal{X}_t \mapsto A + \mathbf{a} \cap \mathcal{X}_t$ com $\varphi(\mathbf{y}) = \mathbf{y} + \mathbf{a}$. Claramente φ está bem definida, pois se $\mathbf{y} \in A \cap \mathcal{X}_t$ então $\mathbf{y} + \mathbf{a} \in A + \mathbf{a}$ e $\sum_{i=1}^N (y_i + a_i) = t$, já que $\sum_{i=1}^N a_i = 0$. Também é imediato que φ é uma bijeção, portanto $A \cap \mathcal{X}_t$ e $A + \mathbf{a} \cap \mathcal{X}_t$ possuem a mesma cardinalidade. Consequentemente,

$$\begin{aligned} Q_{Nt}(A) &= (\#(A \cap \mathcal{X}_t)) \binom{N+t-1}{t}^{-1} = (\#(A + \mathbf{a} \cap \mathcal{X}_t)) \binom{N+t-1}{t}^{-1} \\ &= Q_{Nt}(A + \mathbf{a}). \end{aligned}$$

Como t é arbitrário, segue que $P(A) = P(A + \mathbf{a})$.

$\Leftarrow$ Temos a partir da hipótese que

$$P(X_1 = x_1, \dots, X_N = x_N) = P(X_1 = x_1 + a_1, \dots, X_N = x_N + a_N),$$

para cada $(a_1, \dots, a_N) \in \mathbb{Z}^N$ tal que $\sum_{i=1}^N a_i = 0$ e $(x_1, \dots, x_N) \in \mathbb{N}_0^N$ tal que $(x_1 + a_1, \dots, x_N + a_N) \in \mathbb{N}_0^N$.

Sejam $(x_1, \dots, x_N), (x'_1, \dots, x'_N) \in \mathbb{N}_0^N$ tal que $\sum_{i=1}^N x_i = \sum_{i=1}^N x'_i$.

Notando que $\sum_{i=1}^N x'_i = \sum_{i=1}^N x_i + \sum_{i=1}^N (x'_i - x_i)$ e escolhendo $\mathbf{a} = (x'_1 - x_1, \dots, x'_N - x_N)$ temos a partir da hipótese que

$$P(X_1 = x_1, \dots, X_N = x_N) = P(X_1 = x'_1, \dots, X_N = x'_N) .$$

Consequentemente

$$P(X_1 = x_1, \dots, X_N = x_N) = p_t I_{\{t\}} \left(\sum_{i=1}^N x_i \right) ,$$

onde p_t denota a probabilidade constante de $(x_1, \dots, x_N) \in \mathcal{X}_t$. Assim,

$$P \left(\sum_{i=1}^N X_i = t \right) = (\#\mathcal{X}_t) \cdot p_t = \binom{N+t-1}{t} p_t .$$

A prova segue-se a partir disto e a definição de probabilidade condicional. $\square$

A classe $\mathcal{P}^N$ de medidas de probabilidade aqui definida é um conjunto convexo com pontos extremos $\{Q_{Nt} : t \in \mathbb{N}_0\}$ e mais ainda $\mathcal{P}^N$ é um simplex. Um aspecto interessante desta classe de medidas de probabilidade permutáveis é que as únicas medidas de probabilidade produtos contidas nela são aquelas associadas a N variáveis aleatórias independentes com distribuição comum Geométrica (θ) (Pereira (1980)), Kagan et al (1972)). Notemos também que nem toda coleção finita de variáveis aleatórias permutáveis tomando valores em $\mathbb{N}_0$ admite a representação em (4). Por exemplo, se P_λ^N denota a lei de N variáveis aleatórias independentes com distribuição comum Poisson (λ), então P_λ^N (veja o Exemplo 4, no Capítulo 3) não admite a representação em (4).

Antes de enunciar a versão do Teorema 2 para esta situação, é conveniente notar que se Q_{Ntn} denota a distribuição n -dimensional ($n < N$) de Q_{Nt} , então

$$Q_{Ntn}(x_1, \dots, x_n) = \frac{\binom{N-n+t-t_s-1}{t-t_s}}{\binom{N+t-1}{t}} I_{\{0, \dots, t\}}(t_s) , \quad (5)$$

onde $t_s = \sum_{i=1}^n x_i$, $(x_1, \dots, x_n) \in \mathbb{N}_0^n$. De fato, se existem um total de t_s defeitos em n itens ($n < N$), então o número de maneiras em que $t - t_s$ defeitos podem ser distribuídos nos restantes $N - n$ itens, é exatamente o numerador na expressão (5).

A versão infinita da representação que decorre das identidades (4) e (5) é dada a seguir.

Teorema 5: *Seja $\{X_n : n \in \mathbb{N}\}$ seqüência infinita de variáveis aleatórias tomando valores em $\mathbb{N}_0$, tal que*

$$P\{X_1 = x_1, \dots, X_n = x_n\} = P(X_1 = x_1 + a_1, \dots, X_n = x_n + a_n) ,$$

para cada $n \in \mathbb{N}$, $(x_1, \dots, x_n) \in \mathbb{N}_0^n$ e $(a_1, \dots, a_n) \in \mathbb{Z}^n$ com $\sum_{i=1}^n a_i = 0$ e $(x_1 + a_1, \dots, x_n + a_n) \in \mathbb{N}_0^n$. Suponha que $E(X_1) < \infty$. Então existe uma única medida de probabilidade μ sobre $[0, 1]$ tal que

$$P(X_1 = x_1, \dots, X_n = x_n) = \int_{[0,1]} (1 - \theta)^n \theta^{t_s} d\mu(\theta) ,$$

para cada $n \in \mathbb{N}$, $(x_1, \dots, x_n) \in \mathbb{N}_0^n$.

Prova: A partir da hipótese e da Proposição 7, temos que para cada $N \in \mathbb{N}$

$$X_1, X_2, \dots, X_N \mid \sum_{i=1}^N X_i = t \sim U\{(x_1, \dots, x_N) \in \mathbb{N}_0^N : \sum_{i=1}^N x_i = t\} .$$

Logo, para $N \in \mathbb{N}$ e $n < N$ temos que

$$\begin{aligned} P(X_1 = x_1, \dots, X_n = x_n) &= \sum_{t=0}^{\infty} P(X_1 = x_1, \dots, X_n = x_n \mid \sum_{i=1}^N X_i = t) P(\sum_{i=1}^N X_i = t) \\ &= \sum_{t=t_s}^{\infty} \frac{\binom{N-n+t-t_s}{t-t_s}}{\binom{N+t-1}{t}} P\left(\sum_{i=1}^N X_i = t\right) \end{aligned}$$

Agora,

$$\begin{aligned} \frac{\binom{N-n+t-t_s}{t-t_s}}{\binom{N+t-1}{t}} &= \frac{(N-1) \cdots (N-n) \overbrace{t(t-1) \cdots (t-t_s+1)}^{t_s \text{ fatores}}}{(N+t-1)(N+t-2) \cdots (N+t-t_s-n)} \left(\frac{N}{N}\right)^{t_s+n} \\ &= \frac{\prod_{i=1}^n \left(1 - \frac{i}{N}\right) \prod_{i=0}^{t_s-1} \left(\frac{t}{N} - \frac{i}{N}\right)}{\prod_{i=1}^{t_s+n} \left(1 + \frac{t}{N} - \frac{i}{N}\right)} . \end{aligned}$$

Logo,

$$P(X_1 = x_1, \dots, X_n = x_n) = \prod_{i=1}^n \left(1 - \frac{i}{N}\right) \sum_{i=0}^{\infty} \frac{\prod_{i=0}^{t_s-1} \left(\frac{t}{N} - \frac{i}{N}\right)}{\prod_{i=1}^{t_s+n} \left(1 + \frac{t}{N} - \frac{i}{N}\right)} I_{\{t_s, t_s+1, \dots\}}(t) P\left(\sum_{i=1}^N X_i = t\right).$$

Fazendo $\lambda = \frac{t}{N}$, obtemos

$$P(X_1 = x_1, \dots, X_n = x_n) = \int_{(0, \infty)} \frac{\prod_{i=1}^n \left(1 - \frac{i}{N}\right) \prod_{i=0}^{t_s-1} \left(\lambda - \frac{i}{N}\right)}{\prod_{i=1}^{t_s+n} \left(1 + \lambda - \frac{i}{N}\right)} I_{(\frac{t_s}{N}, +\infty)}(\lambda) d\nu_N(\lambda), \quad (6)$$

onde ν_N é a P -lei de $\frac{1}{N} \sum_{i=1}^N X_i$.

Notando pela lei forte dos grandes números que $\nu_N \mapsto \nu$, onde ν é a lei de $\lim_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N X_i$ (q.c) e que a função no integrando converge uniformemente para $\frac{\lambda^{t_s}}{(1 + \lambda)^{t_s+n}} I_{(0, \infty)}(\lambda)$ quando $N \rightarrow \infty$, temos, tomando o limite em (6), que

$$P(X_1 = x_1, \dots, X_n = x_n) = \int_{[0, \infty)} \frac{\lambda^{t_s}}{(1 + \lambda)^{t_s+n}} d\nu(\lambda).$$

Finalmente, fazendo $\theta = \frac{\lambda}{1+\lambda}$, conclui-se a prova. $\square$

Processos permutáveis com espaço de estado discreto foram caracterizados originalmente por Freedman (1962a), que fornece caracterizações de alguns modelos discretos - tais como geométrico e Poisson - através de resultados que caracterizam processos estacionários como misturas de cadeias de Markov. Estes resultados estão estreitamente relacionados aos desenvolvidos por Lauritzen (1974) e Dynkin (1978), no contexto de modelos de pontos extremos e aos posteriormente fornecidos por Diaconis e Freedman (1981). Todos estes trabalhos envolvem de alguma maneira o conceito de seqüências de estatísticas preditivas suficientes, introduzido originalmente por Lauritzen (1974). Este conceito foi também abordado por Campanino e Spizzichino (1982) como uma maneira de expressar suficiência em termos de distribuições condicionais de observações dados os valores de outras observações, eliminando assim a referência ao conceito de parâmetro de uma população.

Conforme observamos na literatura, a tendência dos trabalhos que vêm sendo publicados nas duas últimas décadas é obter resultados que abrangem uma ampla classe de modelos. Mesmo assim, devido ao nível de sofisticação das ferramentas utilizadas em alguns destes trabalhos (por exemplo Ressel (1985) utiliza resultados recentes de análise harmônica sobre semigrupos), continuaram a aparecer resultados particulares. A explicação para isso é a simplicidade das técnicas utilizadas e a possibilidade de interpretação dos resultados em termos de fundamentos de inferência estatística. Um exemplo disto é o resultado de Heath e Sudderth para processos permutáveis 0-1, apresentado no começo desta seção. Seguindo esta linha e enfatizando os aspectos que descrevem a visão preditivista em inferência estatística, Wechsler (1993) forneceu uma caracterização de processos poissonianos.

Outras caracterizações que exibem modelos normais univariados e multivariados, exponencial etc. foram fornecidos por Kingman (1972), Smith (1981), Dawid (1977), Diaconis e Freedman (1981). Alguns destes resultados serão discutidos na próxima seção, no contexto das formas finitas do teorema de de Finetti.

1.4 Formas finitas do Teorema de de Finetti

Seqüências infinitas de variáveis aleatórias permutáveis podem ser representadas como mistura de seqüências de variáveis aleatórias independentes e identicamente distribuídas. Tal tipo de representação nem sempre é verdadeira quando tratamos com seqüências finitas de variáveis aleatórias permutáveis. O Exemplo 1 da seção 2 ilustra esta situação. As formas finitas do teorema de de Finetti tratam de seqüências finitas de variáveis aleatórias permutáveis de comprimento n , que podem ser estendidas para seqüências finitas de comprimento N ($n < N$). Estima-se a distância de variação total entre a lei P^n da seqüência $\{X_k : 1 \leq k \leq n\}$ e uma lei mistura de variáveis aleatórias independentes e identicamente

distribuídas. Diaconis (1977) forneceu um dos primeiros resultados nesta direção, considerando seqüências finitas de variáveis aleatórias permutáveis tomando valores em $\{0, 1\}$. Posteriormente, Diaconis e Freedman (1980) forneceram resultados para seqüências finitas de variáveis aleatórias permutáveis com espaço de estados finito. Todos estes resultados foram provados utilizando a forma finita exata, que permite representar a lei da seqüência como mistura de processos de urna, onde a seleção é feita sem reposição de uma urna com uma dada composição.

Diaconis e Freedman (1987) forneceram versões finitas do teorema de de Finetti para uma série de modelos particulares: Poisson, Geométrico, Normal e Exponencial.

Diaconis, Eaton e Lauritzen (1988) desenvolveram formas finitas do teorema para modelos lineares normais (univariados e multivariados) através de postulados de simetria expressos em termos de invariância sob a ação de vários grupos de transformações ortogonais. A construção de medidas invariantes sobre grupos topológicos (Nachbin (1965)) foi a principal técnica utilizada por esses autores.

O objetivo desta seção é discutir alguns destes trabalhos. Começaremos com uma breve discussão da noção de N -estendibilidade e das propriedades da distância de variação total.

Definição 6: Uma seqüência finita de variáveis aleatórias permutáveis $\{X_k : 1 \leq k \leq n\}$ é dita N -estendível ($N > n$) se existe uma seqüência $\{Z_k : 1 \leq k \leq N\}$ de variáveis aleatórias permutáveis tal que

$$(X_1, \dots, X_n) \stackrel{D}{=} (Z_1, \dots, Z_n) .$$

O Exemplo 2, da seção 2, mostra um par de variáveis aleatórias permutáveis que pode ser estendida para uma seqüência de comprimento três, mas não para uma seqüência de comprimento quatro. Retomaremos esta discussão no exemplo a seguir.

Exemplo 3: Sejam X_1, X_2 variáveis aleatórias tomando valores em $\{0, 1\}$. Seja $\mathcal{P}_2$ o espaço de todas as possíveis medidas de probabilidade permutável para (X_1, X_2) , isto é,

$$\mathcal{P}_2 = \{(p_{00}, p_{10}, p_{01}, p_{11}) \in [0, 1]^4 : \sum_{i,j=0,1} p_{ij} = 1, \quad p_{10} = p_{01}\},$$

onde $p_{ij} = P(X_1 = i, X_2 = j)$.

Então, a partir da Proposição 5, temos que $\mathcal{P}_2$ é um simplex com pontos extremos $\mathcal{E}_1 = \langle (1, 0, 0, 0), (0, \frac{1}{2}, \frac{1}{2}, 0), (0, 0, 0, 1) \rangle$.

Seja $I_2 = \{(p^2, p(1-p), p(1-p), (1-p)^2) : 0 \leq p \leq 1\}$ o espaço de medidas de probabilidade produto (permutáveis). Claramente, $I_2 \subset \mathcal{P}_2$. Seja $\mathcal{F}_2$ a classe de medidas de probabilidade obtida misturando-se os elementos de I_2 . Então,

i) A seqüência X_1, X_2 é estendível se, e somente se, $P \in \mathcal{F}_2$.

Com efeito, se $P \in \mathcal{F}_2$, então existe μ medida de probabilidade sobre $[0, 1]$ tal que

$$P(X_1 = x_1, X_2 = x_2) = \int_{[0,1]} p^{x_1+x_2} (1-p)^{2-(x_1+x_2)} d\mu(p),$$

$(x_1, x_2) \in \{0, 1\}^2$.

Seja $\{Z_k : k \geq 1\}$ processo permutável cujas distribuições finito dimensionais são definidas por

$$P(Z_1 = z_1, \dots, Z_N = z_N) = \int_{[0,1]} p^{\sum_{i=1}^N z_i} (1-p)^{N-\sum_{i=1}^N z_i} d\mu(p),$$

para cada $N \in \mathbb{N}$ e $(z_1, \dots, z_N) \in \{0, 1\}^N$. Claramente $(X_1, X_2) \stackrel{\mathcal{D}}{=} (Z_1, Z_2)$. Assim, (X_1, X_2) é N -estendível para cada $N > 2$. A recíproca é imediata.

Note que se $P \in \mathcal{P}_2 - \mathcal{F}_2$ então P não é estendível (infinitamente).

Notemos também que os elementos $P \in \mathcal{F}_2$ não necessariamente admitem representação única como mistura de elementos de $\mathcal{P}_2$. De fato, a partir da relação

$$P(X_1 = x_1, X_2 = x_2) = \int_{[0,1]} p^{x_1+x_2} (1-p)^{2-(x_1+x_2)} d\mu(p),$$

$(x_1, x_2) \in \{0, 1\}^2$, temos que $E(\mathbf{p}) = p_{01} + p_{11}$ e $E(\mathbf{p}^2) = p_{11}$, onde $\mathbf{p}$ é uma variável aleatória com lei μ . Assim, P determina os dois primeiros momentos de μ , o qual não é em geral suficiente para definir μ .

ii) Se $P \in \mathcal{P}_2 - \mathcal{F}_2$, então X_1, X_2 é 3-estendível se, e somente se, o sistema de equações

$$\begin{bmatrix} 1 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 1 \\ 1 & 3 & 3 & 1 \end{bmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix} = \begin{pmatrix} p_{00} \\ p_{01} \\ p_{11} \\ 1 \end{pmatrix}, \quad (x_1, x_2, x_3, x_4) \in [0, 1]^4$$

tem solução.

De fato, se X_1, X_2 for 3-estendível então

$$P(X_1 = x_1, X_2 = x_2) = P(X_1 = x_1, X_2 = x_2, X_3 = 0) + P(X_1 = x_1, X_2 = x_2, X_3 = 1)$$

para cada $(x_1, x_2) \in \{0, 1\}^2$.

Seja $p_{ijk} = P(X_1 = i, X_2 = j, X_3 = k)$, então a partir da identidade acima temos que

$$p_{00} = p_{000} + p_{001}$$

$$p_{01} = p_{010} + p_{011}$$

$$p_{11} = p_{110} + p_{111}.$$

Pela simetria temos que $p_{001} = p_{010}$, $p_{011} = p_{110}$ e $p_{000} + 3p_{001} + 3p_{011} + p_{111} = 1$. Assim, $x_1 = p_{000}$, $x_2 = p_{010}$, $x_3 = p_{110}$ e $x_4 = p_{111}$ é solução do sistema de equações acima. A recíproca segue da mesma maneira. $\square$

Os argumentos utilizados no exemplo acima podem ser facilmente estendidos para seqüências de variáveis aleatórias de comprimento maior que 2. Notemos também que o sistema de equações em ii) define um algoritmo para estudar N -estendibilidade de processos permutáveis 0-1 (Diaconis e Freedman (1980)).

No Capítulo 3 deste trabalho serão fornecidos alguns resultados adicionais sobre estendibilidade em conexão com o problema de predição em populações finitas.

O exemplo acima (parte (i)) mostra que seqüências finitas de variáveis aleatórias permutáveis (digamos de comprimento n) nem sempre admitem representação como mistura de seqüências de variáveis aleatórias independentes. Porém, se a seqüência é N -estendível para N suficientemente grande em relação a n , a representação é “aproximadamente” verdadeira (isto pode ser notado ao longo da prova do Teorema 2). A proximidade é medida em termos da distância de variação total, cuja definição damos a seguir.

Definição 7: Sejam P e Q duas medidas de probabilidade definidas sobre o espaço mensurável $(\Omega, \mathcal{A})$. A distância de variação total entre P e Q é definida por

$$\|P - Q\|_{\mathcal{A}} = 2 \sup_{A \in \mathcal{A}} |P(A) - Q(A)| .$$

Propriedades

1. $\|P - Q\|_{\mathcal{A}} = 2 \sup_{A \in \mathcal{A}} |\int \varphi dP - \int \varphi dQ|$, onde o supremo é tomado sobre todas as funções φ , $\mathcal{A}$ -mensuráveis com $0 \leq \varphi \leq 1$.
2. Se $\{P, Q\} \ll \rho$, onde ρ é uma medida σ -finita tal que $\frac{dP}{d\rho} = p$ e $\frac{dQ}{d\rho} = q$, então

$$\begin{aligned} \|P - Q\|_{\mathcal{A}} &= \int |p - q| d\rho = \int |p/q - 1| q d\rho \\ &= 2 \int (p - q)^+ d\rho = 2 \int (p/q - 1)^+ q d\rho , \end{aligned}$$

onde $f^+ = \max\{f, 0\}$ e $f^- = \max\{-f, 0\}$ com f função real.

3. Se Σ é uma subálgebra suficiente de $\mathcal{A}$ no sentido que para cada $A \in \mathcal{A}$

$$P(A|\Sigma) = Q(A|\Sigma) \text{ q.c. } \{P, Q\}$$

então

$$\|P - Q\|_{\Sigma} = \|P - Q\|_{\mathcal{A}} .$$

4. Se $T : (\Omega, \mathcal{A}) \mapsto (\Omega', \mathcal{A}')$ é uma transformação mensurável e 1-1, então

$$\|P_T(A') - Q_T(A')\|_{\mathcal{A}'} = \|P(A) - Q(A)\|_{\mathcal{A}},$$

onde P_T e Q_T são as medidas de probabilidade induzidas por T .

As provas destas propriedades são simples e podem ser encontradas em Diaconis e Freedman (1987).

A seguir, serão abordadas as formas finitas do teorema de de Finetti que envolvem alguns membros da família exponencial. O lema a seguir será de utilidade para estudar os casos que envolvem os modelos multinomial e geométrico. O resultado foi utilizado de maneira indireta por Diaconis e Freedman (1980) e formalmente estabelecido por estes autores em 1987.

Lema 5:

i) Para $0 \leq x < N$, seja $\varphi(x) = \log\left(1 - \frac{x}{N}\right)$. Então $\varphi(0) = 0$, φ é estritamente decrescente, estritamente côncava e $\varphi(N^-) = -\infty$.

ii) Para $0 \leq x < 1$, seja $f(x) = -(1-x)\log(1-x) - x$. Então $f(0) = 0$, f é estritamente decrescente, estritamente côncava e $f(1^-) = -1$.

iii) $\sum_{i=1}^m \varphi(i) \leq \int_0^m \varphi(x) dx$, $1 \leq m < N$.

iv) $\sum_{i=1}^{m-1} \varphi(i) \geq \int_1^m \varphi(x) dx \geq \int_0^m \varphi(x) dx$, $1 \leq m < N$.

v) $\int_0^y \varphi(x) dx = -(N-y)\varphi(y) - y = Nf(y/N)$.

A prova deste lema é simples, portanto será omitida.

O Modelo Multinomial

Sejam $\mathbf{Y}_1, \mathbf{Y}_2, \dots, \mathbf{Y}_N$ vetores aleatórios permutáveis tomando valores em $\{\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_M\}$ base padrão de $\mathbb{R}^M$ e seja $\mathbf{T}_N = \sum_{i=1}^N \mathbf{Y}_i$. Então, de acordo com a Proposição 6 e a identidade (3) da seção anterior, a distribuição de $\mathbf{Y}_1, \dots, \mathbf{Y}_n | \mathbf{T}_N = \mathbf{t}$ ($n < N$) é definida por

$$Q_{Ntn}(y_1, y_2, \dots, y_n) = \frac{(N-n)!}{N!} \prod_{j=1}^M \left\{ \frac{t_j!}{(t_j - t_{sj})!} \right\} I_{\prod_{i=1}^M \{0, \dots, t_i\}}(\mathbf{t}_s), \quad (8)$$

onde $\mathbf{t}_s = \sum_{i=1}^n \mathbf{y}_i = (t_{s1}, \dots, t_{sM})$ com $\sum_{j=1}^M t_{sj} = n$.

De modo que se P^n denota a lei de $\mathbf{Y}_1, \dots, \mathbf{Y}_n$, então para cada $1 \leq n \leq N$

$$P^n(\cdot) = \sum_{\mathbf{t} \in \mathcal{Y}} Q_{Ntn}(\cdot) P^N(\mathbf{T}_N = \mathbf{t}),$$

onde $\mathcal{Y} = \{(y_1, \dots, y_M) \in \{0, 1, \dots, N\}^M : \sum_{i=1}^M y_i = N\}$.

Seja P_θ^n a lei de n vetores aleatórios independentes com distribuição comum Multinomial $(1, \theta_1, \theta_2, \dots, \theta_M)$ (na sua forma singular). Notemos que se $\mathbf{t} = (t_1, t_2, \dots, t_M)$ é um valor dado de $\mathbf{T}_N$, então

$$P_{\mathbf{t}/N}^n(\mathbf{y}_1, \dots, \mathbf{y}_n) = \prod_{i=1}^M \left(\frac{t_i}{N} \right)^{t_{si}},$$

onde $\mathbf{t}_s = \sum_{i=1}^n \mathbf{y}_i = (t_{s1}, \dots, t_{sM})$ com $\mathbf{y}_1, \dots, \mathbf{y}_n$ tomando valores em $\{\mathbf{e}_1, \dots, \mathbf{e}_M\}$ e $\sum_{i=1}^M t_{si} = n$.

Notemos que $P_{\mathbf{t}/N}^n$ é a distribuição associada à seleção com reposição de n bolas de uma urna contendo N bolas com M cores diferentes e $\mathbf{t} = (t_1, \dots, t_M)$ e o vetor do número de bolas de cada cor na urna.

O resultado a seguir, fornecido por Diaconis e Freedman (1980), estabelece um limite para a distância de variação total entre distribuições associadas à seleção com e sem reposição de uma urna com a composição acima especificada.

Teorema 6: Para cada $1 \leq n \leq N$,

$$\|Q_{Ntn} - P_{t/N}^n\| \leq 2M \frac{n}{N} . \quad (9)$$

Prova: Vamos mostrar primeiro que

$$Q_{Ntn}(\mathbf{y}_1, \dots, \mathbf{y}_n) \prod_{i=1}^M \left\{ 1 - \left(\frac{t_{si}}{t_i} \right) \right\} \leq P_{t/N}^n(\mathbf{y}_1, \dots, \mathbf{y}_n) ,$$

supondo sem perda de generalidade que $t_i \geq 1$, para cada $i = 1, \dots, M$. Notando que a desigualdade acima é trivialmente verdadeira se $t_{si} = t_i$ para algum i , podemos também supor que $t_{si} < t_i$ para cada i . Desta forma $P_{t/N}^n(\mathbf{y}_1, \dots, \mathbf{y}_n) > 0$. Assim,

$$\frac{Q_{Ntn}(\mathbf{y}_1, \dots, \mathbf{y}_n)}{P_{t/N}^n(\mathbf{y}_1, \dots, \mathbf{y}_n)} \prod_{i=1}^M \left\{ 1 - \left(\frac{t_{si}}{t_i} \right) \right\} = \frac{\prod_{i=1}^M \left\{ 1 - \left(\frac{t_{si}}{t_i} \right) \right\} \left(\frac{t_i^{-t_{si}} t_i!}{(t_i - t_{si})!} \right)}{N^{-n} N! / (N - n)!} . \quad (10)$$

Agora

$$\log \left\{ \frac{N^{-n} N!}{(N - n)!} \right\} = \sum_{j=1}^{n-1} \log \left\{ 1 - \frac{j}{N} \right\} \geq N f \left(\frac{n}{N} \right) , \quad (11)$$

onde $f(x) = -(1-x) \log(1-x) - x$, $0 \leq x < 1$. Esta última desigualdade segue do Lema 5, partes (iv) e (v).

Também a partir do Lema 5, partes (iii) e (v), temos que

$$\left\{ 1 - \left(\frac{t_{si}}{t_i} \right) \right\} \left[\frac{t_i^{-t_{si}} t_i!}{(t_i - t_{si})!} \right] \leq \exp \left\{ t_i f \left(\frac{t_{si}}{t_i} \right) \right\} ,$$

com f como na identidade (11).

A prova da primeira parte conclui-se a partir do Lema 5, parte (ii), notando que f é côncava. De fato, pela desigualdade de Jensen,

$$\sum_{i=1}^M t_i f \left(\frac{t_{si}}{t_i} \right) \leq N f \left(\frac{n}{N} \right) .$$

Assim, o denominador de (10) é limitado inferiormente por $\exp[N f(\frac{n}{N})]$ e o numerador de (10) é limitado superiormente pela mesma quantidade. Consequentemente,

$$Q_{Ntn}(\mathbf{y}_1, \dots, \mathbf{y}_n) \prod_{i=1}^M \left\{ 1 - \left(\frac{t_{si}}{t_i} \right) \right\} \leq P_{t/N}^n(\mathbf{y}_1, \dots, \mathbf{y}_n) ,$$

e portanto

$$\begin{aligned} P_{\mathbf{t}/N}^n(\mathbf{y}_1, \dots, \mathbf{y}_n) - Q_{N\mathbf{t}n}(\mathbf{y}_1, \dots, \mathbf{y}_n) &\geq -Q_{N\mathbf{t}n}(\mathbf{y}_1, \dots, \mathbf{y}_n) \left\{ 1 - \prod_{i=1}^M \left\{ 1 - \left(\frac{t_{si}}{t_i} \right) \right\} \right\} \\ &\geq -Q_{N\mathbf{t}n}(\mathbf{y}_1, \dots, \mathbf{y}_n) \sum_{i=1}^M \left(\frac{t_{si}}{t_i} \right) . \end{aligned}$$

Logo

$$\left(Q_{N\mathbf{t}n}(\mathbf{y}_1, \dots, \mathbf{y}_n) - P_{\mathbf{t}/N}(\mathbf{y}_1, \dots, \mathbf{y}_n) \right)^+ \leq Q_{N\mathbf{t}n}(\mathbf{y}_1, \dots, \mathbf{y}_n) \sum_{i=1}^M \left(\frac{t_{si}}{t_i} \right) .$$

Agora, pela propriedade 2 da distância de variação total temos que

$$\begin{aligned} \|Q_{N\mathbf{t}n} - P_{\mathbf{t}/N}\| &= 2 \sum_{(\mathbf{y}_1, \dots, \mathbf{y}_n) \in S} (Q_{N\mathbf{t}n}(\mathbf{y}_1, \dots, \mathbf{y}_n) - P_{\mathbf{t}/N}(\mathbf{y}_1, \dots, \mathbf{y}_n))^+ \\ &\leq 2 \sum_{i=1}^M \sum_{(\mathbf{y}_1, \dots, \mathbf{y}_n) \in S} Q_{N\mathbf{t}n}(\mathbf{y}_1, \dots, \mathbf{y}_n) \left(\frac{t_{si}}{t_i} \right) , \end{aligned}$$

onde $t_{si} = \sum_{j=1}^n \mathbf{y}_j \mathbf{e}_i^t$ e $S = \{(\mathbf{z}_1, \dots, \mathbf{z}_n) \in \{\mathbf{e}_1, \dots, \mathbf{e}_M\}^n : (\sum_{j=1}^n \mathbf{z}_j) \mathbf{1}_M = n \text{ e } 0 \leq \sum_{j=1}^n \mathbf{z}_j \mathbf{e}_i^t \leq t_i, i = 1, \dots, M\}$, sendo $\mathbf{1}_M$ o vetor coluna M -dimensional de 1's. Notando que $\frac{n}{N}$ é a $Q_{N\mathbf{t}n}$ esperança de $\frac{t_{si}}{t_i}$, conclui-se a prova. $\square$

Teorema 7: *Sejam $\mathbf{Y}_1, \mathbf{Y}_2, \dots, \mathbf{Y}_N$ vetores aleatórios permutáveis tomando valores em $\{\mathbf{e}_1, \dots, \mathbf{e}_M\}$. Seja P^n a lei de $\mathbf{Y}_1, \mathbf{Y}_2, \dots, \mathbf{Y}_n$ ($1 \leq n \leq N$). Então existe uma medida de probabilidade μ sobre $S = \{(z_1, \dots, z_M) \in [0, 1]^M : \sum_{i=1}^M z_i = 1\}$ tal que para cada $1 \leq n \leq N$*

$$\|P^n - \int_S P_\theta^n d\mu(\theta)\| \leq 2M \frac{n}{N} . \quad (12)$$

Prova: A partir da hipótese temos que

$$P^n(\cdot) = \sum_{\mathbf{t} \in \mathcal{Y}} Q_{N\mathbf{t}n}(\cdot) P^N(\mathbf{T}_N = \mathbf{t}) ,$$

onde Q_{Nt_n} é definida pela identidade 8. Escolhendo μ como a P^N -lei de $\frac{\mathbf{T}_N}{N}$ temos que

$$\int_S P_\theta^n(\cdot) d\mu(\theta) = \sum_{t \in \mathcal{Y}} P_{t/N}^n(\cdot) P^N(\mathbf{T}_N = t) .$$

Logo

$$\begin{aligned} \|P^n - \int_S P_\theta^n d\mu(\theta)\| &= \left\| \sum_{t \in \mathcal{Y}} (Q_{Nt_n} - P_{t/N}) P^N(\mathbf{T}_N = t) \right\| \\ &\leq \sum_{t \in \mathcal{Y}} \|Q_{Nt_n} - P_{t/N}\| P^N(\mathbf{T}_N = t) . \end{aligned}$$

A prova segue a partir desta última desigualdade e do Teorema 6. $\square$

A versão infinita do teorema de de Finetti pode ser extraída a partir deste último resultado. Com efeito, se $\{\mathbf{Y}_k : k \geq 1\}$ é uma seqüência infinita de vetores aleatórios permutáveis tomando valores em $\{\mathbf{e}_1, \dots, \mathbf{e}_M\}$, então para cada $n < N$ e $N \in \mathbb{N}$ vale a desigualdade (12). Fixando n e fazendo N tender para infinito, obtemos, a partir da desigualdade (12) e a lei forte dos grandes números, o resultado estabelecido no Teorema 4. Este procedimento é de fato equivalente ao utilizado no Teorema 4, já que a desigualdade (9) do Teorema 6 garante a convergência uniforme da seqüência $\{f_N, N \geq n\}$ para f , definidas na prova do Teorema 4. Notemos que o Teorema 5 estabelece a conhecida relação entre amostragem aleatória simples com e sem reposição. Algumas aplicações destes resultados serão apresentados na seção 4 do Capítulo 3.

O Modelo Geométrico

Denotemos por Q_{Nt} à distribuição de probabilidade uniforme sobre

$$\mathcal{X}_t^N = \{(x_1, \dots, x_N) \in \mathbb{N}_0^N : \sum_{i=1}^N x_i = t\}, \quad t \in \mathbb{N}_0 ,$$

e seja P_θ^n a distribuição de probabilidade associada a n variáveis aleatórias independentes com distribuição comum Geométrica (θ) , isto é,

$$P_\theta^n(x_1, \dots, x_n) = (1 - \theta)^n \theta^{\sum_{i=1}^n x_i} I_{\mathbb{N}_0^n}(x_1, \dots, x_n) .$$

No que segue mostraremos que as n primeiras coordenadas de um ponto uniformemente distribuído sobre $\mathcal{X}_t^N$ são próximas em distância de variação total de n variáveis aleatórias independentes com distribuição Geométrica. A partir deste resultado deduz-se a versão finita do Teorema 5. Os resultados são devido a Diaconis e Freedman (1987).

Teorema 8: *Sejam Q_{Ntn} a distribuição n dimensional de Q_{Nt} e P_θ^n a distribuição acima definida. Então, para cada $1 \leq n \leq N - 3$,*

$$\|Q_{Ntn} - P_{t/N+t}^n\| \leq 2 \left\{ \frac{N^2}{(N-n-1)(N-n-2)} - 1 \right\}. \quad (13)$$

Prova: A partir da identidade (5) na Seção 3, temos que

$$\begin{aligned} Q_{Ntn}(x_1, \dots, x_n) &= \frac{\binom{N-n+t-t_s-1}{t-t_s}}{\binom{N+t-1}{t}} I_{\{0,1,\dots,t\}}(t_s) \\ &= \frac{\prod_{i=1}^n \left(1 - \frac{i}{N}\right) \prod_{i=0}^{t_s-1} \left(\frac{t}{N} - \frac{i}{N}\right)}{\prod_{i=1}^{t_s+n} \left(1 + \frac{t}{N} - \frac{i}{N}\right)} I_{\{0,1,\dots,t\}}(t_s) \end{aligned}$$

onde $t_s = \sum_{i=1}^n x_i$.

A partir disto e da definição de P_θ^n temos para $0 \leq t_s \leq t$ que

$$\frac{Q_{Ntn}(x_1, \dots, x_n)}{P_{t/N+t}^n(x_1, \dots, x_n)} = \frac{\prod_{i=1}^n \left(1 - \frac{i}{N}\right) \prod_{i=1}^{t_s-1} \left\{1 - \frac{i}{t}\right\}}{\prod_{i=1}^{t_s+n} \left(1 - \frac{i}{N+t}\right)}.$$

Utilizando argumentos análogos aos do Teorema 6, pode ser verificado a partir do Lema 5 que

$$\frac{Q_{Ntn}(x_1, \dots, x_n)}{P_{t/N+t}^n(x_1, \dots, x_n)} \left(1 - \frac{n+1}{N}\right) \left(1 - \frac{n+2}{N}\right) \leq 1.$$

Logo, se $n \leq N - 3$,

$$\left(\frac{Q_{Ntn}(x_1, \dots, x_n)}{P_{t/N+t}^n(x_1, \dots, x_n)} - 1 \right)^+ \leq \frac{N^2}{[N - (n+1)][N - (n+2)]} - 1.$$

Consequentemente,

$$\begin{aligned} \|Q_{Ntn} - P_{t/N+t}^n\| &= 2 \sum_{(x_1, \dots, x_n) \in S} \left(\frac{Q_{Ntn}(x_1, \dots, x_n)}{P_{t/N+t}^n(x_1, \dots, x_n)} - 1 \right)^+ P_{t/N+t}^n(x_1, \dots, x_n) \\ &\leq 2 \left\{ \frac{N^2}{[N - (n+1)](N - (n+2))} - 1 \right\}, \end{aligned}$$

com $S = \{(z_1, \dots, z_n) \in \mathbb{N}_0^n : 0 \leq \sum_{i=1}^n z_i \leq t\}$. Concluindo-se assim a prova. $\square$

Teorema 9: *Sejam $X_1, X_2, \dots, X_N$ variáveis aleatórias permutáveis tomando valores em $\mathbb{N}_0$, tal que $X_1, X_2, \dots, X_N | \sum_{i=1}^N X_i = t \sim U\{(x_1, \dots, x_N) \in \mathbb{N}_0^N : \sum_{i=1}^N x_i = t\}$. Seja P^n a distribuição de probabilidade de $X_1, \dots, X_n$ ($1 \leq n \leq N$) e P_θ^n a distribuição de n variáveis aleatórias independentes com distribuição comum Geométrica (θ). Então existe uma medida de probabilidade μ sobre $[0, 1]$ tal que*

$$\|P^n - \int P_\theta^n d\mu(\theta)\| \leq 2 \left\{ \frac{N^2}{(N - n - 1)(N - n - 2)} - 1 \right\}. \quad (11)$$

Prova: É suficiente escolher μ como a P^N -lei de $\frac{\sum_{i=1}^N X_i}{N + \sum_{i=1}^N X_i}$ e proceder como na prova do Teorema 7. $\square$

O Modelo Exponencial

Diaconis e Freedman (1987) mostraram que as n ($n < N$) primeiras coordenadas de um ponto uniformemente distribuído sobre $\mathcal{X}_t^N = \{(x_1, \dots, x_N) \in \mathbb{R}_+^N : \sum_{i=1}^N x_i = t\}$, $t > 0$, são próximas em distância de variação total de n variáveis aleatórias independentes com distribuição comum exponencial. Seguindo-se a partir deste resultado, as formas finita e infinita do teorema da representação envolvendo o modelo exponencial. Nosso propósito no que segue é discutir estes resultados. Antes disto, são necessárias algumas definições e caracterizações relacionadas com a distribuição uniforme sobre $\mathcal{X}_t^N$.

Definição 8: Seja P uma medida de probabilidade sobre $(\mathbb{R}_+^N, \mathcal{B}_1^+)$. Dizemos que P é uma medida de probabilidade uniforme sobre $\mathcal{X}_t^N$ se $P(\mathcal{X}_t^N) = 1$ e $P(A) = P(A + \mathbf{a})$ para cada $A \in \mathcal{B}_1^+ \cap \mathcal{X}_t^N$, $\mathbf{a} = (a_1, \dots, a_N) \in \mathbb{R}^N$ tal que $\sum_{i=1}^N a_i = 0$ e $A + \mathbf{a} \in \mathcal{B}_1^+ \cap \mathcal{X}_t^N$.

Em outras palavras, P é uma medida de probabilidade uniforme sobre $\mathcal{X}_t^N$ se P concentra sua massa em $\mathcal{X}_t^N$, e é uma invariante sob translações.

Uma útil definição em termos de variáveis aleatórias é dada a seguir.

Proposição 8: *Sejam $X_1, X_2, \dots, X_N$ variáveis aleatórias independentes com distribuição comum Exponencial (1). Seja $U_i = tX_i/T_N$, $i = 1, \dots, N$ com $T_N = \sum_{i=1}^N X_i$. Então o vetor aleatório $\mathbf{U}_N = (U_1, \dots, U_N)$ é uniformemente distribuído sobre $\mathcal{X}_t^N$.*

Prova: Sem perda de generalidade, assumimos $t = 1$. Seja P a lei de $(X_1, \dots, X_N)$ e Q a lei de $\mathbf{U}_N$. É claro a partir da definição de $\mathbf{U}_N$ que $Q(\mathcal{X}_1^N) = 1$. Seja $A = \times_{i=1}^N [b_i, c_i] \cap \mathcal{X}_1^N$ com $A \neq \emptyset$. Sem perda de generalidade, podemos supor $\times_{i=1}^N [b_i, c_i] \subseteq [0, 1]^N$. Seja $\mathbf{a} = (a_1, \dots, a_N) \in \mathbb{R}^N$ tal que $\sum_{i=1}^N a_i = 0$ e $\times_{i=1}^N [b_i + a_i, c_i + a_i] \subseteq [0, 1]^N$. Então

$$\begin{aligned} Q(A) &= P\left(\left(\frac{X_1}{T_N}, \dots, \frac{X_N}{T_N}\right) \in A\right) = P\left(\bigcap_{i=1}^N \left\{b_i \leq \frac{X_i}{T_N} \leq c_i\right\}\right) \\ &= \int_0^\infty P\left(\bigcap_{i=1}^N \left\{b_i \leq \frac{X_i}{t} \leq c_i\right\} \mid T_N = t\right) dF(t). \end{aligned}$$

Agora, se $X_1, X_2, \dots, X_N$ são variáveis aleatórias independentes com distribuição comum Exponencial (λ), então T_N é uma estatística suficiente e completa para λ e $U = \left(\frac{X_1}{T_N}, \dots, \frac{X_N}{T_N}\right)$ é ancilar (Branco (1990), pp. 61-63). Pelos lemas de Basu (1958, 1959), U é independente de T_N . Portanto,

$$Q(A) = \int_0^\infty \prod_{i=1}^N P\left(b_i \leq \frac{X_i}{t} \leq c_i\right) dF(t).$$

Agora

$$\prod_{i=1}^N P(b_i t \leq X_i \leq t c_i) = \prod_{i=1}^N (e^{-b_i t} - e^{-c_i t})$$

$$\begin{aligned}
&= e^{-\sum_{i=1}^N b_i t} \left(1 - e^{-t \sum_{i=1}^N (c_i - b_i)} \right) \\
&= e^{-t \sum_{i=1}^N (b_i + a_i)} \left(1 - e^{-t \sum_{i=1}^N [c_i + a_i - (b_i + a_i)]} \right) \\
&= \prod_{i=1}^N P \left(b_i + a_i \leq \frac{X_i}{t} \leq c_i + a_i \right) .
\end{aligned}$$

Logo

$$\begin{aligned}
Q(A) &= \int P \left(\bigcap_{i=1}^N \left\{ b_i + a_i \leq \frac{X_i}{t} \leq c_i + a_i \right\} \right) dF(t) \\
&= P \left(\bigcap_{i=1}^N \left\{ b_i + a_i \leq \frac{X_i}{T_N} \leq c_i + a_i \right\} \right) \\
&= Q(A + \mathbf{a}) .
\end{aligned}$$

□

Corolário 1: Se $X_1, X_2, \dots, X_N$ são variáveis aleatórias independentes com distribuição Exponencial (λ) e $T_N = \sum_{i=1}^N X_i$, então

$$X_1, \dots, X_N \mid T_N = t \sim U\{\mathcal{X}_t^N\} , \quad t > 0 .$$

Prova: Seja f função mensurável limitada e $\mathbf{U}_N = \left(\frac{X_1}{T_N}, \dots, \frac{X_N}{T_N} \right)$, então

$$E(f(X_1, \dots, X_N) \mid T_N = t) = E(f(\mathbf{U}_N T_N) \mid T_N = t) = E(f(\mathbf{U}_N t)) ,$$

já que $\mathbf{U}_N \perp\!\!\!\perp T_N$. Notando a partir da proposição anterior que $\mathbf{U}_N t \sim U\{\mathcal{X}_t^N\}$, conclui-se a prova. □

Com estas caracterizações torna-se simples obter a distribuição das n primeiras coordenadas (e funções dela) de um ponto uniformemente distribuído sobre $\mathcal{X}_t^N$. A proposição a seguir exhibe a forma de tal distribuição.

Proposição 9: Se o vetor aleatório $\mathbf{U}_N = (U_1, \dots, U_N)$ é uniformemente distribuído sobre $\mathcal{X}_t^N$, então o vetor aleatório $\mathbf{U}_n = (U_1, \dots, U_n)$ ($1 \leq n < N$) tem densidade em relação à medida de Lebesgue definida por

$$f_t(u_1, \dots, u_n) = \frac{\Gamma(N) \left\{ 1 - \frac{\sum_{i=1}^n u_i}{t} \right\}^{N-n-1}}{\Gamma(N-n)t^n} I_{[0,t]} \left(\sum_{i=1}^n u_i \right) I_{[0,\infty)^n}(u_1, \dots, u_n) .$$

Prova: A partir da Proposição 8, é claro que $\mathbf{U}_n = (U_1, \dots, U_n)$ ($n < N$) possui uma densidade em relação à medida de Lebesgue n -dimensional. Pelo Corolário 1, esta densidade é definida pela densidade condicional de $X_1, \dots, X_n | \sum_{i=1}^N X_i = t$, onde $X_1, X_2, \dots, X_n$ são variáveis aleatórias independentes com distribuição comum Exponencial (1). Portanto, se f_1, f_2, f_3 e f são as densidades de $(X_1, \dots, X_n)$, $\sum_{i=n+1}^N X_i$, $\sum_{i=1}^N X_i$ e $(X_1, \dots, X_n) | \sum_{i=1}^n X_i = t$, respectivamente, então

$$f(u_1, \dots, u_n | t) = \frac{f_2(t - \sum_{i=1}^n u_i) f_1(u_1, \dots, u_n)}{f_3(t)} I_{[0,t]} \left(\sum_{i=1}^n u_i \right) I_{(0,\infty)^n}(u_1, \dots, u_n) .$$

Notando que f_2 e f_3 são densidades gamma $(N-n, 1)$ e gamma $(N, 1)$, respectivamente, e fazendo algumas operações algébricas obtém-se o resultado. $\square$

Proposição 10: Se o vetor aleatório $\mathbf{U}_N = (U_1, \dots, U_N) \sim U\{\mathcal{X}_t^N\}$ então, se $n < N$, $\sum_{i=1}^n U_i \stackrel{\mathcal{D}}{=} tY$, onde $Y \sim \text{Beta}(n, N-n)$.

Prova: A partir da Proposição 8 temos que $\sum_{i=1}^n U_i \stackrel{\mathcal{D}}{=} \frac{t \sum_{i=1}^n X_i}{\sum_{i=1}^N X_i}$, com $X_1, X_2, \dots, X_N$ variáveis aleatórias independentes com distribuição comum Exponencial (1).

Seja $V = \sum_{i=1}^n X_i$ e $W = \sum_{k=n+1}^N X_k$, então $V \sim \text{Gamma}(n, 1)$ e $W \sim \text{Gamma}(N-n, 1)$. Além disso, $V \perp\!\!\!\perp W$, portanto $\frac{V}{V+W} \sim \text{Beta}(n, N-n)$, concluindo-se a partir disto a prova.

$\square$

Como tem sido usual ao longo desta seção, Q_{Nt} denotará a lei de um ponto uniforme-

mente distribuído sobre $\mathcal{X}_t^N$ e Q_{Ntn} a distribuição n -dimensional.

Observação 2: Um rápido exame das proposições 9 e 10 mostra que

$$f_t(u_1, \dots, u_n) = g_t \left(\sum_{i=1}^n u_i \right) h(u_1, \dots, u_n) ,$$

onde g_t é a densidade de $\sum_{i=1}^n U_i$ com $(U_1, \dots, U_n) \sim Q_{Ntn}$ e h uma função de $(u_1, \dots, u_n)$ que não depende de t . Em outras palavras, pelo critério da fatoração, $\sum_{i=1}^n U_i$ é uma estatística suficiente para a família $\{Q_{Ntn}, t > 0\}$.

Esta observação e as proposições 9 e 10 serão de utilidade para estabelecer o próximo resultado.

Denotemos por P_λ^n à lei de n variáveis aleatórias independentes com distribuição comum Exponencial (λ).

Teorema 10: Se Q_{Ntn} é a distribuição n -dimensional de um ponto uniformemente distribuído sobre $\mathcal{X}_t^N$ e P_λ^n a distribuição acima definida, então para cada $1 \leq n \leq N - 2$

$$\|Q_{Ntn} - P_{N/t}^n\| \leq \frac{2(n+1)}{(N-n+1)} . \quad (15)$$

Prova: Um esboço da prova é dada a seguir.

A partir da propriedade 4 da distância de variação total (invariância sobre transformações 1-1), temos que

$$\|Q_{Ntn} - P_{N/t}^n\| = \|Q_{NNn} - P_1^n\| .$$

Pela suficiência de $\sum_{i=1}^n U_i$ com relação às famílias $\{Q_{Ntn} : t > 0\}$ e $\{P_\lambda^n, \lambda > 0\}$, temos a partir da propriedade 3 da distância de variação que

$$\|Q_{Ntn} - P_{N/t}^n\| = \|\tilde{Q} - \tilde{P}\| ,$$

onde $\tilde{Q}$ e $\tilde{P}$ são as Q_{NNn} e P_1^n leis de $\sum_{i=1}^n U_i$, respectivamente.

Pela proposição 10, $\tilde{Q}$ é definida pela função de densidade

$$f(x) = \frac{1}{N} \frac{\Gamma(N)}{\Gamma(n)\Gamma(N-n)} \left(\frac{x}{N}\right)^{n-1} \left(1 - \frac{x}{N}\right)^{N-n-1} I_{(0,N]}(x)$$

e $\tilde{P}$ pela função de densidade

$$g(x) = \frac{1}{\Gamma(n)} e^{-x} x^{n-1} I_{(0,\infty)}(x) .$$

Portanto,

$$\|\tilde{Q} - \tilde{P}\| = 2 \int_0^\infty \left(\frac{f(x)}{g(x)} - 1 \right)^+ g(x) dx ,$$

onde

$$\frac{f(x)}{g(x)} = \begin{cases} \frac{1}{N^n} \frac{\Gamma(n)}{\Gamma(N-n)} e^x \left(1 - \frac{x}{N}\right)^{N-n-1} & \text{se } 0 \leq x \leq N \\ 0 & \text{se } x > N . \end{cases}$$

O resultado segue a partir do Lema 5, que permite mostrar que $\left(1 - \frac{n+1}{N}\right) f/g \leq 1$. $\square$

Sejam $X_1, X_2, \dots, X_N$ variáveis aleatórias tomando valores em $\mathbb{R}_+$. Seja $\mathcal{C}_N$ a classe de medidas de probabilidade tal que

$$X_1, X_2, \dots, X_N \mid \sum_{i=1}^N X_i = t \sim U\{\mathcal{X}_t^N\} , \quad t > 0 .$$

Seja P^n a distribuição n -dimensional de $P \in \mathcal{C}_N$, então $(1 \leq n \leq N)$

$$P^n(\cdot) = \int Q_{Ntn}(\cdot) d\mu(t) , \quad (16)$$

onde μ é a P -lei de $\sum_{i=1}^N X_i$. Tal expressão fornece uma representação exata para a classe de variáveis aleatórias permutáveis com lei $P \in \mathcal{C}_N$. Se N é grande em relação a n (ou $n = o(N)$) tal representação é, em virtude da proposição anterior, próxima de mistura de n variáveis aleatórias independentes com distribuição comum Exponencial. O teorema a seguir estabelece formalmente este resultado.

Teorema 11: *Sejam $X_1, X_2, \dots, X_N$ variáveis aleatórias tomando valores em $\mathbb{R}_+$ com lei $P \in \mathcal{C}_N$. Seja P_λ^n a lei de n variáveis aleatórias independentes com distribuição comum*

Exponencial (λ) e P^n a lei de $X_1, X_2, \dots, X_n$. Então existe uma medida de probabilidade μ sobre $(0, \infty)$ tal que para cada $1 \leq n \leq N - 2$

$$\|P^n - \int P_\lambda^n d\mu(\lambda)\| \leq \frac{2(n+1)}{(N-n+1)}. \quad (17)$$

Prova: Segue do Teorema 10, escolhendo μ como a P -lei de $\frac{N}{\sum_{i=1}^N X_i}$. □

A representação infinita segue a partir deste resultado. Com efeito, seja $\{X_n : n \geq 1\}$ sequência infinita de variáveis aleatórias permutáveis tomando valores em $\mathbb{R}_+$ com a propriedade que para cada $N \in \mathbb{N}$ e $t > 0$, $X_1, X_2, \dots, X_N | \sum_{i=1}^N X_i = t \sim U\{\mathcal{X}_t^N\}$, então para cada n dado, $n < N$ e $N \in \mathbb{N}$ vale a desigualdade (17). Fazendo N tender para infinito e utilizando a lei forte dos grandes números temos que para cada $n \in \mathbb{N}$

$$P((X_1, \dots, X_n) \in A) = \int_{(0, \infty)} P_\lambda^n(A) d\mu(\lambda), \quad A \in \mathcal{B}^n, \quad (18)$$

onde μ é a medida de probabilidade associada à variável aleatória $\lim_{N \rightarrow \infty} \left(\frac{N}{\sum_{i=1}^N X_i} \right)$ (q.c.).

O modelo exponencial surge como consequência de supor que para cada $N \in \mathbb{N}$ e $t > 0$, $X_1, X_2, \dots, X_N | \sum_{i=1}^N X_i = t \sim U(\mathcal{X}_t^N)$. Esta condição é equivalente a supor que $P((X_1, \dots, X_N) \in A) = P((X_1, \dots, X_N) \in A + a)$, para cada N e todo subconjunto de Borel $A \subset \mathbb{R}_+^N$ com $a \in \mathbb{R}^N$ tal que $\sum a_i = 0$ e $A + a \in \mathbb{R}_+^N$ (Diaconis e Freedman (1987)). Utilizando-se desta caracterização, Diaconis e Ylvisaker (1985) obtiveram a representação (18) utilizando um argumento simples. De fato, a condição acima é mais forte que permutabilidade. Assim, se a_1, a_2 são dois números positivos então pela hipótese e o teorema de de Finetti temos que

$$\begin{aligned} P(X_1 > a_1 + a_2 | \mathcal{I}) &= P(X_1 > a_1 + a_2, X_2 > 0 | \mathcal{I}) \\ &= P(X_1 > a_1, X_2 > a_2 | \mathcal{I}) \\ &= P(X_1 > a_1 | \mathcal{I}) P(X_1 > a_2 | \mathcal{I}) \end{aligned}$$

onde $\mathcal{I}$ é a σ -álgebra caudal. Esta última igualdade implica que X_1 tem distribuição exponencial.

Estes mesmos autores mostraram que se além deste julgamento de invariância opinamos que $E(X_n|X_1) = a + bX_1$, para cada $n \geq 2$, então a medida na mistura é uma distribuição Gamma. Este resultado fornece uma caracterização completamente preditiva do processo.

O Modelo Normal

Freedman (1962b) e Kingman (1972), mostraram que se $\{X_n : n \geq 1\}$ é um processo estocástico ortogonalmente invariante, abaixo definido, então a lei do processo pode ser expressa como mistura de medidas produto $N(0, \sigma^2)$. A seguir, as versões finitas deste teorema são discutidas. Estes resultados são devido a Eaton (1981, 1989) e Diaconis e Freedman (1987). Começaremos com algumas definições básicas.

No que segue, O_N denotará o grupo de matrizes ortogonais de ordem $N \times N$. $\mathcal{L}(\mathbf{X})$ denotará a lei do vetor aleatório $\mathbf{X}$.

Definição 9: Um vetor aleatório $\mathbf{X}_N = (X_1, X_2, \dots, X_N)$ assumindo valores em $\mathbb{R}^N$ é dito ortogonalmente invariante se $\mathcal{L}(\mathbf{X}_N) = \mathcal{L}(\Gamma \mathbf{X}_N)$ para cada $\Gamma \in O_N$. A sequência infinita $\{X_n : n \geq 1\}$ de variáveis aleatórias assumindo valores em $\mathbb{R}$ é dita ortogonalmente invariante, se para cada $n \in \mathbb{N}$, $\mathbf{X}_n = (X_1, X_2, \dots, X_n)$ é ortogonalmente invariante.

Definição 10: Um vetor aleatório $\mathbf{X}_N = (X_1, \dots, X_N)$ assumindo valores em $\mathbb{R}^N$ é dito uniformemente distribuído sobre $\mathcal{X}_r^N = \{(x_1, \dots, x_N) \in \mathbb{R}^N : \sum_{i=1}^N x_i^2 = r^2\}$, $r > 0$, se $\mathbf{X}_N$ é ortogonalmente invariante e $P(\mathbf{X}_N \in \mathcal{X}_r^N) = 1$.

A proposição a seguir é de utilidade para o tratamento das distribuições acima definidas.

Proposição 11: *Seja $\mathbf{X}_N = (X_1, X_2, \dots, X_N)$ um vetor aleatório com distribuição ortogonalmente invariante. Seja $U_i = r \frac{X_i}{R}$ com $R = \left(\sum_{i=1}^N X_i^2\right)^{1/2}$, $i = 1, \dots, N$. Então o vetor aleatório $\mathbf{U}_N = (U_1, \dots, U_N)$ é uniformemente distribuído sobre $\mathcal{X}_r^N$.*

Prova: Sem perda de generalidade, assumimos $r = 1$. Seja P a lei de $\mathbf{X}_N$ e Q a lei de $\mathbf{U}_N$. É imediato a partir da definição que $Q(\mathcal{X}_r^N) = 1$.

Seja $\psi : \mathbb{R}^N \mapsto \mathbb{R}^N$ tal que $\psi(x_1, \dots, x_N) = \left(\frac{x_1}{\left(\sum_{i=1}^N x_i^2\right)^{1/2}}, \dots, \frac{x_N}{\left(\sum_{i=1}^N x_i^2\right)^{1/2}} \right)$. Então $\mathbf{U}_N = \psi(\mathbf{X}_N)$ e $\psi(\Gamma \mathbf{X}_N) = \Gamma \psi(\mathbf{X}_N)$ para cada $\Gamma \in O_N$.

Logo, para cada $B \in \mathcal{B}_N$, temos que

$$\begin{aligned} Q(B) &= P(\psi(\mathbf{X}_N) \in B) = P(\psi(\Gamma \mathbf{X}_N) \in B) \\ &= P(\Gamma \psi(\mathbf{X}_N) \in B) = P(\psi(\mathbf{X}_N) \in \Gamma^t B) = Q(\Gamma^t B). \end{aligned}$$

Notando que Γ é arbitrário, o resultado segue. $\square$

Observação 3: Eaton (1983, pp. 237) mostra que se $\mathbf{X}_N = (X_1, \dots, X_N)$ é um vetor aleatório ortogonalmente invariante, então $\|\mathbf{X}_N\| \perp \mathbf{U}_N$, onde $\mathbf{U}_N = \left(\frac{X_1}{\|\mathbf{X}_N\|}, \dots, \frac{X_N}{\|\mathbf{X}_N\|} \right)$. A partir disto não é difícil concluir que $X_1, \dots, X_N | \|\mathbf{X}_N\| = r \sim U\{\mathcal{X}_r^N\}$. Com efeito, se f é uma função mensurável e limitada, então

$$E(f(X_1, \dots, X_N) | \|\mathbf{X}_N\| = r) = E(f(\mathbf{U}_N \|\mathbf{X}_N\|) | \|\mathbf{X}_N\| = r) = E(f(r \mathbf{U}_N)).$$

Esta última igualdade decorre da independência entre $\mathbf{U}_N$ e $\|\mathbf{X}_N\|$.

Denotemos por Q_{Nr} à lei de um ponto uniformemente distribuído sobre $\mathcal{X}_r^N$. Seja $\mathcal{P}^N$ a classe de medidas de probabilidade ortogonalmente invariantes em $\mathbb{R}^N$. Então $\mathcal{P}^N$ é um conjunto convexo com pontos extremos $\{Q_{Nr} : r > 0\}$. De fato, pelas observações acima, temos que se $P \in \mathcal{P}^N$, então

$$P(\cdot) = \int Q_{Nr}(\cdot) d\mu(r),$$

onde μ é a P -lei de $(\sum_{i=1}^N X_i^2)^{1/2}$. Assim, se P^n é uma distribuição n -dimensional de P , então para cada $1 \leq n \leq N$

$$P^n(\cdot) = \int Q_{Nrn}(\cdot) d\mu(r), \quad (19)$$

onde Q_{Nrn} é a distribuição n -dimensional de Q_{Nr} .

A igualdade (19) é uma representação finita exata para variáveis aleatórias $X_1, \dots, X_n$ ortogonalmente invariantes que podem ser estendidas para uma seqüência finita de N ($N > n$) variáveis aleatórias com a mesma propriedade. O modelo na mistura não corresponde à lei de seqüências de variáveis aleatórias i.i.d, porém se N é grande em relação a n (ou $n = o(N)$), Q_{Nrn} se comporta quase como produto de n distribuições Normais com média zero e desvio padrão $r/\sqrt{N}$. A proximidade em distância de variação total foi estabelecida por Diaconis e Freedman (1987). Discutiremos o resultado a seguir. Antes disto, algumas propriedades da distribuição Q_{Nrn} são dadas.

Proposição 12: *Se o vetor aleatório $\mathbf{U}_N = (U_1, U_2, \dots, U_N)$ é uniformemente distribuído sobre $\mathcal{X}_r^N$, então o vetor aleatório $\mathbf{U}_n = (U_1, \dots, U_n)$ ($n < N$) tem densidade em relação à medida de Lebesgue dada por*

$$f_r(u_1, \dots, u_n) = \frac{\Gamma(\frac{N}{2})}{\Gamma(\frac{N-n}{2})} \frac{(1 - \sum_{i=1}^n u_i^2/r^2)^{\frac{N-n}{2}}}{(\sqrt{\pi}r^2)^n} I_{[0,r^2]} \left(\sum_{i=1}^n u_i^2 \right).$$

Prova: A partir da Proposição 11 e da Observação 3, temos que é suficiente determinar a lei de $X_1, X_2, \dots, X_n | \sum_{i=1}^N X_i^2 = r^2$ com $\mathbf{X}_N = (X_1, \dots, X_N) \sim N(o_N, I_N)$. Notando agora que o vetor aleatório $(X_1, \dots, X_n, \|\mathbf{X}_N\|^2)$ tem densidade $n+1$ -dimensional e denotando por f_1, f_2, f_3 às densidades de $(X_1, \dots, X_n)$, $\sum_{i=n+1}^N X_i^2$ e $\sum_{i=1}^N X_i^2$, respectivamente, obtemos,

$$f_r(u_1, \dots, u_n) = \frac{f_2(r^2 - \sum_{i=1}^n u_i^2) f_1(u_1, \dots, u_n)}{f_3(r^2)} I_{[0,r^2]} \left(\sum_{i=1}^n u_i^2 \right).$$

Agora, f_1, f_2 e f_3 são densidades $N(o_n, I_n)$, $\mathcal{X}_{N-n}^2$ e $\mathcal{X}_N^2$, respectivamente. Assim, realizando algumas operações algébricas, conclui-se a prova. $\square$

Proposição 13: Se o vetor aleatório $\mathbf{U}_N = (U_1, \dots, U_N) \sim U\{\mathcal{X}_r^N\}$, então, para $1 \leq n < N$, $\sum_{i=1}^n U_i^2 \stackrel{D}{=} r^2 Y$ onde Y é uma variável aleatória com distribuição Beta $\left(\frac{n}{2}, \frac{N-n}{2}\right)$.

Prova: A partir da Proposição 14 temos que $\sum_{i=1}^n U_i^2 \stackrel{D}{=} r^2 \frac{\sum_{i=1}^n X_i^2}{\|\mathbf{X}_N\|^2}$, onde $\mathbf{X}_N \sim N(\overset{\mathbf{O}_N}{\underset{\downarrow}{o_N}}, I_N)$.
A prova conclui-se notando que $\frac{\sum_{i=1}^n X_i^2}{\|\mathbf{X}_N\|^2} \sim \text{Beta}\left(\frac{n}{2}, \frac{N-n}{2}\right)$ (Cramer (1946), Seção 18.4). $\square$

Observação 4: Um rápido exame das proposições 12 e 13 mostra, pelo critério da fatoração, que $\sum_{i=1}^n U_i^2$ é uma estatística suficiente para a família de medidas de probabilidade $\{Q_{Nr n}, r > 0\}$.

Teorema 12: Se $Q_{Nr n}$ é a lei das n -primeiras coordenadas de um ponto uniformemente distribuído sobre $\mathcal{X}_r^N$ e P_σ^n denota a lei da n variáveis aleatórias independentes com distribuição comum $N(0, \sigma^2)$, então para cada $1 \leq n \leq N-4$

$$\|Q_{Nr n} - P_{r/\sqrt{N}}^n\| \leq 2 \frac{(n+3)}{(N-n-3)} . \quad (20)$$

Prova: Damos um esboço da prova.

Seja $T : \mathbb{R}^N \mapsto \mathbb{R}^N$ tal que $T(x_1, \dots, x_N) = \frac{\sqrt{N}(x_1, \dots, x_N)}{r}$, então a partir da propriedade 4 da distância de variação total temos que

$$\|Q_{Nr n} - P_{r/\sqrt{N}}^n\| = \|Q_{N\sqrt{N}n} - P_1^n\| .$$

Agora, pela propriedade 3 da distância de variação, temos que

$$\|Q_{N\sqrt{N}n} - P_1^n\| = \|\tilde{Q} - \tilde{P}\| ,$$

onde $\tilde{Q}$ é a Q_{Nrn} -lei de $\sum_{i=1}^n U_i^2$ e $\tilde{P}$ é a correspondente P_1^n -lei. Logo, pela Proposição 13, $\tilde{Q}$ é definida pela função densidade

$$f(x) = \frac{1}{N} \frac{\Gamma(\frac{N}{2})}{\Gamma(\frac{n}{2})\Gamma(\frac{N-n}{2})} \left(\frac{x}{N}\right)^{\frac{n}{2}-1} \left(1 - \frac{x}{N}\right)^{\frac{N-n}{2}-1} I_{[0,N]}(x) ,$$

e $\tilde{P}$ pela função densidade

$$g(x) = \frac{1}{2^{n/2}\Gamma(\frac{n}{2})} e^{-x/2} x^{\frac{n}{2}-1} I_{(0,\infty)}(x) .$$

Assim, a partir da propriedade 2 da distância de variação total, temos que

$$\|\tilde{Q} - \tilde{P}\| = 2 \int_0^\infty \left(\frac{f(x)}{g(x)} - 1 \right)^+ g(x) dx .$$

O resultado segue mostrando que

$$\log e^{x/2} \left(1 - \frac{x}{N}\right)^{\frac{N-n}{2}-1} \leq \frac{1}{2}(n+2) + \frac{1}{2}(N-n-2) \log \left(1 - \frac{n+2}{N}\right)$$

se $1 \leq n \leq N-3$ e que

$$\log \left(1 - \frac{n+2}{N}\right) \left(\frac{2}{N}\right)^{n/2} \frac{\Gamma(\frac{N}{2})}{\Gamma(\frac{N-n}{2})} \leq \frac{1}{2} [-(N-n-2) \log \left(1 - \frac{n+2}{N}\right) - n-2] . \quad \square$$

A forma finita do teorema no estilo de Finetti é consequência direta deste teorema.

Teorema 13: *Sejam $X_1, X_2, \dots, X_N$ variáveis aleatórias assumindo valores em $\mathbb{R}$, ortogonalmente invariantes. Seja P^n a lei de $X_1, X_2, \dots, X_n$ ($n \leq N$) e P_σ^n a lei de n variáveis aleatórias independentes com distribuição comum $N(0, \sigma^2)$. Então existe uma medida de probabilidade μ sobre $(0, \infty)$ tal que para cada $1 \leq n \leq N-4$*

$$\|P^n - \int P_\sigma^n d\mu(\sigma)\| \leq 2 \frac{(n+3)}{N-n-3} . \quad (21)$$

Prova: Basta observar que $P^n(\cdot) = \int Q_{Nrn}(\cdot) d\nu(r)$, onde ν é a P^N -lei de $(\sum_{i=1}^N X_i^2)^{1/2}$. Escolhendo μ como a ν -lei de $\frac{1}{\sqrt{N}}(\sum_{i=1}^N X_i^2)^{1/2}$ e utilizando o teorema anterior, conclui-se a prova. $\square$

A versão infinita do teorema de de Finetti segue como nos outros casos a partir deste último resultado. Com efeito, se $\{X_k : k \geq 1\}$ é uma seqüência infinita de variáveis aleatórias assumindo valores em $\mathbb{R}$, ortogonalmente invariantes e P^n denota a lei de $X_1, X_2, \dots, X_n$, então para cada $n \leq N - 4$, $N \in \mathbb{N}$, a desigualdade (21) vale. Portanto, fixando n e fazendo N tender para infinito, temos a partir da lei forte dos grandes números que

$$P^n(\cdot) = \int_{(0,\infty)} P_\sigma^n(\cdot) d\mu(\sigma) ,$$

com μ a lei da variável aleatória limite $\left(\frac{1}{N} \sum_{i=1}^N X_i^2\right)^{1/2}$.

Kingman (1972) derivou este resultado utilizando-se do teorema de de Finetti, a partir do fato que invariância sob transformações ortogonais implica em permutabilidade e utilizando-se das propriedades da função característica de variáveis aleatórias ortogonalmente invariantes. Smith (1981), seguindo este mesmo estilo, considerou processos $\{X_k : k \geq 1\}$ assumindo valores em $\mathbb{R}$, com a propriedade que para cada $n \in \mathbb{N}$ a distribuição de $X_1, X_2, \dots, X_n$ é invariante sob o grupo de transformações $O_n(\mathbf{1}_n) = \{\Gamma | \Gamma \in O_n \text{ e } \Gamma \mathbf{1}_n = \mathbf{1}_n\}$, onde $\mathbf{1}_n$ é o vetor n -dimensional de 1's em $\mathbb{R}^n$. Mostrando que se P^n é a lei de $X_1, X_2, \dots, X_n$, então

$$P^n(\cdot) = \int_{\mathbb{R} \times (0,\infty)} P_{\mu,\sigma}(\cdot) d\lambda(\mu, \sigma) ,$$

onde $P_{\mu,\sigma}(\cdot)$ é a lei de n variáveis aleatórias independentes com distribuição comum $N(\mu, \sigma^2)$. A versão finita deste teorema (Eaton (1989), pp. 118-122) é derivada por consideração da distribuição uniforme sobre

$$\mathcal{X}_{r,a}^N = \{(x_1, \dots, x_N) \in \mathbb{R}^N : \sum_{i=1}^N (X_i - a)^2 = r^2 \text{ e } \bar{X} = a\}, \quad r > 0, \quad a \in \mathbb{R} .$$

Tais distribuições são os pontos extremais de distribuições invariantes sob $O_N(1_N)$.

Comentários

Nesta seção foram revisados alguns modelos particulares cuja aplicação será abordada no Capítulo 3, dentro do contexto de populações finitas. Todas as situações abordadas correspondem a classes de processos permutáveis. Contudo, processos permutáveis abrangem apenas uma pequena parte de todos os possíveis campos de opinião que podemos achar na prática. Uma situação comum é quando tratamos com variáveis auxiliares. Nesta direção, Diaconis, Eaton e Lauritzen (1988) desenvolveram uma série de resultados que envolvem processos invariantes sob subgrupos do grupo de transformações ortogonais, caracterizando via teoremas do tipo de Finetti modelos lineares univariados e multivariados.

Numa direção mais aplicada, Barlow e Spizzichino (1992) e Hayakawa (1992) obtiveram formas finitas exatas no contexto de modelos de análise de sobrevivência. Não são fornecidos nesses trabalhos estimativas para a distância de variação total entre a forma finita e a correspondente forma infinita.

Também, conforme observamos na literatura, as formas finitas do teorema de de Finetti têm sido desenvolvidas para distribuições cujo suporte não depende do parâmetro. No próximo capítulo serão desenvolvidos uma série de resultados neste sentido.

1.5 Permutabilidade em populações finitas

Nesta seção faremos uma breve discussão sobre alguns resultados básicos relacionados a modelos de população finita. Dentro deste contexto, discutiremos o papel de permutabilidade.

Considere uma população finita de N unidades identificáveis $\mathcal{P} = \{1, 2, \dots, N\}$, onde N , o tamanho da população, é conhecido. Associada à i -ésima unidade em $\mathcal{P}$, existe uma

quantidade desconhecida X_i . Seja $\mathcal{X}^N$ o conjunto (conhecido) de valores possíveis do vetor aleatório $\mathbf{X}_N = (X_1, X_2, \dots, X_N)$. Com o objetivo de obter informação a respeito de uma quantidade populacional $T(\mathbf{X}_N)$ o estatístico seleciona uma amostra de n unidades ($n < N$) de $\mathcal{P}$.

Normalmente o pesquisador possui informação a priori sobre a estrutura da população relacionada à característica estudada. Na teoria de amostragem clássica essa informação é utilizada para escolher um plano de seleção. O plano amostral é representado pelo par (S, p) , onde $S = \{s | s \subset \mathcal{P}\}$ e p uma função de probabilidade sobre S . O conjunto $d = \{(i, X_i) : i \in s\}$ representa o conjunto dos dados obtidos por este plano amostral. Esses dados compreendem os rótulos, s é as características associadas X_i , $i \in s$.

Na amostragem clássica a informação a priori é utilizada para escolher o plano p de forma a obter-se uma amostra representativa (no sentido de captar a variabilidade presente na população). Dentro desta perspectiva não fica explícita a relação existente entre os elementos da amostra, $(X_i; i \in s)$, e os elementos que não fazem parte dela, $(X_i; i \notin s)$.

Embora os conceitos de estimação e parâmetro estivessem presentes dentro deste enfoque (Cochrane (1963)), a teoria de amostragem clássica e a teoria de inferência estatística eram consideradas até recentemente como áreas diferentes. Basu (1967, 1969, 1971) foi um dos primeiros a formalizar a teoria de amostragem clássica no contexto da estatística paramétrica, notando que o plano amostral induz uma família de distribuições sobre o conjunto $D = \{\{(i, X_i) : i \in s\}; s \in S, (X_1, \dots, X_N) \in \mathcal{X}^N\}$ definida por

$$P(d/\mathbf{X}_N) = \begin{cases} p(s) & \text{se } d \in \mathcal{X}_d^N \\ 0 & \text{se } d \notin \mathcal{X}_d^N \end{cases}, \quad (22)$$

onde, $\mathcal{X}_d^N = \{(Y_1, \dots, Y_N) \in \mathcal{X}^N : Y_i = X_i, i \in s\}$. Assim, a função de verossimilhança assume o valor $p(s)$ para todas as possíveis populações em $\mathcal{X}^N$ que poderiam ter gerado o dado amostral e zero para todas aquelas que não poderiam gerar esse dado. Notemos a partir disto que dentro do conjunto com verossimilhança positiva não é possível diferenciar um membro do outro. Neste sentido, a função de verossimilhança é completamente não

informativa. Consequentemente, depois da amostra ter sido selecionada conhecemos todos os $(X_i : i \in s)$ e nada sobre $(X_i : i \notin s)$. Além disto, foi estabelecido a partir deste modelo a inexistência de estimadores não-viciados de variância uniformemente mínima para o total populacional (Godambe (1955) e Basu (1971)).

A análise convencional do problema, de amostragem falha porque o espaço paramétrico é N -dimensional e o dado é um ponto n -dimensional, com $n < N$. Por isso, a única maneira de obter informação sobre os $N - n$ valores não observados é relacionando estes com os n valores observados através de um modelo matemático. Dentro deste contexto a abordagem de superpopulação (Royall, 1970b, 1971) vem auxiliar a teoria de amostragem.

No enfoque de superpopulação assume-se que a população é uma amostra aleatória de uma população infinita. O modelo de superpopulação na abordagem clássica é expresso em forma paramétrica (por exemplo Normal, Gama etc.), ficando desta maneira as unidades que estão na amostra relacionadas com as unidades fora da amostra. Esta perspectiva coloca a questão de inferência em populações finitas dentro da área de predição estatística convencional.

Neste trabalho não questionamos as falhas do modelo apresentado na expressão (22), enquanto modelo paramétrico, mas a forma de utilização do conhecimento a priori. Esse conhecimento deve representar probabilisticamente a nossa incerteza sobre o vetor populacional $\mathbf{X}_N$. Assim por exemplo, se Q é uma medida de probabilidade a priori para $\mathbf{X}_N$, com função de densidade q temos a partir da função de verossimilhança exibida em (22) que a função de densidade a posteriori é definida por

$$f(\mathbf{X}_N|d) \propto \begin{cases} p(s)q(\mathbf{X}_N) & \text{se } \mathbf{X}_N \in \mathcal{X}_d^N \\ 0 & \text{se } \mathbf{X}_N \notin \mathcal{X}_d^N. \end{cases} \quad (23)$$

Para planejamentos p não informativos $f(\cdot|d)$ fica reduzida à densidade condicional de $\mathbf{X}_N$ dado $\mathbf{X}_i, i \in s$. Por este motivo e de acordo com o princípio da verossimilhança, dentro da perspectiva bayesiana as inferências não dependem da p -distribuição gerada pela aleatorização.

O problema fica agora reduzido a traduzir probabilisticamente o conhecimento que leva ao pesquisador a utilizar um ou outro planejamento amostral. Um caso simples ocorre quando os rótulos são não informativos a priori. Isto pode ser expresso elegantemente através da noção de permutabilidade. Ericson (1969) foi um dos primeiros a explorar estas idéias, resgatando através da noção de permutabilidade resultados obtidos por métodos clássicos. Contudo, em seu trabalho, Ericson assume que a classe de prioris permutáveis pode ser expressa na forma

$$q(\mathbf{X}_N) = \int_{\Theta} \prod_{i=1}^N f(X_i/\theta) d\mu(\theta) .$$

Em outras palavras, assume que existe uma variável aleatória (ou vetor) θ tal que $X_1, X_2, \dots, X_N$ são condicionalmente independentes dado θ . Está implícito em sua suposição que a coleção de quantidades aleatórias $X_1, X_2, \dots, X_N$ é infinitamente estendível. Essa suposição nem sempre é verdadeiro e, além disto é questionável quando consideramos populações finitas.

Diferentemente, podemos tratar a noção de permutabilidade ou similares finitamente e utilizarmos das representações finitas (ou aproximações) para fazermos inferências em populações finitas. Esta postura é discutida no Capítulo 3 e além disso são apresentadas algumas aplicações e conexões com a abordagem bayesiana descrita por Ericson.

CAPÍTULO 2

Formas finitas do teorema de de Finetti para distribuições uniformes univariadas

2.1 Introdução

Conforme observamos no Capítulo 1, o teorema de de Finetti (1931, 1937) para seqüências binárias de variáveis aleatórias permutáveis, tem sido estendido em diversas direções. Resultados relativos a formas infinitas, caracterizando diversas classes de variáveis aleatórias permutáveis, têm sido desenvolvidos em variados contextos e em diferentes graus de generalidade. Por exemplo, Ressel (1985) utilizando-se de resultados de análise harmônica de semigrupos, forneceu um teorema geral a partir do qual muitos resultados concretos são deduzidos, em particular formas infinitas do teorema de de Finetti envolvendo modelos paramétricos uniformes, podem ser derivados a partir deste teorema. No entanto, devido à sofisticação das ferramentas utilizadas neste trabalho, diversos auto-

res (veja, por exemplo, Spizzichino (1988), Wechsler (1993)) continuaram desenvolvendo resultados particulares, utilizando-se de condições sobre quantidades observáveis.

Nos últimos anos, um grande volume de trabalhos em relação à forma finita de teoremas do tipo de Finetti, vem-se produzindo. Uma das razões que motivou o trabalho nesta linha é que as formas finitas referem-se à caracterização de classes de seqüências finitas de variáveis aleatórias permutáveis. Vários modelos particulares dentro deste contexto foram analisados no Capítulo 1. Um outro aspecto interessante é a relativa simplicidade com que versões infinitas de teoremas do tipo de Finetti podem ser derivados, a partir da forma finita correspondente.

Nossa contribuição neste capítulo refere-se à forma finita de teoremas do tipo de Finetti envolvendo distribuições uniformes paramétricas univariadas, dependendo de um e dois parâmetros. A linha de trabalho é similar à apresentada no Capítulo 1, onde geralmente as formas finitas foram obtidas a partir da caracterização da lei das n ($n < N$) primeiras coordenadas de um ponto uniformemente distribuído sobre certa superfície do $\mathbb{R}^N$ (por exemplo, sobre a N -esfera). Como foi mencionado no Capítulo 1, distribuições uniformes sobre tais superfícies já tinham sido caracterizadas na literatura (veja, por exemplo, Eaton (1981)). Assim, o trabalho consistiu essencialmente em estimar a distância de variação total entre a distribuição das n -primeiras coordenadas de tal ponto e o modelo associado à correspondente versão infinita. A forma finita do teorema era uma consequência direta desta estimação.

Dentro do espírito do Capítulo 1, obtivemos os resultados para modelos uniformes, os quais não encontramos na literatura. As superfícies nestes casos, no lugar de hiperesferas e simplexes, são as bordas dos hiperparalelepípedos. No nosso caso não foi possível a utilização de caracterizações já existentes. Assim, toda a construção apresentada parece ser original.

Por motivos didáticos, os resultados estão apresentados em ordem de dificuldade.

Assim, na seção 2, derivamos a forma finita do teorema envolvendo distribuições uniformes discretas. A seção 3 trata este problema para distribuições uniformes contínuas sobre os intervalos da forma $(0, \theta)$, $\theta > 0$. As distribuições uniformes sobre intervalos simétricos são discutidos na seção 4. Para completar o quadro, as distribuições uniformes sobre intervalos da forma (θ_1, θ_2) , com $\theta_1 < \theta_2$, $(\theta_1, \theta_2) \in \mathbb{R}^2$ são estudadas na seção 5.

Os resultados estabelecidos em cada uma destas seções são aproveitados para se obter, através deste caminho, a forma infinita de teoremas do tipo de Finetti, envolvendo distribuições uniformes.

Com isto, classes de seqüências finitas de variáveis aleatórias permutáveis, que podem ser representadas como quase mistura de modelos paramétricos uniformes, são caracterizadas e teoremas de representação envolvendo esta classe de distribuições são derivados em termos de condições sobre quantidades observáveis.

2.2 Distribuição uniforme sobre $\{0, 1, \dots, \theta\}$, $\theta \in \mathbb{N}$

Seja $\mathcal{X}_m^N = \{(x_1, \dots, x_N) \in \mathbb{N}_0^N : \max\{x_1, \dots, x_N\} = m\}$, $m \in \mathbb{N}_0$. No que segue mostraremos que a lei das n ($n < N$) primeiras coordenadas, de um ponto uniformemente distribuído sobre $\mathcal{X}_m^N$, é próxima em distância de variação total da lei de n variáveis aleatórias independentes com distribuição comum $U\{0, 1, \dots, m\}$, $m \in \mathbb{N}_0$. A partir deste resultado, uma classe de seqüências finitas de variáveis aleatórias permutáveis é caracterizada como quase mistura de leis de variáveis aleatórias independentes com distribuição comum $U\{0, 1, \dots, m\}$, $m \in \mathbb{N}_0$. Inicialmente, derivamos uma expressão para a distribuição uniforme sobre $\mathcal{X}_m^N$ e sua distribuição n -dimensional ($n < N$).

Denotemos por ν a medida de contagem e por Q_{Nm} a lei de um ponto uniformemente distribuído sobre $\mathcal{X}_m^N$. Então

$$Q_{Nm}(x_1, \dots, x_N) = (\nu(\mathcal{X}_m^N))^{-1} I_{\{m\}}(\max\{x_1, \dots, x_N\}) . \quad (1)$$

Agora, $\mathcal{X}_m^N = \bigcup_{i=1}^N M_i$, onde

$$M_i = \{(x_1, \dots, x_N) \in \mathcal{X}_m^N : x_i = m\} .$$

Logo

$$\begin{aligned} \nu(\mathcal{X}_m^N) &= \nu\left(\bigcup_{i=1}^N M_i\right) = \sum_{i=1}^N \nu(M_i) - \sum_{1 \leq i < j \leq N} \nu(M_i \cap M_j) \\ &\quad + \sum_{1 \leq i < j < k \leq N} \nu(M_i \cap M_j \cap M_k) + \dots + \\ &\quad + (-1)^{N+1} \nu(M_1 \cap M_2 \cap \dots \cap M_N) . \end{aligned}$$

Mas, $\nu(M_{\ell_1} \cap M_{\ell_2} \cap \dots \cap M_{\ell_k}) = (m+1)^{N-k}$ para $k = 1, \dots, N$ e $\ell_1 < \ell_2 < \dots < \ell_k$, com $\{\ell_1, \dots, \ell_k\} \subset \{1, \dots, N\}$.

Portanto,

$$\nu(\mathcal{X}_m^N) = \sum_{i=1}^N \binom{N}{i} (-1)^{i+1} (m+1)^{N-i} = (m+1)^N - m^N . \quad (2)$$

Substituindo (2) em (1), temos que

$$Q_{Nm}(x_1, \dots, x_N) = [(m+1)^N - m^N]^{-1} I_{\{m\}}(\max(x_1, \dots, x_N)) . \quad (3)$$

Não é difícil ver que Q_{Nm} é uma medida de probabilidade permutável, portanto a distribuição de qualquer conjunto de n ($n < N$) coordenadas é a mesma. Denotemos por $Q_{Nm,n}$ a distribuição n -dimensional de Q_{Nm} , isto é,

$$Q_{Nm,n}(x_1, \dots, x_n) = \sum_{(z_1, \dots, z_{N-n}) \in C} Q_{Nm}(x_1, \dots, x_n, z_1, \dots, z_{N-n}) ,$$

onde $C = \{(x_{n+1}, \dots, x_N) \in \mathbb{N}_0^{N-n} : \max\{x_1, \dots, x_n, x_{n+1}, \dots, x_N\} = m\}$.

Mas, se $\max\{x_1, \dots, x_n\} < m$, então

$$C = \{(x_{n+1}, \dots, x_N) \in \mathbb{N}_0^{N-n} : \max\{x_{n+1}, \dots, x_N\} = m\} ,$$

isto é, $C = \mathcal{X}_m^{N-n}$.

Logo, utilizando-nos de (2) e (3), obtemos

$$Q_{Nmn}(x_1, \dots, x_n) = \frac{(m+1)^{N-n} - m^{N-n}}{(m+1)^N - m^N} \quad \text{se} \quad \max\{x_1, \dots, x_n\} < m .$$

Por outro lado, se $\max\{x_1, \dots, x_n\} = m$, então

$$C = \{(x_{n+1}, \dots, x_N) \in \mathbb{N}_0^{N-n} : x_{n+1} \leq m, \dots, x_N \leq m\} .$$

Portanto,

$$Q_{Nmn}(x_1, \dots, x_n) = \frac{(m+1)^{N-n}}{(m+1)^N - m^N} \quad \text{se} \quad \max\{x_1, \dots, x_n\} = m .$$

Ordenando os termos, temos que

$$Q_{Nmn}(x_1, \dots, x_n) = \begin{cases} \frac{1}{(m+1)^n} \left\{ \frac{1 - \left(\frac{m}{m+1}\right)^{N-n}}{1 - \left(\frac{m}{m+1}\right)^N} \right\} & \text{se} \max\{x_1, \dots, x_n\} < m \\ \frac{1}{(m+1)^n} \left\{ \frac{1}{1 - \left(\frac{m}{m+1}\right)^N} \right\} & \text{se} \max\{x_1, \dots, x_n\} = m . \end{cases} \quad (4)$$

O teorema a seguir fornece um limite para a distância de variação total entre a lei Q_{Nmn} e a lei de n variáveis aleatórias independentes com distribuição comum $U\{0, 1, \dots, m\}$.

Teorema 1: *Seja Q_{Nmn} a lei das primeiras n coordenadas de um ponto uniformemente distribuído sobre $\mathcal{X}_m^N$ e P_m^n a lei de n variáveis aleatórias independentes com distribuição comum $U\{0, 1, \dots, m\}$, então se $1 \leq n \leq N$,*

$$\|Q_{Nmn} - P_m^n\| \leq 2 \frac{n}{N} . \quad (5)$$

Prova: Notando que as medidas Q_{Nmn} e P_m^n são ambas dominadas pela medida de contagem temos, pelas propriedades da distância de variação total que

$$\|Q_{Nmn} - P_m^n\| = \int_{\{0,1,\dots,m\}^n} \left| 1 - \frac{Q_{Nmn}}{P_m^n} \right| P_m^n d\nu$$

$$\begin{aligned}
&= \int_{\{(x_1, \dots, x_n) \in \mathbb{N}_0^n : \max\{x_1, \dots, x_n\} < m\}} \left\{ \frac{\left(\frac{m}{m+1}\right)^{N-n} - \left(\frac{m}{m+1}\right)^N}{1 - \left(\frac{m}{m+1}\right)^N} \right\} \frac{1}{(m+1)^n} d\nu(x_1, \dots, x_n) \\
&\quad + \int_{\{(x_1, \dots, x_n) \in \mathbb{N}_0^n : \max\{x_1, \dots, x_n\} = m\}} \left\{ \frac{\left(\frac{m}{m+1}\right)^N}{1 - \left(\frac{m}{m+1}\right)^N} \right\} \frac{1}{(m+1)^n} d\nu(x_1, \dots, x_n) \\
&= \left\{ \frac{\left(\frac{m}{m+1}\right)^{N-n} - \left(\frac{m}{m+1}\right)^N}{1 - \left(\frac{m}{m+1}\right)^N} \right\} \left(\frac{m}{m+1} \right)^n + \left\{ \frac{\left(\frac{m}{m+1}\right)^N}{1 - \left(\frac{m}{m+1}\right)^N} \right\} \left\{ \frac{(m+1)^n - m^n}{(m+1)^n} \right\} \\
&= 2 \left\{ \frac{\left(\frac{m}{m+1}\right)^N - \left(\frac{m}{m+1}\right)^{N+n}}{1 - \left(\frac{m}{m+1}\right)^N} \right\}. \tag{6}
\end{aligned}$$

O núcleo da prova consiste em estimar (6).

Seja $g(x) = \frac{x^N - x^{N+n}}{1 - x^N}$, $0 \leq x < 1$. Vamos mostrar que $g(x) \leq \frac{n}{N}$ para todo $0 \leq x < 1$,

concluindo a partir disto a prova.

Notando que

$$\begin{aligned}
g(x) &= \frac{x^N(1-x)(1+x+x^2+\dots+x^{n-1})}{(1-x)(1+x+x^2+\dots+x^{N-1})} \\
&= \frac{x^N(1+x+x^2+\dots+x^{n-1})}{(1+x+x^2+\dots+x^{N-1})},
\end{aligned}$$

e que

$$1 + x + x^2 + \dots + x^{n-1} < n \quad \text{e}$$

$$1 + x + x^2 + \dots + x^{N-1} > Nx^{N-1}$$

temos que

$$g(x) < \frac{nx^N}{Nx^{N-1}} = \frac{n}{N}x < \frac{n}{N}.$$

□

Observação 1: Convém observar que a função g no Teorema 1 tem supremo igual a $\frac{n}{N}$.

De fato, não é difícil ver que $\lim_{x \rightarrow 1} g(x) = \frac{n}{N}$ e que $g'(x) \geq 0$, para $0 \leq x < 1$.

O Teorema 1 fornece o resultado básico para caracterizar seqüências finitas de variáveis aleatórias permutáveis cuja lei é obtida misturando-se os elementos da família de distribuições $\{Q_{Nm}, m \in \mathbb{N}_0\}$.

Sejam $X_1, X_2, \dots, X_N$ variáveis aleatórias tomando valores em $\mathbb{N}_0$ e seja $X^{(N)} = \max\{X_1, \dots, X_N\}$. Seja P_N a lei de $\mathbf{X}_N = (X_1, X_2, \dots, X_N)$ e suponha que dado $X^{(N)} = m$, $\mathbf{X}_N$ é uniformemente distribuído sobre $\mathcal{X}_m^N$. Seja P_θ^n a lei de n variáveis aleatórias independentes com distribuição comum $U\{0, 1, \dots, \theta\}$, $\theta \in \mathbb{N}_0$ e $P_{\mu_n} = \int P_\theta^n d\mu(\theta)$, para alguma medida de probabilidade μ sobre $\mathbb{N}_0$.

Teorema 2: Se P_N e P_{μ_n} são as medidas de probabilidade acima definidas, então existe uma medida de probabilidade μ sobre $\mathbb{N}_0$ tal que para cada $1 \leq n \leq N$

$$\|P_n - P_{\mu_n}\| \leq 2 \frac{n}{N}, \quad (7)$$

onde P_n é a lei n -dimensional de P_N .

Prova: A partir da definição de P_n , temos que

$$P_n = \int Q_{Nm_n} d\mu_N(m),$$

onde μ_N é a P_N -lei de $X^{(N)} = \max\{X_1, \dots, X_N\}$. Logo escolhendo $\mu = \mu_N$ e utilizando-nos da convexidade da distância de variação total junto com o Teorema 1, conclui-se a prova. $\square$

O teorema estabelece que a lei das n ($n < N$) primeiras componentes de uma seqüência finita $X_1, X_2, \dots, X_N$ de variáveis aleatórias tomando valores em $\mathbb{N}_0$, com a propriedade que a distribuição condicional de $X_1, X_2, \dots, X_N$ dado $X^{(N)} = m$ é uniforme sobre $\mathcal{X}_m^N$, é quase mistura (com erro de aproximação menor ou igual a $\frac{2n}{N}$) de n variáveis aleatórias independentes com distribuição comum $U\{0, 1, \dots, \theta\}$, $\theta \in \mathbb{N}_0$.

A versão infinita do Teorema 2, no estilo de de Finetti (Ressel (1985), exemplo 6) é derivada a partir deste resultado com relativa simplicidades.

Teorema 3: *Seja $X_1, X_2, \dots$ uma seqüência infinita de variáveis aleatórias tomando valores em $\mathbb{N}_0$. Se para cada $N \in \mathbb{N}$, $m \in \mathbb{N}_0$, a distribuição condicional de $X_1, X_2, \dots, X_N$ dado $X^{(N)} = m$ é uniforme sobre $\mathcal{X}_m^N$, então existe uma medida de probabilidade μ sobre $\mathbb{N}_0$ tal que para cada $n \in \mathbb{N}$:*

$$\Pr \{X_1 = x_1, \dots, X_n = x_n\} = \int_{\mathbb{N}_0} \frac{1}{(\theta + 1)^n} I_{\{0,1,\dots,\theta\}}(\max(x_1, \dots, x_n)) d\mu(\theta) ,$$

com $x_i \in \mathbb{N}_0$, para $i = 1, \dots, n$.

Prova: Seja μ_N a P -lei de $X^{(N)}$. Então pela hipótese e Teorema 2 temos que para cada $n \leq N$ e $N \in \mathbb{N}$

$$\|P_n - P_{\mu_{Nn}}\| \leq 2 \frac{n}{N} ,$$

onde P_n é a lei de $X_1, X_2, \dots, X_n$ e

$$P_{\mu_{Nn}} = \int \frac{1}{(\theta + 1)^n} I_{\{0,1,\dots,\theta\}}(\max(x_1, \dots, x_n)) d\mu_N(\theta) .$$

A prova se segue, tomando limite nesta desigualdade, sempre que mostremos que a seqüência $\{\mu_N\}_{N \in \mathbb{N}}$ é tight. Pela hipótese

$$\begin{aligned} \Pr \{X_1 > k\} &= \int_{\mathbb{N}_0} \Pr \{X_1 > k | X^{(N)} = m\} d\mu_N(m) \\ &= \int_{\mathbb{N}_0} [1 - \Pr \{X_1 \leq k | X^{(N)} = m\}] d\mu_N(m) \\ &= \int_{\mathbb{N}_0} \left[1 - \sum_{x_1=0}^k Q_{Nm1}(x_1) \right] I_{\{k+1,\dots\}}(m) d\mu_N(m) . \end{aligned}$$

Assim, desde (4) temos que

$$\begin{aligned} \Pr \{X_1 > k\} &= \int \left[1 - \frac{k+1}{m+1} \left\{ \frac{1 - \left(\frac{m}{m+1}\right)^{N-1}}{1 - \left(\frac{m}{m+1}\right)^N} \right\} \right] I_{\{k+1,\dots\}}(m) d\mu_N(m) \\ &\geq \int \left[1 - \left(\frac{k+1}{m+1}\right) \right] I_{\{k+1,\dots\}}(m) d\mu_N(m) . \end{aligned} \tag{8}$$

Agora $\Pr \{X_1 > k\} \xrightarrow{k \rightarrow \infty} 0$, assim, dado $\eta > 0$ existe $k_0 \in \mathbb{N}$ tal que se $k \geq k_0$, $\Pr \{X_1 > k\} < \eta$.

Tomando $k = k_0$ em (8), obtemos

$$\eta \geq \int \left[1 - \left(\frac{k_0 + 1}{m + 1} \right) \right] I_{\{k_0+1, \dots\}}(m) d\mu_N(m) .$$

Para k_0 fixado, seja $g(m) = \left[1 - \frac{k_0+1}{m+1} \right] I_{\{k_0+1, \dots\}}(m)$ então $g(m) \xrightarrow{m \rightarrow \infty} 1$. Assim, para $\delta > 0 \exists M \in \mathbb{N}$, $M > k_0(\eta)$ tal que se $m > M$,

$$1 - \delta < g(m) < 1 + \delta .$$

Consequentemente,

$$\int_{(M, +\infty)} (1 - \delta) d\mu_N(m) \leq \int_{(M, +\infty)} \left\{ 1 - \left(\frac{k_0 + 1}{m + 1} \right) \right\} d\mu_N(m) \leq \eta .$$

Logo, $\mu_N([M, +\infty)) < \frac{\eta}{1-\delta} = \varepsilon$, para cada $N \in \mathbb{N}$.

Assim, μ_N é tight, portanto existe uma subsequência $\{\mu_{N_j}\}$ e uma medida de probabilidade μ tal que $\mu_{N_j} \xrightarrow{j \rightarrow \infty} \mu$. Logo, tomando o limite através da subsequência conclui-se a prova. $\square$.

Para finalizar a seção, a questão de estendibilidade implícita nos Teoremas 2 e 3 será discutida em detalhe. O teorema que segue é de utilidade para este propósito. Na prova deste utilizaremos o fato de que se $(X_1, X_2, \dots, X_N)$ dado $X^{(N)} = m$ é uniformemente distribuído sobre $\mathcal{X}_m^N$ então para cada $1 \leq n \leq N$, $(X_1, \dots, X_n)$ dado $X^{(N)} = s$ é uniformemente distribuído sobre $\mathcal{X}_s^n$. Este resultado será provado no Capítulo 3, dentro de um contexto geral.

➤ **Teorema 4:** *Sejam $X_1, X_2, \dots, X_N$ variáveis aleatórias tomando valores em $\mathbb{N}_0$, independentes e identicamente distribuídas. Então a distribuição condicional de $X_1, X_2, \dots, X_N$ dado $X^{(N)} = m$ é uniforme sobre $\mathcal{X}_m^N$ se e somente se $X_1, X_2, \dots, X_N$ possuem distribuição uniforme sobre $\{0, 1, \dots, K\}$ para algum $K \in \mathbb{N}_0$.*

Prova: Primeiro notemos que se $X_1, X_2, \dots, X_N$ são variáveis aleatórias independentes e identicamente distribuídas, então

$$\Pr \{X_1 = x_1, \dots, X_N = x_N | X^{(N)} = m\} = \frac{\prod_{i=1}^N \Pr \{X_i = x_i\} I_{\{m\}}(\max(x_1, \dots, x_N))}{\{\Pr \{X_1 \leq m\}\}^N - \{\Pr \{X_1 \leq m-1\}\}^N} . \quad (9)$$

O denominador à direita de (9) é deduzido como segue:

$$\Pr \{\max\{X_1, \dots, X_N\} = m\} = \Pr \left\{ \bigcup_{i=1}^N \{X_i = m, 0 \leq X_j \leq m, \text{ se } j \neq i, j = 1, \dots, N\} \right\} .$$

Seja $A_i = \{X_i = m, 0 \leq X_j \leq m \text{ se } j \neq i \text{ e } j = 1, \dots, N\}$ então

$$\Pr \{A_{i_1} \cap A_{i_2} \cap \dots \cap A_{i_k}\} = [\Pr \{X_1 = m\}]^k [\Pr \{X_1 \leq m\}]^{N-k}$$

para $\{i_1, i_2, \dots, i_k\} \subset \{1, \dots, N\}$, com $i_l \neq i_j$ para todo $l, j \in \{1, \dots, k\}$.

Logo

$$\begin{aligned} \Pr \{\max\{X_1, \dots, X_N\} = m\} &= \Pr \left\{ \bigcup_{i=1}^N A_i \right\} \\ &= \sum_{k=1}^N \binom{N}{k} (-1)^{k+1} [\Pr \{X_1 = m\}]^k [\Pr \{X_1 \leq m\}]^{N-k} \\ &= \Pr \{X_1 \leq m\}^N - [\Pr \{X_1 \leq m-1\}]^N . \end{aligned}$$

Por outro lado $X_1, X_2, \dots, X_N / X^{(N)} = m$ é uniformemente distribuído sobre $\mathcal{X}_m^N$. Portanto, a partir da expressão (3) no começo desta seção e da identidade (9), obtemos

$$\frac{\prod_{i=1}^N \Pr \{X_i = x_i\} I_{\{m\}}(\max(x_1, \dots, x_N))}{[\Pr \{X_1 \leq m\}]^N - [\Pr \{X_1 \leq m-1\}]^N} = \frac{I_{\{m\}}(\max(x_1, \dots, x_N))}{(m+1)^N - m^N} .$$

Fazendo $N = 2$, $x_1 = x_2 = m$, temos que

$$\frac{[\Pr \{X_1 = m\}]^2}{[\Pr \{X_1 \leq m\}]^2 - [\Pr \{X_1 \leq m-1\}]^2} = \frac{1}{2m+1} . \quad (10)$$

Agora,

$$\begin{aligned} [\Pr \{X_1 \leq m-1\}]^2 &= [\Pr \{X_1 \leq m\} - \Pr \{X_1 = m\}]^2 \\ &= [\Pr \{X_1 \leq m\}]^2 - 2\Pr \{X_1 \leq m\} \Pr \{X_1 = m\} \\ &\quad + [\Pr \{X_1 = m\}]^2 . \end{aligned}$$

Substituindo em (10) obtemos

$$\frac{[\Pr\{X_1 = m\}]^2}{\Pr\{X_1 = m\}[2\Pr\{X_1 \leq m\} - \Pr\{X_1 = m\}]} = \frac{1}{2m+1}.$$

Cancelando termos e desenvolvendo esta última igualdade obtemos

$$\frac{\Pr\{X_1 = m\}}{\Pr\{X_1 \leq m\}} = \frac{1}{m+1}. \quad (11)$$

Avaliando (11) para $m = 0, 1, \dots$, obtemos

$$\Pr\{X_1 = x\} = \Pr\{X_1 = 0\}, \quad x \in \mathbb{N}.$$

Mas P é uma medida de probabilidade, portanto esta igualdade implica que existe $K \in \mathbb{N}$ tal que

$$\Pr\{X_1 > K\} = 0.$$

Estes dois últimos fatos implicam que existe $K \in \mathbb{N}$ tal que

$$\Pr\{X_1 = x\} = \frac{1}{K+1} I_{\{0,1,\dots,K\}}(x).$$

Reciprocamente, se $X_1, X_2, \dots, X_N$ são variáveis aleatórias independentes com distribuição comum uniforme $\{0, 1, \dots, K\}$ temos a partir da identidade (9) que

$$\begin{aligned} \Pr\{X_1 = x_1, \dots, X_N = x_N | X^{(N)} = m\} &= \frac{\frac{1}{(K+1)^N} I_{\{m\}} I(\max(x_1, \dots, x_N))}{\left(\frac{m+1}{K+1}\right)^N - \left(\frac{m}{K+1}\right)^N} \\ &= [(m+1)^N - m^N]^{-1} I_{\{m\}}(\max(x_1, \dots, x_N)). \end{aligned}$$

□

O teorema anterior, além de fornecer uma outra caracterização de variáveis aleatórias independentes com distribuição comum uniforme sobre $\{0, 1, \dots, \theta\}$, permite-nos, em conexão com o Teorema 3, determinar uma subclasse de seqüências finitas de variáveis

aleatórias permutáveis que podem ser estendidas para uma seqüência infinita de variáveis aleatórias admitindo a representação no Teorema 3.

De fato, se $\mathcal{C}_N = \{P_N \text{ sobre } \mathbb{N}_0 : P_N(\cdot) = \int_{\mathbb{N}_0} Q_{Nm}(\cdot) d\mu(m) \text{ para alguma medida de probabilidade } \mu \text{ sobre } \mathbb{N}_0\}$ e ε_N é a classe de medidas produtos em $\mathcal{C}_N$, então pelo Teorema 4, $\varepsilon_N = \{P_\theta^N : \theta \in \mathbb{N}_0\}$, onde $P_\theta^N\{(x_1, \dots, x_N)\} = \frac{1}{(\theta+1)^N} I_{\{0,1,\dots,\theta\}}(\max(x_1, \dots, x_N))$.

Seja $\mathcal{F}_N$ o conjunto de medidas de probabilidade obtidas misturando-se elementos de ε_N e $\mathcal{C} = \{P \text{ sobre } \mathbb{N}_0^\infty : P_N \in \mathcal{C}_N \text{ para cada } N \in \mathbb{N}\}$, com P_N a distribuição N dimensional de P .

Corolário 1: *Sejam $X_1, X_2, \dots, X_N$ variáveis aleatórias tomando valores em $\mathbb{N}_0$ com lei $P_N \in \mathcal{C}_N$. Uma condição necessária e suficiente para que exista uma seqüência infinita de variáveis aleatórias $Y_1, Y_2, \dots$ com lei $P \in \mathcal{C}$ tal que $(Y_1, Y_2, \dots, Y_N) \stackrel{\mathcal{D}}{=} (X_1, X_2, \dots, X_N)$ é que $P_N \in \mathcal{F}_N$.*

Prova: A necessidade é direta a partir do Teorema 3. A suficiência segue por argumentos análogos aos da seção 4 (Exemplo 3) no Capítulo 1. $\square$

Nas próximas seções as versões dos Teoremas 1, 2 e 3, envolvendo distribuições uniformes contínuas sobre a reta, serão fornecidas.

2.3 Distribuição uniforme sobre o intervalo $(0, \theta)$, $\theta > 0$

Na seção anterior uma classe de medidas de probabilidade permutáveis gerada misturando-se os elementos da família $\{Q_{Nm} : m > 0\}$ foi caracterizada em termos de teoremas do tipo de Finetti envolvendo distribuições uniformes sobre $\{0, 1, \dots, \theta\}$, $\theta \in \mathbb{N}_0$. Parece natural pensar que se definíssemos agora medidas de probabilidade uniforme sobre $\mathcal{X}_m^N = \{(x_1, \dots, x_N) \in \mathbb{R}_+^N : \max\{x_1, \dots, x_N\} = m\}$, $m > 0$, obteríamos teoremas do tipo

de Finetti envolvendo distribuições uniformes sobre o intervalo $(0, \theta)$, $\theta > 0$. Porém, neste caso, a construção de uma medida de probabilidade uniforme sobre $\mathcal{X}_m^N$ não é tão natural como no caso anterior. De fato, $\mathcal{X}_m^N$ possui medida de Lebesgue zero como subconjunto do $\mathbb{R}^N$. Antes de entrar em detalhes, consideremos o caso $N = 2$. Neste caso,

$$\mathcal{X}_m^2 = \{(x_1, x_2) \in \mathbb{R}_+^2 : \max(x_1, x_2) = m\}, \quad m > 0.$$

Seja

$$M_1(m) = \{(x_1, x_2) \in \mathbb{R}_+^2 : x_1 = m, x_2 \leq m\}$$

e

$$M_2(m) = \{(x_1, x_2) \in \mathbb{R}_+^2 : x_1 \leq m, x_2 = m\}.$$

Então, $\mathcal{X}_m^2 = M_1(m) \cup M_2(m)$. A figura a seguir descreve estes conjuntos.

Figura 1: Gráfico dos espaços $\mathcal{X}_m^2$, $m > 0$.

Uma maneira de colocar a idéia de equiprobabilidade é por consideração dos comprimentos de cada um dos lados de $\mathcal{X}_m^2$. Isto é, se Q_{2m} denota uma medida de probabilidade sobre a σ -álgebra $\mathcal{A} = \mathcal{X}_m^2 \cap \mathcal{B}_2$, com $\mathcal{B}_2$ a σ -álgebra de Borel sobre $\mathbb{R}^2$, então

$$Q_{2m}(A) = \frac{\text{Compr}(A \cap M_1(m)) + \text{Compr}(A \cap M_2(m))}{\text{Compr}(M_1(m)) + \text{Compr}(M_2(m))}$$

$$= \frac{\text{Compr}(A \cap M_1(m)) + \text{Compr}(A \cap M_2(m))}{2m}, \quad A \in \mathcal{A}.$$

A medida de probabilidade Q_{2m} é determinada pelas probabilidades sobre os conjuntos da forma $\{m\} \times (0, x_2]$ e $(0, x_1] \times \{m\}$, $0 < x_1, x_2 \leq 1$. Em particular, se $m = 1$,

$$\begin{aligned} Q_{21}((0, x_1] \times \{1\}) &= \frac{x_1}{2}, \quad 0 < x_1 \leq 1, \\ \text{e} \quad Q_{21}(\{1\} \times (0, x_2]) &= \frac{x_2}{2}, \quad 0 < x_2 \leq 1. \end{aligned}$$

Notemos que tal distribuição pode ser caracterizada utilizando-nos de variáveis aleatórias com distribuição conhecida. Sejam X e Y duas variáveis aleatórias com distribuição comum uniforme sobre o intervalo $(0, 1]$. Seja $Z : (\mathbb{R}^2, \mathcal{B}^2) \mapsto (\mathcal{X}_1^2, \mathcal{X}_1^2 \cap \mathcal{B}^2)$ definida por $Z(X, Y) = (Z_1(X, Y), Z_2(X, Y))$, com

$$Z_1(X, Y) = \frac{X}{\max(X, Y)} \quad \text{e} \quad Z_2(X, Y) = \frac{Y}{\max(X, Y)}.$$

Notemos que $\max(Z_1, Z_2) = 1$ com probabilidade 1.

Também $\{Z_1 = 1, Z_2 \leq x_2\} = \{X = \max(X, Y) \text{ e } Y \leq x_2 \max(X, Y)\} = \{Y \leq x_2 X, 0 < X \leq 1\}$, se $0 \leq x_2 \leq 1$. Logo

$$\begin{aligned} \Pr\{Z_1 = 1, Z_2 \leq x_2\} &= \Pr\{Y \leq x_2 X\} \\ &= \int_0^1 \Pr\{Y \leq x_2 x\} dx \\ &= \int_0^1 x_2 x dx = \frac{x_2}{2}. \end{aligned}$$

Analogamente,

$$\Pr\{Z_1 \leq x_1, Z_2 = 1\} = \frac{x_1}{2} \quad \text{se } 0 < x_1 \leq 1.$$

Consequentemente, $Z = (Z_1, Z_2)$ tem distribuição uniforme sobre $\mathcal{X}_1^2$, a qual é invariante sob permutações. Note que

$$\Pr\{Z_1 \leq x_1\} = \begin{cases} \frac{x_1}{2} & \text{se } 0 < x_1 < 1 \\ 1 & \text{se } x_1 = 1. \end{cases}$$

Isto é, a distribuição de Z_1 coloca massa uniforme $1/2$ no intervalo $(0, 1)$ e $1/2$ no ponto $\{1\}$.

Este exemplo traduz o espírito desta seção. Dentro deste contexto, teoremas do tipo de Finetti envolvendo distribuições contínuas uniformes, sobre o intervalo $(0, \theta)$, $\theta > 0$, serão derivadas.

Formalmente, seja $\mathcal{X}_m^N = \{(x_1, \dots, x_N) \in \mathbb{R}_+^N : \max\{x_1, \dots, x_N\} = m\}$, $m > 0$. No que segue mostraremos que a lei das n ($n < N$) primeiras coordenadas de um ponto uniformemente distribuído sobre $\mathcal{X}_m^N$ é próxima em distância de variação total da lei de n variáveis aleatórias independentes com distribuição comum uniforme sobre $(0, m)$.

Seja $M_i(m) = \{(x_1, \dots, x_N) \in \mathcal{X}_m^N : x_i = m\}$, $m > 0$ e $i = 1, \dots, N$. É claro que $\mathcal{X}_m^N = \bigcup_{i=1}^N M_i(m)$. Consideremos a função $\varphi^i : \mathbb{R}^N \rightarrow \mathbb{R}^{N-1}$, definida por

$$\varphi^i(x_1, \dots, x_N) = (x_1, \dots, x_{i-1}, x_{i+1}, \dots, x_N) .$$

Seja λ a medida de Lebesgue definida sobre os conjuntos de Borel do $\mathbb{R}^{N-1}$ e seja $\mathcal{A} = \mathcal{X}_m^N \cap \mathcal{B}_N^+$, onde $\mathcal{B}_N^+$ é a σ -álgebra de Borel do $\mathbb{R}_+^N$.

Considere os espaços de medida $\{(\mathcal{X}_m^N, \mathcal{A}, \mu_i)\}_{i=1}^N$, onde $\mu_i : \mathcal{A} \rightarrow [0, \infty)$ é definida por

$$\mu_i(A) = \lambda(\varphi^i(A \cap M_i(m))) , \quad A \in \mathcal{A} .$$

Seja agora $\tilde{Q}_{Nm} : \mathcal{A} \rightarrow [0, 1]$ uma medida de probabilidade definida por,

$$\tilde{Q}_{Nm}(A) = \frac{\sum_{i=1}^N \mu_i(A)}{Nm^{N-1}} , \quad A \in \mathcal{A} .$$

A função $Q_{Nm} : \mathcal{B}_N^+ \rightarrow [0, 1]$ definida por

$$Q_{Nm}(B) = \tilde{Q}_{Nm}(B \cap \mathcal{X}_m^N) , \quad B \in \mathcal{B}_N^+ \tag{12}$$

será denominada de medida de probabilidade uniforme sobre $\mathcal{X}_m^N$.

Esta definição estabelece explicitamente o sentido de distribuição uniforme que utilizaremos aqui.

Propriedades:

i) Q_{Nm} é uma medida de probabilidade permutável.

Prova: Seja $B \in \mathcal{B}_N^+$ e π uma permutação de $\{1, \dots, N\}$. Seja $A = B \cap \mathcal{X}_m^N$, então $\pi A = \pi B \cap \mathcal{X}_m^N$, pois $\mathcal{X}_m^N$ é invariante sob permutações. Logo,

$$\begin{aligned} Q_{Nm}(B) &= \tilde{Q}_{Nm}(\pi B \cap \mathcal{X}_m^N) \\ &= \tilde{Q}_{Nm}(\pi A) \\ &= \frac{1}{Nm^{N-1}} \sum_{i \in I} \mu_i(\pi A \cap M_i(m)) , \end{aligned}$$

onde $I = \{i \in \{1, \dots, N\} : \pi A \cap M_i(m) \neq \emptyset\}$. Agora

$$\begin{aligned} \mu_i(\pi A) &= \lambda(\varphi^i(\pi A \cap M_i(m))) \\ &= \lambda(\varphi^{\pi^{-1}(i)}(A \cap M_{\pi^{-1}(i)}(m))) \\ &= \mu_{\pi^{-1}(i)}(A) , \end{aligned}$$

pois $\pi^{-1}(I) = \{j \in \{1, \dots, N\} : A \cap M_j(m) \neq \emptyset\}$. Logo

$$\tilde{Q}_{Nm}(\pi A) = \frac{1}{Nm^{N-1}} \sum_{j \in J} \mu_j(A) = \tilde{Q}_{Nm}(A) ,$$

onde $J = \pi^{-1}(I)$. □

ii) Considere os espaços mensuráveis $(\mathbb{R}_+^N, \mathcal{B}_N^+)$ e $(\mathbb{R}_+, \mathcal{B}_1^+)$. Seja ψ uma função de $\mathcal{B}_N^+ \times \mathbb{R}_+ \mapsto [0, 1]$ definida por

$$\psi(B, m) = Q_{Nm}(B) , \quad B \in \mathcal{B}_N^+ , \quad m > 0 .$$

Então ψ é uma função de transição de $(\mathbb{R}_+, \mathcal{B}_1^+)$ para $(\mathbb{R}_+^N, \mathcal{B}_N^+)$.

Prova: Segue a partir da definição que, para cada $m > 0$, $Q_{Nm}(\cdot)$ é uma função de probabilidade sobre $(\mathbb{R}_+^N, \mathcal{B}_N^+)$. Para mensurabilidade, é suficiente notar que se $B = \times_{i=1}^{N-1} (0, x_i) \times \{m\}$, então

$$Q_{Nm}(B) = \frac{1}{Nm^{N-1}} \prod_{i=1}^{N-1} x_i ,$$

a qual é uma função contínua em m , para $m > 0$. $\square$

A propriedade (ii) garante a existência de uma sequência finita de variáveis $X_1, X_2, \dots, X_N$ tomando valores em $\mathbb{R}^+$ tal que $Q_N(\cdot)$ é uma versão de probabilidade condicional regular de $X_1, X_2, \dots, X_N$ dado $X^{(N)}$. Basta considerar $P_N(\cdot) = \int_{(0, \infty)} Q_{Nm}(\cdot) d\nu(m)$, como a lei de $X_1, \dots, X_N$ onde ν é uma medida de probabilidade sobre $(0, \infty)$

De maneira análoga ao caso discreto, nosso próximo objetivo é determinar a distribuição n -dimensional de Q_{Nm} . Porém, a expressão (12) não fornece uma definição analiticamente tratável. A proposição a seguir fornece uma útil definição de Q_{Nm} em termos de funções de variáveis aleatórias com distribuição conhecida.

Proposição 1: *Sejam $X_1, X_2, \dots, X_N$ variáveis aleatórias independentes com distribuição comum $U(0, 1)$ e $X^{(N)} = \max\{X_1, \dots, X_N\}$. Seja $Y_i = m \frac{X_i}{X^{(N)}}$, $i = 1, \dots, N$ e $m > 0$. Então o vetor aleatório $\mathbf{Y}_N = (Y_1, \dots, Y_N)$ possui distribuição uniforme sobre $\mathcal{X}_m^N$.*

Prova: Seja P a lei de $(X_1, \dots, X_N)$ e a Q a P -lei de $\mathbf{Y}_N$.

É claro que $Q(\mathcal{X}_m^N) = \Pr\{\mathbf{Y}_N \in \mathcal{X}_m^N\} = 1$. Agora para $A \in \mathcal{A}$,

$$\begin{aligned} Q(A) &= \Pr\{\mathbf{Y}_N \in A\} = \Pr\{\mathbf{Y}_N \in \bigcup_{i=1}^N (A \cap M_i(m))\} \\ &= \sum_{i=1}^N \Pr\{\mathbf{Y}_N \in A \cap M_i(m)\}. \end{aligned} \quad (13)$$

Mas,

$$\begin{aligned} \Pr\{\mathbf{Y}_N \in A \cap M_i(m)\} &= \Pr\{Y_i = m, (Y_1, \dots, Y_{i-1}, Y_{i+1}, \dots, Y_N) \in \varphi^i(A \cap M_i(m))\} \\ &= \Pr\left\{\frac{X_i}{X^{(N)}} = 1, \left(m \frac{X_1}{X^{(N)}}, \dots, m \frac{X_{i-1}}{X^{(N)}}, m \frac{X_{i+1}}{X^{(N)}}, \dots, m \frac{X_N}{X^{(N)}} \in \varphi^i(A \cap M_i(m))\right)\right\} \end{aligned} \quad (14)$$

Agora

$$\{X_i = X^{(N)}, \left(m \frac{X_1}{X^{(N)}}, \dots, m \frac{X_{i-1}}{X^{(N)}}, m \frac{X_{i+1}}{X^{(N)}}, \dots, m \frac{X_N}{X^{(N)}} \in \varphi^i(A \cap M_i(m))\right)\}$$

$$= \{X_j \leq X_i \text{ para } j \neq i, j = 1, \dots, N \text{ e} \\ \left(m \frac{X_1}{X_i}, \dots, m \frac{X_{i-1}}{X_i}, m \frac{X_{i+1}}{X_i}, \dots, m \frac{X_N}{X_i}\right) \in \varphi^i(A \cap M_i(m))\} .$$

Mas $(z_1, \dots, z_{N-1}) \in \varphi^i(A \cap M_i(m)) \Rightarrow 0 < z_j \leq m$ para cada $j = 1, \dots, N-1$, se $A \cap M_i \neq \emptyset$, implicando em

$$\left\{ \left(m \frac{X_1}{X_i}, \dots, m \frac{X_{i-1}}{X_i}, m \frac{X_{i+1}}{X_i}, \dots, m \frac{X_N}{X_i}\right) \in \varphi^i(A \cap M_i(m)) \right\} \\ \subseteq \{X_j \leq X_i \text{ para } j \neq i, j = 1, \dots, N\} .$$

Daqui e da igualdade (14), temos que

$$\begin{aligned} \Pr \{Y_N \in A \cap M_i(m)\} &= \Pr \left\{ \left(m \frac{X_1}{X_i}, \dots, m \frac{X_{i-1}}{X_i}, m \frac{X_{i+1}}{X_i}, \dots, m \frac{X_N}{X_i}\right) \in \varphi^i(A \cap M_i(m)) \right\} \\ &= \int_0^1 \Pr \left\{ \left(m \frac{X_1}{x}, \dots, m \frac{X_{i-1}}{x}, m \frac{X_{i+1}}{x}, \dots, m \frac{X_N}{x}\right) \in \varphi^i(A \cap M_i(m)) \right\} dx \\ &= \int_0^1 \Pr \{(X_1, \dots, X_{i-1}, X_{i+1}, \dots, X_N) \in \frac{x}{m} \varphi^i(A \cap M_i(m))\} dx , \end{aligned} \tag{15}$$

onde para $c \in \mathbb{R}$, $A \subseteq \mathbb{R}^{N-1}$, $cA = \{ca, a \in A\}$.

Mas, $X_1, X_2, \dots, X_N$ são variáveis aleatórias independentes e identicamente distribuídas $U(0, 1)$, logo

$$\begin{aligned} \Pr \{(X_1, \dots, X_{i-1}, X_{i+1}, \dots, X_N) \in \frac{x}{m} \varphi^i(A \cap M_i(m))\} \\ = \lambda\left(\frac{x}{m} \varphi^i(A \cap M_i(m))\right) \\ = \frac{x^{N-1}}{m^{N-1}} \lambda(\varphi^i(A \cap M_i(m))) , \end{aligned} \tag{16}$$

onde λ é a medida de Lebesgue $N-1$ dimensional.

Substituindo (16) em (15), temos que

$$\begin{aligned} \Pr \{Y_N \in A \cap M_i(m)\} &= \int_0^1 \frac{x^{N-1}}{m^{N-1}} \lambda(\varphi^i(A \cap M_i(m))) dx \\ &= \frac{1}{Nm^{N-1}} \mu_i(A) . \end{aligned} \tag{17}$$

Finalmente, substituindo (17) em (13), conclui-se a prova. $\square$

Como antes, Q_{Nmn} denotará a distribuição n -dimensional de Q_{Nm} .

Proposição 2: *Seja $\mathbf{Y}_N = (Y_1, Y_2, \dots, Y_N)$ vetor aleatório uniformemente distribuído sobre $\mathcal{X}_m^N$, então a lei de $\mathbf{Y}_n = (Y_1, Y_2, \dots, Y_n)$, ($n < N$) é definida por*

$$F_m(y_1, \dots, y_n) = \begin{cases} 0 & \text{se } y_i \leq 0 \text{ para algum } i \in \{1, \dots, n\} \\ \frac{N-n}{N} \prod_{i=1}^n \left(\frac{y_i}{m}\right) & \text{se } 0 < y_i < m \text{ para cada } i \in \{1, \dots, n\} \\ \left\{ \frac{N-n+k}{N} \right\} \prod_{i=1}^n \left(\frac{y_i}{m}\right) & \text{se } y_i = m \text{ para } i \in I_k \text{ e } 0 < y_i < m \text{ para } i \in J_k \\ & k = 1, 2, \dots, n \\ 1 & \text{se } y_i > m \text{ para cada } i \in \{1, \dots, n\}, \end{cases}$$

onde $I = \{i_1, \dots, i_k\} \subseteq \{1, \dots, n\}$, $J = \{1, \dots, n\} - I$ e F_m é a função de distribuição de $\mathbf{Y}_N$.

Prova: Tendo em conta a Proposição 1, é suficiente determinar a lei do vetor aleatório $(m \frac{X_1}{X^{(N)}}, \dots, m \frac{X_n}{X^{(N)}})$ onde $X_1, X_2, \dots, X_N$ são variáveis aleatórias independentes com distribuição comum $U(0, 1)$. Sem perda de generalidade, fazemos $m = 1$.

Seja $0 < z_i < 1$ para cada $i = 1, \dots, n$, então

$$\Pr \left\{ \frac{X_i}{X^{(N)}} \leq z_i, i = 1, \dots, n \right\} = \sum_{l=1}^n \Pr \left\{ \frac{X_i}{X^{(N)}} \leq z_i, i = 1, \dots, n \text{ e } X^{(N)} = X_{n+l} \right\}. \quad (18)$$

Mas,

$$\begin{aligned} & \Pr \left\{ \frac{X_i}{X^{(N)}} \leq z_i; i = 1, \dots, n \text{ e } X^{(N)} = X_{n+l} \right\} \\ &= \Pr \left\{ \{X_i \leq z_i X_{n+l}; i = 1, \dots, n\} \text{ e } \{X_{n+i} \leq X_{n+l}; i = 1, \dots, N-n, i \neq l\} \right\} \quad (19) \\ &= \int_0^1 x^n \left(\prod_{i=1}^n z_i \right) x^{N-n-1} dx = \frac{1}{N} \prod_{i=1}^n z_i \end{aligned}$$

Substituindo (19) em (18), vem

$$\Pr \left\{ \frac{X_i}{X^{(N)}} \leq z_i, i = 1, \dots, n \right\} = \frac{N-n}{N} \prod_{i=1}^n z_i. \quad (20)$$

Agora, se $z_i = 1$ e $0 < z_j < 1$ para $j = 1, \dots, n$ e $j \neq i$, temos que

$$\begin{aligned} & \Pr \left\{ \frac{X_i}{X^{(N)}} \leq 1, \left\{ \frac{X_j}{X^{(N)}} \leq z_j; \quad j = 1, \dots, n \text{ e } j \neq i \right\} \right\} \\ &= \Pr \left\{ \frac{X_i}{X^{(N)}} < 1, \left\{ X_j \leq z_j X^{(N)}; \quad j = 1, \dots, n \text{ e } j \neq i \right\} \right\} \\ &+ \Pr \left\{ X_i = X^{(N)}, \left\{ X_j \leq z_j X^{(N)}; \quad j = 1, \dots, n \text{ e } j \neq i \right\} \right\}. \end{aligned} \quad (21)$$

Mas

$$\begin{aligned} & \Pr \left\{ X_i = X^{(N)}, \left\{ X_j \leq z_j X^{(N)}; \quad j = 1, \dots, n \text{ e } j \neq i \right\} \right\} \\ &= \Pr \left\{ \{X_{n+l} < X_i, \quad \text{se } l = 1, \dots, N - n\} \quad \text{e} \quad \{X_j \leq z_j X_i, \quad j = 1, \dots, n \text{ e } j \neq i\} \right\} \\ &= \int_0^1 x^{N-1} \prod_{\substack{j=1 \\ j \neq i}}^n z_j dx = \frac{1}{N} \prod_{\substack{j=1 \\ j \neq i}}^n z_j. \end{aligned} \quad (22)$$

Substituindo (22) em (21) e utilizando (20), temos que

$$\begin{aligned} & \Pr \left\{ \frac{X_i}{X^{(N)}} \leq 1, \left\{ \frac{X_j}{X^{(N)}} \leq z_j; \quad j = 1, \dots, n \text{ e } j \neq i \right\} \right\} \\ &= \frac{N - n}{N} \prod_{\substack{j=1 \\ j \neq i}}^n z_j + \frac{1}{N} \prod_{\substack{j=1 \\ j \neq i}}^n z_j. \end{aligned} \quad (23)$$

Analogamente, se $z_i = z_j = 1$ e $0 < z_l < 1$, $l = 1, \dots, n$; $l \neq i, j$ obtemos

$$\begin{aligned} & \Pr \left\{ \frac{X_i}{X^{(N)}} \leq 1, \frac{X_j}{X^{(N)}} \leq 1 \quad \text{e} \quad \left\{ \frac{X_l}{X^{(N)}} < z_l, \quad l = 1, \dots, n, \quad l \neq i, j \right\} \right\} \\ &= \Pr \left\{ \frac{X_i}{X^{(N)}} < 1, \frac{X_j}{X^{(N)}} \leq 1 \quad \text{e} \quad \left\{ \frac{X_l}{X^{(N)}} < z_l, \quad l = 1, \dots, n, \quad l \neq i, j \right\} \right\} \\ &+ \Pr \left\{ \frac{X_i}{X^{(N)}} = 1, \frac{X_j}{X^{(N)}} \leq 1 \quad \text{e} \quad \left\{ \frac{X_l}{X^{(N)}} < z_l, \quad l = 1, \dots, n, \quad l \neq i, j \right\} \right\} \\ &= \frac{N - n}{N} \prod_{\substack{l=1 \\ l \neq i, j}}^n z_l + \frac{2}{N} \prod_{\substack{l=1 \\ l \neq i, j}}^n z_l. \end{aligned}$$

Este último é consequência das identidades (22) e (23).

Seguindo o mesmo raciocínio, quando $z_j = 1$ para $j \in I_k = \{i_1, \dots, i_k\} \subseteq \{1, \dots, n\}$ e $0 < z_j < 1$ para $j \in \{1, \dots, n\} - I_k = J_k$, obtemos

$$\Pr \left\{ \left\{ \frac{X_j}{X^{(N)}} \leq 1, \quad j \in I_k \right\} \quad \text{e} \quad \left\{ \frac{X_j}{X^{(N)}} < z_j, \quad j \in J_k \right\} \right\} = \left\{ \frac{N - n + k}{N} \right\} \prod_{j \in J_k} z_j.$$

Finalmente, fazendo $y_i = mz_i$ conclui-se a prova. $\square$

Oberservação 2: Notemos que a menos do fator $\left(1 - \frac{n}{N}\right)$ a parte contínua da função de distribuição de $\mathbf{Y}_n$ coincide com a associada a n variáveis aleatórias independentes com distribuição comum $U(0, m)$.

Proposição 3: Se o vetor aleatório $\mathbf{Y}_n$ tem lei Q_{Nmn} , então a distribuição da variável aleatória $Z = \max\{Y_1, \dots, Y_n\}$ é definida por

$$F_m(z) = \begin{cases} 0 & \text{se } z \leq 0 \\ \left(\frac{N-n}{N}\right) \left(\frac{z}{m}\right)^n & \text{se } 0 < z < m \\ 1 & \text{se } z \geq m, \end{cases}$$

onde F_m é a função de distribuição de Z .

Prova: A partir da Proposição 1, temos que a distribuição de Z coincide com a da variável $Y = m \frac{\max\{X_1, \dots, X_n\}}{X^{(N)}}$, onde $X_1, X_2, \dots, X_N$ são variáveis aleatórias independentes com distribuição comum $U(0, 1)$. Como antes, fazemos $m = 1$. Seja $0 < y < 1$, então

$$\begin{aligned} \Pr \left\{ \frac{\max\{X_1, \dots, X_n\}}{X^{(N)}} \leq y \right\} &= \Pr \{X_i \leq yX^{(N)} ; i = 1, \dots, n\} \\ &= \sum_{l=1}^{N-n} \Pr \{ \{X_i \leq yX^{(N)} ; i = 1, \dots, n\} \text{ e } X^{(N)} = X_{n+l} \} . \end{aligned} \quad (24)$$

Mas

$$\begin{aligned} &\Pr \{ \{X_i \leq yX^{(N)} ; i = 1, \dots, n\} \text{ e } X^{(N)} = X_{n+l} \} \\ &= \Pr \{ \{X_i \leq yX_{n+l} ; i = 1, \dots, n\} \text{ e } \{X_{n+i} \leq X_{n+l} ; i = 1, \dots, N-n \text{ e } i \neq l\} \} \\ &= \int_0^1 \Pr \{ \{X_i \leq yx ; i = 1, \dots, n\} \text{ e } \{X_{n+i} \leq x ; i = 1, \dots, N-n \text{ e } i \neq l\} \} dx \\ &= \int_0^1 y^n x^n x^{N-n-1} dx = \frac{1}{N} y^n . \end{aligned} \quad (25)$$

Substituindo (25) em (24) temos que

$$\Pr \left\{ \frac{\max\{X_1, \dots, X_n\}}{X^{(N)}} \leq y \right\} = \frac{N-n}{N} y^n . \quad (26)$$

Finalmente,

$$\begin{aligned} \Pr \left\{ \frac{\max\{X_1, \dots, X_n\}}{X^{(N)}} = 1 \right\} &= \Pr \{ \max\{X_1, \dots, X_n\} = X^{(N)} \} \\ &= \sum_{i=1}^n \Pr \{ \{X_l \leq X_i ; \ l = 1, \dots, N \text{ e } l \neq i\} \} . \end{aligned} \quad (27)$$

Mas

$$\begin{aligned} &\Pr \{X_l \leq X_i ; \ l = 1, \dots, N \text{ e } l \neq i\} \\ &= \int_0^1 \Pr \{X_l < x ; \ l = 1, \dots, N \text{ e } l \neq i\} dx = \int_0^1 x^{N-1} dx = \frac{1}{N} . \end{aligned} \quad (28)$$

Substituindo (28) em (27) temos que

$$\Pr \left\{ \frac{\max\{X_1, \dots, X_n\}}{X^{(N)}} = 1 \right\} = \frac{n}{N} .$$

Finalmente, fazendo $z = my$ conclui-se a prova. $\square$

A Proposição 3 é de utilidade para estabelecer o análogo do Teorema 1, desde que mostremos que $Z = \max\{Y_1, \dots, Y_n\}$ é uma estatística suficiente para a família $\{Q_{Nmn}, m > 0\}$. Este resultado será estabelecido formalmente no Capítulo 3, seção 5. Por enquanto, e para não desviarmos do nosso objetivo, assumiremos este como verdadeiro.

Teorema 5: *Seja Q_{Nmn} a lei das n primeiras coordenadas de um ponto uniformemente distribuído sobre $\mathcal{X}_m^N$ e P_m^n a lei de n variáveis aleatórias independentes com distribuição comum uniforme sobre o intervalo $(0, m)$, então se $1 \leq n \leq N$*

$$\|Q_{Nmn} - P_m^n\| \leq 4 \frac{n}{N} . \quad (29)$$

Prova: Levando em conta a invariância da distância de variação total sob transformações 1-1, temos que

$$\|Q_{Nmn} - P_m^n\| = \|Q_{N1n} - P_1^n\| .$$

Seja $\tilde{Q}_1$ a Q_{N1n} -lei de $Z = \max\{Y_1, \dots, Y_n\}$ e $\tilde{P}_1$ a correspondente P_1^n -lei.

Notando que Z é suficiente para as famílias $\{Q_{Nmn}, m > 0\}$ e $\{P_m^n, m > 0\}$, temos pela propriedade 3 da distância de variação que

$$\|Q_{Nmn} - P_m^n\| = \|\tilde{Q}_1 - \tilde{P}_1\| .$$

Agora, a partir da Proposição 3, temos que para $B \in \mathcal{B}_1$

$$\tilde{Q}_1(B) = \int_{B \cap (0,1)} \left(\frac{N-n}{N} \right) n z^{n-1} dz + \tilde{Q}_1(B \cap \{1\}) ,$$

e

$$\tilde{P}_1(B) = \int_{B \cap (0,1)} n z^{n-1} dz .$$

Logo

$$\begin{aligned} \|\tilde{Q}_1 - \tilde{P}_1\| &= 2 \sup_{B \in \mathcal{B}_1} |\tilde{Q}_1(B) - \tilde{P}_1(B)| \\ &\leq 2 \sup_{B \in \mathcal{B}_1} \left| \int_{B \cap (0,1)} \left(\frac{N-n}{N} \right) n z^{n-1} dz - \int_{B \cap (0,1)} n z^{n-1} dz \right| \\ &\quad + 2 \sup_{B \in \mathcal{B}_1} |\tilde{Q}_1(B \cap \{1\})| \\ &\leq 2 \int_{(0,1)} \frac{n}{N} n z^{n-1} dz + 2 \frac{n}{N} = 4 \frac{n}{N} . \end{aligned}$$

□

Como na seção prévia, os teoremas do tipo de Finetti envolvendo distribuições uniformes contínuas sobre o intervalo $(0, \theta)$ são consequência direta deste teorema.

Sejam $X_1, X_2, \dots, X_N$ variáveis aleatórias tomando valores em $\mathbb{R}_+$. Seja P_N a lei de $X_N = (X_1, X_2, \dots, X_N)$ e suponha que dado $X^{(N)} = m$, X_N é uniformemente distribuído sobre $\mathcal{X}_m^N$. Seja P_θ^n a lei de n variáveis aleatórias independentes com distribuição comum

uniforme sobre o intervalo $(0, \theta)$, $\theta > 0$ e $P_{\mu n} = \int P_\theta^n d\mu(\theta)$, para alguma medida de probabilidade μ concentrada sobre $\mathbb{R}_+$.

Teorema 6: *Se P_N e $P_{\mu n}$ são as medidas de probabilidade acima definidas, então existe uma medida de probabilidade μ concentrada sobre $(0, \infty)$ tal que para cada $1 \leq n \leq N$*

$$\|P_n - P_{\mu n}\| \leq 4 \frac{n}{N}, \quad (30)$$

onde P_n é a lei n -dimensional de P_N .

Prova: Basta escolher $\mu = \mu_N$, a P_N -lei de $X^{(N)}$ e utilizar o Teorema 5. $\square$

A forma infinita do Teorema 6 pode ser derivada com relativa simplicidade utilizando-se deste teorema e os resultados prévios.

Teorema 7: *Seja $X_1, X_2, \dots$ uma seqüência infinita de variáveis aleatórias, tomando valores em $\mathbb{R}^+$ e com lei P . Se para cada $N \in \mathbb{N}$ e $m > 0$, a distribuição condicional de $X_1, X_2, \dots, X_N$ dado $X^{(N)} = m$ é uniforme sobre $\mathcal{X}_m^N$, então existe uma medida de probabilidade μ sobre $(0, \infty)$ tal que para cada $n \in \mathbb{N}$*

$$P(X_n \in B) = \int_{(0, \infty)} P_\theta^n(B) d\mu(\theta), \quad B \in \mathcal{B}_n,$$

onde P_θ^n é a lei de n variáveis aleatórias independentes com distribuição comum $U(0, \theta)$.

Prova: A prova é análoga à do Teorema 3. Seja μ_N a P -lei de $X^{(N)} = \max\{X_1, \dots, X_N\}$, então a partir da hipótese e da Proposição 2,

$$\begin{aligned} \Pr\{X_1 > k\} &= \int_{(k, \infty)} \left(1 - \left(\frac{N-1}{N}\right) \frac{k}{m}\right) d\mu_N(m) \\ &\geq \int_{(k, \infty)} \left(1 - \frac{k}{m}\right) d\mu_N(m). \end{aligned}$$

Notando que $\Pr \{X_1 > k\} \xrightarrow[k \rightarrow \infty]{} 0$ e que para k fixo, $(1 - \frac{k}{m}) I_{(k, +\infty)}(m) \xrightarrow[m \rightarrow \infty]{} 1$, conclui-se a partir da desigualdade acima que μ_N é tight. Utilizando-nos deste fato, da hipótese e do Teorema 6, obtemos o resultado. $\square$

No que segue mostraremos que seqüências de variáveis aleatórias independentes com distribuição comum $U(0, \theta)$, $\theta > 0$, satisfazem as condições do Teorema 6.

Proposição 4: *Se $X_1, X_2, \dots, X_N$ são variáveis aleatórias independentes com distribuição comum $U(0, \theta)$, então a distribuição condicional de $X_1, X_2, \dots, X_N$ dado $X^{(N)} = m$ é uniforme sobre $\mathcal{X}_N^m$.*

Prova: Denote por P_θ^N a lei de $X_1, X_2, \dots, X_N$. É bem conhecido que $X^{(N)} = \max\{X_1, \dots, X_N\}$ é uma estatística suficiente e completa para a família $\{P_\theta^N : \theta > 0\}$. Também, tendo em vista a Proposição 1, não é difícil mostrar que $\mathbf{T}_N = (\frac{X_1}{X^{(N)}}, \dots, \frac{X_N}{X^{(N)}})$ é ancilar, isto é, a distribuição de $\mathbf{T}_N$ é a mesma para cada θ . De fato

$$\mathbf{T}_N = \left(\frac{X_1/\theta}{X^{(N)}/\theta}, \dots, \frac{X_N/\theta}{X^{(N)}/\theta} \right) \stackrel{D}{=} \left(\frac{Y_1}{Y^{(N)}}, \dots, \frac{Y_N}{Y^{(N)}} \right),$$

onde $Y_1, Y_2, \dots, Y_N$ são variáveis aleatórias independentes com distribuição comum $U(0, 1)$ e $Y^{(N)} = \max\{Y_1, Y_2, \dots, Y_N\}$. Consequentemente, pela Proposição 1, $\mathbf{T}_N$ tem distribuição uniforme sobre $\mathcal{X}_1^m$ e portanto é ancilar.

Por outro lado os lemas de Basu (Basu, 1955, 1958; veja também Branco, D.M. (1991)) garantem que $\mathbf{T}_N$ é estatisticamente independente de $X^{(N)}$, isto é, $\mathbf{T}_N \perp\!\!\!\perp X^{(N)}/\theta$. Logo, para f limitada temos que

$$\begin{aligned} E_\theta(f(X_1, \dots, X_N)/X^{(N)} = m) &= E_\theta(f(\mathbf{T}_N X^{(N)})/X^{(N)} = m) \\ &= E_{\mathbf{T}_N}(f(\mathbf{T}_N m)). \end{aligned}$$

Mas, $m\mathbf{T}_N$ têm distribuição uniforme sobre $\mathcal{X}_m^N$, concluindo-se assim a prova. $\square$

Observação 3: Convém notar que o recíproco do Teorema 7 também é válido. Com efeito, se existe uma variável aleatória Θ tal que para cada $N \in \mathbb{N}$, $X_1, \dots, X_N/\Theta = \theta$, são independentes com distribuição comum $U(0, \theta)$, então

$$(X_1, \dots, X_N) \coprod \Theta / X^{(N)} ,$$

pois $X^{(N)}$ é suficiente para $X_1, \dots, X_N$ com relação a θ .

Consequentemente,

$$(X_1, \dots, X_N)/(\Theta, X^{(n)}) = (\theta, m) \stackrel{\mathcal{D}}{=} (X_1, \dots, X_N)/X^{(N)} = m .$$

Portanto, a partir da proposição prévia, conclui-se que $(X_1, \dots, X_N)/X^{(N)} = m$ possui distribuição uniforme sobre $\mathcal{X}_m^N$.

2.4 Distribuição uniforme sobre $(-\theta, \theta)$, $\theta > 0$

Considere a borda de um hipercubo N -dimensional, digamos

$$H_m^N = \{(x_1, \dots, x_N) \in \mathbb{R}^N : \max\{|x_1|, \dots, |x_N|\} = m\}, \quad m > 0 .$$

Por exemplo, se $N = 2$, então

$$\begin{aligned} H_m^2 &= \{(x_1, x_2) \in \mathbb{R}^2 : \max\{|x_1|, |x_2|\} = m\} \\ &= \bigcup_{j=0}^1 \bigcup_{i=1}^2 M_i((-1)^j m) , \end{aligned}$$

onde

$$\begin{aligned} M_1(m) &= \{(x_1, x_2) \in \mathbb{R}^2 : x_1 = m, \quad -m \leq x_2 \leq m\} , \\ M_1(-m) &= \{(x_1, x_2) \in \mathbb{R}^2 : x_1 = -m, \quad -m \leq x_2 \leq m\} , \\ M_2(m) &= \{(x_1, x_2) \in \mathbb{R}^2 : x_2 = m, \quad -m \leq x_1 \leq m\} , \\ M_2(-m) &= \{(x_1, x_2) \in \mathbb{R}^2 : x_2 = -m, \quad -m \leq x_1 \leq m\} . \end{aligned}$$

Assim, H_m^2 é a borda de um quadrado cujos lados são de comprimento m , como ilustra a figura abaixo.

Figura 2: Gráfico dos espaços H_m^2 .

Como antes, podemos traduzir a idéia de equiprobabilidade por consideração de comprimentos. Assim, fazendo uma analogia com o caso anterior, parece natural definir para $B \in \mathcal{B}_2$, $A = B \cap H_m^2$,

$$\Pr \{A\} = \frac{\sum_{j=0}^1 \sum_{i=1}^2 \text{Comprimento}(A \cap M_i((-1)^j m))}{4m},$$

como a nossa medida de probabilidade uniforme sobre H_m^2 .

Seguindo esta idéia, mostraremos na seqüência que as n -primeiras ($n < N$) coordenadas de um ponto uniformemente distribuído sobre H_m^N é próxima (no sentido de distância de variação total) da lei de n variáveis aleatórias independentes com distribuição comum uniforme sobre o intervalo $(-m, m)$. Teoremas do tipo de Finetti envolvendo distribuições paramétricas uniformes sobre o intervalo $(-\theta, \theta)$, $\theta > 0$, serão derivadas a partir deste

resultado.

No que segue, construiremos formalmente a distribuição uniforme sobre H_m^N .

Considere os subconjuntos de H_m^N , definidos por

$$\begin{aligned} M_i(m) &= \{(x_1, \dots, x_N) \in H_m^N : x_i = m\} \quad \text{e} \\ M_i(-m) &= \{(x_1, \dots, x_N) \in H_m^N : x_i = -m\}, \end{aligned}$$

então

$$H_m^N = \bigcup_{j=0}^1 \bigcup_{i=1}^N M_i((-1)^j m).$$

Sejam φ^i , $i = 1, \dots, N$, as projeções definidas no início da seção 3 e $\mathcal{A} = H_m^N \cap \mathcal{B}_N$, onde $\mathcal{B}_N$ é a σ -álgebra de Borel sobre $\mathbb{R}^N$. Seja λ a medida de Lebesgue N -dimensional.

Considere os espaços de medida $\{(H_m^N, \mathcal{A}, \mu_i)\}_{i=1}^N$, onde $\mu_i : \mathcal{A} \mapsto [0, \infty)$ é definida por

$$\mu_i(A) = \lambda(\varphi^i(A \cap M_i(m))) + \lambda(\varphi^i(A \cap M_i(-m))), \quad A \in \mathcal{A}$$

Seja agora $\tilde{Q}_{Nm} : \mathcal{A} \mapsto [0, 1]$ a medida de probabilidade definida por

$$\tilde{Q}_{Nm}(A) = \frac{\sum_{i=1}^N \mu_i(A)}{2N(2m)^{N-1}}, \quad A \in \mathcal{A}.$$

A função $Q_{Nm} : \mathcal{B}_N \mapsto [0, 1]$ definida por

$$Q_{Nm}(B) = \tilde{Q}_{Nm}(B \cap H_m^N), \quad B \in \mathcal{B}_N \quad (31)$$

será denominada de medida de probabilidade uniforme sobre H_m^N .

Seguindo o mesmo raciocínio do caso anterior, não é difícil ver que para cada m , Q_{Nm} é uma medida de probabilidade permutável e que $Q_{Nm}(B)$, vista como função de m e B , é uma função de transição de $(\mathbb{R}^+, \mathcal{B}_1^+)$ para $(\mathbb{R}^N, \mathcal{B}^N)$.

Uma definição de Q_{Nm} em termos de variáveis aleatórias é dada a seguir.

Proposição 5: *Sejam $X_1, X_2, \dots, X_N$ variáveis aleatórias independentes com distribuição comum $U(-1, 1)$. Seja $Y_i = m \frac{X_i}{M}$, onde $M = \max\{|X_1|, \dots, |X_N|\}$, então o vetor aleatório $\mathbf{Y}_N = (Y_1, \dots, Y_N)$ é uniformemente distribuído sobre H_m^N .*

Prova: Dada a similaridade com a Proposição 1, somente um esboço da prova será dado a seguir.

Seja P a lei de $(X_1, \dots, X_N)$ e Q a lei de $\mathbf{Y}_N$. Notando a partir da definição de $\mathbf{Y}_N$ que $\max\{|Y_1|, \dots, |Y_N|\} = m$, temos que

$$Q(H_m^N) = \Pr\{\mathbf{Y}_N \in H_m^N\} = 1.$$

Agora, para $A \in \mathcal{A}$,

$$\begin{aligned} Q(A) &= \Pr\{\mathbf{Y}_N \in A\} \\ &= \Pr\{\mathbf{Y}_N \in \bigcup_{j=0}^1 \bigcup_{i=1}^N (A \cap M_i((-1)^j m))\} \\ &= \sum_{j=0}^1 \sum_{i=1}^N \Pr\{\mathbf{Y}_N \in A \cap M_i((-1)^j m)\}. \end{aligned}$$

Por outro lado,

$$\Pr\{\mathbf{Y}_N \in A \cap M_i(m)\} = \Pr\{Y_i = m, (Y_1, \dots, Y_{i-1}, Y_{i+1}, \dots, Y_N) \in \varphi^i(A \cap M_i(m))\}$$

e

$$\Pr\{\mathbf{Y}_N \in A \cap M_i(-m)\} = \Pr\{Y_i = -m, (Y_1, \dots, Y_{i-1}, Y_{i+1}, \dots, Y_N) \in \varphi^i(A \cap M_i(-m))\}$$

Assim, desenvolvendo estas igualdades de maneira similar à utilizada na Proposição 1, conclui-se que $Q = Q_{Nm}$. $\square$

Notemos a partir desta última proposição que se o vetor aleatório $\mathbf{Y}_N$ possui distribuição uniforme sobre H_m^N , então o vetor $g(\mathbf{Y}_N) = (|Y_1|, \dots, |Y_N|)$ é uniformemente distribuído sobre $\mathcal{X}_m^N$.

Proposição 6: Se o vetor aleatório $\mathbf{Y}_N = (Y_1, \dots, Y_N)$ tem lei Q_{Nm} , então a lei de $\mathbf{Y}_n = (Y_1, \dots, Y_n)$ é definida por

$$F_m(y_1, \dots, y_n) = \begin{cases} 0 & \text{se } y_i = -m \text{ para } i \in I \subset \{1, \dots, n\} \\ & \text{com } \#I \geq 2 \text{ ou se } y_i < -m \text{ para} \\ & \text{algum } i = 1, \dots, n \\ \frac{1}{2N} \prod_{\substack{j=1 \\ j \neq i}}^n \left\{ \frac{m+y_j}{2m} \right\} & \text{se } -m \leq y_j \leq m \text{ para } j = 1, \dots, n \\ & \text{e } y_i = -m \\ \sum_{i=1}^n \frac{1}{2N} \prod_{\substack{j=1 \\ j \neq i}}^n \left\{ \frac{m+y_j}{2m} \right\} \\ + \left(\frac{N-n}{N} \right) \prod_{i=1}^n \left(\frac{m+y_i}{2m} \right) & \text{se } -m < y_i < m \text{ para } i = 1, \dots, n \\ \left(\frac{N-n}{N} + \frac{k}{2N} \right) \prod_{j \in I_k} \left(\frac{m+y_j}{2m} \right) & \text{se } y_i = m \text{ para} \\ & i \in I_k = \{i_1, \dots, i_k\} \subset \{1, \dots, n\} \\ + \sum_{i=1}^n \frac{1}{2N} \prod_{\substack{j=1 \\ j \neq i}}^n \left\{ \frac{m+y_j}{2m} \right\} & \text{e } y_j < m \text{ para } j \in J_k = I_k^c \text{ com } k = 1, \dots, n \\ 1 & \text{se } y_i \geq m \text{ para } i = 1, \dots, n \end{cases}$$

onde F_m é a função de distribuição de $\mathbf{Y}_n$.

Prova: Utilizando-nos da caracterização dada na Proposição 5 e assumindo $m = 1$, temos em primeiro lugar que para $i \neq j$

$$\Pr \{Y_i = -1, Y_j = -1, \{Y_\ell < y_\ell \text{ para } \ell = 1, \dots, n \text{ e } \ell \neq j, i\}\} = 0$$

$$\text{e} \quad \Pr \{Y_i = -1, Y_j = 1, \{Y_\ell < y_\ell \text{ para } \ell = 1, \dots, n \text{ e } \ell \neq j, i\}\}$$

$$= \Pr \{X_i = -M, X_j = M, \{X_\ell < y_\ell \text{ para } \ell = 1, \dots, n \text{ e } \ell \neq j, i\}\} = 0,$$

pois $X_1, \dots, X_N$ são variáveis aleatórias contínuas.

Seja agora, $-1 \leq z_j \leq 1$, $j \neq i$, então

$$\begin{aligned} & \Pr \left\{ \frac{X_i}{M} = -1 \text{ e } \left\{ \frac{X_j}{M} \leq z_j; j = 1, \dots, n \text{ e } j \neq i \right\} \right\} \\ = & \Pr \{X_i < 0 \text{ e } \{X_i \leq X_j \leq z_j(-X_i); j = 1, \dots, n \text{ e } j \neq i\} \\ & \text{e } \{-X_i \leq X_{n+\ell} \leq X_i; \ell = 1, \dots, N-n\}\} \end{aligned}$$

$$\begin{aligned}
&= \int_0^1 \Pr \{ \{-x \leq X_j \leq z_j x; j = 1, \dots, n \text{ e } j \neq i\} \text{ e } \{-x \leq X_{n+\ell} \leq x; \ell = 1, \dots, N-n\} \} \frac{1}{2} dx \\
&= \prod_{\substack{j=1 \\ j \neq i}}^n \left(\frac{z_j + 1}{2} \right) \int_0^1 x^{n-1} x^{N-n} \frac{1}{2} dx = \frac{1}{2N} \prod_{\substack{j=1 \\ j \neq i}}^n \left(\frac{z_j + 1}{2} \right) .
\end{aligned}$$

Analogamente, se $-1 \leq z_j \leq 1$, $j \neq i$

$$\Pr \left\{ \frac{X_i}{M} = 1 \text{ e } \left\{ \frac{X_j}{M} \leq z_j; j = 1, \dots, n \text{ e } j \neq i \right\} \right\} = \frac{1}{2N} \prod_{\substack{j=1 \\ j \neq i}}^n \left\{ \frac{z_j + 1}{2} \right\} .$$

Finalmente, se $-1 < z_j < 1$, $j \neq i$, temos

$$\Pr \left(\{-1 < \frac{X_i}{M} < z_i; i = 1, \dots, n\} \right) = \sum_{\ell=n+1}^N \Pr \{ \{-1 < \frac{X_i}{M} < z_i; i = 1, \dots, n\} \text{ e } M = |X_\ell| \} . \quad (32)$$

Mas, se $\ell \in \{n+1, \dots, N\}$,

$$\begin{aligned}
&\Pr \{ \{-1 < \frac{X_i}{M} < z_i; i = 1, \dots, n\} \text{ e } M = |X_\ell| \} \\
&= \Pr \{ \{-|X_\ell| < X_i < |X_\ell| z_i; i = 1, \dots, n\} \text{ e } \{|X_j| < |X_\ell|; j = n+1, \dots, N \text{ e } j \neq \ell\} \} \\
&= \int_0^1 \Pr \{ \{-x < X_i < x z_i; i = 1, \dots, n\} \text{ e } \{-x < X_j < x; j \neq \ell, j = n+1, \dots, N\} \} dx \\
&= \prod_{i=1}^n \left(\frac{1 + z_i}{2} \right) \int_0^1 x^{N-1} dx = \frac{1}{N} \prod_{i=1}^n \left(\frac{1 + z_i}{2} \right) . \quad (33)
\end{aligned}$$

Substituindo (33) em (32) temos que

$$\Pr \left(\left\{ -1 < \frac{X_i}{M} < z_i, i = 1, \dots, n \right\} \right) = \frac{N-n}{N} \prod_{i=1}^n \left(\frac{z_i + 1}{2} \right) .$$

Fazendo algumas operações algébricas para o cálculo das probabilidades necessárias para obter F_m , conclui-se a prova. $\square$

Como antes, Q_{Nmn} denotará a distribuição n -dimensional de Q_{Nm} . Como na seção anterior pode ser mostrado que $W = \max\{|Y_1|, \dots, |Y_n|\}$ é suficiente para a família $\{Q_{Nmn}, m > 0\}$ (veja seção 5 do capítulo 3). A distribuição de W é dada a seguir.

Proposição 7: *Seja $\mathbf{Y}_n = (Y_1, \dots, Y_n)$ vetor aleatório com lei Q_{Nmn} , então a variável aleatória $W = \max\{|Y_1|, \dots, |Y_n|\}$ tem lei definida por*

$$F_m(w) = \begin{cases} 0 & \text{se } w \leq 0, \\ \left(\frac{N-n}{N}\right) \left(\frac{w}{m}\right)^n & \text{se } 0 < w < m, \\ 1 & \text{se } w \geq m, \end{cases}$$

onde F_m é a função de distribuição de W .

Prova: A partir da Proposição 5, temos que a lei de W é igual à lei do vetor aleatório $(m \frac{|X_1|}{M}, \dots, m \frac{|X_n|}{M})$, onde $M = \max\{|X_1|, \dots, |X_N|\}$ e $X_1, X_2, \dots, X_N$ são variáveis aleatórias independentes com distribuição comum $U(-1, 1)$. Mas, $Z_i = |X_i|$ tem distribuição uniforme sobre o intervalo $(0, 1)$. Consequentemente, a distribuição de W coincide com a deduzida na Proposição 3. $\square$

As proposições acima e os teoremas a seguir permitem-nos estabelecer as versões finita e infinita do teorema de de Finetti envolvendo distribuições uniformes sobre intervalos simétricos em torno de zero.

Teorema 8: *Seja Q_{Nmn} a lei das primeiras n ($n < N$) coordenadas de um ponto uniformemente distribuído sobre H_m^N e P_m^n a lei de n variáveis aleatórias independentes com distribuição comum uniforme sobre o intervalo $(-m, m)$, então se $N \geq 2$, $1 \leq n \leq N$*

$$\|Q_{Nmn} - P_m^n\| \leq 4 \frac{n}{N}. \quad (34)$$

Prova: Levando em conta a invariância da distância de variação total sob transformações 1-1, temos que

$$\|Q_{Nmn} - P_m^n\| = \|Q_{N1n} - P_1^n\|.$$

Seja $\tilde{Q}_1$ a Q_{N1n} lei de $W = \max\{|Y_1|, \dots, |Y_n|\}$ e $\tilde{P}_1$ a correspondente P_1^n -lei.

Utilizando o fato de que W é uma estatística suficiente para as famílias $\{P_m^n, m > 0\}$ e $\{Q_{Nmn}, m > 0\}$, temos pela propriedade 3, da distância de variação, que

$$\|Q_{N1n} - P_1^n\| = \|\tilde{Q}_1 - \tilde{P}_1\|.$$

Mas o termo à direita nesta igualdade é exatamente igual ao termo à direita na segunda identidade da prova do Teorema 5. Consequentemente,

$$\|\tilde{Q}_1 - \tilde{P}_1\| \leq 4 \frac{n}{N}.$$

Concluindo-se assim a prova. $\square$

A forma finita do teorema do tipo de Finetti envolvendo distribuições paramétricas é consequência deste resultado.

Sejam $X_1, X_2, \dots, X_N$ variáveis aleatórias tomando valores em $\mathbb{R}$. Seja P_N a lei de $\mathbf{X}_N = (X_1, \dots, X_N)$ e suponha que dado $M = \max\{|X_1|, \dots, |X_N|\} = m$, $\mathbf{X}_N$ é uniformemente distribuído sobre H_m^N . Seja P_θ^n a lei de n variáveis aleatórias independentes com distribuição comum uniforme sobre o intervalo $(-\theta, \theta)$ e $P_{\mu_n} = \int P_\theta^n d\mu(\theta)$ para alguma medida de probabilidade μ sobre $\mathbb{R}$. Denotemos por P_n a lei n -dimensional de P_N .

Teorema 9: *Se P_N e P_{μ_n} são as medidas de probabilidade acima definidas, então existe uma medida de probabilidade μ sobre $\mathbb{R}$ tal que para cada $1 \leq n \leq N$, $N \geq 2$,*

$$\|P_n - P_{\mu_n}\| \leq 4 \frac{n}{N}. \quad (35)$$

Prova: Basta escolher μ como a P_N -lei de $M = \max\{|X_1|, \dots, |X_N|\}$ e usar o Teorema 8. $\square$

O teorema caracteriza uma classe de seqüências finitas de variáveis aleatórias permutáveis como quase mistura de variáveis aleatórias independentes com distribuição comum uniforme sobre $(-m, m)$. A versão infinita do teorema é dada a seguir.

Teorema 10: *Sejam $X_1, X_2, \dots$ seqüência infinita de variáveis aleatórias, tomando valores em $\mathbb{R}^+$ com lei P . Suponha que para cada $N \in \mathbb{N}$ e $m > 0$ ($N \geq 2$) a distribuição condicional de $\mathbf{X}_N = (X_1, X_2, \dots, X_N)$ dado $M = \max\{|X_1|, \dots, |X_N|\} = m$ é uniforme sobre H_m^N , então existe uma medida de probabilidade μ sobre $\mathbb{R}$ tal que para cada $n \in \mathbb{N}$,*

$$\Pr \{\mathbf{X}_n \in B\} = \int_{\mathbb{R}} P_{\theta}^n(B) d\mu(\theta), \quad B \in \mathcal{B}_n,$$

onde P_{θ}^n é a lei de n variáveis aleatórias independentes com distribuição comum $U(-\theta, \theta)$, $\theta > 0$.

Prova: A prova será omitida, pois é similar à do Teorema 7. Somente alguns cálculos adicionais são necessários. $\square$

No que segue, mostraremos que seqüências de variáveis aleatórias independentes com distribuição comum $U(-\theta, \theta)$, $\theta > 0$, satisfazem as condições do Teorema 10.

Proposição 8: *Se $X_1, X_2, \dots, X_N$ são variáveis aleatórias independentes com distribuição comum uniforme sobre o intervalo $(-\theta, \theta)$, então a distribuição condicional de $X_1, X_2, \dots, X_N \mid M = m$, com $M = \max\{|X_1|, \dots, |X_N|\}$, é uniforme sobre H_m^N .*

Prova: Denotemos por P_{θ}^N a lei de $X_1, X_2, \dots, X_N$. É sabido que $M = \max\{|X_1|, \dots, |X_N|\}$ é uma estatística suficiente e completa para a família $\{P_{\theta}^N : \theta > 0\}$. Também, em virtude da Proposição 5, $\mathbf{T}_N = \left(\frac{X_1}{M}, \dots, \frac{X_N}{M}\right)$ é uma estatística ancilar. Basta observar que

$$\mathbf{T}_N = \left(\frac{X_1/\theta}{M/\theta}, \dots, \frac{X_N/\theta}{M/\theta}\right) \stackrel{D}{=} \left(\frac{Y_1}{M^*}, \dots, \frac{Y_N}{M^*}\right),$$

onde $M^* = \max\{|Y_1|, \dots, |Y_N|\}$ e $Y_1, Y_2, \dots, Y_N$ são variáveis aleatórias independentes com distribuição comum $U(-1, 1)$. Seguindo agora o mesmo procedimento utilizado na Proposição 4, conclui-se a prova. $\square$

Observação 4: O recíproco do Teorema 10 pode ser estabelecido utilizando-se de ar-

gumentos de suficiência. De fato, se $M = \max\{|X_1|, \dots, |X_N|\}$ e existe uma variável aleatórias Θ tal que, para cada $N \in \mathbb{N}$, $X_1, X_2, \dots, X_N/\Theta = \theta$ são independentes com distribuição $U(-\theta, \theta)$, então

$$(X_1, \dots, X_N)/(\Theta, M) = (\theta, m) \stackrel{D}{=} (X_1, \dots, X_N)/M = m ,$$

a qual, pela proposição prévia, é uniforme sobre H_m^N .

Os resultados anteriores caracterizam uma classe de processos permutáveis que podem ser expressos como quase mistura (seqüência finita) ou mistura (seqüência infinita) de variáveis aleatórias independentes com distribuição comum uniforme sobre intervalos simétricos em torno de zero. A extensão destes resultados para distribuições uniformes simétricas em torno de um valor $\lambda \in \mathbb{R}$, conhecido, constituem uma aplicação direta deste resultado. Basta definir

$$H_m^N(\lambda) = \{(x_1, \dots, x_N) \in \mathbb{R}^N : \max\{|x_1 - \lambda|, \dots, |x_N - \lambda|\} = m\}, \quad m > 0$$

e proceder como antes.

A questão natural que pode ser levantada neste ponto é: que classe de processos permutáveis podem ser representados como quase mistura de seqüências finitas (seqüências infinitas) de variáveis aleatórias independentes com distribuição comum $U(\theta_1, \theta_2)$ (onde $\theta_1, \theta_2 \in \mathbb{R}$, $\theta_1 < \theta_2$ e tanto θ_1 como θ_2 são desconhecidos)? Esta questão é resolvida na próxima seção.

2.5 Distribuição uniforme sobre (θ_1, θ_2)

Consideremos o espaço

$$\mathcal{X}_{m_1, m_2}^N = \{(x_1, \dots, x_N) \in \mathbb{R}^N : \min\{x_1, \dots, x_N\} = m_1, \quad \max\{x_1, \dots, x_N\} = m_2\}$$

com $m_1, m_2 \in \mathbb{R}$, $m_1 < m_2$ e $N \geq 2$.

No que segue, a construção de uma medida de probabilidade uniforme será definida. Antes disso, convém ilustrar as idéias com alguns casos particulares. Por exemplo, se $N = 2$, então $\mathcal{X}_{m_1, m_2}^2 = \{(m_1, m_2), (m_2, m_1)\}$, em cujo caso a distribuição uniforme é definida naturalmente colocando massa $1/2$ em cada ponto. Se $N = 3$, então

$$\begin{aligned}\mathcal{X}_{m_1, m_2}^3 &= \{(x_1, x_2, x_3) \in \mathbb{R}^3 : \min\{x_1, x_2, x_3\} = m_1 \text{ e } \max\{x_1, x_2, x_3\} = m_2\} \\ &= \bigcup_{\substack{i \neq j \\ i, j \in \{1, 2, 3\}}} M_{ij}(m_1, m_2),\end{aligned}$$

onde

$$\begin{aligned}M_{ij}(m_1, m_2) &= \{(x_1, x_2, x_3) \in \mathbb{R}^3 : x_i = m_1, x_j = m_2 \\ &\text{e } m_1 \leq x_\ell \leq m_2, \quad i \neq j \neq \ell\}, \quad m_1 < m_2.\end{aligned}$$

A figura a seguir exhibe $\mathcal{X}_{m_1, m_2}^3$ para $m_1 = 0, m_2 = 1$.

Figura 3: Gráfico do espaço $\mathcal{X}_{m_1, m_2}^3$ com $m_1 = 0, m_2 = 1$.

O espaço tem medida de Lebesgue zero como subconjunto de $\mathbb{R}^3$. Também, qualquer projeção de M_{ij} sobre $\mathbb{R}^2$ tem medida de Lebesgue zero como subconjunto de $\mathbb{R}^2$.

Assim, novamente, nossa idéia de equiprobabilidade é traduzida por consideração dos comprimentos de cada $M_{ij}(m_1, m_2)$.

Formalmente, seja

$$M_{ij}(\mathbf{m}) = \{(x_1, \dots, x_N) \in \mathcal{X}_{\mathbf{m}}^N : x_i = m_1, x_j = m_2, \text{ para } i \neq j \text{ e } i, j \in \{1, \dots, N\}\},$$

onde $\mathbf{m} = (m_1, m_2)$, então

$$\mathcal{X}_{\mathbf{m}}^N = \bigcup_{\substack{i, j \in \{1, \dots, N\} \\ i \neq j}} M_{ij}(\mathbf{m}).$$

Considere as projeções $\varphi^{ij} : \mathbb{R}^N \mapsto \mathbb{R}^{N-2}$ definidas por

$$\varphi^{ij}(x_1, \dots, x_i, \dots, x_j, \dots, x_N) = (x_1, \dots, x_{i-1}, x_{i+1}, \dots, x_{j-1}, x_{j+1}, \dots, x_N),$$

com $i, j \in \{1, \dots, N\}$ e $i \neq j$.

Seja agora λ a medida de Lebesgue $N - 2$ dimensional e $\mathcal{A} = \mathcal{X}_{\mathbf{m}}^N \cap \mathcal{B}_N$, com $\mathcal{B}_N$ a σ -álgebra de Borel sobre $\mathbb{R}^N$.

Considere os espaços de medida $\{(\mathcal{X}_{\mathbf{m}}^N, \mathcal{A}, \mu_{ij})\}_{\substack{i, j=1 \\ i \neq j}}^N$, onde $\mu_{ij} : \mathcal{A} \mapsto [0, \infty)$ é definida por

$$\mu_{ij}(A) = \lambda(\varphi^{ij}(A \cap M_{ij}(\mathbf{m}))), \quad A \in \mathcal{A}.$$

Seja agora $\tilde{Q}_{N\mathbf{m}} : \mathcal{A} \mapsto [0, 1]$ definida por

$$\tilde{Q}_{N\mathbf{m}}(A) = \sum_{\substack{i \neq j \\ i, j \in \{1, \dots, N\}}} \frac{\mu_{ij}(A)}{N(N-1)(m_2 - m_1)^{N-2}}, \quad A \in \mathcal{A}.$$

A função $Q_{N\mathbf{m}} : \mathcal{B}_N \mapsto [0, 1]$ definida por

$$Q_{N\mathbf{m}}(B) = \tilde{Q}_{N\mathbf{m}}(B \cap \mathcal{X}_{\mathbf{m}}^N), \quad B \in \mathcal{B}_N \quad (35)$$

será denominada de medida de probabilidade uniforme sobre $\mathcal{X}_{\mathbf{m}}^N$.

Não é difícil ver a partir desta definição que $Q_{N\mathbf{m}}$ é uma medida de probabilidade permutável e que $Q_{N\mathbf{m}}(B)$ vista como função de $\mathbf{m}$ e B é uma função de transição.

Mostraremos no que segue que a lei das primeiras n coordenadas de um ponto uniformemente distribuído sobre $\mathcal{X}_m^N$ é próxima (no sentido de distância de variação total) da lei de n variáveis aleatórias independentes com distribuição comum uniforme sobre o intervalo (m_1, m_2) . Os passos a serem seguidos são idênticos aos das seções prévias. Porém, estes exigem mais trabalho e requerem alguns cuidados adicionais, especialmente na derivação da forma infinita do teorema do tipo de Finetti envolvendo distribuições uniformes dependendo de dois parâmetros.

Os resultados são apresentados mantendo a mesma estrutura dos casos anteriores.

Assim, uma definição de $Q_{N\mathbf{m}}$ em termos de variáveis aleatórias é dada a seguir.

Proposição 9: *Sejam $X_1, X_2, \dots, X_N$ variáveis aleatórias independentes com distribuição comum $U(0, 1)$. Seja*

$$Y_i = (m_2 - m_1) \left\{ \frac{X_i - M_{1N}}{M_{2N} - M_{1N}} \right\} + m_1 ,$$

onde $M_{1N} = \min\{X_1, X_2, \dots, X_N\}$ e $M_{2N} = \max\{X_1, \dots, X_N\}$ e $m_1, m_2 \in \mathbb{R}$, com $m_1 < m_2$. Então o vetor aleatório $\mathbf{Y}_N = (Y_1, Y_2, \dots, Y_N)$ possui distribuição uniforme sobre $\mathcal{X}_{(m_1, m_2)}^N$.

Prova: Seja P a lei de $(X_1, \dots, X_N)$ e Q a P -lei de $\mathbf{Y}_N$. Notando a partir da definição de $\mathbf{Y}_N$ que $\min\{Y_1, \dots, Y_N\} = m_1$, $\max\{Y_1, \dots, Y_N\} = m_2$, temos que

$$Q(\mathcal{X}_{m_1, m_2}^N) = \Pr \{Y_N \in \mathcal{X}_{m_1, m_2}^N\} = 1 .$$

Agora, para $A \in \mathcal{A}$,

$$Q(A) = \Pr \{\mathbf{Y}_N \in A\} = \Pr \left\{ \mathbf{Y}_N \in \bigcup_{\substack{i, j \in \{1, \dots, N\} \\ i \neq j}} (A \cap M_{ij}(\mathbf{m})) \right\} = \sum_{\substack{i, j \in \{1, \dots, N\} \\ i \neq j}} \Pr \{\mathbf{Y}_N \in A \cap M_{ij}(\mathbf{m})\} . \quad (36)$$

Mas para $i \neq j$, $i, j \in \{1, \dots, N\}$,

$$\begin{aligned}
& \Pr \{ \mathbf{Y}_N \in A \cap M_{ij}(\mathbf{m}) \} \\
&= \Pr \{ Y_i = m_1, Y_j = m_2, \varphi^{ij}(\mathbf{Y}_N) \in \varphi^{ij}(A \cap M_{ij}(\mathbf{m})) \} \\
&= \Pr \{ X_i = M_1, X_j = M_2, \varphi^{ij}(\mathbf{Y}_N) \in \varphi^{ij}(A \cap M_{ij}(\mathbf{m})) \} \\
&= \int_0^1 \int_{x_1}^1 \Pr \{ (m_2 - m_1)(X_1, \dots, X_{i-1}, X_{i+1}, \dots, X_{j-1}, X_{j+1}, \dots, X_N) \in \\
&\quad \{(x_2 - x_1)\{\varphi^{ij}(A \cap M_{ij}(\mathbf{m})) - m_1\} + x_1\} \} dx_2 dx_1 \\
&= \int_0^1 \int_{x_1}^1 \lambda \left\{ \frac{1}{m_2 - m_1} \left\{ (x_2 - x_1) \left\{ \varphi^{ij}(A \cap M_{ij}(\mathbf{m})) - m_1 \right\} + x_1 \right\} \right\} dx_2 dx_1 \\
&= \frac{1}{N(N-1)} \frac{\lambda(\varphi^{ij}(A \cap M_{ij}(\mathbf{m})))}{(m_2 - m_1)^{N-2}}.
\end{aligned} \tag{37}$$

Substituindo (37) em (36) obtemos

$$Q(A) = \frac{1}{N(N-1)} \sum_{\substack{i, j \in \{1, \dots, N\} \\ i \neq j}} \frac{\lambda(\varphi^{ij}(A \cap M_{ij}(\mathbf{m})))}{(m_2 - m_1)^{N-2}} = \tilde{Q}_{N\mathbf{m}}(A).$$

Concluindo-se assim a prova. $\square$.

Proposição 10: Se o vetor aleatório $\mathbf{Y}_N = (Y_1, \dots, Y_N)$ tem lei $Q_{N\mathbf{m}}$ ($N \geq 3$) então a lei de $\mathbf{Y}_n = (Y_1, \dots, Y_n)$ ($n \leq N$) é definida por

$$F_{m_1, m_2}(y_1, \dots, y_n) = \begin{cases} 0 & \begin{array}{l} \text{se } y_i < m_1 \text{ para } i \in I \subset \{1, \dots, n\} \\ \text{com } \#I \geq 1 \text{ ou} \\ \text{se } y_i = m_1 \text{ para } i \in I, \#I \geq 2 \end{array} \\ \frac{N-n}{N(N-1)} \prod_{\substack{\ell=1 \\ \ell \neq i}}^n \left(\frac{y_\ell - m_1}{m_2 - m_1} \right) & \begin{array}{l} \text{se } y_i = m_1 \text{ para algum} \\ i = 1, \dots, n \text{ e } m_1 \leq y_\ell < m_2 \\ \text{para } \ell = 1, \dots, n \text{ e } \ell \neq i \end{array} \end{cases}$$

$$F_{m_1, m_2}(y_1, \dots, y_n) = \begin{cases} \frac{(N-n)(N-n-1)}{N(N-1)} \prod_{i=1}^n \left(\frac{y_i - m_1}{m_2 - m_1} \right) & \text{se } m_1 < y_i < m_2 \\ + \sum_{\ell=1}^n \frac{N-n}{N(N-1)} \prod_{\substack{i=1 \\ i \neq \ell}}^n \left(\frac{y_i - m_1}{m_2 - m_1} \right) & \text{para } i = 1, \dots, n \\ \frac{(N-n)(N-n+2k-1) + k(k-1)}{N(N-1)} \prod_{i=1}^n \left(\frac{y_i - m_1}{m_2 - m_1} \right) & \text{se } y_i = m_2 \text{ para} \\ & i \in I_k \subset \{i_1, \dots, i_k\} \subset \{1, \dots, n\} \\ + \sum_{j \in J_k} \frac{N-n+k}{N(N-1)} \prod_{\substack{i \in I_k \\ i \neq j}} \left(\frac{y_i - m_1}{m_2 - m_1} \right) & \text{e } m_1 < y_i < m_2 \text{ se } i \in J = I_k^c \\ & \text{com } k = 1, \dots, n \\ 1 & \text{se } y_i \geq m_2 \text{ para } i = 1, \dots, n, \end{cases}$$

onde F_{m_1, m_2} é a função de distribuição de $\mathbf{Y}_N$.

Prova: Utilizando-nos da caracterização dada na Proposição 9 e fazendo $m_1 = 0$, $m_2 = 0$ temos em primeiro lugar que

$$\Pr \{Y_i < 0\} = \Pr \{X_i < \min\{X_1, \dots, X_N\}\} = 0$$

para qualquer $i \in \{1, \dots, n\}$.

Também para $i, j = 1, \dots, n$ e $i \neq j$,

$$\Pr \{Y_i = 0, Y_j = 0\} = \Pr \{X_i = M_{1N}, X_j = M_{1N}\} = 0,$$

já que $X_1, \dots, X_N$ são variáveis aleatórias contínuas.

Agora, se $0 \leq z_\ell < 1$ para $\ell \neq i$, temos

$$\begin{aligned} & \Pr \left\{ \frac{X_i - M_{1N}}{M_{2N} - M_{1N}} = 0, \left\{ \frac{X_\ell - M_{1N}}{M_{2N} - M_{1N}} \leq z_\ell; \ell = 1, \dots, n \text{ e } \ell \neq i \right\} \right\} \\ &= \Pr \left\{ M_{1N} = X_i, \left\{ \frac{X_\ell - M_{1N}}{M_{2N} - M_{1N}} \leq z_\ell; \ell = 1, \dots, n \text{ e } \ell \neq i \right\} \right\} \\ &= \sum_{j=1}^{N-n} \Pr \left\{ M_{1N} = X_i, M_{2N} = X_{n+j}, \left\{ \frac{X_\ell - M_{1N}}{M_{2N} - M_{1N}} \leq z_\ell; \ell = 1, \dots, n; \ell \neq i \right\} \right\}. \end{aligned} \tag{38}$$

Mas para $j = 1, \dots, N - n$

$$\begin{aligned}
& \Pr \left\{ M_{1N} = X_i, M_{2N} = X_{n+j}, \left\{ \frac{X_\ell - M_{1N}}{M_{2N} - M_{1N}} \leq z_\ell; \ell = 1, \dots, n \text{ e } \ell \neq i \right\} \right\} \\
&= \Pr \left\{ \{X_i \leq X_\ell \leq z_\ell(X_{n+j} - x_i) + X_i, \ell = 1, \dots, n, \ell \neq i\} \right. \\
&\quad \left. \text{e } \{X_i \leq X_{n+\ell} < X_{n+j}; \ell = 1, \dots, N - n \text{ e } \ell \neq j\} \right. \\
&\quad \left. \text{e } \{X_i < X_{n+j}\} \right\} \\
&= \int_0^1 \int_{x_1}^1 \Pr \left\{ \{x_1 \leq X_\ell \leq z_\ell(x_2 - x_1) + x_1, \ell = 1, \dots, n \text{ e } \ell \neq i\} \right. \\
&\quad \left. \text{e } \{x_1 \leq X_{n+\ell} < x_2, \ell = 1, \dots, n \text{ e } \ell \neq j\} \right\} dx_2 dx_1 \\
&= \frac{1}{N(N-1)} \prod_{\substack{\ell=1 \\ \ell \neq i}}^n z_\ell .
\end{aligned} \tag{39}$$

Substituindo (39) em (38) temos que

$$\Pr \left\{ \frac{X_i - M_{1N}}{M_{2N} - M_{1N}} = 0, \left\{ \frac{X_\ell - M_{1N}}{M_{2N} - M_{1N}} \leq z_\ell; \ell = 1, \dots, n, \ell \neq i \right\} \right\} = \frac{N-n}{N(N-1)} \prod_{\substack{\ell=1 \\ \ell \neq i}}^n z_\ell , \tag{40}$$

produzindo assim a segunda parte de F_{m_1, m_2} .

Agora, se $0 < z_\ell < 1$ para $\ell = 1, \dots, n$,

$$\begin{aligned}
& \Pr \left\{ \left\{ \frac{X_\ell - M_{1N}}{M_{2N} - M_{1N}} \leq z_\ell; \ell = 1, \dots, n \right\} \right\} \\
&= \sum_{i=1}^n \Pr \left\{ X_i = M_{1N}, \left\{ \frac{X_\ell - M_{1N}}{M_{2N} - M_{1N}} \leq z_\ell, \ell = 1, \dots, n \text{ e } \ell \neq i \right\} \right\} \\
&\quad + \Pr \left\{ \left\{ 0 < \frac{X_\ell - M_{1N}}{M_{2N} - M_{1N}} \leq z_\ell; \ell = 1, \dots, n \right\} \right\} \\
&= \sum_{i=1}^n \frac{N-n}{N(N-1)} \prod_{\substack{\ell=1 \\ \ell \neq i}}^n z_\ell + \Pr \left\{ \left\{ 0 < \frac{X_\ell - M_{1N}}{M_{2N} - M_{1N}} \leq z_\ell; \ell = 1, \dots, n \right\} \right\} ,
\end{aligned} \tag{41}$$

este último pela identidade (40).

Mas

$$\begin{aligned}
& \Pr \left\{ \left\{ 0 < \frac{X_\ell - M_{1N}}{M_{2N} - M_{1N}} \leq z_\ell; \ell = 1, \dots, n \right\} \right\} \\
&= \sum_{j,k \in \{1, \dots, N-n\}} \Pr \left\{ \left\{ 0 < \frac{X_\ell - M_{1N}}{M_{2N} - M_{1N}} \leq z_\ell; \ell = 1, \dots, n \right\} \text{ e } X_{n+j} = M_{1N}, X_{n+k} = M_{2N} \right\}
\end{aligned} \tag{42}$$

e

$$\begin{aligned}
& \Pr \left\{ \left\{ 0 < \frac{X_\ell - M_{1N}}{M_{2N} - M_{1N}} \leq z_\ell; \ell = 1, \dots, n \right\} \text{ e } X_{n+j} = M_{1N}, X_{n+k} = M_{2N} \right\} \\
&= \Pr \left\{ \left\{ X_{n+j} < X_\ell \leq z_\ell(X_{n+k} - X_{n+j}) + X_{n+j}, \ell = 1, \dots, n \right\} \right. \\
&\quad \text{e } \left\{ X_{n+j} \leq X_\ell \leq X_{n+k}, \ell = n+1, \dots, N \text{ e } \ell \neq n+j, n+k \right\} \\
&\quad \left. \text{e } \{X_{n+j} < X_{n+k}\} \right\} \\
&= \int_0^1 \int_{x_1}^1 \Pr \left\{ \left\{ x_1 \leq X_\ell \leq z_\ell(x_2 - x_1) + x_1; \ell = 1, \dots, n \right\} \right. \\
&\quad \left. \text{e } \{x_1 < X_\ell < x_2, \ell = n+1, \dots, N \text{ e } \ell \neq n+j, n+k\} \right\} dx_2 dx_1 \\
&= \frac{1}{N(N-1)} \prod_{i=1}^n z_i .
\end{aligned} \tag{43}$$

Substituindo (43) em (42) dá

$$\Pr \left\{ \left\{ 0 < \frac{X_\ell - M_{1N}}{M_{2N} - M_{1N}} \leq z_\ell; \ell = 1, \dots, n \right\} \right\} = \frac{(N-n)(N-n-1)}{N(N-1)} \prod_{i=1}^n z_i . \tag{44}$$

Substituindo (44) em (41), obtemos a terceira parte F_{m_1, m_2} . Para obtermos a quarta parte, consideremos as seguintes probabilidades: se $0 < z_\ell < 1, i, j \in \{1, \dots, n\}$

$$\begin{aligned}
& \Pr \left\{ \frac{X_i - M_{1N}}{M_{2N} - M_{1N}} = 0, \frac{X_j - M_{1N}}{M_{2N} - M_{1N}}, \left\{ \frac{X_\ell - M_{1N}}{M_{2N} - M_{1N}} \leq z_\ell; \ell = 1, \dots, n, \ell \neq i, j \right\} \right\} \\
&= \Pr \left\{ \left\{ \frac{X_\ell - X_i}{X_j - X_i} \leq z_\ell; \ell = 1, \dots, n, \ell \neq i, j \right\} \text{ e } \{X_i < X_\ell < X_j; \ell = 1, \dots, N, \ell \neq i, j\} \right\}
\end{aligned}$$

$$\begin{aligned}
& e \{X_i < X_j\} \\
& = \Pr \{ \{X_i \leq X_\ell \leq z_\ell(X_j - X_i) + X_i; \ell = 1, \dots, n \text{ e } \ell \neq i, j\} \\
& \quad e \{X_i < X_\ell < X_j; \ell = 1, \dots, N, \ell \neq i, j\} e \{X_i < X_j\} \} \\
& = \int_0^1 \int_{x_1}^1 \Pr \{ \{x_1 < X_\ell < z_\ell(x_2 - x_1) + x_1; \ell = 1, \dots, n \text{ e } \ell \neq i, j\} \\
& \quad e \{x_1 < X_\ell < x_2; \ell = n+1, \dots, N\} \} dx_2 dx_1 \\
& = \prod_{\substack{i=1 \\ \ell \neq i, j}}^n z_\ell \int_0^1 \int_{x_1}^1 (x_2 - x_1)^{N-2} dx_2 dx_1 = \frac{1}{N(N-1)} \prod_{\substack{\ell=1 \\ \ell \neq i, j}}^n z_\ell
\end{aligned} \tag{45}$$

e para $i \in \{1, \dots, n\}$,

$$\begin{aligned}
& \Pr \left\{ \frac{X_i - M_{1N}}{M_{2N} - M_{1N}} = 1, \left\{ 0 < \frac{X_\ell - M_{1N}}{M_{2N} - M_{1N}} \leq z_\ell; \ell = 1, \dots, n, \ell \neq i \right\} \right\} \\
& = \Pr \left\{ M_{2N} = X_i, \left\{ 0 < \frac{X_\ell - M_{1N}}{M_{2N} - M_{1N}} \leq z_\ell; \ell = 1, \dots, n \text{ e } \ell \neq i \right\} \right\} \\
& = \sum_{j=1}^{N-n} \Pr \left\{ M_{2N} = X_i, M_{1N} = X_{n+j} \text{ e } \left\{ 0 < \frac{X_\ell - M_{1N}}{M_{2N} - M_{1N}} \leq z_\ell; \ell = 1, \dots, n; \ell \neq i \right\} \right\} .
\end{aligned} \tag{46}$$

Mas para $j = 1, \dots, N - n$

$$\begin{aligned}
& \Pr \left\{ M_{2N} = X_i, M_{1N} = X_{n+j}; \left\{ \frac{X_\ell - M_{1N}}{M_{2N} - M_{1N}} \leq z_\ell; \ell = 1, \dots, n; \ell \neq i \right\} \right\} \\
& \Pr \{ \{X_{n+j} \leq X_\ell \leq z_\ell(X_i - X_{n+j}) + X_{n+j}; \ell = 1, \dots, n; \ell \neq i\} \\
& \quad e \{X_{n+j} \leq X_{n+\ell} \leq X_i; \ell = 1, \dots, N - n \text{ e } \ell \neq j\} \text{ e } \{X_{n+j} < X_i\} \} \\
& = \int_0^1 \int_{x_1}^1 \Pr \{ \{x_1 \leq X_\ell \leq z_\ell(x_2 - x_1) + x_1; \ell = 1, \dots, n \text{ e } \ell \neq i\} \\
& \quad e \{x_1 \leq X_{n+\ell} \leq x_2; \ell = 1, \dots, N - n, \ell \neq j\} \} dx_2 dx_1 \\
& = \int_0^1 \int_{x_1}^1 \prod_{\substack{\ell=1 \\ \ell \neq i, j}}^n z_\ell (x_2 - x_1)^{N-2} dx_2 dx_1 = \frac{1}{N(N-1)} \prod_{\substack{\ell=1 \\ \ell \neq i, j}}^n z_\ell .
\end{aligned} \tag{47}$$

Substituindo (47) em (46) obtemos

$$\Pr \left\{ \frac{X_i - M_{1N}}{M_{2N} - M_{1N}} = 1, \left\{ 0 < \frac{X_\ell - M_{1N}}{M_{2N} - M_{1N}} \leq z_\ell; \ell = 1, \dots, n; \ell \neq i \right\} \right\} = \frac{N-n}{N(N-1)} \prod_{\substack{\ell=1 \\ \ell \neq i, j}}^n z_\ell. \quad (48)$$

Também para $i \in \{1, \dots, n\}$ e $z_\ell < 1$, $\ell = 1, \dots, n$ e $\ell \neq i$

$$\begin{aligned} & \Pr \{Y_i = 1, \{Y_\ell \leq z_\ell; \ell = 1, \dots, n \text{ e } \ell \neq i\}\} \\ &= \Pr \{Y_i = 1, \{0 < Y_\ell \leq z_\ell; \ell = 1, \dots, n \text{ e } \ell \neq i\}\} \\ &+ \sum_{\substack{j=1 \\ j \neq i}}^n \Pr \{Y_i = 1, Y_j = 0, \{0 < Y_\ell < z_\ell; \ell = 1, \dots, n; \ell \neq i, j\}\} \\ &= \frac{N-n}{N(N-1)} \prod_{\substack{\ell=1 \\ \ell \neq i}}^n z_\ell + \sum_{\substack{j=1 \\ j \neq i}}^n \frac{1}{N(N-1)} \prod_{\substack{\ell=1 \\ \ell \neq i, j}}^n z_\ell. \end{aligned} \quad (49)$$

Esta última identidade decorre das igualdades (45) e (48).

Agora, se $I_k = \{i_1, i_2, \dots, i_k\} \subset \{1, \dots, n\}$ e $J_k = I_k^c$, temos a partir de (44) e (49) que

$$\begin{aligned} & \Pr \{\{Y_i \leq 1, i \in I_k\} \text{ e } \{Y_j \leq z_j, j \in J_k\}\} = \Pr \{\{Y_i < 1, i \in I_k\} \text{ e } \{Y_j \leq z_j, j \in J_k\}\} \\ &+ \sum_{i \in I_k} \Pr \{Y_i = 1, \{Y_\ell < 1, \ell \in I_k \text{ e } \ell \neq i\} \text{ e } \{Y_j < z_j, j \in J_k\}\} \\ &= \frac{(N-n)(N-n-1)}{N(N-1)} \prod_{j \in J_k} (z_j) + \sum_{j \in J_k} \frac{N-n}{N(N-1)} \prod_{\substack{\ell \in J \\ \ell \neq j}} z_\ell + \sum_{i \in I_k} \frac{N-n}{N(N-1)} \prod_{\ell \in J_k} z_\ell \\ &+ \sum_{i \in I_k} \sum_{\substack{j=1 \\ j \neq i}}^n \frac{1}{N(N-1)} \prod_{\substack{\ell \in J \\ \ell \neq j}} z_\ell + \sum_{i \in I_k} \frac{N-n}{N(N-1)} \prod_{\ell \in J_k} z_\ell. \end{aligned}$$

Arrumando os termos nesta identidade, obtemos a quarta parte de F_{m_1, m_2} , concluindo-se assim a prova. $\square$

Observação 5: Note que se na Proposição 10 deixamos n , m_1 e m_2 fixos e fazemos tender N para infinito, a função de distribuição F_{m_1, m_2} converge para a função de distribuição de n variáveis aleatórias independentes com distribuição comum uniforme sobre o intervalo

(m_1, m_2) . Como é usual neste capítulo, $Q_{N\mathbf{m}}$ denotará a distribuição n -dimensional de $Q_{N\mathbf{m}}$. Não é difícil ver que $Z = (\min\{Y_1, \dots, Y_n\}, \max\{Y_1, \dots, Y_n\})$ é suficiente para a família $\{Q_{N\mathbf{m}n} : m_1 < m_2, m_1, m_2 \in \mathbb{R}\}$. A proposição a seguir determina a lei de Z , acima definida.

Proposição 11: *Se o vetor aleatório $\mathbf{Y}_n = (Y_1, \dots, Y_n)$ tem lei $Q_{N\mathbf{m}n}$, então a lei de $Z = (\min\{Y_1, \dots, Y_n\}, \max\{Y_1, \dots, Y_n\})$ é definida por;*

$$F_{m_1, m_2}(z_1, z_2) = \begin{cases} 0 & \text{se } z_1 < m_1 \text{ ou } z_2 < m_1 \\ \frac{n(N-n)}{N(N-1)} \left(\frac{z_2 - m_1}{m_2 - m_1} \right)^{n-1} & \text{se } z_1 = m_1 \text{ e } m_1 < z_2 < m_2 \\ \frac{(N-n)(N-n-1)}{N(N-1)} \left\{ \left(\frac{z_2 - m_1}{m_2 - m_1} \right)^n - \left(\frac{z_2 - z_1}{m_2 - m_1} \right)^n \right\} + \frac{n(N-n)}{N(N-1)} \left(\frac{z_2 - m_1}{m_2 - m_1} \right)^{n-1} & \text{se } m_1 \leq z_1 < z_2 < m_2 \\ \frac{(N-n)(N-n-1)}{N(N-1)} \left(\frac{z_2 - m_1}{m_2 - m_1} \right)^n + \frac{n(N-n)}{N(N-1)} \left(\frac{z_2 - m_1}{m_2 - m_1} \right)^{n-1} & \text{se } z_1 \geq z_2, m_1 < z_1, z_2 < m_2 \\ \frac{(N-n)(N-n-1)}{N(N-1)} \left\{ 1 - \left(\frac{m_2 - z_1}{m_2 - m_1} \right)^n \right\} + \frac{n(N-n)}{N(N-1)} \left\{ 1 - \left(\frac{m_2 - z_1}{m_2 - m_1} \right)^{n-1} \right\} + \frac{n}{N} & \text{se } z_2 = m_2, m_1 < z_1 < m_2 \\ 1 & \text{se } z_1, z_2 \geq m_2. \end{cases}$$

Prova: A partir da Proposição 9, temos que é suficiente determinar a lei do vetor aleatório

$\mathbf{W} = (W_1, W_2)$, onde

$$W_1 = (m_2 - m_1) \left\{ \frac{\min\{X_1, \dots, X_n\} - M_{1N}}{M_{2N} - M_{1N}} \right\} + m_1$$

e

$$W_2 = (m_2 - m_1) \left\{ \frac{\max\{X_1, \dots, X_n\} - M_{1N}}{M_{2N} - M_{1N}} \right\} + m_1,$$

com $X_1, X_2, \dots, X_N$ são variáveis aleatórias independentes com distribuição comum $U(0, 1)$ e $M_{1N} = \min\{X_1, \dots, X_N\}$, $M_{2N} = \max\{X_1, \dots, X_N\}$. Sem perda de generalidade fazemos $m_1 = 0$, $m_2 = 1$. Notemos em primeiro lugar que a parte nula de F_{m_1, m_2} segue da definição de $\mathbf{W} = (W_1, W_2)$. Para obtermos o resto é suficiente determinar as probabilidades dos seguintes conjuntos,

$$A_1 = \{W_1 = 0, W_2 = 1\}$$

$$A_2 = \{W_1 \leq \omega_1, W_2 = 1\} ; 0 < \omega_1 < 1$$

$$A_3 = \{W_1 = 0, W_2 \leq \omega_2\} ; \omega_2 < 1$$

$$A_4 = \{0 < W_1 \leq \omega_1, 0 < W_2 \leq \omega_2\} ; 0 < \omega_1 \leq \omega_2 < 1 .$$

A figura a seguir descreve estes conjuntos,

Figura 3

A seguir determinaremos as probabilidades sobre estes conjuntos.

$$\begin{aligned}
\text{i) } \Pr \{A_1\} &= \Pr \{W_1 = 0, W_2 = 1\} \\
&= \Pr \{\min\{X_1, \dots, X_n\} = M_{1N}, \max\{X_1, \dots, X_n\} = M_{2N}\} \\
&= \sum_{\substack{i \neq j \\ i, j \in \{1, \dots, n\}}} \Pr \{M_{1N} = X_i, M_{2N} = X_j\} .
\end{aligned}
\tag{50}$$

Mas, para $i, j \in \{1, \dots, n\}$ e $i \neq j$,

$$\begin{aligned}
&\Pr \{M_{1N} = X_i, M_{2N} = X_j\} \\
&= \Pr \{\{X_i < X_\ell < X_j; \ell = 1, \dots, N \text{ e } \ell \neq i, j\} \text{ e } \{X_i \leq X_j\}\} \\
&= \int_0^1 \int_{x_1}^1 (x_2 - x_1)^{N-2} dx_2 dx_1 = \frac{1}{N(N-1)} .
\end{aligned}
\tag{51}$$

Substituindo (51) em (50) obtemos

$$\Pr \{W_1 = 0, W_2 = 1\} = \frac{n(n-1)}{N(N-1)} .
\tag{52}$$

ii) Se $0 < w_1 < 1$, então

$$\begin{aligned}
\Pr \{A_2\} &= \Pr \{W_1 \leq \omega_1, W_2 = 1\} \\
&= \Pr \{W_1 \leq \omega_1, \max_{1 \leq i \leq n} \{X_i\} = M_{2N}\} \\
&= \sum_{\ell=n+1}^N \Pr \{W_1 \leq \omega_1, \max_{1 \leq i \leq n} \{X_i\} = M_{2N}, M_{1N} = X_\ell\} .
\end{aligned}
\tag{53}$$

Mas, para $\ell \in \{n+1, \dots, N\}$,

$$\begin{aligned}
&\Pr \{W_1 \leq \omega_1, \max_{1 \leq i \leq n} \{X_i\} = M_{2N}, M_{1N} = X_\ell\} \\
&= \sum_{j=1}^n \Pr \{W_1 \leq \omega_1, M_{2N} = X_j, M_{1N} = X_\ell, \ell \in \{n+1, \dots, N\}\} .
\end{aligned}
\tag{54}$$

Agora para $\ell \in \{n+1, \dots, N\}$ e $j \in \{1, \dots, n\}$,

$$\begin{aligned}
&\Pr \{W_1 \leq \omega_1, M_{2N} = X_j, M_{1N} = X_\ell\} \\
&= \Pr \{\{\min_{1 \leq i \leq n} \{X_i\} \leq \omega_2(X_j - X_\ell) + X_\ell\} \text{ e } \{X_j < X_h < X_\ell; h = 1, \dots, N; h \neq j, \ell\} \\
&\quad \text{e } \{X_j < X_\ell\}\} .
\end{aligned}
\tag{55}$$

A probabilidade em (55) pode ser decomposta segundo mostra a figura abaixo

$i \neq j$ e $k = 1, \dots, n-1$.

Dando

$$\begin{aligned} \sum_{k=1}^{n-1} \binom{n-1}{k} \Pr \{ \{X_j < X_i < \omega_1(X_j - X_\ell) + X_j \text{ para } kX'_{is}, i \neq j\} \\ \text{e } \{ \omega_1(X_j - X_\ell) + X_j < X_i < X_\ell \text{ para } (n-1-k)X'_{is} \text{ } i \neq j\} \\ \text{e } \{X_j < X_h < X_\ell; h = n+1, \dots, N, h \neq \ell\} \text{ e } \{X_j < X_\ell\} \} , \end{aligned}$$

Mas as probabilidades nesta somatória são iguais a

$$\int_0^1 \int_{x_1}^1 \omega_1^2 (1 - \omega_1)^{n-k-1} (x_2 - x_1)^{n-1} (x_2 - x_1)^{N-n-1} dx_2 dx_1 = \frac{1}{N(N-1)} \omega_1^k (1 - \omega_1)^{n-k-1} .$$

Substituindo estas obtemos para $\ell \in \{n+1, \dots, N\}$ e $j \in \{1, \dots, n\}$

$$\begin{aligned} \Pr \{W_1 \leq \omega_1, M_{2N} = X_j, M_{1N} = X_\ell\} &= \sum_{k=1}^{n-1} \binom{n-1}{k} \omega_1^k (1 - \omega_1)^{n-k-1} \frac{1}{N(N-1)} \\ &= \frac{1}{N(N-1)} \{1 - (1 - \omega_1)^{n-1}\} . \end{aligned}$$

Substituindo agora em (54) com $\ell \in \{n+1, \dots, N\}$ temos que,

$$\Pr \{W_1 \leq \omega_1, \max_{1 \leq i \leq n} \{X_i\} = M_{2N}, M_{1N} = X_\ell\} = \frac{n}{N(N-1)} \{1 - (1 - \omega_1)^{n-1}\} .$$

Finalmente, substituindo este último termo em (53), obtemos

$$\Pr \{W_1 \leq \omega_1, W_2 = 1\} = \frac{n(N-n)}{N(N-1)} \{1 - (1 - \omega_1)^{n-1}\} . \quad (56)$$

iii) Se $0 < \omega_2 < 1$

$$\Pr \{A_3\} = \Pr \{ \min_{1 \leq i \leq n} \{X_i\} = M_{1N}, W_2 \leq \omega_2 \} = \sum_{j=1}^n \Pr \{X_j = M_{1N}, W_2 \leq \omega_2\} . \quad (57)$$

Mas para $j \in \{1, \dots, n\}$,

$$\begin{aligned}
& \Pr \{X_j = M_{1N}, W_2 \leq \omega_2\} \\
&= \Pr \{\{X_j = M_{1N}\} \text{ e } \{X_i \leq \omega_2(M_{2N} - M_{1N}) + M_{1N}; i = 1, \dots, n, i \neq j\}\} \\
&= \sum_{\ell=n+1}^N \Pr \{\{X_i \leq \omega_2(M_{2N} - M_{1N}) + M_{1N}; i = 1, \dots, n \text{ e } i \neq j\} \\
&\quad \text{e } \{X_j = M_{1N}, X_\ell = M_{2N}\}\} .
\end{aligned} \tag{58}$$

Por outro lado, se $\ell \in \{n+1, \dots, N\}$ e $j \in \{1, \dots, n\}$

$$\begin{aligned}
& \Pr \{\{X_i \leq \omega_2(M_{2N} - M_{1N}) + M_{1N}; i = 1, \dots, n \text{ e } i \neq j\} \text{ e } \{X_j = M_{1N}, X_\ell = M_{2N}\}\} \\
&= \int_0^1 \int_{x_1}^1 \omega_2^{\lambda-1} (x_2 - x_1)^{N-2} dx_2 dx_1 = \frac{1}{N(N-1)} W_2^{n-1} .
\end{aligned} \tag{59}$$

Substituindo (59) em (58), obtemos para $j \in \{1, \dots, n\}$,

$$\Pr \{X_j = M_{1N}, W_2 \leq \omega_2\} = \frac{N-n}{N(N-1)} \omega_2^{n-1} ,$$

e substituindo esta última igualdade em (57), obtemos

$$\Pr \{A_3\} = \frac{n(N-n)}{N(N-1)} \omega_2^{n-1} .$$

iv) Se $0 < \omega_1 \leq \omega_2 < 1$ temos que,

$$\begin{aligned}
\Pr \{A_4\} &= \Pr \{0 < W_1 \leq \omega_1, 0 < W_2 \leq \omega_2\} \\
&= \sum_{\substack{h, \ell \in \{n+1, \dots, N\} \\ \ell \neq h}} \Pr \{0 < W_1 \leq \omega_1, 0 < W_2 \leq \omega_2, M_1 = X_\ell, M_2 = X_h\}
\end{aligned} \tag{60}$$

Agora, para $\ell, h \in \{n+1, \dots, N\}$ e $\ell \neq h$,

$$\begin{aligned}
& \Pr \{0 < W_1 \leq \omega_1, 0 < W_2 \leq \omega_2, M_{1N} = X_\ell, M_{2N} = X_h\} \\
&= \Pr \{\{\min_{1 \leq i \leq n} \{X_i\} \leq \omega_1(X_h - X_\ell) + X_\ell \text{ e } \max_{1 \leq i \leq n} \{X_i\} \leq \omega_2(X_h - X_\ell) + X_\ell\} \\
&\quad \text{e } \{X_\ell \leq X_j \leq X_h; j = 1, \dots, N \text{ e } j \neq \ell, h\} \text{ e } \{X_\ell < X_h\}\}
\end{aligned}$$

$$\begin{aligned}
&= \Pr \left\{ \min_{1 \leq i \leq n} \{X_i\} \leq \omega_1(X_h - X_\ell) + X_\ell \right\} \text{ e } \{X_\ell < X_i < \omega_2(X_h - X_\ell) + X_\ell; i = 1, \dots, n\} \\
&\text{ e } \{X_\ell \leq X_j \leq X_h; j \in \{h+1, \dots, N\} \text{ e } j \neq \ell, h\} \text{ e } \{X_\ell < X_h\} \\
&= \sum_{k=1}^n \binom{n}{k} \Pr \left\{ \{X_\ell < X_i < \omega_1(X_h - X_\ell) + X_\ell \text{ para } kX'_{is}\} \text{ e } \right. \\
&\quad \left. \{\omega_1(X_h - X_\ell) + X_\ell < X_i < \omega_2(X_h - X_\ell) + X_\ell \text{ para } (n-k)X'_{is}\} \text{ e } \right. \\
&\quad \left. \{X_\ell \leq X_j < X_h; j \in \{n+1, \dots, N\} \text{ e } j \neq \ell, h\} \text{ e } \{X_\ell < X_h\} \right\} \\
&= \sum_{k=1}^n \binom{n}{k} \frac{\omega_1^k (\omega_2 - \omega_1)^{n-k}}{N(N-1)} = \frac{1}{N(N-1)} \{\omega_2^n - (\omega_2 - \omega_1)^n\}.
\end{aligned} \tag{61}$$

Finalmente, substituindo (61) em (60), obtemos

$$\Pr \{A_4\} = \frac{(N-n)(N-n-1)}{N(N-1)} \{\omega_2^n - (\omega_2 - \omega_1)^n\}.$$

Assim a segunda parte de F_{m_1, m_2} é obtida a partir da probabilidade em iii). A terceira parte é obtida somando-se a esta a probabilidade em iv). (58) e (61). A quarta parte é obtida desta última fazendo $\omega_1 = \omega_2$. Finalmente, a quinta parte é obtida acrescentando-se a esta, com $\omega_2 = m_2$, as identidades em (52) e (56). $\square$

A versão finita do teorema do tipo de Finetti envolvendo distribuições uniformes dependendo de dois parâmetros é uma consequência do teorema a seguir. Seja $P_{(\theta_1, \theta_2)}^n$ a lei de n variáveis aleatórias independentes com distribuição comum $U(\theta_1, \theta_2)$ e $Q_{N\mathbf{m}}$ como na identidade (35).

Teorema 11: *Se $Q_{N\mathbf{m}n}$ denota a lei n -dimensional de $Q_{N\mathbf{m}}$ e $P_{(m_1, m_2)}^n$ como foi definida acima, então para cada $2 \leq n \leq N$, $N \geq 3$, tem-se que*

$$\|Q_{N\mathbf{m}n} - P_{(m_1, m_2)}^n\| \leq 2 \frac{n(4N - n - 3)}{N(N-1)}. \tag{62}$$

Prova: Utilizando-nos como antes das propriedades da distância de variação total, temos que

$$\|Q_{N\mathbf{m}n} - P_{(m_1, m_2)}^n\| = \|Q_{N(0,1)n} - P_{(0,1)}^n\| .$$

Seja $\tilde{Q}_1$ a $Q_{N(0,1)n}$ -lei de $Z = (\min(Y_1, \dots, Y_n), \max(Y_1, \dots, Y_n))$ e P_1 a correspondente $P_{(0,1)}^n$ -lei. Utilizando o fato de que Z é uma estatística suficiente para as famílias $\{P_{(\theta_1, \theta_2)}^n : \theta_1 < \theta_2, \theta_1, \theta_2 \in \mathbb{R}\}$ e $\{Q_{N\mathbf{m}n} : m_1 < m_2, m_1, m_2 \in \mathbb{R}\}$ obtemos

$$\|Q_{N\mathbf{m}n} - P_{(m_1, m_2)}^n\| = \|\tilde{Q}_1 - \tilde{P}_1\| .$$

Esta última igualdade segue da propriedade 3 da distância de variação total. Agora, a $P_{(0,1)}^n$ -lei de Z é definida pela densidade

$$g(z_1, z_2) = \begin{cases} n(n-1)(z_2 - z_1)^{n-2} & \text{se } 0 < z_1 < z_2 < 1 \\ 0 & \text{em caso contrário} \end{cases}$$

e a $Q_{N(0,1)n}$ -lei de Z é definida por

$$\tilde{Q}_1(B) = \sum_{i=1}^4 \int_{B \cap B_i} d\tilde{Q}_1, \quad B \in \mathcal{B}_1,$$

onde

$$B_1 = \{0\} \times \{1\}, \quad B_2 = (0, 1] \times \{1\}, \quad B_3 = \{0\} \times [0, 1]$$

e

$$B_4 = \{(z_1, z_2) \in \mathbb{R}^2 : 0 < z_1 < z_2 < 1\} .$$

Portanto,

$$\begin{aligned} \|\tilde{Q}_1 - \tilde{P}_1\| &= 2 \sup_{B \in \mathcal{B}_1} \left| \int_B d\tilde{Q}_1 - \int_B d\tilde{P}_1 \right| \\ &= 2 \sup_{B \in \mathcal{B}_1} \left| \sum_{i=1}^4 \int_{B \cap B_i} d\tilde{Q}_1 - \int_{B \cap B_4} n(n-1)(z_2 - z_1)^{n-2} dz_2 dz_1 \right| . \end{aligned}$$

Mas, a partir da Proposição 11, decorre que a função de distribuição de $\tilde{Q}_1$ é contínua sobre B_4 , portanto,

$$\int_{B \cap B_4} d\tilde{Q}_1 = \frac{(N-n)(N-n-1)}{N(N-1)} \int_{B \cap B_4} n(n-1)(z_2 - z_1)^{n-2} dz_2 dz_1 .$$

Consequentemente

$$\begin{aligned} \|\tilde{Q}_1 - \tilde{P}_1\| &= 2 \sup_{B \in \tilde{B}_1} \left| \frac{2n}{N} \int_{B \cap B_4} n(n-1)(z_2 - z_1)^{n-2} dz_2 dz_1 - \sum_{i=1}^3 \int_{B \cap B_i} d\tilde{Q}_1 \right| \\ &\leq 2 \left\{ \frac{2n}{N} + \sum_{i=1}^3 \int_{B_i} d\tilde{Q}_1 \right\}. \end{aligned}$$

Utilizando novamente a Proposição 11, obtemos

$$\int_{B_1} d\tilde{Q}_1 = \frac{n(n-1)}{N(N-1)} \quad \text{e} \quad \int_{B_2} d\tilde{Q}_1 = \int_{B_3} d\tilde{Q}_1 = \frac{n(N-n)}{N(N-1)}.$$

Logo,

$$\|\tilde{Q}_1 - \tilde{P}_1\| \leq 2 \left\{ \frac{2n}{N} + \frac{n(n-1)}{N(N-1)} + \frac{2n(N-n)}{N(N-1)} \right\} = \frac{2n(4N-n-3)}{N(N-1)},$$

concluindo-se assim a prova. $\square$

Com este resultado a forma finita de um teorema do tipo de Finetti pode ser derivada na maneira usual.

Sejam $X_1, X_2, \dots, X_N$ variáveis aleatórias tomando valores em $\mathbb{R}$ e suponha que dado $M_{1N} = m_1$, $M_{2N} = m_2$ (com M_{1N} e M_{2N} definidos como na Proposição 9 e $m_1 < m_2$, $m_1, m_2 \in \mathbb{R}$), $\mathbf{X}_N = (X_1, \dots, X_N)$ tem distribuição uniforme sobre $\mathcal{X}_{m_1, m_2}^N$. Denotemos por P_N a lei de $\mathbf{X}_N$ e por P_{θ_1, θ_2}^n a lei de n variáveis aleatórias independentes com distribuição comum $U(\theta_1, \theta_2)$ e $P_{\mu_n} = \int P_{\theta_1, \theta_2}^n d\mu(\theta_1, \theta_2)$, para alguma medida de probabilidade μ concentrada em $S = \{(x, y) \in \mathbb{R}^2 : x < y\}$.

Teorema 12: *Se P_N e P_{μ_n} são as medidas de probabilidade acima definidas, então existe uma medida de probabilidade μ concentrada sobre S , tal que para cada $2 \leq n < N$,*

$$\|P_n - P_{\mu_n}\| \leq 2 \frac{n(4N-n-3)}{N(N-1)}, \quad (63)$$

com P_n a lei n -dimensional de P_N .

Prova: Basta escolher μ como a P_N -lei de $\mathbf{M}_N = (M_{1N}, M_{2N})$ e usar o Teorema 11. $\square$

A versão infinita deste teorema é derivada de maneira similar à dos teoremas 3, 7 e 10. Porém, a condição de tightness usualmente verificada requer de mais trabalhos e cuidados neste caso.

Teorema 13: *Seja $X_1, X_2, \dots$ seqüência infinita de variáveis aleatórias tomando valores em $\mathbb{R}$ e seja P a lei da seqüência. Se para cada $N \in \mathbb{N}$ ($N \geq 3$) e $m_1, m_2 \in \mathbb{R}$ ($m_1 < m_2$) a distribuição condicional de $X_1, X_2, \dots, X_N$ dado $M_{1N} = m_1, M_{2N} = m_2$ é uniforme sobre $\mathcal{X}_{m_1, m_2}^N$, então existe uma medida de probabilidade μ concentrada em S tal que para cada $n \in \mathbb{N}$*

$$\Pr \{ \mathbf{X}_n \in B \} = \int_S P_{\theta_1, \theta_2}^n(B) d\mu(\theta_1, \theta_2), \quad B \in \mathcal{B}_n,$$

onde P_{θ_1, θ_2}^n é a lei de n variáveis aleatórias independentes com distribuição comum $U(\theta_1, \theta_2)$.

Prova: Como nos casos anteriores, basta mostrar que a seqüência $\{\mu_N\}$ é tight, onde μ_N é a P -lei de $\mathbf{M}_N = (M_{1N}, M_{2N})$.

Seja F_{m_1, m_2} a função de distribuição de X_1 dado $M_{1N} = m_1, M_{2N} = m_2$, então fazendo $n = 1$ na Proposição 10, obtemos que

$$F_{m_1, m_2}(y) = \begin{cases} 0 & \text{se } y < m_1 \\ \frac{N-2}{N} \left(\frac{y-m_1}{m_2-m_1} \right) + \frac{1}{N} & \text{se } m_1 \leq y < m_2 \\ 1 & \text{se } y \geq m_2 \end{cases} \quad (64)$$

Agora,

$$\Pr \{|X_1| > K\} = \int_{-\infty < m_1 < m_2 < \infty} \Pr \{|X_1| > K/M_{1N} = m_1, M_{2N} = m_2\} d\mu_N(m_1, m_2).$$

Mas, $\Pr \{|X_1| > K\} \xrightarrow{K \rightarrow \infty} 0$, portanto, dado $\eta > 0$, $\exists A > 0$ tal que $\Pr \{|X_1| > K\} < \eta$ sempre que $K > A$.

Tomando $K_0 > A > 0$ e utilizando (64), temos que

$$\eta \geq \Pr \{|X_1| > K_0\} = \int_{-\infty < m_1 < m_2 < \infty} \Pr \{|X_1| > K_0/M_{1N} = m_1, M_{2N} = m_2\} d\mu_N(m_1, m_2)$$

$$\begin{aligned}
&= \int_{-\infty < m_1 < m_2 < \infty} \{1 - \Pr \{|X_1| \leq K_0/M_{1N} = m_1, M_{2N} = m_2\}\} d\mu_N(m_1, m_2) \\
&= \int_{\substack{m_2 < -K_0 \\ m_1 < m_2}} 1 d\mu_N(m_1, m_2) + \int_{\substack{m_1 > K_0 \\ m_1 < m_2}} 1 d\mu_N(m_1, m_2) \\
&\quad + \int_{\substack{m_2 = -K_0 \\ m_1 < m_2}} \left(1 - \frac{1}{N}\right) d\mu_N(m_1, m_2) + \int_{\substack{m_1 = K_0 \\ m_2 < m_1}} \left(1 - \frac{1}{N}\right) d\mu_N(m_1, m_2) \\
&\quad + \int_{m_1 < -K_0 < m_2 \leq K_0} \left\{ \frac{N-2}{N} \left(\frac{-K_0 - m_1}{m_2 - m_1} \right) \right\} + \frac{1}{N} d\mu_N(m_1, m_2) \\
&\quad + \int_{-K_0 \leq m_1 < K_0 < m_2} \left\{ 1 - \left\{ \frac{N-2}{N} \left(\frac{K_0 - m_1}{m_2 - m_1} \right) + \frac{1}{N} \right\} \right\} d\mu_N(m_1, m_2) \\
&\quad + \int_{\substack{m_1 < -K_0, m_2 > K_0 \\ m_1 < m_2}} \left\{ 1 - \frac{N-2}{N} \left(\frac{2K_0}{m_2 - m_1} \right) \right\} d\mu_N(m_1, m_2) \\
&\geq \underbrace{\int_{\substack{m_2 < -K_0 \text{ ou } m_1 > K_0 \\ m_1 < m_2}} 1 d\mu_N(m_1, m_2)}_{(1)} + \underbrace{\int_{\substack{m_2 = -K_0 \text{ ou } m_1 = K_0 \\ m_2 < m_1}} \left(1 - \frac{1}{N}\right) 1 d\mu_N(m_1, m_2)}_{(2)} \\
&\quad + \underbrace{\frac{N-2}{N} \int_{m_1 < -K_0 < m_2 \leq K_0} \left(\frac{-K_0 - m_1}{m_2 - m_1} \right) d\mu_N(m_1, m_2)}_{(3)} \\
&\quad + \underbrace{\int_{-K_0 \leq m_1 < K_0 < m_2} \left(1 - \left(\frac{K_0 - m_1}{m_2 - m_1} \right) - \frac{1}{N} \right) d\mu_N(m_1, m_2)}_{(4)} \\
&\quad + \underbrace{\int_{\substack{m_1 < -K_0, m_2 > K_0 \\ m_1 < m_2}} \left(1 - \frac{2K_0}{m_2 - m_1} \right) d\mu_N(m_1, m_2)}_{(5)} .
\end{aligned}$$

Agora

$$(2) = \int_{\substack{m_2 = -K_0 \text{ ou } m_1 = K_0 \\ m_2 < m_1}} \left(1 - \frac{1}{N}\right) d\mu_N(m_1, m_2) .$$

Escolhendo $\varepsilon_1 = 1/2$, temos que $\frac{1}{N} < \varepsilon_1$ se $N > 2$ e considerando o conjunto

$$A(\eta) = \{(m_1, m_2) \in \mathbb{R}^2 : m_2 = -K_0 \text{ ou } m_1 = K_0, m_1 \leq m_2\}, \text{ obtemos}$$

$$(2) \geq \frac{1}{2} \mu_N(A(\eta)) .$$

Por outro lado, se

$$B(\eta) = \{(m_1, m_2) \in \mathbb{R}^2 : m_2 < -K_0 \text{ ou } m_1 > K_0, m_1 \leq m_2\}$$

então

$$(1) = \mu_N(B(\eta)) .$$

Também

$$(3) = \frac{N-2}{N} \int_{m_1 < -K_0 < m_2 < K_0} \left[\frac{-K_0 - m_1}{m_2 - m_1} \right] d\mu_N(m_1, m_2) .$$

Seja $g(m_1, m_2) = \frac{-K_0 - m_1}{m_2 - m_1}$ se $m_1 < -K_0 < m_2 < K_0$, então para cada $m_2 \leq K_0$, $m_2 > m_1$, $g(m_1, m_2) \xrightarrow{m_1 \rightarrow -\infty} 1$, isto é, dado m_2 , $0 < \eta_1 < 1$, $\exists L_1 > 0$ tal que, $\frac{-K_0 - m_1}{m_2 - m_1} > 1 - \eta_1$ sempre que $m_1 < -L_1$. Em particular, se $m_1 < \frac{-K_0 - (1 - \eta_1)m_2}{\eta_1}$, $m_2 \leq K_0$ então $\frac{-K_0 - m_1}{m_2 - m_1} > 1 - \eta_1$.

$$\text{Seja } C(\eta, \eta_1) = \{(m_1, m_2) \in \mathbb{R}^2 : m_1 < \frac{-K_0 - (1 - \eta_1)m_2}{\eta_1}, -K_0 \leq m_2 < K_0, m_1 \leq m_2\},$$

então

$$\begin{aligned} (3) &\geq \frac{N-2}{N} \int_{C(\eta, \eta_1)} (1 - \eta_1) d\mu_N(m_1, m_2) \\ &= \frac{N-2}{N} (1 - \eta_1) \mu_N(C(\eta, \eta_1)) . \end{aligned}$$

Agora para $0 < \varepsilon_2 < 1$, existe $N(\varepsilon_1)$ tal que $\frac{N-2}{N} > 1 - \varepsilon_2$ se $N \geq N(\varepsilon_1)$. Escolhendo $\varepsilon_2 = 1/2$ temos que

$$(3) \geq \frac{1}{2} (1 - \eta_1) \mu_N(C(\eta, \eta_1)) \text{ se } N > 2 .$$

Também

$$(4) = \int_{-K_0 \leq m_1 < K_0 < m_2} \left\{ \left(\frac{m_2 - K_0}{m_2 - m_1} \right) - \frac{1}{N} \right\} d\mu_N(m_1, m_2) .$$

Seja agora $g(m_1, m_2) = \frac{m_2 - K_0}{m_2 - m_1}$ se $-K_0 \leq m_1 < K_0 < m_2$ então, para

$$m_1 > -K_0, g(m_1, m_2) \xrightarrow{m_2 \rightarrow -\infty} 1.$$

Assim, dado $m_1 \geq -K_0$, $0 < \eta_2 < 1/2$, existe $L_2 > 0$ tal que $\frac{m_2 - K_0}{m_2 - m_1} > 1 - \eta_2$ sempre que $m_2 > L_2$. Em particular, se $m_2 > \frac{K_0 - (1 - \eta_2)m_1}{\eta_2}$, $m_1 > m_2$, tem-se que $\frac{m_2 - K_0}{m_2 - m_1} > 1 - \eta_2$.

Seja

$$D(\eta, \eta_2) = \{(m_1, m_2) \in \mathbb{R}^2 : m_2 > \frac{K_0 - (1 - \eta_2)m_1}{\eta_2}, -K_0 \leq m_1 < K_0, m_1 \leq m_2\},$$

então

$$(4) \geq \int_{D(\eta, \eta_2)} (1 - \eta_2 - 1/N) d\mu_N(m_1, m_2).$$

Agora, para $0 < \varepsilon_3$, existe $N(\varepsilon_3)$ tal que $\frac{1}{N} < \varepsilon_3$ sempre que $N > N(\varepsilon_3)$.

Escolhendo $\varepsilon_3 = 1/2$ temos que

$$(4) \geq (1/2 - \eta_2)\mu_N(D(\eta, \eta_2)) \quad \text{se } N > 2.$$

Finalmente,

$$(5) = \int_{\substack{m_1 < -K_0, m_2 > K_0 \\ m_1 < m_2}} \left(1 - \frac{2K_0}{m_2 - m_1}\right) d\mu_N(m_1, m_2).$$

Seja agora $g(m_1, m_2) = 1 - \frac{2K_0}{m_2 - m_1}$, se $m_1 < -K_0$, $m_2 > K_0$, então

$$g(m_1, m_2) \xrightarrow{m_1 \rightarrow -\infty, m_2 \rightarrow \infty} 1.$$

Definindo

$$E(\eta, \eta_3) = \{(m_1, m_2) \in \mathbb{R}^2 : m_2 - m_1 > \frac{2K_0}{\eta_3}, m_1 < -K_0, m_2 > K_0\},$$

temos que

$$(5) \geq (1 - \eta_3)\mu_N(E(\eta, \eta_3)).$$

Daqui conclui-se que

$$\begin{aligned} \eta &\geq \frac{1}{2}\mu_N(A(\eta)) + \mu_N(B(\eta)) + \frac{1}{2}(1 - \eta_1)\mu_N(C(\eta, \eta_1)) \\ &\quad + \frac{1}{2}(1 - 2\eta_2)\mu_N(D(\eta, \eta_2)) + (1 - \eta_3)\mu_N(E(\eta, \eta_3)) \\ &\geq \frac{1}{2}\mu_N(A(\eta)) + \frac{1}{2}\mu_N(B(\eta)) + \frac{1}{2}(1 - \eta_1)\mu_N(C(\eta, \eta_1)) \\ &\quad + \frac{1}{2}(1 - 2\eta_2)\mu_N(D(\eta, \eta_2)) + (1 - \eta_3)\frac{1}{2}\mu_N(E(\eta, \eta_3)), \end{aligned}$$

se $N > 2$. Todos os termos à direita da desigualdade acima são positivos, implicando em $\mu_N(A(\eta)) \leq 2\eta$, $\mu(B(\eta)) \leq 2\eta$, $\mu_N(C(\eta_1, \eta)) \leq \frac{2\eta}{1-\eta}$, $\mu_N(D(\eta, \eta_2)) \leq \frac{2\eta}{1-2\eta_2}$, $\mu_N(E(\eta, \eta_3)) \leq \frac{2\eta}{1-\eta_3}$, se $N \geq 3$. Seja $\varepsilon = 2\eta(2 + \frac{1}{1-\eta_1} + \frac{1}{1-2\eta_2} + \frac{1}{1-\eta_3})$ e

$$A(\varepsilon) = A(\eta) \cup B(\eta) \cup C(\eta_1, \eta) \cup D(\eta, \eta_2) \cup E(\eta, \eta_3) ,$$

então as desigualdades acima implicam em

$$\mu_N(A(\varepsilon)) \leq \varepsilon \quad \text{se } N \geq 3 .$$

A figura a seguir mostra que $A(\varepsilon)$ é complementar de um compacto.

Figura 4: Gráfico de $A(\varepsilon)$

Assim, existe uma subsequência $\{\mu_{N_j}\}$ tal que $\mu_{N_j} \xrightarrow{j \rightarrow \infty} \mu$, com μ medida de probabilidade sobre S . Agora, tomando o limite através da subsequência e utilizando-se a desigualdade (63) do Teorema 12, conclui-se a prova. $\square$

Como tem sido feito nos outros casos mostraremos no que segue que seqüências de

variáveis aleatórias independentes com distribuição comum $U(\theta_1, \theta_2)$, $\theta_1 < \theta_2$ e $\theta_1, \theta_2 \in \mathbb{R}$, satisfazem as condições do Teorema 13.

Proposição 12: *Se $X_1, X_2, \dots, X_N$ são variáveis aleatórias independentes com distribuição comum $U(\theta_1, \theta_2)$, então a distribuição condicional de $X_1, X_2, \dots, X_N$ dado $M_{1N} = m_1$, $M_{2N} = m_2$ ($m_1 < m_2$, $m_1, m_2 \in \mathbb{R}$) é uniforme sobre $\mathcal{X}_{m_1, m_2}^N$.*

Prova: Denote por P_{θ_1, θ_2}^N a lei de $X_1, X_2, \dots, X_N$. Não é difícil ver que $\mathbf{M}_N = (M_{1N}, M_{2N})$ é suficiente e completo para a família $\{P_{\theta_1, \theta_2}^N : \theta_1 < \theta_2, \theta_1, \theta_2 \in \mathbb{R}\}$. Também, levando em consideração a Proposição 9 e utilizando-nos de um raciocínio análogo ao da Proposição 4, podemos concluir que o vetor aleatório

$$\mathbf{T}_N = \left(\frac{X_1 - M_{1N}}{M_{2N} - M_{1N}}, \dots, \frac{X_N - M_{1N}}{M_{2N} - M_{1N}} \right),$$

é uma estatística ancilar.

Consequentemente, pelos lemas de Basu (1955, 1958) $\mathbf{M}_N \perp\!\!\!\perp \mathbf{T}_N | \theta_1, \theta_2$.

Por outro lado,

$$\mathbf{X}_N = (X_1, X_2, \dots, X_N) = (M_{2N} - M_{1N})\mathbf{T}_N + M_{1N}.$$

Portanto, para f limitada, temos que

$$\begin{aligned} & E_{\theta}(f(\mathbf{X}_N) | M_{1N} = m_1, M_{2N} = m_2) = \\ & E_{\theta}(f((M_{1N} - M_{2N})\mathbf{T}_N + M_{1N}) | M_{1N} = m_1, M_{2N} = m_2) \\ & = E_{\mathbf{T}_N}(f((m_1 - m_2)\mathbf{T}_N + m_1)) \end{aligned}$$

Mas, pela Proposição 9, $(m_1 - m_2)\mathbf{T}_N + m_1$ têm distribuição uniforme sobre $\mathcal{X}_{m_1, m_2}^N$, concluindo-se a partir disto a prova. $\square$

Observação 6: A Proposição 12 permite-nos estabelecer o recíproco do teorema. Com efeito, se existe um vetor aleatório $\Theta = (\Theta_1, \Theta_2)$, tal que para cada $N \in \mathbb{N}$, $X_1, X_2, \dots$,

$X_N/\Theta_1 = \theta_1$, $\Theta_2 = \theta_2$ são independentes com distribuição comum $U(\theta_1, \theta_2)$, então

$$(X_1, \dots, X_N) \coprod \Theta / \mathbf{M}_N ,$$

pois $\mathbf{M}_N$ é suficiente para $X_1, \dots, X_N$ com relação a (θ_1, θ_2) .

Consequentemente

$$(X_1, \dots, X_N) / \Theta = (\theta_1, \theta_2) , \quad \mathbf{M}_N = (m_1, m_2) \stackrel{\mathcal{D}}{=} (X_1, \dots, X_N) / \mathbf{M}_N = (m_1, m_2)$$

Portanto, a partir da proposição prévia, conclui-se que $(X_1, \dots, X_N) / \mathbf{M}_N = (m_1, m_2)$ possui distribuição uniforme sobre $\mathcal{X}_{m_1, m_2}^N$.

CAPÍTULO 3

Abordagem Preditivista para Problemas de Inferência em Populações Finitas

3.1 Introdução

Na Seção 5 do Capítulo 1 foram discutidas diversas abordagens para inferências em populações finitas. O papel do planejamento amostral foi discutido dentro destes enfoques. Na teoria de amostragem clássica (modelo de população fixada) o planejamento amostral é a principal fonte de aleatorização. Contudo, o modelo induzido pelo planejamento não expressa relação alguma entre as unidades amostradas e não amostradas, a não ser pelo fato que estas últimas poderiam ter sido selecionadas com probabilidade positiva. Em outras palavras, o conhecimento a priori utilizado pelo pesquisador para escolher um determinado planejamento amostral não fica refletido no modelo probabilístico induzido por tal planejamento. Assim tal relação não parece ser suficiente para que as inferências sobre a parte desconhecida, guardem relação com a parte observada.

Na abordagem bayesiana, o conhecimento a priori que o pesquisador possui é uti-

lizado para expressar (probabilisticamente) incertezas a respeito das relações entre as unidades da população. Neste contexto, e de acordo com o princípio da verossimilhança, o planejamento amostral é irrelevante para a inferência sobre o vetor de características populacionais, após a realização da amostra.

A inferência bayesiana em populações finitas tem sido desenvolvida dentro do enfoque de superpopulações. Um dos trabalhos pioneiros que emprega esta visão é o de Ericson (1969), onde um modelo bayesiano baseado em prioris permutáveis é introduzido. A classe de prioris permutáveis considerada neste trabalho, são aquelas que podem ser representadas como misturas de modelos paramétricos. Como o julgamento de permutabilidade nem sempre é suficiente para a especificação completa do modelo, a especificação da distribuição a priori do vetor de características é completada através de dois estágios:

- i) descrição de relações entre elementos da população (modelo paramétrico na mistura);
- ii) expressão da incerteza a respeito dos parâmetros que são utilizados para descrever as relações de (i) (especificação da medida na mistura).

Do ponto de vista preditivista devido a de Finetti (1937,1938), tanto a priori quanto a verossimilhança são componentes igualmente relevantes da representação que expressa a opinião preditiva, não existindo a assimetria usualmente percebida entre elas. Na verdade a opinião preditiva se expressa como uma distribuição conjunta de todas as quantidades envolvidas no estudo. De Finetti (1938,1959) em quase todos os seus trabalhos considera que o objetivo primordial dos estatísticos é a predição de fatos desconhecidos, mas observáveis, dada a ocorrência de outros fatos, que eles julgam ser relacionados com aqueles de interesse.

Na abordagem bayesiana de superpopulação, a análise inicia-se sob esta perspectiva, através da consideração de prioris permutáveis (ou fortalecimentos, como por exemplo

permutabilidade parcial), no entanto em suas análises posteriores (estágios (i) e (ii) descritos anteriormente), fica implícito que o vetor aleatório de quantidades populacionais faz parte de uma sequência infinita de variáveis (ou vetores) aleatórios permutáveis, cuja lei quase sempre admite uma representação como mistura de modelos paramétricos.

Além disso, como foi discutido ao longo do capítulo 1, (seção 3) a medida da mistura expressa incerteza a respeito de um limite desconhecido, associado à sequência infinita de variáveis aleatórias permutáveis. Como veremos neste capítulo, no contexto de populações finitas a suposição de estendibilidade nem sempre é válida. Além disso o conhecimento a priori do pesquisador está mais associado a quantidades da própria população, do que a quantidades de populações infinitas. Tendo isto em vista, é razoável formular modelos probabilísticos que se utilizem do conhecimento a priori a respeito de quantidades da população finita envolvida no estudo. Tal ponto de vista foi utilizado recentemente por Barlow e Mendel (1992), Hayakawa (1992) no contexto de modelos de sobrevivência em populações finitas, com mais ênfase na sua construção do que propriamente em aspectos de inferência.

Na seção 2 discutimos as idéias acima através de um exemplo. Esta discussão está baseada essencialmente nas idéias apresentadas nos trabalhos de Cifarelli, Regazzini (1982) e Wechsler (1993). A sistematização apresentada na seção 3, é abordada pela primeira vez aqui. A análise de dados discretos em populações finitas é apresentada na seção 4, através de uma classe de modelos baseados em distribuições condicionalmente uniformes. As relações com a formulação bayesiana usual de superpopulação são discutidas através do conceito de estendibilidade. Nessa seção além dos aspectos formais e conexões com outras áreas da estatística, são introduzidos os resultados que estabelecem condições de estendibilidade para representação por modelos paramétricos conhecidos.

A seção conclui-se com algumas aplicações relativas à predição do total populacional e o valor máximo da característica estudada na população.

A seção 5 é dedicada ao estudo de modelos não discretos condicionalmente uniformes. A distribuição de Pareto exhibe propriedades interessantes enquanto distribuição a priori para quantidades populacionais de interesse aqui consideradas. Outros modelos em relação ao uso de variáveis auxiliares de distribuições condicionais não uniformes são comentados na seção 6.

3.2 Preditivismo em Populações Finitas

Nesta seção a população $\mathcal{P}$ de interesse consiste de N unidades identificadas por rótulos $1, 2, \dots, N$, com N conhecido. Associada à i -ésima unidade populacional existe uma quantidade (ou vetor de quantidades populacionais) desconhecida X_i . O objetivo é fazer inferências a respeito de uma quantidade (ou vetor de quantidades) desconhecida $\mathbf{T}_N = T(\mathbf{X}_N)$, onde $\mathbf{X}_N = (X_1, X_2, \dots, X_N)$. Com este propósito, uma amostra $s \subset \mathcal{P}$, de tamanho n ($n < N$) é selecionada.

O objetivo desta seção é descrever o processo de formulação de modelos probabilísticos que levam em consideração a informação a priori sobre as características da população finita $\mathcal{P}$. Como foi dito no início deste capítulo, no trabalho com populações finitas o pesquisador possui informação a priori a respeito da quantidade populacional de interesse e de variáveis auxiliares relacionadas. Esse conhecimento não está restrito apenas à especificação de uma distribuição a priori para $\mathbf{T}_N$.

Incertezas a respeito dos possíveis valores de $\mathbf{X}_N$ que podem gerar um determinado valor t de $\mathbf{T}_N$, é obviamente descrita por uma distribuição condicional $P(\cdot/\mathbf{T}_N = t)$, que depende de t . Por exemplo, pode ser do conhecimento do pesquisador que para um certo valor t de $\mathbf{T}_N$, cada um dos possíveis valores de $\mathbf{X}_N$ que geram t são igualmente prováveis, enquanto que para um outro valor t' de $\mathbf{T}_N$, algumas das configurações de $\mathbf{X}_N$ que geram t' , são mais prováveis que outras. Sob esta perspectiva, o modelo probabilístico

que descreve a incerteza a respeito de $\mathbf{X}_N$ é acessado em dois estágios:

- i) Uma distribuição a priori para $\mathbf{T}_N$;
- ii) Um modelo que expressa as relações entre os componentes de $\mathbf{X}_N$, dado um valor t de $\mathbf{T}_N$.

Notemos então que estará feita a especificação da distribuição conjunta de todas as quantidades envolvidas no estudo. Não há nesta abordagem quantidade alguma que não seja, ao menos conceitualmente observável (de Finetti, 1949). Neste sentido a abordagem proposta para modelos de população finita é preditivista. O exemplo a seguir ilustra estas idéias.

Exemplo 1: Considere uma população finita, $\mathcal{P} = \{1, 2, \dots, N\}$. Suponhamos que seja de interesse fazer inferência para o total de indivíduos que possuem uma determinada característica C . A quantidade de interesse é $T_N = \sum_{i=1}^N X_i$, onde X_i é o indicador da característica C no i -ésimo indivíduo da população.

Suponhamos a título de ilustração, que na opinião do pesquisador T_N possui distribuição uniforme em $\{0, 1, 2, \dots, N\}$.

Por outro lado $\mathcal{X}^N = \{0, 1\}^N$ é o espaço de valores possíveis para o vetor $\mathbf{X}_N = (X_1, X_2, \dots, X_N)$. Seja $\mathcal{X}_t^N = T^{-1}(\{t\})$ e $\Pi = \{\mathcal{X}_t^N : t = 0, 1, \dots, N\}$, isto é, Π é a partição sobre $\mathcal{X}^N$ induzida por T .

Suponhamos também que seja razoável admitir que os rótulos são não-informativos. Ou em outras palavras, que o rótulo i nada informa sobre a presença ou ausência da característica C no indivíduo i . Esta última suposição implica que

$$(X_1, \dots, X_N)/T_N = t \sim U\{\mathcal{X}_t^N\},$$

isto é,

$$P_r\{\mathbf{X}_N = \mathbf{x}_N / T_N = t\} = \binom{N}{t}^{-1} I_{\{t\}} \left(\sum_{i=1}^N x_i \right). \quad (1)$$

Com o objetivo de obter informações adicionais sobre T_N , uma amostra de tamanho n ($n < N$) é selecionada de $\mathcal{P}$ e os valores amostrais $\mathbf{x}_n = (x_{i_1}, \dots, x_{i_n})$ são observados, onde $s = \{i_1, \dots, i_n\}$ e o conjunto de rótulos dos indivíduos selecionados de $\mathcal{P}$.

Notando que T_N é uma função simétrica de X_N , e que a distribuição a priori de $\mathbf{X}_N$ é permutável, temos que a distribuição a posteriori de T_N não depende dos rótulos dos indivíduos selecionados de $\mathcal{P}$. Assim, sem perda de generalidade, podemos assumir que $s = \{1, \dots, n\}$. Levando isto em consideração, temos

$$P_r\{T = t / X_1 = x_1, \dots, X_n = x_n\} \propto \frac{\binom{N-n}{t-t_s}}{\binom{n}{t}} \cdot \frac{1}{N+1} I_{\{t_s, \dots, N-n+t_s\}}(t)$$

onde $t_s = \sum_{i=1}^n x_i$.

□

O exemplo evidencia vários aspectos que desejamos enfatizar:

- i) Em muitas situações experimentais que tratam com respostas binárias, a verossimilhança resultante de utilizar o modelo paramétrico Bernoulli, com repetições independentes e propensão comum θ , é usado como base para a análise de dados binários. Se, no exemplo anterior for possível mergulhar o vetor aleatório de quantidades populacionais em uma sequência infinita de variáveis aleatórias permutáveis, tal modelo é em virtude do teorema de Finetti (Teorema 3, Capítulo 1) uma consequência desta avaliação. Embora seja um modelo útil e conveniente, não é apropriado quando apenas um número finito de observações podem ser potencialmente realizadas.

Na abordagem bayesiana usual de superpopulação com este modelo, uma distribuição a priori para a propensão θ é especificada, onde θ representa o limite quase certo da sequência $\{T_N, N \geq 1\}$. As distribuições preditivas requeridas são obtidas posteriormente por integração. No entanto parece muito mais simples pensar a respeito da distribuição a priori das quantidades observáveis, digamos o número de indivíduos

na população que possuem a característica C , em lugar de dar atenção a uma distribuição a priori para θ , o qual perde o sentido quando tratamos com populações finitas.

- ii) A hipótese de permutabilidade introduzida através da identidade (1), vem da suposição a priori de que os indivíduos na população não sejam distinguíveis em relação à característica estudada. Isto como enfatiza Ericson (1969) é o que os pesquisadores têm em mente ao utilizarem a amostragem aleatória simples.

A seguir os aspectos formais desta abordagem serão estudados.

3.3 O Modelo de Predição

O objetivo desta seção é sistematizar o processo descrito na seção anterior. O modelo de predição implícito neste processo é estabelecido de maneira similar à formulação usual bayesiana. (Veja por exemplo Branco (1991), Paulino (1988) e Pereira (1980)).

Como é usual $\mathcal{X}$ denota o conjunto de valores possíveis da característica populacional estudada (que é um subconjunto de $\mathbb{R}^p$, para algum $p = 1, 2, \dots$). O conjunto de valores possíveis de $\mathbf{X}_N = (X_1, X_2, \dots, X_N)$ é $\mathcal{X}^N = \times_{i=1}^N \mathcal{X}$. Seja $\mathcal{A}$ a σ -álgebra de Borel sobre $\mathcal{X}$ e $\mathcal{A}^N$ a correspondente σ -álgebra produto sobre $\mathcal{X}^N$.

Seja $T : (\mathcal{X}^N, \mathcal{A}^N) \longrightarrow (\mathcal{Y}_N, \mathcal{B})$ uma função mensurável (a quantidade populacional de interesse) e $q(N, \cdot)$ uma medida de probabilidade (distribuição a priori para T) sobre $(\mathcal{Y}_N, \mathcal{B})$. Seja $Q_N(\cdot, \cdot) : \mathcal{Y}_N \times \mathcal{A}^N \longrightarrow [0, 1]$ uma função de transição de $(\mathcal{Y}_N, \mathcal{B})$ para $(\mathcal{X}^N, \mathcal{A}^N)$, isto é:

- i) Para cada $t \in \mathcal{Y}_N$, $Q_N(t, \cdot)$ é uma medida de probabilidade sobre $(\mathcal{X}^N, \mathcal{A}^N)$.
- ii) Para cada $A \in \mathcal{A}^N$, $Q_N(\cdot, A)$ é $\mathcal{B}$ -mensurável.

Como tem sido dito ao longo das seções anteriores a especificação de uma medida preditiva a priori é feita em dois estágios. No segundo estágio o espaço amostral $\mathcal{X}^N$ é particionado segundo T , e uma medida de probabilidade é definida sobre cada uma destas partes. A função implícita nesta construção deve ser uma função de transição, a fim de garantir que esta represente uma probabilidade condicional regular. A proposição a seguir estabelece este resultado.

Observação 1: No que segue não faremos distinção entre a função mensurável $T : (\mathcal{X}^N, \mathcal{A}^N) \rightarrow (\mathcal{Y}_N, \mathcal{B})$ e $\mathbf{T}_N = T_0 \mathbf{X}_N$.

Proposição 1: *Sejam $Q_N(\cdot, \cdot)$ e T as funções acima definidas e Q_N uma medida de probabilidade sobre $(\mathcal{X}^N, \mathcal{A}^N)$ definida por*

$$Q_N(\cdot) = \int_{\mathcal{Y}_N} Q_N(t, \cdot) dq(N, t)$$

Seja $C = \{t \in \mathcal{Y}_N : Q_N(t, T^{-1}(\{t\})) = 1\}$. Se $q(N, C) = 1$, então

- i) $q(N, \cdot)$ é a Q_N -lei de T , isto é, Q_N determina $q(N, \cdot)$.
- ii) $Q_N(\cdot, \cdot)$ é uma Q_N probabilidade condicional regular de $\mathbf{X}_N$ dado $\mathbf{T}_N$.

Prova: Para mostrar (i), notemos primeiro que para cada $A \in \mathcal{A}^N$ e $t \in C$, temos que

$$Q_N(t, A) = Q_N(t, A \cap T^{-1}(\{t\})) .$$

Por outro lado, se $B \in \mathcal{B}$,

$$T^{-1}(B) \cap T^{-1}(\{t\}) = \begin{cases} \emptyset & \text{se } t \notin B \\ T^{-1}(\{t\}) & \text{se } t \in B \end{cases}$$

A partir disto e da hipótese, concluímos que

$$Q_N(T^{-1}(B)) = \int_{\mathcal{Y}_N} Q_N(t, T^{-1}(B)) dq(N, t) = q(B).$$

A prova de (ii) segue da definição de probabilidade condicional regular (veja por exemplo, Ash, (1972), pp. 262-266). $\square$

Observação 2: As hipóteses na proposição 1, garantem uma correspondência biunívoca entre a medida preditiva a priori de $\mathbf{X}_N$ e a medida preditiva a priori $q(N, \cdot)$ de $\mathbf{T}_N$. Esta última conclusão, aparentemente óbvia, nem sempre é verdadeira quando medidas de probabilidade são definidas a partir de mistura. Isto é ilustrado através do exemplo (3), no Capítulo 1.

Com estes elementos, a informação a priori pode ser idealizada no trio $(\mathcal{X}^N, \mathcal{A}^N, Q_N)$, a partir do qual a distribuição condicional de $\mathbf{T}_N$ dado $\mathbf{X}_n = \mathbf{x}_n$ (com $X_n = (X_{i1}, \dots, X_{in})$, $\{i_1, \dots, i_n\} \subset \{1, \dots, N\}$) é obtida. Porém, a forma explícita de Q_N nem sempre é simples de se obter, pois a própria função de transição requerida para nosso estudo, nem sempre pode ser expressa explicitamente. Um exemplo disto são as distribuições uniformes construídas no Capítulo 2 (Seções 3 e 4). Estas foram definidas em termos abstratos e posteriormente caracterizadas através da distribuição de funções de variáveis aleatórias com distribuição conhecida. Tal caracterização permitiu determinar explicitamente as distribuições n -dimensionais, que é exatamente o que precisamos para determinar a distribuição a posteriori de $\mathbf{T}_N$ após a observação de uma amostra de tamanho n ($n < N$).

A proposição a seguir é útil para este propósito.

Proposição 2: Para cada $t \in \mathcal{Y}_N$, seja $Q_{Nn}(t, \cdot)$ uma distribuição n -dimensional de $Q_N(t, \cdot)$. Então $Q_{Nn}(\cdot, \cdot)$ é uma função de transição de $(\mathcal{Y}, \mathcal{B})$ para $(\mathcal{X}^n, \mathcal{A}^n)$.

Prova: Por hipótese para cada t fixo, $Q_{Nn}(t, \cdot)$ é uma medida de probabilidade sobre $\mathcal{A}^n$. Para mensurabilidade note que $Q_N(\cdot, A)$ é $\mathcal{B}$ -mensurável para cada $A \in \mathcal{A}^N$, em particular para $A = C \times \mathcal{X}^{N-n}$, $C \in \mathcal{A}^n$. $\square$

Tendo em conta estes resultados (de caráter técnico) o modelo de predição pode ser

estabelecido formalmente em analogia à formulação usual bayesiana.

Seja Π a medida de probabilidade sobre $(\mathcal{X}^n \times \mathcal{Y}_N, \mathcal{A}^n \times \mathcal{B})$ definida por,

$$\Pi(A \times B) = \int_B Q_{Nn}(t, A) dq(N, t), \quad A \in \mathcal{A}^n, \quad B \in \mathcal{B}$$

O espaço de probabilidade $(\mathcal{X}^n \times \mathcal{Y}_N, \mathcal{A}^n \times \mathcal{B}, \Pi)$ será denominado de **modelo de predição para $\mathbf{T}_N$** . Notemos a partir da definição de Π e da Proposição 2 que $Q_{Nn}(\cdot, \cdot)$ é uma Π -probabilidade condicional de $\mathbf{X}_n$ dado $\mathbf{T}_N$.

Também se existir uma função mensurável $\mathbf{F}_n : (\mathcal{X}^n, \mathcal{A}^n) \rightarrow (\mathcal{Y}_n, \mathcal{C})$ tal que $\mathbf{X}_n \amalg \mathbf{T}_N / \mathbf{F}_n$, diremos que $\mathbf{F}_n$ é **suficiente preditiva** para $\mathbf{T}_N$.

O modelo acima definido fornece o elemento básico para obtermos a nossa ferramenta de trabalho: a distribuição a posteriori de $\mathbf{T}_N$ dado $\mathbf{X}_n = \mathbf{x}_n$. Notemos que para tal propósito somente precisamos da distribuição n -dimensional de $Q_N(t, \cdot)$ para cada $t \in \mathcal{Y}_N$, e de $q(N, \cdot)$.

Dentro deste contexto, alguns modelos particulares serão abordados. A ênfase é dada aos modelos que são gerados a partir do julgamento de uma distribuição uniforme para $\mathbf{X}_N$ dado $\mathbf{T}_N = t$.

3.4 Modelos Discretos Localmente Uniformes

Aqui consideramos $\mathcal{X}$ como um subconjunto enumerável de $\mathbb{R}^p$ (algum $p = 1, 2, \dots$) e $\mathbf{T}_N$ como uma função simétrica de $\mathbf{X}_N$ tal que cada parte da partição induzida por $\mathbf{T}_N$ é finita. Os modelos sob consideração são aqueles que especificam distribuição uniforme sobre cada uma destas partes.

Assim se para cada $t \in \mathcal{Y}_N$, $Q_N(t, \cdot)$ denota a medida de probabilidade uniforme concentrada sobre $\mathbf{T}_N^{-1}\{t\}$, então

$$Q_N(t, (x_1, \dots, x_N)) = \frac{1}{N(t)} I_{\{\mathbf{T}_N^{-1}(t)\}}((x_1, \dots, x_N)) , \quad (2)$$

onde $N(t) = \#\{(x_1, \dots, x_N) \in \mathcal{X}^N : T(x_1, \dots, x_N) = t\}$.

Notemos que neste caso a função $Q_N(\cdot, \cdot)$ de $\mathcal{Y}_N \times \mathcal{A}^N$ em $[0, 1]$ definida a partir de (2) é uma função de transição, pois neste caso as σ -álgebras definidas na seção prévia correspondem às σ -álgebras de todas as partes.

Se $Q_N(\cdot) = \int_{\mathcal{Y}_N} Q_N(t, \cdot) dq(N, t)$ é a medida preditiva a priori definida a partir de (2), então, $Q_N(\cdot)$ é uma medida de probabilidade permutável, como mostra a proposição a seguir.

Proposição 3: Q_N é uma medida de probabilidade permutável.

Prova: $Q_N(t, (x_1, \dots, x_N)) = Q_N(t, (x_{\pi(1)}, \dots, x_{\pi(N)}))$ para todo $\pi \in S^N$, pois $\mathbf{T}_N$ é função simétrica de $\mathbf{X}_N$ (hipótese no começo da seção) e $Q_N(t, \cdot)$ é constante sobre $\mathbf{T}_N^{-1}\{t\}$ e zero fora dali. A permutabilidade de Q_N é direta a partir disto. $\square$

Consequentemente, o estudo nesta seção refere-se a medidas preditivas a priori permutáveis. Note, a partir da Proposição 2 que cada medida a priori $q(N, \cdot)$ determina uma e somente uma medida Q_N na classe de medidas de probabilidade permutáveis definidas a partir de misturas de (2).

Agora pela permutabilidade de $Q_N(t, \cdot)$ temos que a distribuição de qualquer n coordenadas é definida por

$$Q_{Nn}(t, (x_1, \dots, x_n)) = \frac{N(t, \mathbf{x}_n)}{N(t)} I_S((x_1, \dots, x_n)) , \quad (3)$$

onde $N(t, \mathbf{x}_n) = \#\{(x_{n+1}, \dots, x_N) \in \mathcal{X}^{N-n} : T(x_1, \dots, x_n, x_{n+1}, \dots, x_N) = t\}$ e $S = \{(z_1, \dots, z_n) \in \mathcal{X}^n : T(z_1, \dots, z_n, x_{n+1}, \dots, x_N) = t \text{ para algum } (x_{n+1}, \dots, x_N) \in \mathcal{X}^{N-n}\}$.

Consequentemente a distribuição a posteriori para $\mathbf{T}_n$ dado $\mathbf{X}_n = \mathbf{x}_n$ é definida por

$$\prod(t/\mathbf{x}_n) \propto \frac{N(t, \mathbf{x}_n)}{N(t)} q(N, t) I_{S^*}(t) ,$$

onde $S^* = \{t^* \in \mathcal{Y}_N : T(x_1, \dots, x_n, x_{n+1}, \dots, x_N) = t^* \text{ para algum } (x_{n+1}, \dots, x_N) \in \mathcal{X}^{N-n}\}$.

Exemplo 2: Seja $\mathbf{T}_N = \max\{X_1, \dots, X_N\}$ e $\mathcal{X} = \mathbb{N}_0$, então de acordo com a identidade 2, da seção 2 no Capítulo 2, temos que

$$N(t) = \#\{(x_1, \dots, x_N) \in \mathbb{N}_0^N : \max\{x_1, \dots, x_N\} = t\} = (t+1)^N - t^N$$

e

$$N(t, \mathbf{x}_n) = \begin{cases} (t+1)^{N-n} & \text{se } \max\{x_1, \dots, x_n\} = t \\ (t+1)^{N-n} - t^{N-n} & \text{se } \max\{x_1, \dots, x_n\} < t \end{cases}$$

Logo

$$\prod(t/\mathbf{x}_n)\alpha \begin{cases} \frac{(t+1)^{N-n}}{(t+1)^N - t^N} q(N, t) & \text{se } t = \max\{x_1, \dots, x_n\} \\ \frac{(t+1)^{N-n} - t^{N-n}}{(t+1)^N - t^N} \cdot q(N, t) & \text{se } t > \max\{x_1, \dots, x_n\} \text{ } t \in \mathbb{N}_0 \\ 0 & \text{em caso contrário.} \end{cases}$$

A seguir uma classe de quantidades populacionais é caracterizada em termos de suficiência para os modelos tratados nesta seção. Seja $\mathbf{T}_N = T(x_1, \dots, x_N)$ definida de $\mathcal{X}^N$ em $\mathcal{Y}_N$ a quantidade populacional de interesse e $\mathbf{T}_n = T(x_1, \dots, x_n)$ a restrição de T a $\mathcal{X}^n$, cuja imagem denotaremos por $\mathcal{Y}_n$.

Suponhamos que para cada $1 \leq n \leq N$, existe uma função φ_{Nn} ,

$$\varphi_{Nn} : \mathcal{Y}_n \times \mathcal{X}^{N-n} \rightarrow \mathcal{Y}_N$$

tal que $T(x_1, \dots, x_N) = \varphi_{Nn}(T(x_1, \dots, x_n), x_{n+1}, \dots, x_N)$

Proposição 4: *Sejam $X_1, X_2, \dots, X_N$ quantidades aleatórias tomando valores em $\mathcal{X}$, tal que a distribuição condicional de $X_1, X_2, \dots, X_N$ dado $\mathbf{T}_N = t$ é uniforme sobre $\mathbf{T}_N^{-1}(\{t\})$. Suponha que $\{\mathbf{T}_n : 1 \leq n \leq N\}$ satisfaz a condição acima mencionada, então $(X_1, X_2, \dots, X_n) \amalg \mathbf{T}_N / \mathbf{T}_n$.*

Prova: É suficiente mostrar que

$$\Pr \{X_1 = x_1, \dots, X_n = x_n / \mathbf{T}_N = t, \mathbf{T}_n = t_s\} = \Pr \{X_1 = x_1, \dots, X_n = x_n / \mathbf{T}_n = t_s\}$$

para cada $(x_1, \dots, x_n) \in \mathcal{X}^n$ e $(t_s, t) \in \mathcal{Y}_n \times \mathcal{Y}_N$ tal que $\Pr \{\mathbf{T}_N = t, \mathbf{T}_n = t_s\} > 0$.

Agora

$$\begin{aligned} \Pr \{X_1 = x_1, \dots, X_n = x_n / \mathbf{T}_N = t, \mathbf{T}_n = t_s\} &= \\ \frac{\Pr \{X_1 = x_1, \dots, X_n = x_n / \mathbf{T}_N = t\} \Pr \{\mathbf{T}_N = t\}}{\Pr \{\mathbf{T}_N = t, \mathbf{T}_n = t_s\}} I_{\{\mathbf{T}_N^{-1}(\{t_s\})\}}(x_1, \dots, x_n) &= \\ \frac{\frac{N(t, \mathbf{x}_n)}{N(t)} \cdot I_{\{\mathbf{T}_N^{-1}(\{t_s\})\}}(x_1, \dots, x_n)}{\Pr \{\mathbf{T}_n = t_s / \mathbf{T}_N = t\}} . \end{aligned}$$

Esta última igualdade decorre da hipótese e da identidade (3). Mas

$$\begin{aligned} \Pr \{\mathbf{T}_n = t_s / \mathbf{T}_N = t\} &= \sum_{(z_1, \dots, z_n) \in \mathbf{T}_N^{-1}(\{t_s\})} \Pr \{X_1 = z_1, \dots, X_n = z_n / \mathbf{T}_N = t\} \\ &= \sum_{(z_1, \dots, z_n) \in \mathbf{T}_N^{-1}(\{t_s\})} \frac{N(t, \mathbf{z}_n)}{N(t)} . \end{aligned}$$

Porém,

$$\begin{aligned} N(t, \mathbf{z}_n) &= \#\{(x_{n+1}, \dots, x_N) \in \mathcal{X}^{N-n} : T(z_1, \dots, z_n, x_{n+1}, \dots, x_N) = t\} \\ &= \#\{(x_{n+1}, \dots, x_N) \in \mathcal{X}^{N-n} : \varphi_{Nn}(t_s, x_{n+1}, \dots, x_N) = t\} \\ &= N(t, \mathbf{x}_n) , \end{aligned}$$

se $\mathbf{x}_n \in \mathbf{T}_N^{-1}(\{t_s\})$.

Logo

$$\Pr \{\mathbf{T}_n = t_s / \mathbf{T}_N = t\} = N(t_s) \frac{N(t, \mathbf{x}_n)}{N(t)} ,$$

onde, $N(t_s) = \#\{(x_1, \dots, x_n) \in \mathcal{X}^n : T(x_1, \dots, x_n) = t_s\}$.

Consequentemente

$$\Pr \{X_1 = x_1, \dots, X_n = x_n / \mathbf{T}_N = t, \mathbf{T}_n = t_s\} = \frac{1}{N(t_s)} I_{\{\mathbf{T}_N^{-1}(\{t_s\})\}}(x_1, \dots, x_n) .$$

Por outro lado

$$\Pr \{X_1 = x_1, \dots, X_n = x_n / \mathbf{T}_n = t_s\} = \frac{\Pr \{X_1 = x_1, \dots, X_n = x_n\}}{\Pr \{\mathbf{T}_n = t_s\}} I_{\{\mathbf{T}_n^{-1}(\{t_s\})\}}(x_1, \dots, x_n) .$$

Mas

$$\Pr \{X_1 = x_1, \dots, X_n = x_n\} = \sum_{t \in \mathcal{Y}_N} \frac{N(t, \mathbf{x}_n)}{N(t)} \Pr \{\mathbf{T}_N = t\} ,$$

e

$$\begin{aligned} \Pr \{\mathbf{T}_n = t_s\} &= \sum_{(z_1, \dots, z_n) \in \mathbf{T}_n^{-1}(\{t_s\})} \Pr \{X_1 = z_1, \dots, X_n = z_n\} \\ &= \sum_{(z_1, \dots, z_n) \in \mathbf{T}_n^{-1}(\{t_s\})} \sum_{t \in \mathcal{Y}_N} \Pr \{X_1 = z_1, \dots, X_n = z_n / \mathbf{T}_N = t\} P\{\mathbf{T}_N = t\} \\ &= \sum_{(z_1, \dots, z_n) \in \mathbf{T}_n^{-1}(\{t_s\})} \sum_{t \in \mathcal{Y}_N} \frac{N(t, \mathbf{z}_n)}{N(t)} \Pr \{\mathbf{T}_N = t\} , \end{aligned}$$

Agora, pela hipótese, $N(t, \mathbf{z}_n)$ é constante em $\mathbf{z}_n$, para cada $\mathbf{z}_n \in \mathbf{T}_n^{-1}(\{t_s\})$. Consequentemente

$$\Pr \{\mathbf{T}_n = t_s\} = N(t_s) \sum_{t \in \mathcal{Y}_N} \frac{N(t, \mathbf{x}_n)}{N(t)} \Pr(\mathbf{T}_N = t) .$$

Concluindo-se a partir disto, a prova. □

O resultado estabelecido na proposição acima, está estritamente relacionado aos conceitos de estatísticas preditivas e de S -estrutura associados a sequências infinitas de variáveis aleatórias, estudados por Lauritzen, (1974, 1988). Veja também, Freedman, (1962). O resultado estabelecido constitui uma adaptação destes conceitos, no contexto deste capítulo.

Observação 3: Notemos que, sob as condições acima mencionadas, o resultado garante que $\mathbf{T}_n$ é suficiente preditiva para $\mathbf{T}_N$. Além disso, da demonstração da proposição acima segue que a distribuição condicional de $X_1, \dots, X_n$ dado $\mathbf{T}_n = t_s$, é uniforme sobre $\mathbf{T}_n^{-1}(\{t_s\})$, para cada $1 \leq n \leq N$.

Exemplo 3: Seja $T_N = \max\{X_1, \dots, X_N\}$ e $\mathcal{X} = \mathbb{N}_0$. Então $\{T_n : 1 \leq n \leq N\}$, satisfaz as condições da Proposição 4. Com efeito se $n \leq N$,

$$\max\{x_1, \dots, x_N\} = \max\{\max\{x_1, \dots, x_n\}, x_{n+1}, \dots, x_N\}$$

Logo, se a distribuição condicional de $X_1, X_2, \dots, X_N$ dado $\max\{X_1, \dots, X_N\} = m$ é uniforme sobre $\mathcal{X}_m^N = \{(x_1, \dots, x_N) \in \mathbb{N}_0^N : \max\{x_1, \dots, x_n\} = m\}$, $m \in \mathbb{N}_0$, então

$$(X_1, \dots, X_n) \prod \max\{X_1, \dots, X_N\} / \max\{X_1, \dots, X_n\} .$$

E ainda, para cada $1 \leq n \leq N$ a distribuição condicional de $(X_1, \dots, X_n)$ dado $\max\{X_1, \dots, X_n\} = m_s$ é uniforme sobre $\mathcal{X}_{m_s}^n$. $\square$

O exemplo acima prova o resultado que foi utilizado para estabelecer a Proposição 4, apresentada no Capítulo 2.

Notemos que a especificação do modelo discreto em discussão nesta seção, depende essencialmente de resultados de análise combinatória. No entanto certos conceitos associados a modelos paramétricos permitem-nos encontrar a forma das identidades (2) e (3), sem ter que recorrer às técnicas de contagem. Um de tais conceitos, associado a modelos discretos em geral, é o de estatística de resumo (Barndoff (1973) e Lauritzen (1974)). A seguir, abordaremos esta noção e conceitos relacionados em conexão com os modelos discretos localmente uniformes.

3.4.1 Conexões com Estatísticas Suficientes e de Resumo.

Seja X uma quantidade aleatória sobre um espaço discreto E e T uma função de E em outro espaço discreto F .

Definição 1 : Uma medida de probabilidade P sobre E é dita resumida por T , se para cada $x, x' \in E$,

$$T(x) = T(x') \Rightarrow P(X = x) = P(X = x') .$$

Observação 4 : Com frequência falaremos que T é uma estatística de resumo para a variável aleatória X relativa a P , onde P é a lei de X .

O conceito pode ser estendido para uma família $\mathcal{P}$ de medidas de probabilidade.

Definição 2: T é dita resumir $\mathcal{P}$, se cada $P \in \mathcal{P}$ é resumida por T .

Exemplo 3: Seja $(\mathbb{N}^n, 2^{\mathbb{N}^n}, \mathcal{P})$, um modelo estatístico, onde $\mathcal{P} = \{P_\theta : 0 \leq \theta \leq 1\}$, com

$$P_\theta(x_1, \dots, x_n) = \theta^n (1 - \theta)^{\sum_{i=1}^n x_i} I_{\mathbb{N}^n}(x_1, \dots, x_n) .$$

Seja $T : \mathbb{N}^n \rightarrow \mathbb{N}$ definida por $T(x_1, \dots, x_n) = \sum_{i=1}^n x_i$, então a partir da definição de P_θ , é imediato que T é uma estatística de resumo para cada $P \in \mathcal{P}$.

Observação 5: O conceito de estatística de resumo é essencialmente restrito a espaços discretos.

Proposição 5: Se T resume P , então a distribuição de X dado $T = t$ é uniforme sobre $T^{-1}(\{t\})$.

Prova: Seguindo Lauritzen, (1974), temos que se $P(T = t) > 0$, então

$$P(X = x / T = t) = \frac{P(X = x, T = t)}{P(T = t)} = \frac{P(X = x)}{\sum_{z \in T^{-1}(\{t\})} P(X = z)} I_{(T^{-1}(\{t\}))}(x)$$

Se $T(x) = t$, temos que

$$P(X = z) = P(X = x) \text{ para cada } z \in T^{-1}(\{t\})$$

pois T resume P . Logo

$$P(T = t) = \sum_{z \in T^{-1}(\{t\})} P(X = z) = P(X = x) N(t)$$

onde $N(t)$ é o número de x'_s tal que $T(x) = t$ e $0 < N(t) < \infty$, pois $P(T = t) > 0$. Logo

$$P(X = x/T = t) = \frac{1}{N(t)} I_{(T^{-1}(\{t\}))}(x) .$$

□

A recíproca da proposição 5 também é verdadeira. Além disto, não é difícil ver a partir desta proposição que se T é uma estatística de resumo para uma família $\mathcal{P}$, então T é suficiente para $\mathcal{P}$. A recíproca nem sempre é verdadeira, como mostra o exemplo a seguir.

Exemplo 4: Considere o espaço estatístico $(\mathbb{N}_0^2, 2^{\mathbb{N}^2}, \mathcal{P})$ com $\mathcal{P} = \{P_\lambda : 0 < \lambda\}$, onde

$$P_\lambda(x_1, x_2) = \frac{\lambda^{x_1+x_2} e^{-\lambda}}{x_1! x_2!} I_{(\mathbb{N}^2)}(x_1, x_2) .$$

Aqui $T(x_1, x_2) = x_1 + x_2$ é uma estatística suficiente mínima para $\mathcal{P}$, porém T não resume $\mathcal{P}$. □

Bardonnff, Nielsen (1973), estudou condições sob as quais estes conceitos são equivalentes. A noção de família universal de medidas de probabilidade (sobre espaços discretos) é útil para tal propósito.

Definição 3: Uma família $\mathcal{P}$ de medidas de probabilidade é dita universal, se para cada $x \in E$, existe $P \in \mathcal{P}$, tal que

$$P(X = x) \geq P(X = y), \text{ para cada } y \in E .$$

Por exemplo, a família de distribuições binomial com o número n de ensaios fixo, é universal.

Proposição 6: Se $\mathcal{P}$ é universal e T é suficiente para $\mathcal{P}$, então T resume $\mathcal{P}$.

Prova: Seguindo Bardonnff, Nielsen (1973), sejam $x, y \in E$ tal que $T(x) = T(y)$, então

$$P(X = x) = P(T = t)P(X = x/T = t) \tag{4}$$

e

$$P(X = y) = P(T = t)P(X = y/T = t) . \quad (5)$$

Selecione agora $P \in \mathcal{P}$, tal que

$$P(X = x) \geq P(X = z) , \text{ para cada } z \in E \text{ (hipótese de Universalidade)} .$$

Então a partir das igualdades (4) e (5), temos que

$$P(X = x/T = t) \geq P(X = y/T = t) .$$

Pela suficiência de T , temos que $P(X = x/T = t)$ e $P(X = y/T = t)$ são constantes em P . Logo a desigualdade acima, é verdadeira para cada $P \in \mathcal{P}$.

Escolhendo agora, $P_1 \in \mathcal{P}$ tal que $P_1(X = y) \geq P_1(X = z)$, para cada $z \in E$, e combinando com as identidades (4) e (5), conclui-se que $P(X = x) = P(X = y)$, para cada $P \in \mathcal{P}$. $\square$

Exemplo 5: Considere o modelo de N ensaios Beurnoulli independentes com parâmetro θ , $0 < \theta < 1$, isto é $E = \{0, 1\}^N$ e

$$P_\theta(x_1, \dots, x_N) = \theta^{\sum_{i=1}^N x_i} (1 - \theta)^{N - \sum_{i=1}^N x_i} I_E(x_1, \dots, x_N) .$$

A família $\mathcal{P} = \{P_\theta : 0 < \theta < 1\}$ é universal e a estatística T definida por $T(x_1, \dots, x_N) = \sum_{i=1}^N x_i$ é suficiente. Consequentemente a distribuição condicional de $X_1, \dots, X_N \mid T = t \sim U\{T^{-1}(\{t\})\}$. $\square$

3.4.2 Conexões com a abordagem bayesiana usual de superpopulação: Estendibilidade

Ao longo desta seção, o conhecimento a priori sobre $X_1, X_2, \dots, X_N$ foi expresso probabilisticamente admitindo primeiro como razoável que a distribuição condicional de $X_1, X_2, \dots, X_N \mid T_N = t$ é uniforme sobre $\mathcal{X}_t = \{(x_1, \dots, x_N) \in \mathcal{X}^N : T(x_1, \dots, x_N) = t\}$,

onde T_N é uma função simétrica nos seus argumentos. A atribuição de uma distribuição a priori é completada após a especificação de uma distribuição para T_N . A distribuição a priori resultante é permutável.

Se bem nossa filosofia é fazer a especificação tendo em mente características de população finita, existem situações onde a distribuição a priori P cai dentro de uma classe de medidas de probabilidade permutáveis que admitem representação como mistura de modelos paramétricos conhecidos. Nestes casos a solução do problema de predição particular é tecnicamente equivalente à solução obtida utilizando-se da abordagem bayesiana de superpopulação. No que segue estudamos condições sobre a distribuição a priori da quantidade populacional de interesse para alguns dos modelos abordados nesta seção.

O Modelo de Bernoulli

Aqui $\mathcal{X} = \{0, 1\}$, $T_N = \sum_{i=1}^N X_i$ e a distribuição condicional de $X_1, X_2, \dots, X_N$ dado $T_N = t$ é uniforme sobre $\mathcal{X}_t = \{(x_1, \dots, x_N) : \sum_{i=1}^N x_i = t\}$. Neste caso pela Proposição 5 do Capítulo 1 a classe $\{Q_{Nt} : t \in \{0, 1, \dots, N\}\}$ são pontos extremos da classe das medidas de probabilidade permutáveis definidas sobre $\{0, 1\}^N$.

Proposição 7: *A seqüência finita de variáveis aleatórias permutáveis tomando valores em $\{0, 1\}$ é estendível se e somente se existe uma medida de probabilidade μ sobre $[0, 1]$ tal que*

$$q(N, t) = \int_{[0,1]} \binom{N}{t} \theta^t (1 - \theta)^{N-t} d\mu(\theta), \quad (6)$$

para cada $t = 0, 1, \dots, N$, onde $q(N, t) = P\{T_N = t\}$ com P a lei de $X_1, X_2, \dots, X_N$.

Prova: Denotemos por P_θ^N a lei de n variáveis aleatórias com distribuição comum $B(1, \theta)$, $\theta \in [0, 1]$. Suponhamos que existe μ medida de probabilidade tal que a iden-

tidade (6) vale, então

$$q(N, t) = \int_{[0,1]} P_{\theta}^N(T_N = t) d\mu(\theta).$$

Agora a partir do exemplo 4, temos que relativa a lei P_{θ}^N , T_N é uma estatística de resumo para $X_1, \dots, X_N$. Portanto

$$\begin{aligned} P\{X_1 = x_1, \dots, X_N = x_N\} &= P\{X_1 = x_1, \dots, X_N = x_N \mid T_N = t\}P(T_N = t) \\ &= \int_{[0,1]} P\{X_1 = x_1, \dots, X_N = x_N \mid T_N = t\}P_{\theta}^N(T_N = t)d\mu(\theta) \\ &= \int_{[0,1]} P_{\theta}^N\{X_1 = x_1, \dots, X_N = x_N \mid T_N = t\}P_{\theta}^N(T_N = t)d\mu(\theta) \\ &= \int_{[0,1]} P_{\theta}^N\{X_1 = x_1, \dots, X_N = x_N\}I_{\{t\}}\left(\sum_{i=1}^n x_i\right)d\mu(\theta) \end{aligned}$$

Logo

$$P\{X_1 = x_1, \dots, X_N = x_N\} = \int_{[0,1]} \theta^t(1 - \theta)^{N-t}d\mu(\theta),$$

onde $t = \sum_{i=1}^n x_i$.

Tomando agora $Z_1, Z_2, \dots$ sequência infinita de variáveis aleatórias permutáveis assumindo valores em $\{0, 1\}$ tal que

$$P\{Z_1 = z_1, \dots, Z_N = z_N\} = \int_{[0,1]} \theta^t(1 - \theta)^{N-t}d\mu(\theta) ,$$

para cada $N \in \mathbb{N}$, $t = \sum_{i=1}^N z_i$, temos que

$$(X_1, \dots, X_N) \stackrel{\mathcal{D}}{=} (Z_1, Z_2, \dots, Z_N) ,$$

e portanto a sequência é estendível.

Reciprocamente, suponhamos que a sequência é estendível, então existe uma sequência infinita de variáveis aleatórias permutáveis $Z_1, Z_2, \dots$ tomando valores em $\{0, 1\}$ tal que

$$(X_1, \dots, X_N) \stackrel{\mathcal{D}}{=} (Z_1, Z_2, \dots, Z_N) .$$

Mas pelo teorema de Finetti (Teorema 3 do Capítulo 1) existe uma medida μ sobre $[0, 1]$ tal que

$$P\{Z_1 = z_1, \dots, Z_N = z_N\} = \int_{[0,1]} \theta^t (1 - \theta)^{N-t} d\mu(\theta)$$

para cada $N \in \mathbb{N}$, $t = \sum_{i=1}^N z_i$, $z_i \in \{0, 1\}$, $i = 1, \dots, N$. Portanto

$$P\{X_1 = x_1, \dots, X_N = x_N\} = \int_{[0,1]} \theta^t (1 - \theta)^{N-t} d\mu(\theta),$$

com $t = \sum_{i=1}^N x_i$.

Consequentemente

$$\begin{aligned} P\{T_N = t\} &= \sum_{(x_1, \dots, x_N) \in T_N^{-1}\{t\}} P\{X_1 = x_1, \dots, X_N = x_N\} \\ &= \sum_{(x_1, \dots, x_N) \in T_N^{-1}\{t\}} \int_{[0,1]} \theta^t (1 - \theta)^{N-t} d\mu(\theta) \\ &= \int \binom{N}{t} \theta^t (1 - \theta)^{N-t} d\mu(\theta). \end{aligned} \quad \square$$

O aspecto importante da proposição anterior é que a distribuição de T_N dado $\mathbf{X}_n = \mathbf{x}_n$, derivada a partir da suposição que $X_1, X_2, \dots, X_N/\theta$ são condicionalmente independentes com distribuição comum $B(1, \theta)$, coincide com a obtida a partir da suposição que $X_1, \dots, X_N|T_N = t \sim U\{\mathcal{X}_t\}$ quando a distribuição a priori para T_N satisfaz a identidade (6).

Por exemplo: se a distribuição a priori para T_N é definida por

$$q(N, t) = \frac{1}{N+1} \frac{\beta(a+t, b+n-t)}{\beta(t+1, N-t+1)B(a, b)} I_{\{0,1,\dots,N\}}(t),$$

isto é, T_N tem distribuição beta binomial, então escolhendo μ definida pela densidade $\text{Beta}(a, b)$, temos que identidade (6) é satisfeita.

Em tais situações os resultados não constituem novidade, a não ser pelo fato que a especificação de uma distribuição a priori é feita levando em consideração características de população finita. Tal distribuição permite-nos as vezes, por razões técnicas, resolver

o problema de predição utilizando-nos dos resultados usuais da abordagem bayesiana de superpopulação. Por exemplo a forma do preditor que minimiza o risco sob função de perda quadrática, derivada por Ericson (1969). Contudo dois pontos importantes devem ser enfatizados:

- i) Existe um número infinito de distribuições a priori para T_N que fazem com que a sequência finita $X_1, \dots, X_N$ não seja estendível. Em outras palavras, sequências finitas para as quais não existe uma medida que satisfaz identidade (6).
- ii) A menos dos casos conhecidos não é uma questão simples encontrar a medida μ que satisfaz identidade (6), se esta existir.

Nosso propósito no que segue é estabelecer condições sobre $q(N, \cdot)$ de modo que exista uma medida μ que satisfaça identidade (6). Antes disto um simples resultado decorrente da proposição 7 será apresentado.

Corolário 1: *Sejam $X_1, X_2, \dots, X_N$ variáveis aleatórias permutáveis tomando valores em $\{0, 1\}$. Seja $q(N, t) = P_r\{T_N = t\}$. Se existe $t_1 \in \{0, 1, \dots, N\}$ tal que $q(N, t_1) = 0$ e $0 < q(N, t) < 1$ para algum $t \in \{0, 1, \dots, N\}$ com $t \neq t_1$, então a sequência finita de variáveis aleatórias $X_1, X_2, \dots, X_N$ não é estendível.*

Prova: Suponha por absurdo que existe uma medida μ sobre $[0, 1]$ tal que

$$q(N, t) = \int_{[0,1]} \binom{N}{t} \theta^t (1 - \theta)^{N-t} d\mu(\theta), \quad t = 0, 1, \dots, N$$

Então, se $t = t_1$

$$0 = \int_{[0,1]} \binom{N}{t_1} \theta^{t_1} (1 - \theta)^{N-t_1} d\mu(\theta)$$

implicando em $\theta = 0$ ou $\theta = 1$. Se $\theta = 0$, então $q(N, 0) = 1$ e $q(N, t) = 0$ para cada $t \in \{1, \dots, N\}$. Se $\theta = 1$, então $q(N, N) = 1$ e $q(N, t) = 0$ para cada $t \in \{0, \dots, N-1\}$ em contradição com a hipótese. $\square$

O corolário acima exhibe uma classe infinita de medidas de probabilidade a priori para T_N que fazem com que a sequência finita de variáveis aleatórias $X_1, X_2, \dots, X_N$ sob consideração não seja estendível. Notemos que se $q(N, 0) = 1$ ($q(N, N) = 1$) então a sequência finita é determinística, isto é, $X_n \equiv 0$ para cada $n = 1, \dots, N$ ($X_n \equiv 1$ para cada $n = 1, \dots, N$), em cujo caso é trivialmente estendível, com μ medida de probabilidade degenerada em zero (degenerada em 1). Assim, se excluímos os casos triviais, as únicas medidas a priori $q(N, \cdot)$ candidatas a satisfazer identidade (6), para alguma medida de probabilidade μ definida sobre $[0, 1]$ são aquelas para as quais, $0 < q(N, t) < 1$ para cada $t \in \{0, 1, \dots, N\}$.

Os resultados a seguir serão de utilidade para estudar estendibilidade nestes casos.

Lema 1: *Seja $A = (a_{ij})$ matriz de ordem N definida por*

$$a_{ij} = \begin{cases} \binom{i-1}{i-j} (-1)^{i+j} & \text{se } i \geq j \\ 0 & \text{se } i < j \end{cases} \quad i, j = 1, \dots, N.$$

Então $A^{-1} = B = (b_{ij})$ é definida por

$$b_{ij} = \begin{cases} \binom{i-1}{i-j} & \text{se } i \geq j \\ 0 & \text{se } i < j \end{cases} \quad i, j = 1, \dots, N.$$

Prova: Seja $C = A \cdot B$ com $C = (c_{ik})$, onde $c_{ik} = \sum_{j=1}^N a_{ij} b_{jk}$. Então pela definição de A e B temos que

$$c_{ik} = \begin{cases} \sum_{j=k}^i a_{ij} b_{jk} & \text{se } k \leq i \\ 0 & \text{se } k > i \end{cases}.$$

Agora, se $k = i$, então $c_{ii} = 1$ e se $k < i$, temos que

$$\begin{aligned} c_{ik} &= \sum_{j=k}^i a_{ij} b_{jk} = \sum_{\ell=0}^{i-k} a_{i, \ell+k} b_{\ell+k, k} \\ &= \sum_{\ell=0}^{i-k} \binom{i-1}{i-\ell-k} \binom{\ell+k-1}{\ell} (-1)^{\ell+k+i} \end{aligned}$$

Mas, $\binom{i-1}{i-\ell-k} \binom{\ell+k-1}{\ell} = \binom{i-1}{k-1} \binom{i-k}{\ell}$. Logo

$$c_{ik} = \binom{i-1}{k-1} (-1)^{k+i} \sum_{\ell=0}^{i-k} \binom{i-k}{\ell} (-1)^\ell = 0,$$

concluindo-se assim a prova. $\square$

Corolário 2: *Sejam $X_1, X_2, \dots, X_N$ variáveis aleatórias permutáveis tomando valores em $\{0, 1\}$ e seja*

$$c_k = \sum_{t=k}^N \frac{\binom{N-k}{t-k}}{\binom{N}{t}} q(N, t), \quad k = 1, \dots, N,$$

onde $q(N, t) = P(T_N = t)$, com P a lei de $X_1, \dots, X_N$. Então a sequência finita $X_1, X_2, \dots, X_N$ é estendível se, e somente se, a sequência de números $\{c_k\}_{k=1}^N$ correspondem a N momentos de uma distribuição de probabilidade concentrada sobre $[0, 1]$.

Prova: Se a sequência finita $X_1, \dots, X_N$ é estendível, então para cada $k = 1, \dots, N$

$$\Pr \{X_1 = 1, \dots, X_k = 1\} = \int_{[0,1]} \theta^k d\mu(\theta),$$

para alguma medida de probabilidade μ definida sobre $[0, 1]$. Seja θ uma variável aleatória com lei μ , então

$$\begin{aligned} E(\theta^k) &= \Pr \{X_1 = 1, \dots, X_k = 1\} \\ &= \sum_{t=k}^N \frac{\binom{N-t}{t-k}}{\binom{N}{t}} q(N, t), \quad k = 1, \dots, N. \end{aligned}$$

Reciprocamente, suponhamos que $\{c_k\}_{k=1}^N$ correspondem aos N primeiros momentos de alguma distribuição de probabilidade definida sobre $[0, 1]$, e, seja θ uma variável aleatória tomando valores em $[0, 1]$ com tal distribuição. Então

$$E(\theta^k) = \sum_{t=k}^N \frac{\binom{N-t}{t-k}}{\binom{N}{t}} q(N, t), \quad k = 1, \dots, N.$$

Considere os vetores linha,

$$\begin{aligned}\mathbf{E}_N &= (E(\theta^N), E(\theta^{N-1}), \dots, E(\theta)), \\ \mathbf{a}_N &= (a(N, N), a(N, N-1), \dots, a(N, 1)), \text{ onde} \\ a(N, t) &= \frac{q(N, t)}{\binom{N}{t}}, \quad t = 1, \dots, N.\end{aligned}$$

E seja $B = (b_{ij})$ matriz de ordem N definida por

$$b_{ij} = \begin{cases} \binom{i-1}{i-j} (-1)^{i+j} & \text{se } i \geq j \\ 0 & \text{se } i < j \end{cases} \quad i, j = 1, \dots, N.$$

Não é difícil ver que o sistema de equações acima pode ser reescrito na forma $B\mathbf{a}_N = \mathbf{E}_N$.

Logo, a partir do lema 1, temos que o sistema tem solução única dada por $\mathbf{a}_N = A\mathbf{E}_N$.

Consequentemente,

$$\begin{aligned}a(N, t) &= \sum_{j=1}^{N-t+1} \binom{N-t}{N-t+1-j} (-1)^{N-t+j+1} E(\theta^{N+1-j}) \\ &= \sum_{j=1}^{N-t+1} \binom{N-t}{j-1} (-1)^{N-t+j+1} E(\theta^{N+1-j})\end{aligned}$$

Fazendo $\ell = j - 1$, obtemos

$$\begin{aligned}a(N, t) &= \sum_{\ell=0}^{N-t} \binom{N-t}{\ell} (-1)^{N-t-\ell} E(\theta^{N-\ell}) \\ &= \sum_{\ell=0}^{N-t} \binom{N-t}{\ell} (-1)^{N-t-\ell} E(\theta^{N-\ell}) \\ &= \int_{(0,1)} \sum_{\ell=0}^{N-t} \binom{N-t}{\ell} (-1)^{N-t-\ell} \theta^{N-\ell} d\mu(\theta)\end{aligned}$$

onde μ é a lei de θ , na hipótese.

Fazendo algumas operações algébricas obtemos

$$q(N, t) = \int_{[0,1]} \binom{N}{t} \theta^t (1 - \theta)^{N-t} d\mu(\theta), \quad t = 1, \dots, N.$$

Finalmente,

$$\begin{aligned}q(N, 0) &= 1 - \sum_{t=1}^N q(N, t) = \int_{[0,1]} \left\{ 1 - \sum_{t=1}^N \binom{N}{t} \theta^t (1 - \theta)^{N-t} \right\} d\mu(\theta) \\ &= \int_{[0,1]} (1 - \theta)^N d\mu(\theta),\end{aligned}$$

concluindo-se assim a prova. $\square$

O problema agora é determinar condições para que uma sequência finita de números correspondam a N momentos de uma distribuição de probabilidade definida sobre $[0, 1]$. Este problema faz parte da teoria clássica de momentos (Shohat e Tamarkin (1943)) iniciada por Stieljes no fim do século passado e que se encaixa dentro dos denominados problemas de momentos reduzidos formulados por P. Tchebycheff e resolvidos por A. Markoff. Os resultados têm sido apresentados em Shohat e Tamarkin (1943). Estes autores consideraram o seguinte problema:

(MR): Dadas $N + 1$ constantes $\mu_0, \mu_1, \dots, \mu_N$, determinar condições para que

$$\int_{-1}^1 t^k d\psi(t) = \mu_k, \quad k = 0, 1, \dots, N$$

admita uma solução em ψ dentro da classe de funções de distribuição.

Mostrando que uma condição necessária e suficiente para que o problema (MR) tenha solução é que quando $N = 2m$, ambas formas quadráticas

$$\sum_{i,j=0}^m \mu_{i+j} x_i x_j \quad \text{e} \quad \sum_{i,j=0}^{m-1} (\mu_{i+j} - \mu_{i+j+2}) x_i x_j$$

sejam não negativas, e que quando $N = 2m - 1$, ambas formas quadráticas

$$\sum_{i,j=0}^{m-1} (\mu_{i+j} + \mu_{i+j+1}) x_i x_j \quad \text{e} \quad \sum_{i,j=0}^{m-1} (\mu_{i+j} - \mu_{i+j+1}) x_i x_j$$

sejam não negativas.

A solução do problema sobre um intervalo (a, b) qualquer $(-\infty < a < b < \infty)$ pode ser deduzido a partir deste resultado, através de mudança de variáveis, em particular quando $a = 0, b = 1$, temos pelo teorema de mudança de variáveis que

$$\int_{-1}^1 \left(\frac{t+1}{2} \right)^k d\psi(t) = \int_0^1 \theta^k d\varphi(\theta), \quad (7)$$

onde $T : (-1, 1) \longrightarrow (0, 1)$ com $T(t) = \frac{t+1}{2}$ e $\varphi = \psi_{T^{-1}}$.

Consequentemente o problema de momentos

$$c_k = \int_0^1 \theta^k d\varphi(\theta), \quad k = 0, 1, \dots, N \quad (8)$$

tem solução em φ , para constantes dadas $c_0, c_1, \dots, c_N$ se somente se

$$\int_{-1}^1 \left(\frac{t+1}{2} \right)^k d\psi = c_k, \quad k = 0, 1, \dots, N$$

tem solução em φ .

Por outro lado,

$$\int_{-1}^1 \left(\frac{t+1}{2} \right)^k d\psi(t) = \sum_{r=0}^k \binom{r}{k} 2^{-k} \int_{-1}^1 t^k d\psi(k), \quad (9)$$

portanto, se o problema (MR) tem solução em ψ para constantes $\mu_0, \mu_1, \dots, \mu_N$ definidas pela relação

$$\sum_{r=0}^k \binom{r}{k} \mu_r = 2^k c_k, \quad k = 0, 1, \dots, N \quad (10)$$

então o problema de momentos em (8), tem solução em φ para as constantes $c_0, c_1, \dots, c_N$ e vice-versa.

Estes resultados e o lema a seguir serão de utilidade para determinar condições sob as quais as constantes definidas no Corolário 2, correspondem aos N primeiros momentos de alguma distribuição de probabilidade definida sobre $[0, 1]$.

Lema 2: *Seja $A = (a_{ij})$ matriz de ordem N definida por*

$$a_{ij} = \begin{cases} \binom{i}{j} & \text{se } i \leq j \\ 0 & \text{se } i > j \end{cases}, \quad i, j = 1, \dots, N$$

então $B = A^{-1}$ é definida por

$$b_{ij} = \begin{cases} \binom{i}{j} (-1)^{i+j} & \text{se } i \leq j \\ 0 & \text{se } i > j \end{cases}, \quad i, j = 1, \dots, N$$

Prova: Seja $C = A \cdot B$, onde $C = (c_{ik})$. Então a partir da definição de A e B temos que

$$c_{ik} = \begin{cases} \sum_{j=k}^i a_{ij} b_{jk} & \text{se } k \leq i \\ 0 & \text{se } k > i \end{cases}$$

Se $k = i$, então $c_{ii} = 1$ e se $k < i$, temos que

$$\begin{aligned} c_{ik} &= \sum_{j=k}^i \binom{i}{j} \binom{j}{k} (-1)^{k+j} \\ &= \sum_{l=0}^{i-k} \binom{i}{l+k} \binom{l+k}{k} (-1)^l \end{aligned}$$

Mas, $\binom{i}{l+k} \binom{l+k}{k} = \binom{i-k}{l} \binom{i}{k}$. Logo

$$\begin{aligned} c_{ik} &= \sum_{l=0}^{i-k} \binom{i-k}{l} \binom{i}{k} (-1)^l \\ &= \binom{i}{k} \sum_{l=0}^{i-k} \binom{i-k}{l} (-1)^l = 0 \end{aligned}$$

concluindo-se assim a prova. $\square$

Sejam $\{c_k\}_{k=1}^N$ as constantes definidas no corolário 2 e seja $c_0 = 1$. Sejam $\{\mu_k\}_{k=1}^N$ outra sequência de constantes definidas por

$$\mu_k = \sum_{r=1}^k \binom{k}{r} (-1)^{k+r} (2^r c_r - 1)$$

para $k = 1, \dots, N$ e $\mu_0 = c_0$.

Corolário 3: *Sejam $X_1, X_2, \dots, X_N$ variáveis aleatórias permutáveis tomando valores em $\{0, 1\}$. Uma condição necessária e suficiente para que a sequência finita seja estendível é que as constantes $\{\mu_k\}_{k=0}^N$ acima definidas satisfaçam as condições que resolvem o problema (MR).*

Prova: Se a sequência finita é estendível então pelo corolário 2,

$$c_k = \int \theta^k d\varphi(\theta), \quad k = 0, 1, \dots, N$$

para alguma função de distribuição de probabilidade φ . Agora a partir das identidades (8) a (11), $\{\mu_k\}_{k=0}^N$ definidas pelas equações

$$\sum_{r=0}^k \binom{r}{k} \mu_r = 2^k c_k, \quad k = 0, 1, \dots, N$$

tem solução em ψ para o problema de momentos (MR) . Mas, $\mu_0 = c_0 = 1$, portanto

$$\sum_{r=1}^k \binom{r}{k} \mu_r = 2^k c^k - 1, \quad k = 1, \dots, N$$

Pelo lema 2, o sistema de equações acima tem solução para $\{\mu_r\}_{r=1}^N$ em termos de c_r dada por

$$\mu_k = \sum_{r=1}^k \binom{k}{r} (-1)^{k+r} (2^r c_r - 1), \quad k = 1, \dots, N.$$

Logo $\{\mu_k\}_{k=0}^N$ com $\mu_0 = 1$ satisfazem as condições que resolvem o problema de momentos (MR) .

Reciprocamente se $\{\mu_k\}_{k=0}^N$ resolve o problema de momentos (MR) , então pela identidade (9)

$$c_k = \sum_{r=0}^k \binom{r}{k} 2^{-k} \mu_r, \quad k = 0, 1, \dots, N$$

resolve o problema de momentos na identidade (9). Dado que $\mu_0 = 1$, a solução é uma função de distribuição de probabilidade, portanto a sequência finita é estendível. $\square$

Exemplo 6: Sejam X_1, X_2 variáveis aleatórias permutáveis tomando valores em $\{0, 1\}$ e suponha que $q(2, t) = P(X_1 + X_2 = t) = \frac{1}{3}$ para $t = 0, 1, 2$. Neste caso

$$c_k = \sum_{t=k}^2 \frac{\binom{2-t}{t-k}}{\binom{2}{t}} \cdot \frac{1}{3}, \quad k = 1, 2,$$

produzindo as soluções $c_1 = \frac{1}{2}$, $c_2 = \frac{1}{3}$. A partir disto, obtem-se, $\mu_0 = c_0 = 1$, $\mu_1 = 0$ e $\mu_2 = \frac{1}{3}$. Agora a matriz

$$A = \begin{bmatrix} 1 & 0 \\ 0 & \frac{1}{3} \end{bmatrix}$$

é definida positiva, consequentemente a sequência X_1, X_2 é estendível. $\square$

Exemplo 7: Sejam X_1, X_2 variáveis aleatórias permutáveis tomando valores em $\{0, 1\}$ e suponha que $q(2, t) = P(X_1 + X_2 = t) = \frac{9}{19} \left(\frac{2}{3}\right)^t$, $t = 0, 1, 2$. Neste caso

$c_1 = \frac{7}{19}$, $c_2 = \frac{4}{19}$ e a partir disto, $\mu_1 = \frac{-5}{19}$ e $\mu_2 = -1$, portanto, a sequência finita X_1, X_2 não é estendível. $\square$

O Modelo Geométrico

Aqui $\mathcal{X} = \mathbb{N}_0$, $T_N = \sum_{i=1}^N X_i$ e a distribuição condicional de $X_1, X_2, \dots, X_N \mid T_N = t$ é uniforme sobre $\mathcal{X}_t = \{(x_1, \dots, x_N) \in \mathbb{N}_0^N : \sum_{i=1}^N x_i = t\}$. Denotemos por P_θ a lei de uma sequência infinita de variáveis aleatórias independentes com distribuição comum Geométrica(θ). Seja $\mathcal{C}$ a classe de medidas de probabilidade P sobre $\mathbb{N}_0^\infty$ tal que, $P(\cdot) = \int_{[0,1]} P_\theta(\cdot) d\mu(\theta)$.

A seguir discutimos condições sob as quais a sequência $X_1, X_2, \dots, X_N$ pode ser mergulhada em uma sequência infinita de variáveis aleatórias permutáveis com lei $P \in \mathcal{C}$.

Proposição 8: *A sequência finita de variáveis aleatórias $X_1, X_2, \dots, X_N$ é estendível para uma sequência infinita de variáveis aleatórias permutáveis com lei $P \in \mathcal{C}$ se e somente se existe uma medida de probabilidade μ sobre $[0, 1]$ tal que*

$$q(N, t) = \int_{[0,1]} \binom{N+t-1}{t} (1-\theta)^N \theta^t d\mu(\theta), \quad t = 0, 1, \dots, \quad (11)$$

onde $q(N, \cdot)$ é a lei de T_N .

Prova: Suponhamos que existe uma medida de probabilidade μ tal que a identidade (11) vale, então

$$\begin{aligned} P_r\{X_1 = x_1, \dots, X_N = x_N\} &= P_r\{X_1 = x_1, \dots, X_N = x_N \mid T_N = t\} q(N, t) = \\ &= \int_{[0,1]} P_r\{X_1 = x_1, \dots, X_N = x_N \mid T_N = t\} \binom{N+t-1}{t} (1-\theta)^N \theta^t d\mu(\theta) \end{aligned}$$

Agora, não é difícil ver que T_N é uma estatística de resumo para $X_1, X_2, \dots, X_N$ relativo a P_θ . Logo

$$P_r\{X_1 = x_1, \dots, X_N = x_N\} =$$

$$= \int_{[0,1]} P_\theta \{X_1 = x_1, \dots, X_N = x_N \mid T_N = t\} \binom{N+t-1}{t} (1-\theta)^N \theta^t d\mu(\theta)$$

Mas,

$$P_\theta \{T_N = t\} = \binom{N+t-1}{t} (1-\theta)^N \theta^t \quad e$$

$$P_\theta \{X_1 = x_1, \dots, X_N = x_N\} = (1-\theta)^N \theta^t ,$$

onde $t = \sum_{i=1}^N x_i$. Logo

$$P \{X_1 = x_1, \dots, X_N = x_N\} = \int_{[0,1]} (1-\theta)^N \theta^t d\mu(\theta) .$$

Consequentemente, $(X_1, X_2, \dots, X_N) \stackrel{\mathcal{D}}{=} (Z_1, \dots, Z_N)$ onde $Z_1, Z_2, \dots, Z_N, \dots$ é uma sequência infinita de variáveis aleatórias permutáveis com lei P definida por $P(\cdot) = \int_{[0,1]} P_\theta(\cdot) d\mu(\theta)$.

Reciprocamente, suponhamos que $X_1, X_2, \dots, X_N$ é estendível para uma sequência infinita de variáveis aleatórias permutáveis com lei $P \in \mathcal{C}$, então existe uma medida μ sobre $[0, 1]$ tal que

$$P_r \{X_1 = x_1, \dots, X_N = x_N\} = \int_{[0,1]} \theta^t (1-\theta)^N d\mu(\theta) ,$$

onde $t = \sum_{i=1}^N x_i$. Logo

$$\begin{aligned} P(T_N = t) &= \sum_{(x_1, \dots, x_N) \in T_N^{-1}(t)} P_r \{X_1 = x_1, \dots, X_N = x_N\} = \int_{[0,1]} \sum_{(x_1, \dots, x_N) \in T_N^{-1}(t)} \theta^t (1-\theta)^N d\mu(\theta) \\ &= \int_{[0,1]} \binom{N+t-1}{t} \theta^t (1-\theta)^N d\mu(\theta), \end{aligned}$$

concluindo-se assim a prova. $\square$

Como no caso anterior a questão de interesse agora é determinar condições sobre $q(N, \cdot)$ para que exista uma medida de probabilidade μ sobre $[0, 1]$ que satisfaça a identidade (11). Tais condições estão também relacionadas a solução dos problemas de momentos,

em particular ao famoso problema de momentos de Hausdorff (1921). Enunciamos este a seguir.

Teorema 1 (Hausdorff): *Uma condição necessária e suficiente para que o problema de momentos*

$$\mu_n = \int_0^1 t^n d\phi(t), \quad n = 0, 1, 2, \dots$$

tenha uma solução em ϕ , função de distribuição é que

$$\Delta^k \mu_\nu \geq 0, \quad k, \nu = 0, 1, 2, \dots$$

onde

$$\Delta^0 \mu_\nu = \mu_\nu \quad \Delta^k \mu_\nu = \sum_{r=0}^k \binom{k}{r} (-1)^r \mu_{\nu+r}.$$

Prova: Ver Shohat e Tamarkin (1943) e Feller (1966, Vol 2). □

Este teorema e o lema a seguir serão utilizados na sequência para estabelecermos condições sobre $q(N, \cdot)$ que garantam estendibilidade da sequência $X_1, \dots, X_N$.

Lema 3: *Sejam $\{c(N, t)\}_{t=0}^N$ constantes dadas e $\{c(t)\}_{t=0}^N$ definidas pela relação*

$$c(N, t) = \sum_{r=0}^N \binom{N}{r} (-1)^r c_{r+t}, \quad t = 0, 1, \dots$$

então o problema

$$c(t) = \int_0^1 \theta^t d\phi(\theta), \quad t = 0, 1, \dots, \tag{12}$$

tem solução em ϕ , função de distribuição, sobre $[0, 1]$ se somente se, o problema

$$c(N, t) = \int_0^1 (1 - \theta)^N \theta^t d\phi(\theta), \quad t = 0, 1, \dots, \tag{13}$$

tem solução em ϕ , função de distribuição sobre $[0, 1]$.

Prova: Seja ϕ uma solução de (12), então

$$\begin{aligned} c(N, t) &= \sum_{r=0}^N \binom{N}{r} (-1)^r c_{r+t} = \int_{[0,1]} \sum_{r=0}^N \binom{N}{r} (-1)^r \theta^{r+t} d\phi(\theta) \\ &= \int_0^1 (1-\theta)^N \theta^t d\phi(\theta), \quad t = 0, 1, \dots \end{aligned}$$

Portanto ϕ é solução de (13). Suponhamos agora que ϕ é solução de (13). Da definição de $c(N, t)$ temos que $c(N, t) = \Delta^N c(t)$ onde $\Delta = 1 - B$, com B o operador forward, isto é, $Bc(t) = c(t+1)$. A partir disto obtemos

$$c(N, t) = \Delta^N c(t) = \Delta \left(\Delta^{N-1} c(t) \right) = \Delta (c(N-1, t)) = c(N-1, t) - c(N-1, t+1)$$

Agora

$$\sum_{k=0}^{t-1} \Delta c(N-1, k) = c(N-1, 0) - c(N-1, t).$$

Logo

$$\sum_{k=0}^{t-1} c(N, k) = c(N-1, 0) - c(N-1, t).$$

Consequentemente

$$\int_0^1 (1-\theta)^N \sum_{k=0}^{t-1} \theta^k d\phi(\theta) = c(N-1, 0) - c(N-1, t).$$

Mas, $\sum_{k=0}^{t-1} \theta^k = \frac{1-\theta^t}{(1-\theta)}$, portanto

$$\int_0^1 (1-\theta)^{N-1} d\phi(\theta) - \int_0^1 (1-\theta)^{N-1} \theta^t d\phi(\theta) = c(N-1, 0) - c(N-1, t)$$

Logo ϕ é solução do problema

$$c(N-1, t) = \int_0^1 (1-\theta)^{N-1} \theta^t d\phi(\theta), \quad t = 0, 1, \dots$$

e a partir daqui ϕ é solução do problema

$$c(0, t) = c(t) = \int_0^1 (1-\theta)^{N-1} \theta^t d\phi(\theta), \quad t = 0, 1, \dots$$

Concluindo-se assim a prova. □

Corolário 4: A sequência de variáveis aleatórias $X_1, X_2, \dots, X_N$ é estendível para uma sequência infinita de variáveis aleatórias permutáveis com lei $P \in \mathcal{C}$, se, e somente se

$$\Delta^k a(N, t) \geq 0, \quad k = 1, 2, \dots \quad t = 0, 1, \dots$$

onde $a(N, t) = \frac{q(N, t)}{\binom{N+t-1}{t}}$ e $\Delta = 1 - B$, com B o operador forward.

Prova: A partir da Proposição 8 temos que se a sequência finita $X_1, X_2, \dots, X_N$ é estendível, então existe ϕ função de distribuição de probabilidade concentrada no intervalo $[0, 1]$ tal que

$$a(N, t) = \int_0^1 (1 - \theta)^N \theta^t d\phi(\theta), \quad t = 0, 1, \dots$$

A partir disto e do lema (3), conclui-se que ϕ é solução do problema de momentos

$$c(t) = \int_0^1 \theta^t d\phi(\theta), \quad t = 0, 1, \dots$$

onde $\{c(t)\}_{t=0}^N$ é uma sequência de constantes satisfazendo a relação $a(N, t) = \Delta^N c(t)$, $t = 0, 1, \dots$. Logo, pelo teorema de Hausdorff

$$\Delta^m c(t) \geq 0, \quad t = 0, 1, \dots \quad m = 0, 1, \dots$$

Mas $\Delta^k a(N, t) = \Delta^{N+k} c(t)$, $k = 0, 1, \dots \quad t = 0, 1, \dots$

Consequentemente, $\Delta^k a(N, t) \geq 0$ para $k = 1, 2, \dots$ e $t = 0, 1, \dots$, concluindo-se assim a primeira parte da prova.

Para provar a recíproca, notemos primeiro que se $\{c(t)\}_{t=0}^N$ é uma sequência de constantes definida pela relação

$$a(N, t) = \Delta^N c(t), \quad t = 0, 1, \dots$$

então, $\Delta^{N-1} c(t) = a(N-1, t)$. Com efeito, $\Delta^N c(t) = \Delta(\Delta^{N-1} c(t)) = a(N, t)$ implica em

$$\Delta^{N-1} c(t) = \sum_{\ell=0}^{\infty} a(N, t + \ell) = \sum_{\ell=0}^{\infty} \frac{P(T_N = t + \ell)}{\binom{N+t+\ell-1}{t+\ell}}.$$

Agora

$$P(T_{N-1} = t / T_N = t + l) = \frac{\binom{N-1+t-1}{t}}{\binom{N+t+l-1}{t+l}}$$

Logo,

$$\Delta^{N-1}c(t) = \binom{N-1+t-1}{t}^{-1} P(T_{N-1} = t) = a(N-1, t)$$

Consequentemente, $\Delta^k c(t) = a(k, t)$, $k = 1, \dots, N$ e $t = 0, 1, \dots$

Também $\Delta c(t) = a(1, t)$ implica em

$$\Delta c(t) = \sum_{l=0}^{\infty} P(T_N = t + l) = P(T_N \geq t).$$

Portanto, $\Delta^k c(t) \geq 0$, $k = 0, 1, \dots, N$ e $c_0 = 1$.

Agora, $\Delta^{N+k} c(t) = \Delta^k a(N, t)$, $k = 1, 2, \dots$ e $t = 0, 1, \dots$

Mas, por hipótese $\Delta^k a(N, t) \geq 0$, $k = 1, 2, \dots$. Consequentemente pelo teorema de Hausdorff, concluímos que existe ϕ função de distribuição de probabilidade concentrada sobre $[0, 1]$ (já que $c_0 = 1$) tal que

$$c(t) = \int_0^1 \theta^t d\phi(\theta), \quad t = 0, 1, 2, \dots$$

Logo, pela definição de $c(t)$ junto ao lema anterior temos que ϕ é solução do problema

$$a(N, t) = \int_0^1 (1 - \theta)^N \theta^t d\phi(\theta), \quad t = 0, 1, \dots$$

A estendibilidade de $X_1, \dots, X_N$ segue-se a partir disto e da proposição 8. □

O exemplo a seguir ilustra este resultado.

Exemplo 8: Sejam $X_1, X_2, \dots, X_N$ variáveis aleatórias tomando valores em $\mathbb{N}_0$ e $T_N = \sum_{i=1}^N X_i$. Se $X_1, X_2, \dots, X_N / T_N = t \sim U\{(x_1, \dots, x_N) \in \mathbb{N}_0^N : \sum_{i=1}^N x_i = t\}$ e $\Pr\{T_N = t\} = \frac{1}{M+1} I_{\{0,1,\dots,M\}}(t)$, para algum $M \in \mathbb{N}$, então a sequência finita $X_1, X_2, \dots, X_N$, não é estendível para uma sequência infinita de variáveis aleatórias permutáveis com lei

$P \in \mathcal{C}$. Com efeito, se $t = M - 1$,

$$\begin{aligned}\Delta^k a(N, t) &= a(N, t) - ka(N, t + 1) \\ &= \frac{1}{M + 1} \binom{N + t - 1}{t}^{-1} \left\{ 1 - k \left(\frac{t + 1}{N + t} \right) \right\}\end{aligned}$$

Portanto, $\Delta^k a(N, t) \geq 0$ para $k = 1, 2, \dots$ se, e somente se, $1 - \{k \cdot \frac{M}{N+M-1}\} \geq 0$ ou equivalentemente, se $M \leq \frac{N-1}{k-1}$, $k = 2, 3, \dots$. Isto é, se $M = 0$. Consequentemente, se $M \in \mathbb{N}$ temos, a partir do Corolário 4, que a sequência finita $X_1, X_2, \dots, X_N$ não é estendível.

Notemos que se, no exemplo acima, $M = 0$, então a sequência é determinística com $X_k = 0$, $k = 1, \dots, N$, em cujo caso é trivialmente estendível. Também não é difícil ver a partir deste exemplo que qualquer distribuição a priori para T_N concentrando sua massa num subconjunto finito de $\mathbb{N}_0$, determina uma sequência finita de variáveis aleatórias $X_1, X_2, \dots, X_N$, não estendível.

O Modelo Uniforme

Aqui $\mathcal{X} = \mathbb{N}_0$, $T_N = \max\{X_1, X_2, \dots, X_N\}$ e a distribuição condicional de $X_1, \dots, X_N / T_N = m$ é uniforme sobre $\mathcal{X}_m = \{(x_1, \dots, x_N) \in \mathbb{N}_0^N : \max\{x_1, \dots, x_N\} = m\}$. No que segue vamos apresentar resultados que estabelecem condições sobre $q(N, \cdot)$ que permitem que a sequência finita, $X_1, X_2, \dots, X_N$, possa ser mergulhada numa sequência infinita de variáveis aleatórias com lei P sobre $\mathbb{N}_0^\infty$ tal que $P(\cdot) = \int_{\mathbb{N}_0} P_\theta(\cdot) d\mu(\theta)$, para alguma medida de probabilidade de μ sobre $\mathbb{N}_0$, com P_θ a lei de uma sequência infinita de variáveis aleatórias independentes com distribuição comum uniforme sobre $\{0, 1, \dots, \theta\}$, $\theta \in \mathbb{N}_0$. Denotemos por $\mathcal{M}$ a classe de medidas de probabilidade P sobre $\mathbb{N}_0^\infty$ obtida misturando-se elementos da família $\{P_\theta : \theta \in \mathbb{N}_0\}$.

Proposição 9: A sequência finita de variáveis aleatórias $X_1, X_2, \dots, X_N$ é estendível para

uma sequência infinita de variáveis aleatórias permutáveis com lei $P \in \mathcal{M}$ se, e somente se, existe uma medida de probabilidade μ sobre $\mathbb{N}_0$, tal que

$$q(N, m) = \int_{\theta \geq m} \frac{(m+1)^N - m^N}{(\theta+1)^N} d\mu(\theta), \quad (14)$$

para cada $m \in \mathbb{N}_0$, onde

$$q(N, m) = \Pr \{ \max\{X_1, \dots, X_N\} = m \} .$$

Prova: Se a sequência finita $X_1, \dots, X_N$ é estendível então existe uma medida de probabilidade μ sobre $\mathbb{N}_0$ tal que

$$\Pr \{X_1 = x_1, \dots, X_N = x_N\} = \int_{\theta \geq \max\{x_1, \dots, x_N\}} \frac{1}{(\theta+1)^N} d\mu(\theta), \quad (x_1, \dots, x_N) \in \mathbb{N}_0^N.$$

Agora

$$\begin{aligned} \Pr \{ \max\{X_1, \dots, X_N\} = m \} &= \sum_{(x_1, \dots, x_N) \in T_N^{-1}(\{m\})} P\{X_1 = x_1, \dots, X_N = x_N\} \\ &= \#T_N^{-1}(\{m\}) \cdot \Pr \{X_1 = x_1, \dots, X_N = x_N\} \end{aligned}$$

Mas, $\#T_N^{-1}(\{m\}) = (m+1)^N - m^N$, logo

$$\Pr \{ \max\{X_1, \dots, X_N\} = m \} = \int_{\theta \geq m} \frac{(m+1)^N - m^N}{(\theta+1)^N} d\mu(\theta) .$$

Para ver a recíproca notemos primeiro que T_N é uma estatística suficiente para $X_1, \dots, X_N$ relativo à família $\mathcal{P} = \{\mathcal{P}_\theta : \theta \in \mathbb{N}\}$. Também não é difícil ver que a família $\mathcal{P}$ é universal, portanto a partir da Proposição 6 conclui-se que

$$P_\theta \{X_1 = x_1, \dots, X_N = x_N / T_N = m\} = \frac{1}{(m+1)^N - m^N} I_{\{m\}} \max\{x_1, \dots, x_N\}.$$

Logo, se existe uma medida de probabilidade μ sobre $\mathbb{N}_0$, tal que identidade (14) vale para cada $m \in \mathbb{N}_0$, temos que

$$\Pr \{X_1 = x_1, \dots, X_N = x_N\} =$$

$$\begin{aligned}
&= \Pr \{X_1 = x_1, \dots, X_N = x_N / T_N = m\} \Pr \{T_N = m\} \\
&= \int_{\theta \geq m} P_\theta \{X_1 = x_1, \dots, X_N = x_N / T_N = m\} \left\{ \frac{(m+1)^N - m^N}{(\theta+1)^N} \right\} d\mu(\theta) \\
&= \int_{\theta \geq m} \frac{P_\theta \{X_1 = x_1, \dots, X_N = x_N\}}{P_\theta \{T_N = m\}} \left\{ \frac{(m+1)^N - m^N}{(\theta+1)^N} \right\} d\mu(\theta)
\end{aligned}$$

onde $m = \max\{x_1, \dots, x_N\}$. Mas, pela identidade (9), no capítulo 2, temos que

$$P_\theta \{T_N = m\} = \frac{(m+1)^N - m^N}{(\theta+1)^N} I_{\{0,1,\dots,\theta\}}(m).$$

Consequentemente

$$\Pr \{X_1 = x_1, \dots, X_N = x_N\} = \int_{\theta \geq m} \frac{1}{(\theta+1)^N} d\mu(\theta),$$

com $m = \max\{x_1, \dots, x_N\}$. A estendibilidade de $X_1, \dots, X_N$ segue-se a partir disto de maneira análoga aos casos anteriores. $\square$

Corolário 5: A sequência finita de variáveis aleatórias $X_1, X_2, \dots, X_N$ é estendível para uma sequência infinita de variáveis aleatórias com lei $P \in \mathcal{M}$ se, e somente se, a função $f : \mathbb{N}_0 \rightarrow \mathbb{R}$ definida por

$$f(m) = (m+1)^N \left\{ \frac{q(N, m)}{(m+1)^N - m^N} - \frac{q(N, m+1)}{(m+2)^N - (m+1)^N} \right\}$$

$m \in \mathbb{N}_0$, define uma medida de probabilidade sobre $\mathbb{N}_0$.

Prova: Se a sequência finita de variáveis aleatórias $X_1, X_2, \dots, X_N$ é estendível, então, pela Proposição 9, temos que existe uma medida de probabilidade μ sobre $\mathbb{N}_0$, tal que

$$\begin{aligned}
q(N, m) &= \Pr \{\max\{X_1, \dots, X_N\} = m\} = \int_{\theta \geq m} \left\{ \frac{(m+1)^N - m^N}{(\theta+1)^N} \right\} d\mu(\theta) \\
&= \sum_{\theta=m}^{\infty} \left\{ \frac{(m+1)^N - m^N}{(\theta+1)^N} \right\} \mu(\{\theta\})
\end{aligned}$$

para cada $m \in \mathbb{N}_0$.

Notando agora que

$$q(N, m+1) = \sum_{\theta=m+1}^{\infty} \left\{ \frac{(m+2)^N - (m+1)^N}{(\theta+1)^N} \right\} \mu(\{\theta\}), \quad m \in \mathbb{N}_0,$$

obtemos, subtraindo este termo do anterior,

$$\mu(\{m\}) = (m+1)^N \left\{ \frac{q(N, m)}{(m+1)^N - m^N} - \frac{q(N, m+1)}{(m+2)^N - (m+1)^N} \right\}$$

Portanto, f define uma medida de probabilidade sobre $\mathbb{N}_0$.

Reciprocamente, suponhamos que f define uma medida de probabilidade μ_f sobre $\mathbb{N}_0$, isto é,

$$\mu_f(\{m\}) = f(m), \quad m \in \mathbb{N}_0.$$

Seja $Z_1, Z_2, \dots$ uma sequência infinita de variáveis aleatórias permutáveis com lei P definida por

$$P(Z_1 = z_1, \dots, Z_N = z_N) = \int_{\theta \geq \max\{z_1, \dots, z_N\}} \frac{1}{(\theta+1)^N} d\mu_f(\theta),$$

$N \in \mathbb{N}$, $(z_1, \dots, z_N) \in \mathbb{N}_0^N$. Por construção, $P \in \mathcal{M}$ e

$$\begin{aligned} P(\max\{Z_1, \dots, Z_N\} = m) &= \int_{\theta \geq m} P_\theta(\max\{Z_1, \dots, Z_N\} = m) d\mu(\theta) \\ &= \sum_{\theta=m}^{\infty} \left\{ \frac{(m+1)^N - m^N}{(\theta+1)^N} \right\} f(\theta) \\ &= \{(m+1)^N - m^N\} \sum_{\theta=m}^{\infty} \left\{ \frac{q(N, \theta)}{(\theta+1)^N - \theta^N} - \frac{q(N, \theta+1)}{(\theta+2)^N - (\theta+1)^N} \right\} \\ &= q(N, m) \end{aligned}$$

Logo, existe uma medida de probabilidade μ sobre $\mathbb{N}_0$ tal que

$$q(N, m) = \int_{\theta \geq m} \left\{ \frac{(m+1)^N - m^N}{(\theta+1)^N} \right\} d\mu(\theta),$$

$m \in \mathbb{N}_0$, basta tomar $\mu = \mu_f$. A estendibilidade da sequência finita de variáveis aleatórias $X_1, \dots, X_N$ decorre da proposição 9. $\square$

Se bem que este caso é mais simples do que os anteriores, o resultado é mais forte, no sentido que o corolário acima exhibe a forma da medida na mistura, no caso da sequência finita ser estendível. O exemplo a seguir ilustra estes resultados.

Exemplo 9: Sejam $X_1, X_2, \dots, X_N$ variáveis aleatórias tomando valores em $\mathbb{N}_0$ e seja $T_N = \max\{X_1, \dots, X_N\}$.

Se $X_1, X_2, \dots, X_N/T_N = m \sim U\{(x_1, \dots, x_N) \in \mathbb{N}_0^N : \max\{x_1, \dots, x_N\} = m\}$ e se $\Pr\{T_N = M\} = 0$ para algum $M \in \mathbb{N}$, $\Pr\{T_N = m\} > 0$ para cada $m \neq M$, então a sequência finita $X_1, X_2, \dots, X_N$ não é estendível para uma sequência infinita de variáveis aleatórias permutáveis com lei $P \in \mathcal{M}$. De fato,

$$f(M) = \left\{ -\frac{q(N, M+1)}{(M+2)^N - (M+1)^N} \right\} (M+1)^N < 0,$$

portanto f não define uma medida de probabilidade. Consequentemente, temos a partir do Corolário 5, que a sequência finita $X_1, \dots, X_N$ não é estendível. $\square$

Em geral não é difícil ver a partir deste exemplo que se $q(N, m) = 0$ para $m \in S \subset \mathbb{N}_0$ $q(N, m) > 0$ para $m \notin S$, onde $\min S > 0$ e S limitado, então a sequência finita $X_1, X_2, \dots, X_N$ não é estendível para uma sequência infinita com lei $P \in \mathcal{M}$.

3.4.3 Aplicações

No que segue exemplificaremos o procedimento descrito nesta seção resolvendo alguns problemas de predição particulares. Com este objetivo a distribuição preditiva a posteriori será derivada em cada situação, utilizando-se das conexões com estatísticas de resumo em lugar de utilizar técnicas de contagem requeridas na identidade (4). Isto último com o propósito de fazer comparações com a abordagem que será apresentada no próximo capítulo.

Predição do Total Populacional para Dados Binários

Sejam $X_1, X_2, \dots, X_N$ quantidades aleatórias assumindo valores em $\mathcal{X} = \{0, 1\}$. Interesse fazer inferências a respeito de $T_N = \sum_{i=1}^N X_i$ baseados numa amostra de tamanho n

($n < N$) selecionada da população. Supondo a priori que

$$X_1, X_2, \dots, X_N / T_N = t \sim U\{(x_1, \dots, x_N) \in \{0, 1\}^N : \sum_{i=1}^N x_i = t\}, \quad t \in \{0, 1, \dots, N\}.$$

Sejam $x_1, x_2, \dots, x_n$ os valores observados na amostra e $lik(t/x_1, \dots, x_n)$ a verossimilhança de t dado $x_1, \dots, x_n$. Como antes, $q(N, \cdot)$ denota a distribuição a priori para T_N .

Distribuição a posteriori de T_N

Utilizando o fato que T_N é uma estatística de resumo para $X_1, \dots, X_N$ relativo a P_θ^N - a lei de N variáveis aleatórias independentes com distribuição $B(1, \theta)$, $\theta \in (0, 1)$, temos que

$$\begin{aligned} lik(t/x_1, \dots, x_n) &= \Pr\{X_1 = x_1, \dots, X_n = x_n / T_N = t\} \\ &= P_\theta^N\{X_1 = x_1, \dots, X_n = x_n / T_N = t\} \\ &= \frac{P_\theta^N\{X_1 = x_1, \dots, X_n = x_n, \sum_{i=n+1}^N X_i = t - t_s\}}{P_\theta^N\{T_N = t\}}, \end{aligned}$$

onde, $t_s = \sum_{i=1}^n x_i$ e $t \in \{t_s, t_{s+1}, \dots, N - n + t_s\}$.

Agora

$$P_\theta^N\{T_N = t\} = \binom{N}{t} \theta^t (1 - \theta)^{N-t}.$$

e

$$P_\theta^N\{X_1 = x_1, \dots, X_n = x_n, \sum_{i=n+1}^N X_i = t - t_s\} = \binom{N-n}{t-t_s} \theta^{t-t_s} (1 - \theta)^{N-t} I_{\{t_s, t_{s+1}, \dots, N-n+t_s\}}(t).$$

Logo

$$lik(t/x_1, \dots, x_n) = \frac{\binom{N-n}{t-t_s}}{\binom{N}{t}} I_{\{t_s, t_{s+1}, \dots, N\}}(t).$$

Consequentemente,

$$\Pi(t/x_1, \dots, x_n) \propto \frac{\binom{N-n}{t-t_s}}{\binom{N}{t}} q(N, t) I_{\{t_s, t_{s+1}, \dots, N-n+t_s\}}(t).$$

Suponhamos agora que a distribuição a priori para T_N é uniforme sobre $\{0, 1, \dots, N\}$, isto é,

$$q(N, t) = \frac{1}{N+1} I_{\{0,1,\dots,N\}}(t) .$$

Vamos derivar neste caso o preditor de máxima verossimilhança para T_N , isto é, o valor $\hat{t}$ de t tal que

$$lik(\hat{t}/x_1, \dots, x_n) \geq lik(\hat{t} - 1/x_1, \dots, x_n)$$

e

$$lik(\hat{t}/x_1, \dots, x_n) \geq lik(\hat{t} + 1/x_1, \dots, x_n)$$

Notando que

$$lik(\hat{t} + 1/x_1, \dots, x_n) = \frac{N - n - (\hat{t} - t_s)}{\hat{t} - t_s + 1} \cdot \frac{\hat{t} + 1}{N - \hat{t}} \cdot lik(\hat{t}/x_1, \dots, x_n) .$$

temos que

$$\begin{aligned} lik(\hat{t} + 1/x_1, \dots, x_n) &\leq lik(\hat{t}/x_1, \dots, x_n) \iff \\ (N - n - (\hat{t} - t_s))(\hat{t} + 1) &\leq (\hat{t} - t_s + 1)(N - \hat{t}) \iff \\ \hat{t} &\geq \frac{N+1}{n} t_s - 1 . \end{aligned}$$

Analogamente

$$lik(\hat{t} - 1/x_1, \dots, x_n) = \frac{(\hat{t} - t_s)(N - \hat{t} + 1)}{\hat{t}(N - n - (\hat{t} - t_s - 1))} lik(\hat{t}/x_1, \dots, x_n) .$$

Logo

$$\begin{aligned} lik(\hat{t} - 1/x_1, \dots, x_n) &\leq lik(\hat{t}/x_1, \dots, x_n) \iff \\ (\hat{t} - t_s)(N - \hat{t} + 1) &\leq \hat{t}(N - n - \hat{t} + t_s + 1) \iff \\ \hat{t} &\leq \frac{N+1}{n} t_s . \end{aligned}$$

Combinando estas duas desigualdades, temos que um preditor de máxima verossimilhança é um inteiro no intervalo

$$\left[\frac{N+1}{n} t_s - 1, \frac{N+1}{n} t_s \right] .$$

Por exemplo, se $N = n + k$ e $\frac{t_s}{n} \leq \frac{1}{k+1}$, então $\hat{t} = t_s$, em particular se $k = 1$, temos que $\hat{t} = t_s$ se $\frac{t_s}{n} < \frac{1}{2}$ e $\hat{t} = t_s + 1$ se $\frac{t_s}{n} \geq \frac{1}{2}$.

A título de ilustração vamos derivar neste caso o preditor que minimiza o risco associado à função de perda quadrática, diretamente do modelo II. Para isto notemos que

$$\sum_{t=k}^{N-n+k} \frac{\binom{N-n}{t-k}}{\binom{N}{t}} = \frac{N+1}{n+1} \cdot \binom{n}{k}^{-1}, \quad k = 0, 1, \dots, n. \quad (16)$$

Logo

$$\Pr \{X_1 = x_1, \dots, X_n = x_n\} = \sum_{t=t_s}^{N-n+t_s} \frac{\binom{N-n}{t-t_s}}{\binom{N}{t}} \cdot \frac{1}{N+1} = \frac{\binom{n}{t_s}^{-1}}{n+1}$$

Consequentemente,

$$\Pi(t/x_1, \dots, x_n) = \frac{\binom{n}{t_s} \binom{N-n}{t-t_s}}{\binom{N}{t}} \frac{n+1}{N+1} I_{\{t_s, \dots, N-n+t_s\}}(t).$$

Agora

$$\begin{aligned} E(T_N + 1/x_1, \dots, x_n) &= \sum_{t=t_s}^{N-n+t_s} \frac{(t+1) \cdot \binom{N-n}{t-t_s} \binom{n}{t_s}}{(N+1) \binom{N}{t}} \cdot (n+1) \\ &= \sum_{t=t_s}^{N-n+t_s} (n+1) \frac{\binom{N-n}{t-t_s} \binom{n}{t_s}}{\binom{N+1}{t+1}} \end{aligned}$$

Fazendo $u = t + 1$ e utilizando a identidade (16), obtemos

$$E(T_N + 1/x_1, \dots, x_n) = (n+1) \binom{n}{t_s} \sum_{u=t_s+1}^{N-n+t_s+1} \frac{\binom{N+1-(n+1)}{u-(t_s+1)}}{\binom{N+1}{u}} = \frac{N+2}{n+2} (t_s + 1).$$

Portanto,

$$E(T_N/x_1, \dots, x_n) = \frac{N+2}{n+2} t_s + \frac{N-n}{n+2}. \quad (17)$$

Do mesmo modo

$$\begin{aligned} E((T_N + 1)^2/x_1, \dots, x_n) &= \sum_{t=t_s}^{N-n+t_s} (t+1)^2 \frac{\binom{N-n}{t-t_s} \binom{n}{t_s}}{\binom{N}{t}} \cdot \frac{n+1}{N+1} \\ &= \frac{(n+1)(N+2) \binom{n}{t_s}}{(n+2) \binom{n+1}{t_s+1}} \cdot \sum_{u=t_s+1}^{N-n+t_s+1} u \frac{\binom{N+1-(n+1)}{u-(t_s+1)} \binom{n+1}{t_s+1}}{\binom{N+1}{u}} \cdot \frac{n+2}{N+2} \\ &= \frac{N+2}{n+2} (t_s + 1) \left\{ \frac{N+3}{n+3} (t_s + 2) - 1 \right\}. \end{aligned}$$

Esta última igualdade segue da identidade (16). A partir disto e utilizando novamente a igualdade (17), obtemos

$$\begin{aligned} E(T_N^2/x_1, \dots, x_n) &= \frac{N+2}{n+2}(t_s+1) \left\{ \frac{N+3}{n+3}(t_s+2) - 3 \right\} + 1, \\ V(T_N/x_1, \dots, x_n) &= \frac{N+2}{n+2}(t_s+1) \left\{ \frac{N+3}{n+3}(t_s+2) - 1 - \left(\frac{N+2}{n+2} \right) (t_s+1) \right\}. \end{aligned}$$

A escolha de uma distribuição a priori uniforme para T_N , foi feita propositalmente, já que neste caso a Proposição 7 garante a estendibilidade da sequência finita de variáveis aleatórias $X_1, X_2, \dots, X_N$. Com efeito

$$\frac{1}{N+1} = \int_{(0,1)} \theta^t (1-\theta)^{N-t} d\theta, \quad t = 0, 1, \dots, N.$$

Assim, os resultados anteriormente expostos, coincidem com aqueles que teriam sido obtidos se tivéssemos formulado um modelo de superpopulação, onde

$$\Pr \{X_1 = x_1, \dots, X_N = x_N\} = \int_{(0,1)} \theta^t (1-\theta)^{N-t} d\theta$$

com $t = \sum_{i=1}^N x_i$, e $x_i \in \{0, 1\}$ para $i = 1, \dots, N$.

Notemos uma vez mais que dentro do contexto deste trabalho a coincidência é de natureza técnica, pois o experimentador expressou probabilisticamente sua opinião a priori sobre quantidades de população finita em lugar de quantidades de população infinita como é feito na usual formulação bayesiana de superpopulação. A vantagem de tal associação é a utilização de resultados já estabelecidos. Porém, a menos de casos conhecidos, nem sempre é simples verificar estendibilidade através dos resultados estabelecidos na subseção 3.4.2, os quais nos dois primeiros casos, só garantem a existência de uma sequência infinita na qual a sequência finita pode ser mergulhado. Em termos práticos é mais simples calcular diretamente a distribuição a posteriori, a qual requer de pouco esforço computacional, como é mostrado a seguir.

Suponhamos que a distribuição a priori para o total populacional é dada por

$$q(N, t) = \frac{a}{N+1} \left(\frac{N}{N+1} \right)^t I_{\{0,1,\dots,N\}}(t)$$

com $a = \left\{1 - \left(\frac{N}{N+1}\right)^{N+1}\right\}^{-1}$. Notemos a partir do exemplo 7 que neste caso a sequência finita $X_1, \dots, X_N$ não é estendível.

A Figura 1 abaixo mostra o comportamento de $q(N, \cdot)$ quando $N = 30$.

Figura 1: Distribuição a priori para o total populacional quando $N = 30$

A distribuição a posteriori para $N = 30$ e alguns valores de n e t_s é mostrada nas figuras a seguir. Também é exibida a distribuição a posteriori aproximada, obtida ao utilizarmos a aproximação decorrente do Teorema 6, no Capítulo 1. Se denotarmos tal aproximação por Π^* , então

$$\Pi^*(t|x_1, \dots, x_n) \propto \left(\frac{t}{N}\right)^{t_s} \left(1 - \frac{t}{N}\right)^{n-t_s} q(N, t) I_{\{t_s, \dots, N-n+t_s\}}(t) .$$

Figura 2: Distribuição a posteriori para o total populacional quando $N = 30$, $t_s = 0$, $n = 5$

Figura 3: Distribuição a posteriori para o total populacional quando $N = 30$, $t_s = 0$, $n = 10$

Figura 4: Distribuição a posteriori para o total populacional quando $N = 30$, $t_s = 0$, $n = 20$

Figura 5: Distribuição a posteriori para o total populacional quando $N = 30$, $t_s = 0$, $n = 29$

Pode ser notado a partir das Figuras 2 até 5 que a aproximação funciona bem (conforme estabelecido no Teorema 6 do Capítulo 1) quando n é pequeno em relação a N .

Predição do total populacional para dados tomando valores inteiros.

Neste caso $X_1, X_2, \dots, X_N$ são quantidades aleatórias assumindo valores em $\mathcal{X} = \mathbb{N}_0$. Interessa fazer inferências a respeito de $T_N = \sum_{i=1}^N X_i$ baseados numa amostra de tamanho n ($n < N$) selecionada da população, supondo a priori que

$$X_1, X_2, \dots, X_N \mid T_N = t \sim U\{(x_1, \dots, x_N) \in \mathbb{N}_0^N : \sum_{i=1}^N x_i = t\}, \quad t \in \mathbb{N}_0.$$

Distribuição a posteriori de T_N

Seja P_θ^N a lei de N variáveis aleatórias independentes com distribuição comum Geométrica (θ), isto é

$$P_\theta^N(x_1, \dots, x_N) = (1 - \theta)^N \theta^t I_{\mathbb{N}_0^N}(x_1, \dots, x_N), \quad t = \sum_{i=1}^N x_i.$$

É imediato a partir desta expressão que T_N é uma estatística de resumo para $X_1, X_2, \dots, X_N$ relativo a P_θ^N . Logo pela proposição 5, temos que se $n \leq N$

$$\begin{aligned} \text{lik}(t|x_1, \dots, x_n) &= P_r\{X_1 = x_1, \dots, X_n = x_n \mid T_N = t\} = \\ &= P_\theta^N \frac{\{X_1 = x_1, \dots, X_n = x_n \text{ e } \sum_{i=n+1}^N X_i = t - t_s\}}{P_\theta^N(T_N = t)} \\ &= \frac{(1 - \theta)^n \theta^{t_s} \binom{N-n+(t-t_s)-1}{t-t_s} \theta^{t-t_s} (1 - \theta)^{N-n}}{\binom{N+t-1}{t} \theta^t (1 - \theta)^N} I_{\{t_s, t_{s+1}, \dots\}}(t) \\ &= \frac{\binom{N-n+(t-t_s)-1}{t-t_s}}{\binom{N+t-1}{t}} I_{\{t_s, t_{s+1}, \dots\}}(t) \end{aligned}$$

onde $t_s = \sum_{i=1}^n x_i$, e $(x_1, \dots, x_N) \in \mathbb{N}_0^N$.

Portanto

$$\Pi(t \mid x_1, \dots, x_n) \propto \frac{\binom{N-n+(t-t_s)-1}{t-t_s}}{\binom{N+t-1}{t}} q(N, t) I_{\{t_s, t_s+1, \dots\}}(t)$$

Somente com o propósito de estabelecer conexões com a abordagem que será apresentada no próximo capítulo, consideraremos uma distribuição a priori imprópria de forma que

$$\Pi(t \mid x_1, \dots, x_n) \propto l_i k(t \mid (x_1, \dots, x_n),$$

o qual é possível, pois, a série $\sum_{t=t_s}^{\infty} l_i k(t \mid x_1, \dots, x_n)$ é convergente (decorre da desigualdade 13, no Teorema 8 do Capítulo 1). Em tal caso um preditor de máxima verossimilhança é um valor $\hat{t}$ de t tal que

$$l_i k(\hat{t} \mid x_1, \dots, x_n) \geq l_i k(\hat{t} + 1 \mid x_1, \dots, x_n)$$

e

$$l_i k(\hat{t} \mid x_1, \dots, x_n) \geq l_i k(\hat{t} - 1 \mid x_1, \dots, x_n).$$

Para resolver-mos estas desigualdades notemos primeiro que

$$l_i k(t + 1 \mid x_1, \dots, x_n) = \frac{t + 1}{t - t_s + 1} \frac{N - n + t - t_s}{N + t} l_i k(t \mid x_1, \dots, x_n)$$

e

$$l_i k(t - 1 \mid x_1, \dots, x_n) = \frac{N + t - 1}{N - n + (t - t_s)} \cdot \frac{t - t_s}{t} l_i k(t \mid x_1, \dots, x_n)$$

Substituindo estas desigualdades nas desigualdades acima, obtemos que $\hat{t}$ é um inteiro no intervalo

$$\left[\frac{N-1}{n} t_s - 1, \frac{N-1}{n} t_s \right]$$

Por exemplo se $N = n + 1$ então $\hat{t} = t_s$, pois, $t \geq t_s$. Notemos a partir do Exemplo 8 que com tal escolha a sequência finita $X_1, X_2, \dots, X_N$ não é estendível.

Alguns resultados particulares quando $N = 30$ e distribuição a priori uniforme sobre $\{0, 1, \dots, 100\}$ para o total populacional são mostradores a seguir.

Os gráficos abaixo mostram o comportamento da distribuição a posteriori para alguns valores de n e t_s . Os resultados são comparados com os obtidos ao utilizarmos da aproximação decorrente do Teorema 8 do Capítulo 1. Se denotamos por Π^* tal aproximação, então

$$\Pi^*(t/t_s) \propto \left(1 - \frac{t}{N+t}\right)^n \left(\frac{t}{N+t}\right)^{t_s} I_{\{t_s, t_s+1, \dots\}}(t)$$

Figura 6: Distribuição a posteriori para o total populacional quando $N = 30$, $t_s = 30$, $n = 5$

Figura 7: Distribuição a posteriori para o total populacional quando $N = 30$, $t_s = 30$, $n = 10$

Figura 8: Distribuição a posteriori para o total populacional quando $N = 30$, $t_s = 30$, $n = 20$

Figura 9: Distribuição a posteriori para o total populacional quando $N = 30$, $t_s = 30$, $n = 29$

Como no caso anterior, a aproximação funciona bem quando n é pequeno em relação a N .

Predição do valor máximo na população para dados tomando valores inteiros.

Na seqüência $X_1, X_2, \dots, X_N$ são quantidades aleatórias assumindo valores em $\mathcal{X} = \mathbb{N}_0$. Interessa fazer inferências a respeito de $T_N = \max\{X_1, \dots, X_N\}$ baseados numa amostra de tamanho n selecionada da população, supondo a priori que a distribuição de $X_1, X_2, \dots, X_N \mid \max\{X_1, \dots, X_N\} = m$ é uniforme sobre o conjunto $\mathcal{X}_t^N = \{(x_1, \dots, x_N) \in \mathbb{N}_0^N : \max\{x_1, \dots, x_N\} = m\}$, $m \in \mathbb{N}_0$.

Distribuição a posteriori de T_N

Sejam $x_1, \dots, x_n$ os valores observados na amostra e $lik(m/x_1, \dots, x_n)$ a verossimilhança de $T_N = m$ dado $x_1, \dots, x_n$.

Seja P_θ^N a lei de N variáveis aleatórias independentes com distribuição comum $U\{0, 1, \dots, \theta\}$, $\theta \in \mathbb{N}_0$, isto é

$$P_\theta^N(x_1, \dots, x_N) = \frac{1}{(\theta + 1)^N} I_{\{0, 1, \dots, \theta\}}(\max(x_1, \dots, x_N))$$

Novamente não é difícil ver a partir desta expressão que T_N é uma estatística de resumo para $X_1, \dots, X_N$ relativo a P_θ^N .

Logo, para $n \leq N$ temos que

$$\begin{aligned} lik(m/x_1, \dots, x_n) &= P_r\{X_1 = x_1, \dots, X_n = x_n \mid T_N = m\} \\ &= P_\theta^N\{X_1 = x_1, \dots, X_n = x_n \mid T_N = m\} \\ &= \frac{P_\theta^N\{X_1 = x_1, \dots, X_n = x_n, \max\{m_s, X_{n+1}, \dots, X_N\} = m\}}{P_\theta^N\{\max\{X_1, \dots, X_N\} = m\}} \end{aligned}$$

onde $m_s = \max\{x_1, \dots, x_n\}$.

Se $m = m_s$, então

$$\begin{aligned} P_r\{X_1 = x_1, \dots, X_n = x_n \mid T_N = m\} &= \\ &= \frac{P_\theta^N\{X_1 = x_1, \dots, X_n = x_n, X_{n+1} \leq m, \dots, X_{N+1} \leq m\}}{P_\theta^N\{\max\{X_1, \dots, X_N\} = m\}}. \end{aligned}$$

Agora

$$P_\theta^N\{\max\{X_1, \dots, X_k\} = m\} = \left(\frac{m+1}{\theta+1}\right)^k - \left(\frac{m}{\theta+1}\right)^k.$$

Logo

$$P_r\{X_1 = x_1, \dots, X_n = x_n \mid T_N = m\} = \frac{(m+1)^{N-n}}{(m+1)^N - m^N}.$$

Analogamente se $m_s < m$

$$P_r\{X_1 = x_1, \dots, X_n = x_n \mid T_N = m\}$$

$$\begin{aligned}
&= \frac{P_r\{X_1 = x_1, \dots, X_n = x_n, \max\{X_{n+1}, \dots, X_N\} = m/\theta\}}{P_r\{\max\{X_1, \dots, X_N\} = m/\theta\}} \\
&= \frac{(m+1)^{N-n} - m^{N-n}}{(m+1)^N - m^N}.
\end{aligned}$$

Consequentemente

$$\Pi(m/x_1, \dots, x_n) \propto \begin{cases} \frac{1}{(m+1)^n} \left\{ \frac{1}{1 - \left(\frac{m}{m+1}\right)^N} \right\} q(N, m) & \text{se } m = \max\{x_1, \dots, x_n\} \\ \frac{1}{(m+1)^n} \left\{ \frac{1 - \left(\frac{m}{m+1}\right)^{N-n}}{1 - \left(\frac{m}{m+1}\right)^N} \right\} q(N, m) & \text{se } m > \max\{x_1, \dots, x_n\}. \end{cases}$$

Com o propósito de estabelecer conexões com a abordagem que será apresentada no próximo capítulo, ilustraremos os resultados utilizando uma distribuição a priori imprópria para T_N , de forma que

$$\Pi(m/x_1, \dots, x_n) \propto l_k(m/x_1, \dots, x_n).$$

Para obtermos o preditor de máxima verossimilhança notemos primeiro que $g(x) = \frac{1 - x^{N-n}}{1 - x^N}$, $0 < x < 1$ é decrescente.

Com efeito,

$$\begin{aligned}
g'(x) &= \frac{x^{N-n-1} \{n(1 - x^N) - N(1 - x^n)\}}{(1 - x^N)^2} \\
&= \frac{x^{N-n-1}(1 - x) \{n(x^{n+1} + \dots + x^{N-1}) - (N - n)(1 + x^2 + \dots + x^{n-1})\}}{(1 - x^N)^2}.
\end{aligned}$$

Mas

$$n(x^n + \dots + x^{N-1}) < n(N - n)x^{n-1},$$

e

$$(N - n)(1 + x^2 + \dots + x^{n-1}) > (N - n)nx^{n-1}.$$

Logo, $g'(x) < 0$. Agora $h(m) = \frac{m}{m+1}$ é crescente, portanto se $m > m_s$

$$l_k(m/m_s) > l_k(m+1/m_s).$$

Por outro lado, $g(m) = \frac{(m+1)^{N-n}}{(m+1)^N - m^N}$, é também decrescente. De fato

$$\begin{aligned} g'(m) &= \frac{(m+1)^{N-m-1} \{Nm^{N-1} + nm^N - n(m+1)^N\}}{((m+1)^N - m^N)^2} \\ &= \frac{Nm^{N-1}(1-n) - n \sum_{k=0}^{N-2} \binom{N}{k} m^k}{((m+1)^N - m^N)^2} < 0. \end{aligned}$$

Logo,

$$\begin{aligned} \frac{1}{(m_s+1)^n} \left\{ \frac{1}{1 - \left(\frac{m_s}{m_s+1}\right)^n} \right\} &> \frac{1}{(m_s+2)^n} \left\{ \frac{1}{1 - \left(\frac{m_s+1}{m_s+2}\right)^N} \right\} \\ &> \frac{1}{(m_s+2)} \left\{ \frac{1 - \left(\frac{m_s+1}{m_s+2}\right)^{N-n}}{1 - \left(\frac{m_s+1}{m_s+2}\right)^N} \right\} \end{aligned}$$

Consequentemente o preditor de máxima verossimilhança é

$$\widehat{m} = \max\{x_1, \dots, x_n\},$$

o qual também maximiza a função de probabilidade a posteriori para qualquer distribuição a priori positiva e decrescente.

Alguns resultados particulares para alguns valores de N , n e m_s e distribuição a priori dada por

$$q(N, m) \propto \frac{1}{(m+1)^2} I_{\{0,1,\dots,20\}}(m)$$

são mostrados nas figuras a seguir. Os resultados são comparados com os obtidos ao utilizarmos da aproximação fornecida pelo Teorema 1 do Capítulo 2. Se denotarmos por Π^* a distribuição a posteriori derivada de tal aproximação, então

$$\Pi^*(m|m_s) \propto \frac{1}{(m+1)^{n+2}} I_{\{(m_s+1),\dots,20\}}$$

Como será notado nas figuras abaixo, a aproximação funciona bem em geral, a menos de situações onde N , n e m_s são pequenos.

Figura 10: Distribuição a posteriori exata e aproximada quando $N = 4$, $m_s = 2$, $n = 2$

Figura 11: Distribuição a posteriori exata e aproximada quando $N = 3$, $m_s = 1$, $n = 1$

Figura 12: Distribuição a posteriori exata e aproximada quando $N = 30, m_s = 5, n = 20$

Figura 13: Distribuição a posteriori exata e aproximada quando $N = 30, m_s = 5, n = 29$

3.5 Modelos não discretos localmente uniformes

No caso anterior a construção de medidas de probabilidade uniformes dependeram unicamente da medida de contagem. Quando $A \subset \mathbb{R}^N$ é uma região tal que $VOL(A) = \int_A \cdots \int dx_1 \cdots dx_n > 0$, então uma medida de probabilidade uniforme P sobre A é definida de maneira natural por

$$P(B) = \int \cdots \int \frac{1}{VOL(A)} I_B(x_1, \dots, x_N) dx_1 \cdots dx_N = \frac{VOL(A \cap B)}{VOL(A)}, \quad B \in \mathcal{B}^N.$$

Isto expressa em certa forma que todos os pontos em A são "equiprováveis", pois, tem o mesmo peso relativo representado pela densidade, $\frac{1}{VOL(A)}$. Se $VOL(A) = 0$, a definição acima não é aplicável, por exemplo se $A_1 = \{(x_1, x_2) \in \mathbb{R}_+^2 : x_1 + x_2 = 1\}$. Neste caso uma medida de probabilidade uniforme sobre $\mathbb{R}^2$ concentrada em A pode ser definida pela integral de linha. De fato, utilizando-nos de representação paramétrica

$$\begin{aligned} \delta : [0, 1] &\longrightarrow \mathbb{R}^2 \\ t &\longrightarrow (t, 1 - t), \end{aligned}$$

temos que

$$\text{Comprimento}(A) = \int_{[0,1]} \sqrt{\delta_1'(t)^2 + \delta_2'(t)^2} dt = \sqrt{2},$$

onde, $\delta_1(t) = t$, $\delta_2(t) = 1 - t$. Assim, a função $P : \mathcal{B}^2 \rightarrow [0, 1]$ definida por

$$P(B) = \frac{\text{Comprimento}(A \cap B)}{\text{Comprimento}(A)} = \int_{\{t: \delta(t) \in B \cap (0,1)\}} dt, \quad B \in \mathcal{B}^2$$

é uma medida de probabilidade sobre $(\mathbb{R}^2, \mathcal{B}^2)$.

Em geral, se $A_s = \{(x_1, x_2) \in \mathbb{R}_+^2 : x_1 + x_2 = s\}$, $s > 0$, podemos entender como medida de probabilidade uniforme sobre A_s a função P_s definida por

$$P_s(B) = \frac{1}{s} \int_{\{t: \delta(t) \in B \cap (0,s)\}} dt, \quad B \in \mathcal{B}^2$$

Não é difícil ver que a função $Q(\cdot, \cdot) : \mathbb{R}^+ \times \mathcal{B}^2 \longrightarrow [0, 1]$, definida por $Q(s, B) = P_s(B)$ é uma função de transição de $(\mathbb{R}^+, \mathcal{B}_1^+)$ para $(\mathbb{R}^2, \mathcal{B}^2)$. Analogamente se $S_r = \{(x_1, x_2) \in \mathbb{R}^2 : x_1^2 + x_2^2 = r^2\}$, então a função P_r definida por

$$P_r(B) = \frac{1}{2\pi} \int_{\{t: \delta(t) \in B \cap [0, 2\pi]\}} dt, \quad B \in \mathcal{B}^2$$

onde, $\delta : [0, 2\pi] \longrightarrow \mathbb{R}^2$, é uma $t \longrightarrow (\cos t, \sin t)$ medida de probabilidade uniforme sobre S_r .

As funções de probabilidade acima mencionadas correspondem às definidas na Seção 4 do Capítulo 1. Uma analogia similar pode ser feita com as distribuições que foram definidas no Capítulo 2. Em geral, para definirmos medidas de probabilidade uniforme neste tipo de situações, podemos utilizar-nos de integrais de variedades (para aplicações veja Pereira (1985)). Uma outra maneira de proceder é através da teoria de medidas invariantes sobre grupos topológicos (Nachbin, (1960), Eaton, (1983), Diaconis, Eaton e Lauritzen (1988)). Porém, a menos de situações conhecidas, como por exemplo o grupo de transformações ortogonais e subgrupos destes (veja por exemplo, Eaton, (1983), Capítulos 6-7)), nem sempre é simples resolver o problema por este caminho. Por outro lado, como é mostrado ao longo do Capítulo 2, é possível resolver o problema utilizando-nos de ferramentas menos sofisticadas. Nesta seção abordamos o problema de predição desde o ponto de vista operacional. Os modelos estudados na Seção 4 do Capítulo 1 e da Seção 3 do Capítulo 2, são utilizados para fazer inferências a respeito das quantidades populacionais envolvidas. O principal resultado é o papel da distribuição de Pareto, enquanto distribuição a priori para a quantidade populacional de interesse. Mostraremos no que segue que tal distribuição permite-nos obter expressões explícitas para a distribuição a posteriori na maioria dos casos estudados. A ênfase é colocada na predição do valor máximo da característica estudada na população.

3.5.1 Predição do valor máximo da característica estudada na população

Consideramos aqui $\mathcal{X} = \mathbb{R}^+$ e $T_N = \max\{X_1, \dots, X_N\}$. A partição induzida por T sobre $\mathcal{X}$ é $\{\mathcal{X}_m^N; m > 0\}$, onde

$$\mathcal{X}_m^N = \{(x_1, \dots, x_N) \in \mathbb{R}_+^N : \max\{x_1, \dots, x_N\} = m\} .$$

A priori $X_1, X_2, \dots, X_N$ são quantidades aleatórias tomando valores em $\mathbb{R}^+$, com a propriedade que $X_1, X_2, \dots, X_N | T_N = m \sim U(\mathcal{X}_m^N)$. A distribuição uniforme sobre $\mathcal{X}_m^N$ foi definida e caracterizada na seção 3, do capítulo 3.

A seguir determinamos a distribuição a posteriori de T_N dado $\mathbf{x}_n = (x_1, \dots, x_n)$, supondo a priori que T_N possui distribuição de Pareto. Antes disto, vamos mostrar que $T_n = \max\{X_1, \dots, X_n\}$ é uma estatística suficiente preditiva para T_N , isto é, $(X_1, \dots, X_n) \amalg T_N \setminus T_n$. Para ver isto, seja P_θ^k a lei de k variáveis aleatórias independentes com distribuição comum $U(0, \theta)$ e lembremos a partir da observação 3 no Capítulo 2, que relativo a P_θ^k , a distribuição condicional de $X_1, \dots, X_k$ dado $T_k = m$ é uniforme sobre $\mathcal{X}_m^k$.

Seja f uma função mensurável, limitada então

$$E(f(X_1, \dots, X_n) | T_N = m, T_n = m_s) = E_\theta(f(X_1, \dots, X_n) | T_N = m, T_n = m_s) ,$$

onde E_θ denota o operador esperança relativo a P_θ^N .

Mas

$$\begin{aligned} & E_\theta(f(X_1, \dots, X_n) | T_N = m, T_n = m_s) \\ &= E_\theta(f(X_1, \dots, X_n) / \max\{m_s, X_{n+1}, \dots, X_N\} = m, T_n = m_s) \\ &= E_\theta(f(X_1, \dots, X_n) | T_n = m_s) . \end{aligned} \tag{18}$$

Esta última igualdade decorre da independência de $X_1, \dots, X_N$ relativo a P_θ^N . Logo

$$E(f(X_1, \dots, X_n) | T_N, T_n) = h(T_n) .$$

Consequentemente,

$$E(f(X_1, \dots, X_n) | T_n) = E(E(f(X_1, \dots, X_n) | T_n, T_N) | T_n) = h(T_n) .$$

Portanto, $(X_1, \dots, X_n) \parallel T_N | T_n$, e além disso, a partir da identidade (18), temos que $X_1, \dots, X_n | T_n = m_s \sim U(\mathcal{X}_{m_s}^n)$. Consequentemente, para obtermos a distribuição a posteriori de T_N dado $\mathbf{x}_n = (x_1, \dots, x_n)$ é suficiente desenvolver a distribuição de T_N dado $T_n = m_s$, onde $m_s = \max\{x_1, \dots, x_n\}$. No que segue, $T_N \sim \text{Pareto}(a, m_0)$, $a > 0$ e $m_0 > 0$, isto é, a distribuição de T_N é definida pela densidade

$$g_{m_0, a}(m) = a \frac{m_0^a}{m^{a+1}} I_{(m_0, +\infty)}^{(m)}.$$

Também, a partir da Proposição 3, no Capítulo 2 temos que a distribuição condicional de T_N dado $T_n = m$ é definida pela função de distribuição

$$F_m(m_s) = \begin{cases} 0 & \text{se } m_s < 0 \\ \left(1 - \frac{n}{N}\right) \left(\frac{m_s}{m}\right)^n & \text{se } 0 < m_s < m \\ 1 & \text{se } m_s \geq m. \end{cases} \quad (19)$$

Distribuição marginal de T_n

Para obtermos a distribuição de T_n , consideramos dois casos possíveis:

i) Se $m_s > m_0$, então

$$\begin{aligned} P(T_n \leq m_s) &= \int_{m_s < m} \left(1 - \frac{n}{N}\right) \left(\frac{m_s}{m}\right)^n g_{m_0, a}(m) dm + \int_{m \leq m_s} g_{m_0, a}(m) dm \\ &= \left(1 - \frac{n}{N}\right) a m_s^n m_0^a \int_{m_s}^{\infty} \frac{1}{m^{n+a+1}} dm + a m_0^a \int_{m_0}^{m_s} \frac{1}{m^{a+1}} dm \\ &= \left(1 - \frac{n}{N}\right) \frac{a}{n+a} \left(\frac{m_0}{m_s}\right)^a + \left(\frac{m_s^a - m_0^a}{m_s^a}\right) \end{aligned}$$

ii) Se $m_s \leq m_0$

$$P(T_n \leq m_s) = \int_{m > m_s} \left(1 - \frac{n}{N}\right) \left(\frac{m_s}{m_0}\right)^n g_{m_0, a}(m) dm + \int_{m \leq m_s} g_{m_0, a}(m) dm.$$

O segundo termo à direita na igualdade acima se anula, pois $g_{m_0, a}(m) = 0$, se $m < m_0$.

Logo

$$P(T_n \leq m_s) = \left(\frac{N-n}{N}\right) \frac{a}{n+a} \left(\frac{m_s}{m_0}\right)^n.$$

A partir disto temos que a distribuição marginal de T_n é definida pela função de distribuição

$$F_{m_0,a}(m_s) = \begin{cases} 0 & \text{se } m_s \leq 0 \\ \left(\frac{N-n}{N}\right) \frac{a}{n+a} \left(\frac{m_s}{m_0}\right)^n & \text{se } 0 < m_s \leq m_0 \\ \frac{N-n}{N} \frac{a}{n+a} \left(\frac{m_0}{m_s}\right)^a + \left(\frac{m_s^a - m_0^a}{m_s^a}\right) & \text{se } m_s > m_0 \end{cases}$$

Agora, F é contínua, portanto a distribuição marginal é definida pela densidade

$$f_{m_0,a}(m_s) = \begin{cases} \frac{N-n}{N} \frac{a}{n+a} n \frac{m_s^{n-1}}{m_0^n} & \text{se } 0 < m_s \leq m_0 \\ \frac{n}{N} \frac{N+a}{n+a} a \frac{m_0^a}{m_s^{a+1}} & \text{se } m_s > m_0 \\ 0 & \text{em caso contrário.} \end{cases} \quad (20)$$

Observação 6: Se $X_1, X_2, \dots, X_n, M$ são variáveis aleatórias tomando valores no R^+ tal que $X_1, X_2, \dots, X_N$, dado $M = m$ são condicionalmente independentes com distribuição comum $U(0, m)$ e M com distribuição de Pareto (a, m_0) , então a distribuição de $T_n = \max\{X_1, \dots, X_n\}$ é definida pela função de densidade

$$h_{m_0,a}(m_s) = \begin{cases} \frac{a}{n+a} n \frac{m_s^{n-1}}{m_0^n} & \text{se } 0 < m_s < m_0 \\ \frac{n}{n+a} a \frac{m_0^a}{m_s^{a+1}} & \text{se } m_s \geq m_0 \end{cases} \quad (21)$$

Notemos que a menos dos fatores $\left(1 - \frac{n}{N}\right)$ e $\left(1 + \frac{a}{N}\right)$, h coincide com f definida na identidade (20).

Esta relação não constitui novidade, tendo em conta a estimativa para a distância de variação total estabelecida nos Teoremas 5 e 6 do Capítulo 2. Porém, se consideramos a classe

$$\mathcal{U}_a = \{\mu \text{ medida de probabilidade sobre } \mathbb{R} : \mu \ll \lambda, \frac{d\mu}{d\lambda} = g_{m_0,a}, \text{ algum } m_0 > 0\}$$

e Q_{Nmn}, P_m^n como no Teorema 5 do Capítulo 2, então o valor exato para a distância de variação entre $P = \int P_m^n d\mu(m)$ e $Q = \int Q_{Nmn} d\mu(m)$, $u \in \mathcal{U}(a)$, pode ser encontrado. A proposição a seguir estabelece este valor.

Proposição 10: *Seja $Q = \int Q_{Nmn} d\mu(m)$, $P = \int P_m^n d\mu(m)$, $\mu \in \mathcal{U}_a$, a fixado. Então para cada $\mu \in \mathcal{U}_a$, $1 \leq n \leq N$*

$$\|P - Q\| = 2 \frac{a}{n+a} \frac{n}{N} .$$

Prova: Seja $T : \mathbb{R}_+^n \rightarrow \mathbb{R}$ tal que $T(x_1, \dots, x_n) = \max\{x_1, \dots, x_n\}$. Já que T é suficiente para $\{P, Q\}$, temos a partir das propriedades 2 e 3, da distância de variação que

$$\|P - Q\| = \|PT^{-1} - QT^{-1}\| = 2 \int (h_{m_0,a} - f_{m_0,a})^+ d\lambda ,$$

onde $f_{m_0,a}$ e $h_{m_0,a}$ são as densidades definidas nas identidades (20) e (21) respectivamente.

Mas

$$(h_{m_0,a}(m_s) - f_{m_0,a}(m_s))^+ = \begin{cases} \frac{n}{N} h_{m_0,a}(m_s) & \text{se } 0 < m_s < m_0 \\ 0 & \text{se } m_s \geq m_0 \end{cases}$$

Logo

$$\begin{aligned} \|P - Q\| &= 2 \int_0^{m_0} \frac{n}{N} h_{m_0,a}(m_s) dm_s \\ &= 2 \frac{a}{N} \frac{n}{n+a} \int_0^{m_0} n \frac{m_s^{n-1}}{m_0^n} dm_s = 2 \frac{n}{N} \frac{a}{n+a} , \end{aligned}$$

concluindo-se assim a prova. □

Convém observar que se deixamos variar a livremente ($a > 0$), então

$$\|P - Q\| \leq 2 \frac{n}{N} .$$

Este limite na verdade é um limite uniforme quando a medida μ na Proposição 10 é qualquer medida de probabilidade sobre $\mathbb{R}^+$, absolutamente contínua com relação à medida de Lebesgue. A proposição a seguir (análoga ao Teorema 6 do Capítulo 2) estabelece este resultado.

Proposição 11: Sejam Q, P duas medidas de probabilidade sobre $\mathbb{R}_+^n$ tal que

$Q = \int Q_{Nmn} d\mu(m)$ e $P = \int P_m^n d\mu(m)$. Se μ é uma medida de probabilidade sobre $\mathbb{R}^+$ absolutamente contínua com relação à medida de Lebesgue, então para $1 \leq n < N$,

$$\|P - Q\| \leq \frac{2n}{N}.$$

Prova: Seja $T : \mathbb{R}_+^n \rightarrow \mathbb{R}$ tal que $T(x_1, \dots, x_n) = \max\{x_1, \dots, x_n\}$. Denotemos por F e H as funções de distribuição de T relativas a Q e P , respectivamente, e por G a função de distribuição associada a μ tal que $G(m_0) = 0$, para algum $m_0 \geq 0$. Seja $L_m(\cdot)$ a função de distribuição de T relativa a P_m^n , isto é,

$$L_m(m_s) = \begin{cases} \left(\frac{m_s}{m}\right)^n & \text{se } 0 \leq m_s < m \\ 1 & \text{se } m \leq m_s. \end{cases}$$

Então

$$H(m_s) = \begin{cases} \int_{m > m_s} L_m(m_s) dG(m) & \text{se } 0 \leq m_s \leq m_0 \\ \int_{m > m_s} L_m(m_s) dG(m) + (G(m_s) - G(m_0)) & \text{se } m_s > m_0. \end{cases}$$

A partir disto e da identidade (19), temos que

$$F(m_s) = \begin{cases} \left(1 - \frac{n}{N}\right) H(m_s) & \text{se } 0 \leq m_s \leq m_0 \\ H(m_s) - \frac{n}{N} \int_{m > m_s} L_m(m_s) dG(m) & \text{se } m_s > m_0. \end{cases}$$

Agora, H e F são contínuas e T é suficiente para $\{P, Q\}$, então a partir das propriedades da distância de variação temos que

$$\|P - Q\| = \|PT^{-1} - QT^{-1}\| = \int |h - f| d\lambda = 2 \int (h - f)^+ d\lambda,$$

onde h e f são as derivadas de H e F , respectivamente.

Por outro lado,

$$f(m_s) = \begin{cases} \left(1 - \frac{n}{N}\right) h(m_s) & \text{se } 0 \leq m_s \leq m_0 \\ h(m_s) - \frac{n}{N} l(m_s) & \text{se } m_s > m_0, \end{cases}$$

onde $l(m_s) = \left[\int_{m > m_s} L_m(m_s) dG(m) \right]'$. Logo

$$\|P - Q\| = 2 \int (h - f)^+ d\lambda = 2 \frac{n}{N} \int_{m_s \leq m_0} h(m_s) d\lambda(m_s) \leq 2 \frac{n}{N} . \quad \square$$

A proposição acima é uma versão do Teorema 6 no Capítulo 2, quando $\mu \ll \lambda$. A estimativa para a distância de variação total reduziu-se à metade. Porém, quando $n = o(N)$, tal diferença não tem relevância.

Distribuição a posteriori de T_N

Como no caso da distribuição marginal de T_n , temos duas situações a considerar:

Caso 1: $m_0 \geq m_s$, ($n < N$).

Neste caso a distribuição a posteriori é absolutamente contínua com relação à medida de Lebesgue, já que F_m é contínua para $m_s < m$, logo

$$f(m|m_s) = \frac{f(m_s/m)g_{m_0,a}(m)}{f(m_s)} I_{(m_0,+\infty)}(m) = \frac{(n+a)m_0^{n+a}}{m^{n+a+1}} I_{(m_0,+\infty)}(m) .$$

Isto é, se $m_s \leq m_0$, então $T_N|T_n = m_s \sim \text{Pareto}(m_0, n+a)$ e a partir disto temos que

$$\begin{aligned} E(T_N|T_n = m_s) &= \frac{n+a}{n+a-1} m_0 , \\ V(T_N|T_n = m_s) &= \frac{n+a}{(n+a-1)^2(n+a-2)} m_0^2 , \end{aligned}$$

Também não é difícil ver que em $m = m_0$ a densidade a posteriori atinge o máximo. Os resultados acima mostram que se o máximo valor da amostra é menor do que o menor valor ao qual atribuímos densidade positiva a priori (a qual é decrescente) então a densidade a posteriori decresce mais lentamente (com relação à distribuição a priori), dando mais peso a valores em torno de m_0 quanto maior n .

Caso 2: $m_0 < m_s$

Para $m > m_s$ a distribuição de probabilidade a posteriori sobre intervalos da forma $(m_s, m]$ pode ser calculada através da função de densidade

$$\begin{aligned} f(m/m_s) &= \frac{f(m_s/m)g_{m_0,a}(m)}{f(m_s)} I_{(m_s,+\infty)}^{(m)} \\ &= \frac{(n+a)(N-n)}{N+a} \frac{m_s^{n+a}}{m^{n+a+1}} I_{(m_s,+\infty)}^{(m)}, \end{aligned}$$

já que F_m é contínua para $m > m_s$.

A partir disto temos que

$$\Pr \{T_N = m_s | T_n = m_s\} = \frac{n+a}{N+a}.$$

Consequentemente a função de distribuição a posteriori é definida por

$$F_{m_s}(m) = \begin{cases} 0 & \text{se } m < m_s \\ \frac{n+a}{N+a} & \text{se } m = m_s \\ \frac{n+a}{N+a} + \frac{N-n}{N+a} \left(1 - \left(\frac{m_s}{m}\right)^{n+a}\right) & \text{se } m > m_s, \end{cases} \quad (22)$$

onde $F_{m_s}(\cdot)$ denota a função de distribuição de $T_N | T_n = m_s$.

Após alguns cálculos obtemos também que

$$E(T_N | T_n = m_s) = \frac{n+a}{n+a-1} \frac{N+a-1}{N+a} m_s$$

e

$$V(T_N | T_n = m_s) = \frac{n+a}{N+a} \left(\frac{N+a-2}{n+a-2} - \frac{n+a}{(n+a-1)^2} \frac{(N+a-1)^2}{N+a} \right).$$

Como é de esperar, quanto mais próximo está n de N , a distribuição a posteriori concentra sua massa em m_s , porém se n é pequeno em relação a N , a massa concentra-se em valores maiores que m_s .

Notemos também, a partir da expressão (22) que se N é grande em relação a n , a distribuição a posteriori de T_N se aproxima da distribuição de Pareto $(m_s, n+a)$. A proposição a seguir estabelece formalmente este fato.

Proposição 12: *Seja P_{m_s} a lei definida por $F_{m_s}(\cdot)$ na expressão (22) e Q a lei de uma variável aleatória com distribuição de Pareto $(m_s, n + a)$. Então, se $n < N$*

$$||P_{m_s}(\cdot) - Q(\cdot)|| \leq 4 \frac{(n + a)}{N + a}$$

Prova:

$$P_{m_s}(A) = \int_{A \cap \{m > m_s\}} dP_{m_s}(m) + \int_{A \cap \{m = m_s\}} dP_{m_s}(m), \quad A \in \mathcal{B}_1.$$

Agora, se $m_s \notin A$, então

$$\begin{aligned} P_{m_s}(A) &= \int_{A \cap \{m > m_s\}} \frac{(n + a)(N - n)}{N + a} \frac{m_s^{n+a}}{m^{n+a+1}} dm \\ &= \frac{N - n}{N + a} \int_{A \cap \{m > m_s\}} (n + a) \frac{m_s^{n+a}}{m^{n+a+1}} dm = \frac{N - n}{N + a} Q(A). \end{aligned}$$

Se $m_s \in A$, então

$$P_{m_s}(A) = \frac{N - n}{N + a} Q(A) + \frac{n + a}{N + a}.$$

Logo, se $m_s \notin A$

$$|P_{m_s}(A) - Q(A)| = |Q(A) \left(1 - \frac{N - n}{N + a}\right)| \leq \frac{n + a}{N + a},$$

e se $m_s \in A$

$$|P_{m_s}(A) - Q(A)| = \left|Q(A) \frac{n + a}{N + a} + \frac{n + a}{N + a}\right| \leq 2 \frac{n + a}{N + a}.$$

Consequentemente,

$$|P_{m_s}(A) - Q(A)| \leq 2 \frac{n + a}{N + a}, \quad A \in \mathcal{B}_1.$$

Concluindo-se assim a prova. □

3.5.2 Predição de outras quantidades populacionais

Na seção 4 do Capítulo 1, foram discutidas as distribuições uniformes sobre os conjuntos

$$\text{i) } \mathcal{X}_t^N = \{(x_1, \dots, x_N) \in \mathbb{R}^N : \sum_{i=1}^N x_i^2 = t\}, t > 0$$

$$\text{ii) } \mathcal{X}_t^N = \{(x_1, \dots, x_N) \in \mathbb{R}_+^N : \sum_{i=1}^N x_i = t\}, t > 0$$

A seguir consideramos soluções para inferências sobre T_N em ambos os casos, baseados numa amostra de tamanho n ($n < N$) selecionada da população e supondo a priori que $X_1, \dots, X_N | T_N = t \sim U\{\mathcal{X}_t^N\}$ e $T_N \sim \text{Pareto}(a, m_0)$.

Predição do total populacional

Neste caso $X_1, X_2, \dots, X_N$ são quantidades aleatórias assumindo valores em $\mathbb{R}^+$. Interessa fazer inferências para T_N sob as condições acima mencionadas. Antes disto, e para simplificar os cálculos, convém notar que $T_n = \sum_{i=1}^n X_i$ é suficiente preditivo para T_N . Com efeito, segundo foi mostrado no Corolário 1 do Capítulo 1, se P_λ^N denota a lei de N variáveis aleatórias independentes com distribuição comum Exponencial (λ), então $X_1, \dots, X_N | T_N = t \sim U(\mathcal{X}_t^N)$. Assim, utilizando argumentos análogos aos do caso anterior, não é difícil ver que $T_n = \sum_{i=1}^n X_i$ ($n < N$) é suficiente preditivo para T_N . Portanto, para inferências sobre T_N , é suficiente obtermos a distribuição condicional de $T_N | T_n = t_s$, onde $t_s = \sum_{i=1}^n x_i$ é o total observado numa amostra de tamanho n . Agora, a partir da Proposição 10 do Capítulo 1, temos que a distribuição de $T_n | T_N = t$ tem densidade definida por

$$f(t_s/t) = \frac{\Gamma(N)}{\Gamma(N-n)\Gamma(n)} \left(1 - \frac{t_s}{t}\right)^{N-n-1} \left(\frac{t_s}{t}\right)^{n-1} \frac{1}{t} I_{(0,t)}(t_s) .$$

Isto é, $T_n | T_N = t \stackrel{\mathcal{D}}{=} tX$, onde $X \sim \text{Beta}(n, N-n)$. Logo, se g é uma densidade a priori para T_N , temos que a densidade a posteriori é definida por

$$f(t/t_s) \propto \left(1 - \frac{t_s}{t}\right)^{N-n-1} \left(\frac{t_s}{t}\right)^{n-1} \frac{1}{t} g(t) I_{[t_s, \infty)}(t) .$$

Seja g é a função de densidade Pareto (a, m_0). Neste caso temos duas situações a considerar:

Caso 1: Se $t_s \geq m_0$, então

$$f(t/t_s) \propto \left(1 - \frac{t_s}{t}\right)^{N-n-1} \left(\frac{t_s}{t}\right)^{n-1} \frac{1}{t^{a+1}} I_{[t_s, \infty)}(t) .$$

A constante de proporcionalidade, digamos K , é dada por

$$K = \int_{t \geq t_s} \left(1 - \frac{t_s}{t}\right)^{N-n-1} \left(\frac{t_s}{t}\right)^{n-1} \frac{1}{t^{a+1}} dt .$$

Fazendo $u = \frac{t_s}{t}$, obtemos

$$K = \frac{1}{t_s^a} \int_0^1 (1-u)^{N-n-1} u^{n+a-2} du = \frac{1}{t_s^a} \frac{\Gamma(N-n)(\Gamma(n+a-1))}{\Gamma(N+a-1)}$$

Logo

$$f(t/t_s) = \frac{\Gamma(n+a-1)}{\Gamma(N-n)\Gamma(n+a-1)} \left(1 - \frac{t_s}{t}\right)^{N-n-1} \left(\frac{t_s}{t}\right)^{n+a-1} \frac{1}{t} I_{(t_s, \infty)}(t) .$$

Notemos que $T_N/T_n = t_s \stackrel{D}{=} \frac{t_s}{U}$ onde $U \sim \text{Beta}(n+a, N-n)$. Assim, um intervalo de credibilidade com caudas simétricas e probabilidade $1 - \alpha$ é da forma

$$\left[\frac{t_s}{B_{1-\alpha/2}}, \frac{t_s}{B_{\alpha/2}} \right] ,$$

onde $B_\gamma > 0$ é tal que $P(U \leq B_\gamma) = \gamma$, $0 < \gamma < 1$. O preditor de máxima verossimilhança é dado por $\hat{t} = \frac{N+a-1}{n+a} t_s$ e o preditor que minimiza o risco em relação à perda quadrática é

$$\tilde{t} = E(T_N|t_s) = \frac{N+a-1}{n+a-1} t_s$$

com risco dado por

$$V(T_N|t_s) = \frac{(N+a-1)}{(n+a-1)^2} \frac{(N-n)}{(n+a-2)} t_s^2 .$$

Note que se $a = 1$, $m_0 > 0$, então $\tilde{t} = \frac{N}{n} t_s$ é o usual preditor do total populacional e $V(T_N|t_s) = N^2 \left(1 - \frac{n}{N}\right) \frac{t_s^2}{n-1}$.

Caso 2: Se $t_s < m_0$, então

$$f(t/t_s) \propto \left(1 - \frac{t_s}{t}\right)^{N-n-1} \left(\frac{t_s}{t}\right)^{n-1} \frac{1}{t^{a+1}} I_{(m_0, \infty)}(t) .$$

A constante de proporcionalidade é dada por

$$K = \int_{t \geq m_0} \left(1 - \frac{t_s}{t}\right)^{N-n-1} \left(\frac{t_s}{t}\right)^{n-1} \frac{1}{t^{a+1}} dt .$$

Fazendo $u = \frac{t_s}{t}$, obtemos

$$\begin{aligned} K &= \frac{1}{t_s^a} \int_0^{t_s/m_0} (1-u)^{N-n-1} u^{n+a-2} du \\ &= \frac{1}{t_s^a} \sum_{k=0}^{N-n-1} \binom{N-n-1}{k} (-1)^k \left(\frac{t_s}{m_0}\right)^{n+a+k-1} \frac{1}{n+a+k-1} \\ &= \frac{1}{t_s^a} \left(\frac{t_s}{m_0}\right)^{n+a-1} \sum_{k=0}^{N-n-1} \binom{N-n-1}{k} (-1)^k \left(\frac{t_s}{m_0}\right)^k \frac{1}{n+a+k-1} . \end{aligned}$$

Logo

$$f(t/t_s) = \frac{\left(1 - \frac{t_s}{t}\right)^{N-n-1} \left(\frac{t_s}{t}\right)^{n+a-1} \frac{1}{t}}{\left(\frac{t_s}{m_0}\right)^{n+a-1} \sum_{k=0}^{N-n-1} \binom{N-n-1}{k} (-1)^k \left(\frac{t_s}{m_0}\right)^k \frac{1}{n+a+k-1}} I_{(t_s, \infty)}(t) .$$

A partir disto temos que

$$E(T_N/t_s) = \frac{t_s \sum_{k=0}^{N-n-1} \binom{N-n-1}{k} (-1)^k \left(\frac{t_s}{m_0}\right)^{k-1} \frac{1}{n+a+k-2}}{\sum_{k=0}^{N-n-1} \binom{N-n-1}{k} (-1)^k \left(\frac{t_s}{m_0}\right)^k \frac{1}{n+a+k-1}}$$

e

$$E(T_N^2/t_s) = \frac{t_s^2 \sum_{k=0}^{N-n-1} \binom{N-n-1}{k} (-1)^k \left(\frac{t_s}{m_0}\right)^{k-2} \frac{1}{n+a+k-3}}{\sum_{k=0}^{N-n-1} \binom{N-n-1}{k} (-1)^k \left(\frac{t_s}{m_0}\right)^k \frac{1}{n+a+k-1}} .$$

O risco do preditor $\tilde{t} = E(T_N/t_s)$ é obtido a partir destas últimas expressões.

Observação 7: A condição $\mathcal{X} = \mathbb{R}^+$ não é restritiva. De fato, se $X_1, X_2, \dots, X_N$ são quantidades aleatórias assumindo valores em $\mathcal{X} = [a, +\infty)$, com $a \in \mathbb{R}$ conhecido, basta definir $T_{na} = \sum_{i=1}^n Y_i$ com $Y_i = X_i - a$ e proceder como antes. A distribuição preditiva a posteriori para $T_N = \sum_{i=1}^N X_i$ é obtida via translação.

Predição de somas de quadrados populacionais

Interessa fazer inferências a respeito de $T_{Na} = \sum_{i=1}^N (X_i - a)^2$, onde $a \in \mathbb{R}$, conhecido, supondo a priori que $X_1, X_2, \dots, X_N$ são quantidades aleatórias assumindo valores em $\mathbb{R}$ e

que $X_1, X_2, X_N/T_N = t \sim U(\mathcal{X}_t^N)$, onde $\mathcal{X}_t^N = \{(x_1, \dots, x_N) \in \mathbb{R}^N : \sum_{i=1}^N (x_i - a)^2 = t\}$, $t > 0$. Sem perda de generalidade assumimos $a = 0$. Também neste caso não é difícil ver que $T_n = \sum_{i=1}^n X_i^2$ ($n < N$) é suficiente preditiva para $T_N = \sum_{i=1}^N X_i^2$.

Agora, a partir da Proposição 13 do Capítulo 1, temos que a densidade de $T_n/T_N = t$ é definida por

$$f(t_s/t) = \frac{\Gamma\left(\frac{N}{2}\right)}{\Gamma\left(\frac{N-n}{2}\right)\Gamma\left(\frac{n}{2}\right)} \left(1 - \frac{t_s}{t}\right)^{\frac{N-n}{2}-1} \left(\frac{t_s}{t}\right)^{\frac{n}{2}-1} \frac{1}{t} I_{(0,t)}(t) .$$

Isto é, $T_n/T_N = t \stackrel{\mathcal{D}}{=} tX$, com $X \sim \text{Beta}\left(\frac{N-n}{2}, \frac{n}{2}\right)$. A partir disto temos que a densidade a posteriori de $T_N/T_n = t_s$ é dada por

$$f(t/t_s) \propto \frac{\Gamma\left(\frac{N}{2}\right)}{\Gamma\left(\frac{N-n}{2}\right)\Gamma\left(\frac{n}{2}\right)} \left(1 - \frac{t_s}{t}\right)^{\frac{N-n}{2}-1} \left(\frac{t_s}{t}\right)^{\frac{n}{2}-1} \frac{1}{t} g(t) I_{(0,t)}(t) ,$$

onde g é a densidade a priori para T_N . Se g é a distribuição de Pareto (a, m_0) temos dois casos a considerar.

Caso 1: Se $t_s \geq m_0$, então

$$f(t/t_s) = \frac{\Gamma\left(\frac{N}{2} + a\right)}{\Gamma\left(\frac{N-n}{2}\right)\Gamma\left(\frac{n}{2} + a\right)} \left(1 - \frac{t_s}{t}\right)^{\frac{N-n}{2}-1} \left(\frac{t_s}{t}\right)^{\frac{n}{2}+a-1} \frac{1}{t} I_{(t_s,+\infty)}(t) .$$

Notemos que $T_N/T_n = t_s \stackrel{\mathcal{D}}{=} \frac{t_s}{U}$, com $U \sim \text{Beta}\left(\frac{n}{2} + a, \frac{N-n}{2}\right)$. Assim, um intervalo de credibilidade com caudas simétricas e probabilidade $1 - \alpha$ é da forma

$$\left[\frac{t_s}{B_{1-\alpha/2}}, \frac{t_s}{B_{\alpha/2}} \right] ,$$

$B_\gamma > 0$ é tal que $P(U \leq B_\gamma) = \gamma$, $0 < \gamma < 1$.

O preditor que minimiza o risco com relação à função de perda quadrática e seu risco são dados por

$$E(T_N/t_s) = \frac{\Gamma\left(\frac{n}{2} + a - 1\right)}{\Gamma\left(\frac{n}{2} + a\right)} \frac{\Gamma\left(\frac{N}{2} + a\right)}{\Gamma\left(\frac{N}{2} + a - 1\right)} t_s$$

e

$$V(T_N/t_s) = t_s^2 \frac{\Gamma\left(\frac{N}{2} + a\right)}{\Gamma\left(\frac{n}{2} + a\right)} \left\{ \frac{\Gamma\left(\frac{n}{2} + a - 2\right)}{\Gamma\left(\frac{N}{2} + a - 2\right)} - \frac{\Gamma\left(\frac{N}{2} + a\right)}{\Gamma\left(\frac{n}{2} + a\right)} \frac{\Gamma^2\left(\frac{n}{2} + a - 1\right)}{\Gamma^2\left(\frac{N}{2} + a - 1\right)} \right\} .$$

E o preditor que maximiza a densidade a posteriori é dado por

$$\hat{t} = \frac{\frac{N}{2} + a}{\frac{n}{2} + a - 1} t_s .$$

Notemos que se n, N são números pares então

$$E(T_N/t_s) = \frac{\frac{N}{2} + a - 1}{\frac{n}{2} + a - 1} t_s$$

e

$$V(T_N/t_s) = \frac{\left(\frac{N}{2} + a - 1\right)}{\left(\frac{n}{2} + a - 1\right)^2} \frac{\left(\frac{N-n}{2}\right)}{\left(\frac{n}{2} + a - 2\right)} t_s^2 .$$

Caso 2: Se $t_s < m_0$, então

$$f(t/t_s) = \frac{\left(1 - \frac{t_s}{t}\right)^{\frac{N-n}{2}-1} \left(\frac{t_s}{t}\right)^{\frac{n}{2}+a} \frac{1}{t}}{B(t_s/m_0, \frac{N-n}{2}, \frac{n}{2} + a)} I_{(m_0, +\infty)}(t) ,$$

onde

$$B\left(\frac{t_s}{m_0}, p, q\right) = \int_0^{t_s/m_0} (1-u)^{p-1} u^{q-1} du .$$

Notemos que se $p \in \mathbb{N}$, então

$$B\left(\frac{t_s}{m_0}, p, q\right) = \sum_{k=0}^{p-1} \binom{p-1}{k} (-1)^k \left(\frac{t_s}{t}\right)^{q+k} \frac{1}{q+k} .$$

Neste caso o preditor que minimiza o risco em relação à função de perda quadrática é dado por

$$E(T_N/t_s) = t_s \frac{B\left(\frac{t_s}{m_0}, \frac{N-n}{2}, \frac{n}{2} + a - 1\right)}{B\left(\frac{t_s}{m_0}, \frac{N-n}{2}, \frac{n}{2} + a\right)}$$

e seu risco por

$$V(T_N/t_s) = t_s^2 \left\{ \frac{B\left(\frac{t_s}{m_0}, \frac{N-n}{2}, \frac{n}{2} + a - 2\right)}{B\left(\frac{t_s}{m_0}, \frac{N-n}{2}, \frac{n}{2} + a\right)} - \frac{B^2\left(\frac{t_s}{m_0}, \frac{N-n}{2}, \frac{n}{2} + a - 1\right)}{B^2\left(\frac{t_s}{m_0}, \frac{N-n}{2}, \frac{n}{2} + a\right)} \right\} .$$

Comentários

1. Os dois problemas de predição abordados aqui possuem a mesma estrutura. Em ambos os casos a densidade de $T_n/T_N = t$ (verossimilhança) é da forma

$$f(t_s/t) = \frac{\Gamma(p+q)}{\Gamma(p)\Gamma(q)} \left(1 - \frac{t_s}{t}\right)^{p-1} \left(\frac{t_s}{t}\right)^{q-1} \frac{1}{t} I_{(0,t)}(t_s) .$$

Isto é, $T_n/T_N = t \stackrel{D}{=} tX$, com $X \sim \text{Beta}(q, p)$. Se $T_N = \sum_{i=1}^N X_i$, então $p = N - n$, $q = n$. E se $T_N = \sum_{i=1}^N X_i^2$, então $p = \frac{N-n}{2}$, $q = \frac{n}{2}$. Assim, se $T_N \sim \text{Pareto}(a, m_0)$, e se $t_s \geq m_0$, então

$$f(t/t_s) = \frac{\Gamma(p+q+a)}{\Gamma(p)\Gamma(q+a)} \left(1 - \frac{t_s}{t}\right)^{p-1} \left(\frac{t_s}{t}\right)^{q+a} \frac{1}{t} I_{(t_s, +\infty)}(t) ,$$

Neste caso $T_N/T_n = t_s \stackrel{D}{=} \frac{t_s}{U}$, com $U \sim \text{Beta}(p, q+a+1)$. E se $t_s < m_0$ então

$$f(t_s/t) = \frac{\left(1 - \frac{t_s}{t}\right)^{p-1} \left(\frac{t_s}{t}\right)^{q+a} \frac{1}{t} I_{(m_0, +\infty)}(t)}{B\left(\frac{t_s}{m_0}, p, q+a\right)} .$$

2. Nestes casos utilizar as aproximações fornecidas pelos Teoremas 10 e 12 do Capítulo 1, não facilita os cálculos. Com efeito, se no primeiro caso tivéssemos assumido que a lei de $X_1, \dots, X_n | T_N = t$ pode ser bem aproximada por $P_{N/t}^n$ (onde P_λ^n denota a lei de n variáveis aleatórias independentes com distribuição comum Exponencial (λ)), então

$$f(t_s/t) \approx \frac{N^n}{\Gamma(n)} \left(\frac{t_s}{t}\right)^{n-1} \frac{1}{t} e^{-\frac{N}{t}t_s} I_{(0,t)}(t_s) .$$

Se $T_N \sim \text{Pareto}(a, m_0)$, então

$$f(t_s/t) \approx \frac{N^n}{\Gamma(n)} \left(\frac{t_s}{t}\right)^{n-1} \frac{1}{t^{a+2}} e^{-\frac{N}{t}t_s} I_{(t_s, +\infty)}(t_s) ,$$

em cujo caso a constante de proporcionalidade não tem fórmula explícita. É claro que podemos utilizar aproximações para o cálculo da função gama incompleta,

porém tal cálculo não se justifica já que com as distribuições exatas podemos obter a forma explícita da distribuição a posteriori, quando a distribuição a priori é Pareto.

Em termos de tratabilidade matemática, a distribuição gama invertida

$(\frac{N}{T_N} \sim \text{Gama}(\alpha, \beta))$ pode resultar atrativa. Neste caso

$$f(t/t_s) \propto \frac{N^n}{\Gamma(n)} \left(\frac{t_s}{t}\right)^{n-1} \frac{1}{t^{\alpha+2}} \frac{(\beta N)^\alpha}{\Gamma(\alpha)} e^{-\frac{N}{t}(t_s+\beta)} I_{(t_s,+\infty)}(t) .$$

Isto é, $\frac{N}{T_N} | T_n = t_s$ tem distribuição Gama $(n + \alpha, t_s + \beta)$ truncada. Porém, mesmo neste caso a distribuição a posteriori exata é dada por

$$f(t/t_s) = \frac{\left(1 - \frac{t_s}{t}\right)^{N-n-1} \left(\frac{\beta N}{t}\right)^{n+\alpha} \frac{1}{t} e^{-\beta N/t}}{\sum_{k=0}^{N-n-1} \binom{N-n-1}{k} (-1)^k \left(\frac{t_s}{\beta N}\right)^k P\left(n + \alpha + k, \frac{\beta N}{t_s}\right) \Gamma(n + \alpha + k)} I_{(t_s,+\infty)}(t) ,$$

onde

$$P\left(n + \alpha + k, \frac{\beta N}{t_s}\right) = P(X \geq n + \alpha + k) ,$$

com $X \sim \text{Poisson}\left(\frac{\beta N}{t_s}\right)$.

3.6 Comentários finais

Embora na Seção 3.3 tenha sido formulado um modelo geral de predição cujas análises ficaram restritas a modelos condicionalmente uniformes , induzidos por quantidades populacionais unidimensionais, o mesmo tipo de análise pode ser feito para outros modelos mencionados nos capítulos anteriores. Por exemplo, suponhamos que cada indivíduo na população pertence a uma das M categorias mutuamente exclusivas e interessa fazer inferências a respeito do vetor de frequências populacionais. Se na nossa opinião todas as possíveis configurações que podem gerar um valor dado do vetor de frequências populacionais são igualmente prováveis, então o modelo descrito na Proposição 6 do Capítulo 1

expressa tal opinião. A função de verossimilhança do vetor de frequências populacionais, dado o vetor de frequências amostrais, está implícita na identidade (3) do Capítulo 1. A discussão de estendibilidade associada a este modelo depende de soluções de problemas de momentos multidimensionais (Shohat, J.A. e Tamarkin, J.D. (1943)).

Com relação a problemas envolvendo variáveis auxiliares conhecidas, os modelos mencionados no final da seção 4 do Capítulo 1 fornecem a base para a predição de coeficientes de regressão populacionais. Tais modelos podem também ser estudados em termos de propriedades referentes a distribuições a priori que permitam a obtenção de expressões analíticas gerais. A análise não está restrita apenas a modelos localmente uniformes. Um exemplo interessante que fica fora deste contexto é o modelo Poissoniano (Wechsler, (1993)).

CAPÍTULO 4

Verossimilhança Preditiva em Populações Finitas. Uma Abordagem Não Bayesiana

4.1 Introdução

Como já foi mencionado na Seção 5 do Capítulo 1 e na introdução do capítulo anterior, predição em populações finitas tem sido estudado através da teoria de amostragem clássica e a abordagem de superpopulação.

Godambe (1966) e Basu (1969), responsáveis pelos desenvolvimentos formais da teoria de amostragem clássica (modelo de população fixada), mostraram que a estrutura do modelo clássico é fraca como para justificar os estimadores comumente usados. Uma das restrições do modelo clássico é que o modelo probabilístico utilizado não está relacionado diretamente à quantidade populacional de interesse.

Dentro da abordagem clássica de superpopulação a solução é fundamentalmente diferente. Nesta perspectiva (veja, por exemplo, Royall (1970b, 1971), Cassel, Sarda e Wretman (1977), Smith (1976)), as quantidades populacionais são consideradas como re-

alizações de variáveis aleatórias com distribuição conjunta F indexada por um parâmetro (ou vetor de parâmetros) θ , usualmente denominado de superparâmetro. Uma vez que a amostra é selecionada, a população (de tamanho N) é particionada naturalmente em duas partes: i) A quantidade a ser observada que representaremos como é usual e sem perda de generalidade, por $\mathbf{X}_n = (X_1, X_2, \dots, X_n)$, onde $n < N$, com $\mathbf{x}_n$ o valor observado de $\mathbf{X}_n$. ii) A quantidade não observada que será representada por $\mathbf{X}_{N-n} = (X_{n+1}, \dots, X_N)$. O interesse é fazer predições com relação a uma função de $\mathbf{X}_N$, digamos $T(\mathbf{X}_N)$, baseados nos valores observados $\mathbf{x}_n$ de $\mathbf{X}_n$ e o modelo F . Por exemplo, se F especifica que $X_1, X_2, \dots, X_N$ são independentes e $T(\mathbf{X}_N)$ é o total populacional, então a predição produz-se via inferências sobre θ a partir da parte de $\mathbf{X}_N$ que será observada. As predições são feitas normalmente em termos de preditores pontuais e propriedades destes.

Como vimos na Seção 1 do Capítulo 3, na abordagem bayesiana de superpopulação, uma distribuição a priori para o superparâmetro é especificada. A distribuição preditiva de $T(\mathbf{X}_N)$ dado $\mathbf{X}_n = \mathbf{x}_n$ é deduzida a partir disto e o modelo F . Por exemplo, se $p(\theta)$ é uma densidade a priori, então a função de densidade preditiva no ponto $T(\mathbf{X}_N) = t$ é

$$f(t/\mathbf{x}_n) = \int f(t/\mathbf{x}_n, \theta) p(\theta/\mathbf{x}_n) d\theta, \quad (1)$$

onde $p(\theta/\mathbf{x}_n)$ é a densidade a posteriori de θ . Em outras palavras, a função de densidade preditiva é uma média de $f(t/\mathbf{x}_n, \theta)$ sob a distribuição a posteriori.

Soluções do tipo acima mencionado, para predição em populações finitas sob o enfoque clássico, não parecem óbvias. De fato, as distribuições que poderiam ser utilizadas para predição $f(t|\theta)$, $f(t|\mathbf{x}_n, \theta)$ usualmente envolvem θ . Uma tentativa poderia ser substituir nestas funções o parâmetro θ por sua estimação de máxima verossimilhança. Tal solução assume implicitamente que o verdadeiro valor de θ é sua estimação e poderia não levar em conta a incerteza a respeito de θ .

Outra alternativa é considerar a distribuição condicional de $\mathbf{X}_n$ dado $T(\mathbf{X}_N) = t$, para cada valor t de T e fazer t agir como parâmetro. Dito de outra forma, se $f(\cdot|t, \theta)$ denota

a densidade condicional de $\mathbf{X}_n$ dado $T(\mathbf{X}_N) = t$, a família $\{f(\cdot|t, \theta) : t \in \mathcal{Y}, \theta \in \Theta\}$, com $\mathcal{Y}$ a imagem de T , pode ser considerado como um modelo, no sentido clássico, indexado pela quantidade de interesse. Consequentemente, procedimentos de inferência clássica podem ser utilizados. O objetivo deste capítulo é fornecer soluções não bayesianas para problemas de inferência em populações finitas a partir desta família de distribuições. Neste contexto, a quantidade populacional de interesse age como parâmetro e os dados $\mathbf{x}_n$ são considerados como gerados por esta distribuição condicional. A idéia está estreitamente relacionada com as desenvolvidas por Lauritzen (1974), Hinkley (1979) e Butler (1986). Estes autores abordaram o problema de predição sob modelos paramétricos, introduzindo funções de verossimilhança preditiva que não requerem estimação de θ . De uma maneira geral, estes trabalhos consideram a distribuição condicional das quantidades aleatórias dada uma estatística suficiente, definindo a partir desta a verossimilhança das quantidades a serem preditas dadas as quantidades observadas. Em particular, quando a quantidade a ser predita é uma estatística suficiente para o parâmetro, a verossimilhança preditiva, dadas as quantidades observadas, definida por Lauritzen (1974) e Hinkley (1979), é a deduzida pela função de densidade condicional das quantidades que serão observadas dada a estatística suficiente (a ser predita). Assim, sob condições de suficiência de $T(\mathbf{X}_N)$ para o superparâmetro, o modelo que será estudado neste capítulo constitui uma aplicação, no contexto de populações finitas, dos resultados estabelecidos por esses autores.

Também, se a quantidade populacional de interesse $T(\mathbf{X}_N)$ é suficiente para o superparâmetro, o modelo derivado da distribuição condicional de $\mathbf{X}_N$ dado $T(\mathbf{X}_N)$ pode ser considerado como a contraparte clássica do modelo de predição deduzido a partir da fórmula (1), sob a abordagem bayesiana de superpopulação. De fato, a densidade preditiva em (1) pode ser reescrita como

$$f(t|\mathbf{x}_n) \propto f(t)f(\mathbf{x}_n|t), \quad (2)$$

que pode ser interpretada como a operação bayesiana quando t é visto como parâmetro e

$f(\mathbf{x}_n|t)$ a verossimilhança de t . Notemos a partir da suficiência de T em relação a θ , que $f(\mathbf{x}_n|t)$ é obtida sem necessidade de integração e define também a função de densidade condicional de $\mathbf{X}_n$ dado $T(\mathbf{X}_N) = t$, quando o superparâmetro θ é considerado como uma constante desconhecida. A operação explicitada na fórmula (2) é também a operação realizada a partir do modelo de predição introduzido na Seção 3 do Capítulo 3, onde não existem parâmetros de superpopulação envolvidos. Mesmo assim, o modelo deduzido sob a perspectiva deste capítulo pode ser considerado, em um sentido técnico, como a contraparte clássica do modelo de predição definido no Capítulo 3, quando a quantidade populacional de interesse é suficiente para o superparâmetro. Isto é consequência dos resultados estabelecidos nas Seções 3.4.1 e 3.4.2 do Capítulo 3. Porém, quando a quantidade populacional de interesse não é suficiente para o superparâmetro, o modelo derivado por condicionamento não guarda relação alguma com o enfoque descrito no Capítulo 3 pois neste caso o modelo deduzido a partir da distribuição condicional é indexado tanto pela quantidade populacional de interesse quanto pelo superparâmetro.

Vários aspectos do problema de predição sob a perspectiva deste capítulo serão discutidos nas Seções 2 e 3, apresentando soluções para alguns modelos particulares de superpopulação. Os resultados, quando possível, serão contrastados com as situações tratadas no Capítulo 3.

Mostraremos que sob o enfoque deste capítulo o modelo normal padrão produz os mesmos resultados que são obtidos através da teoria de amostragem clássica. Estes resultados são comentados na Seção 4.

Vamos utilizar resultados obtidos em Iglesias, Sandoval e Pereira (1991).

4.2 Predição

Sejam $X_1, X_2, \dots, X_N$ quantidades aleatórias com distribuição conjunta F_θ , $\theta \in \Theta$.

Seja $\mathbf{X}_n = (X_1, X_2, \dots, X_n)$ a parte que será observada, onde $n < N$ e $\mathbf{X}_{N-n} = (X_{n+1}, \dots, X_N)$ a parte não observada. Interessa fazer predições com relação a $T(\mathbf{X}_N)$, utilizando as observações $\mathbf{x}_n$ de $\mathbf{X}_n$, a partir do modelo deduzido da distribuição condicional de $\mathbf{X}_n$ dado $T(\mathbf{X}_N)$. Com o objetivo de enfatizar o fato que a quantidade populacional age como parâmetro no modelo deduzido da distribuição condicional, utilizaremos letras gregas para denotar um valor particular da quantidade de interesse. Assim, τ denotará um valor particular de T e $f(\cdot|\tau, \theta)$ denotará a função de densidade condicional de $X_1, X_2, \dots, X_n$ dado $T(\mathbf{X}_N) = \tau$.

O novo modelo fica agora indexado por dois parâmetros τ e θ . Desta maneira e de acordo com o propósito deste estudo, o problema envolve inferências com relação a τ na presença de parâmetros “nuisance” (Basu (1977)).

Notemos que o processo de condicionar não é mais que um artifício para construir um modelo indexado por τ e procedermos dentro do contexto clássico para fazer inferências com relação a τ .

O parâmetro do modelo modificado será denotado por $\eta = (\tau, \theta)$ e a verossimilhança derivada deste modelo por $l(\eta/\mathbf{x}_n)$. As definições a seguir colocam o problema de predição como um problema de inferência clássica.

Definição 1: A função $l(\eta/\mathbf{x}_n)$ será denominada de verossimilhança preditiva. Seguindo o princípio de suficiência, denominamos também de função de predição a qualquer função proporcional a esta.

Definição 2: Se existe um ponto $\hat{\eta} = (\hat{\tau}, \hat{\theta})$ tal que

$$\sup_{\eta} l(\eta/\mathbf{x}_n) = l(\hat{\eta}/\mathbf{x}_n) ,$$

então $\hat{\tau}$ é o preditor de máxima verossimilhança de $T(\mathbf{X}_N)$.

Definição 3: Se existem funções $T_1 = T_1(\mathbf{X}_n)$ e $T_2 = T_2(\mathbf{X}_n)$ com valores observados t_1

e t_2 tal que $T_1 < T_2$ q.c e $\Pr \{(T_1, T_2) \ni \tau/\eta\} = \gamma$, então (t_1, t_2) é denominado intervalo preditivo de $T(\mathbf{X}_N)$ com $100\gamma\%$ de confiança.

O resultado a seguir caracteriza o papel de $T(\mathbf{X}_N)$, quando esta é uma estatística suficiente minimal sob o modelo original F_θ .

No que segue, utilizaremos a notação geral f ou g para funções de densidade. Por simplicidade, não faremos diferenças em dimensão nem em distribuição, pois estas estarão implícitas em cada caso.

Lema 1: *Se existem funções $T_s = T_s(\mathbf{X}_n)$ e $U = U(\mathbf{X}_{N-n})$ tal que (com relação a θ) T_s , U e $T(\mathbf{X}_N)$ são reduções suficientes mínimas de $\mathbf{X}_n$, $\mathbf{X}_{N-n}$ e $\mathbf{X}_N$, respectivamente, então*

i) $f(\mathbf{x}_n/\eta)$ depende de η só através de τ ;

ii) $f(\mathbf{x}_n/\eta) = g(t_s/\tau).h(\mathbf{x}_n)$, onde $g(t_s/\tau)$ é a densidade de $T_s/T(\mathbf{X}_N) = \tau$ no ponto t_s .

Prova: Segue por argumentos de suficiência e simples operações algébricas com densidades condicionais. □

Note que se neste lema $\mathbf{X}_n$ e $\mathbf{X}_{N-n}$ são estatisticamente independentes no modelo original e U é unicamente definida por T_s e $T(\mathbf{X}_N)$, então g define a verossimilhança preditiva de Hinkley (1979). Isto é, a verossimilhança preditiva de Hinkley é igual a $g(t_s/\tau)$ no ponto observado t_s . E mais ainda, sob as condições do lema, o preditor de máxima verossimilhança coincide com o preditor definido por Lauritzen (1974). Também neste caso não existe parâmetro a ser eliminado.

Outro aspecto interessante da verossimilhança modificada quando $f(\mathbf{X}_n/\eta) = f(\mathbf{X}_n/\tau)$ é que sob o ponto de vista bayesiano a fórmula 2 pode ser imediatamente deduzida sem

necessidade de integração. Mais ainda, nestas situações, será possível estabelecer comparações entre os resultados obtidos sob esta perspectiva e os obtidos no Capítulo 3.

O exemplo a seguir ilustra as idéias apresentadas até aqui.

Exemplo 1: Sejam $X_1, X_2, \dots, X_N$ variáveis aleatórias independentes com distribuição exponencial com parâmetro λ e média $1/\lambda$. O total populacional $T(\mathbf{X}_N) = \sum_{i=1}^N X_i$ e o total amostral $T_s(\mathbf{X}_n) = \sum_{i=1}^n X_i$ são reduções suficientes minimais de $\mathbf{X}_N$ e $\mathbf{X}_n$ em relação ao modelo original. Então, em virtude do Lema 1, o modelo modificado consiste da família de distribuição de probabilidade definida por

$$f(\mathbf{x}_n|\boldsymbol{\tau}) \propto f(t_s|\boldsymbol{\tau}) ,$$

onde

$$f(t_s|\boldsymbol{\tau}) = \frac{(\boldsymbol{\tau} - t_s)^{N-n-1}}{\boldsymbol{\tau}^{N-1}} t_s^{n-1} \frac{\Gamma(N)}{\Gamma(N-n)\Gamma(n)} I_{(0,\boldsymbol{\tau})}(s) \quad (3)$$

que não depende de λ . A verossimilhança preditiva pode ser escrita então como

$$l(\boldsymbol{\eta}|\mathbf{x}_n) \propto l(\boldsymbol{\tau}|\mathbf{x}_n) \propto \frac{(\boldsymbol{\tau} - t_s)^{N-n-1}}{\boldsymbol{\tau}^{N-1}} t_s^{n-1} I_{(0,\boldsymbol{\tau})}(s) . \quad (4)$$

O preditor de máxima verossimilhança deduzido a partir de (4) é $\hat{\boldsymbol{\tau}} = \frac{(N-1)}{n} \sum_{i=1}^N X_i$. Para obter o intervalo preditivo, notemos que sob o modelo modificado $Z = \frac{T_i}{\boldsymbol{\tau}}$ é uma quantidade pivotal com distribuição Beta($n, N-n$). A partir disto, um intervalo preditivo com $100\gamma\%$ de confiança é da forma

$$\left[\frac{t_s}{b}, \frac{t_s}{a} \right] \quad (5)$$

onde a e b são escolhidos de tal forma que $\Pr(a < Z < b) = \gamma$. Em particular, se $N = n + 1$, o intervalo preditivo de menor comprimento (no sentido de escolher a e b que façam $(1/a) - (1/b)$ mínimo) é

$$\left[t_s, \frac{t_s}{(1-\gamma)^{1/n}} \right] .$$

O exemplo acima ilustra o procedimento definido nesta seção e permite estabelecer comparações com os resultados derivados no Capítulo 3. De fato, a fórmula em (3) é exatamente a deduzida no primeiro caso estudado na seção 3.5.2, onde supomos a priori que

$$X_1, X_2, \dots, X_N | T(\mathbf{X}_N) = t \sim U\{(x_1, \dots, x_N) \in \mathbb{R}_+^N : \sum_{i=1}^N x_i = t\}, t > 0.$$

Nesse exemplo foi considerada uma distribuição a priori Pareto (a, m_0) , e quando o total observado na amostra resulta ser maior que m_0 , o valor que maximiza a densidade a posteriori é $\hat{t} = \frac{N+a-1}{n+a} t_s$. Quando $a = 0$, isto é, se aceitarmos uma distribuição a priori imprópria para $T(\mathbf{X}_N)$, ambos preditores coincidem. O intervalo de credibilidade de caudas simétricas deduzido na seção 3.5.2 é da forma

$$\left[\frac{t_s}{B_{1-\gamma/2}}, \frac{t_s}{B_{\gamma/2}} \right],$$

onde B_γ é tal que $P(U \leq B_\gamma) = \gamma$, com $U \sim \text{Beta}(n+1, N-n)$. Não é difícil ver que este intervalo tem menor comprimento (quando $a > 0$) que o correspondente intervalo de caudas simétricas em (5).

Como foi mencionado no início deste capítulo, nem sempre a quantidade de interesse $T(\mathbf{X}_N)$ é uma redução suficiente mínima de $\mathbf{X}_N$ em relação a θ . Como consequência, inferências na presença de parâmetros “nuisance” deverão ser feitas. Contudo, resultados interessantes podem ser obtidos sob outros tipos de simplificações. A seguir discutimos uma situação comum que permite obter soluções simples.

Suponhamos que o superparâmetro tem a representação $\theta = (\theta_S, \theta_U)$ e que para cada θ_U fixado, $T(\mathbf{X}_N)$ é uma redução suficiente de $\mathbf{X}_N$ em relação a θ_S , em outras palavras, $T(\mathbf{X}_N)$ é suficiente específico para θ_S (Basu (1977)). Neste caso, o número de parâmetros “nuisance” fica reduzido ao número de componentes de θ_U . Assim, sob suficiência específica o modelo modificado depende do parâmetro $\eta = (\tau, \theta)$ só através de $\eta^* = (\tau, \theta_U)$. Como antes, representaremos o preditor de máxima verossimilhança por $\hat{\tau}$.

A situação dual é quando existe uma redução suficiente minimal de $\mathbf{X}_n$ (em relação a θ), $\lambda = h(\mathbf{X}_N)$ tal que $h(\mathbf{X}_N) = (T(\mathbf{X}_N), V(\mathbf{X}_N))$. A função $l(\lambda|\mathbf{x}_n) = f(\mathbf{x}_n|\lambda)$ pode ser considerada como a verossimilhança preditiva onde, $\gamma = g(\mathbf{X}_N)$ é o parâmetro “nuisance” e $\tau = T(\mathbf{X}_N)$ a quantidade a ser estimada. O preditor de máxima verossimilhança será representada por $\tilde{\tau}$.

O próximo exemplo é padrão e incorpora as duas situações acima descritas.

Exemplo 2: Sejam $X_1, X_2, \dots, X_N$ variáveis aleatórias independentes com distribuição comum Normal com média e variância desconhecidas, representadas por $\theta_S = \mu$ ($\in \mathbb{R}$) e $\theta_U = \sigma^2$ ($\in \mathbb{R}^+$), respectivamente. Aqui, consideramos $T_s(\mathbf{X}_n) = \sum_{i=1}^n X_i$, $V_s(\mathbf{X}_n) = \sum_{i=1}^n X_i^2$ e $T(\mathbf{X}_N) = \sum_{i=1}^N X_i$, $V(\mathbf{X}_N) = \sum_{i=1}^N X_i^2$.

Notando que $\tau = T(\mathbf{X}_N)$ é suficiente específica em relação a μ sob o modelo original, podemos obter a primeira verossimilhança preditiva. Neste caso, a distribuição de $\mathbf{X}_n|\boldsymbol{\eta}^*$ é normal n -variada com média $\left(\frac{\tau}{N}\mathbf{1}_n\right)$ e matriz de covariâncias $\sigma^2\left(I_n - \frac{1}{N}J_n\right)$, onde $\mathbf{1}_n$ é o vetor n -variado com todas as componentes iguais à unidade, I_n a matriz identidade de ordem n e $J_n = \mathbf{1}_n\mathbf{1}_n^t$. Sob este modelo o preditor de máxima verossimilhança é $\hat{\tau} = \frac{N}{n} \sum_{i=1}^n X_i$.

Agora, dado que $h(\mathbf{X}_n) = \left(\sum_{i=1}^n X_i, \sum_{i=1}^n X_i^2\right)$ é uma redução suficiente minimal de $\mathbf{X}_N$ em relação a (μ, σ^2) , o modelo modificado alternativo é baseado na distribuição condicional de $T_s(\mathbf{X}_n)$, $V_s(\mathbf{X}_n)/T(\mathbf{X}_N) = \tau$, $V(\mathbf{X}_N) = \gamma$, pois (Lema 1) $(T(\mathbf{X}_N), V(\mathbf{X}_N))$ é uma redução suficiente para $\mathbf{X}_n$ (sob o modelo original) em relação a (μ, σ^2) . Depois de alguns cálculos, obtemos a verossimilhança de λ dado o valor observado (t_s, v_s)

$$l(\lambda|(t_s, v_s)) \propto \left(\gamma - v_s - \frac{(\tau - t_s)^2}{N - n}\right)^{\frac{(N-n-3)}{2}} \left(\gamma - \frac{\tau^2}{N}\right)^{\frac{-(N-3)}{2}}. \quad (6)$$

O preditor de máxima verossimilhança, $\tilde{\tau}$, coincide com $\hat{\tau} = \frac{N}{n} \sum_{i=1}^n X_i$. Se $V(\mathbf{X}_N)$ é

substituído por

$$V^*(\mathbf{X}_N) = \sum_{i=1}^N (X_i - \bar{X}_N)^2 ,$$

então obteríamos o mesmo preditor para τ , pois existe uma correspondência um a um entre $(T(\mathbf{X}_N), V^*(\mathbf{X}_N))$ e $(T(\mathbf{X}_N), V(\mathbf{X}_N))$.

O lema a seguir caracteriza a verossimilhança preditiva para o total populacional τ quando o modelo de superpopulação é da família exponencial.

Lema 2: *Sejam $X_1, X_2, \dots, X_N$ variáveis aleatórias estatisticamente independentes com função de densidade comum*

$$f(x/\theta) \propto \exp \left\{ \frac{\theta x - b(\theta)}{a(\phi)} + c(x, \phi) \right\} ,$$

onde as funções $a(\cdot)$, $b(\cdot)$, $c(\cdot)$ e ϕ são conhecidas. Então a verossimilhança preditiva do total populacional, τ dado $\mathbf{x}_n$ ($n < N$), é definida por

$$l(\tau/\mathbf{x}_n) \propto \exp\{h_1(t_s, \phi) + h_2(\tau - t_s, \phi) - h_3(\tau, \phi)\} , \quad (7)$$

onde t_s é o valor observado de $T_s(\mathbf{X}_n) = \sum_{i=1}^n X_i$

Prova: Segue do fato que $T(\mathbf{X}_N) = \sum_{i=1}^N X_i$ é uma redução suficiente para $\mathbf{X}_N$ em relação a θ sob o modelo original e de que $T_s(\mathbf{X}_n) = \sum_{i=1}^n X_i$ e $T(\mathbf{X}_N) - T_s(\mathbf{X}_n)$ são estatisticamente independentes sob o modelo original. $\square$

Notemos que se no lema acima ϕ é desconhecido, então $T(\mathbf{X}_N) = \sum_{i=1}^N X_i$ é suficiente específico para $\mathbf{X}_N$ em relação a θ sob o modelo original. Assim, o modelo modificado é indexado por $\eta = (\tau, \phi)$. Neste caso, os argumentos utilizados na prova do lema acima continuam valendo e

$$l(\tau, \phi/\mathbf{x}_n) \propto \exp\{h_1(t_s, \phi) + h_2(\tau - t_s, \phi) - h_3(\tau, \phi)\} .$$

Exemplos que envolvem a família exponencial são dados na próxima seção.

4.3 Predição do total populacional

Nesta seção apresentamos uma série de exemplos para predição do total populacional. Nestes exemplos modelos particulares de superpopulação serão modificados sob a perspectiva deste capítulo.

Como tem sido usual, a seguinte notação será utilizada ao longo desta seção.

$$T_s(\mathbf{X}_n) = X_1 + X_2 + \cdots + X_n \quad \text{e} \quad T(\mathbf{X}_N) = X_1 + X_2 + \cdots + X_N ,$$

onde $n < N$. O valor observado de T_s é representado por t_s .

Exemplo 3: Sejam $X_1, X_2, \dots, X_N$ variáveis aleatórias estatisticamente independentes com distribuição comum Normal com média μ desconhecida e variância conhecida c^2 . A distribuição condicional de $T_s / \sum_{i=1}^N X_i = \tau$ é normal com média $\left(\frac{n}{N}\right)\tau$ e variância $\left(1 - \frac{n}{N}\right)nc^2$. Consequentemente,

$$l(\tau/t_s) \propto \exp \left\{ \frac{-N}{2c^2n(N-n)} \left(t_s - \frac{n}{N}\tau \right)^2 \right\} . \quad (8)$$

A partir disto o preditor de máxima verossimilhança para o total populacional é $\hat{\tau} = \frac{N}{n}t_s$ e o intervalo preditivo para τ pode ser obtido notando que sob o modelo modificado

$$Z = \frac{T_s - \frac{n}{N}\tau}{c\sqrt{n\left[1 - \frac{n}{N}\right]}} ,$$

tem distribuição normal padrão, isto é, Z é uma quantidade pivotal para τ . Logo, o intervalo preditivo é da forma

$$\hat{\tau} \pm Nzc\sqrt{\frac{1}{n} - \frac{1}{N}} ,$$

onde z é tal que $P(-z < Z < z) = \gamma$.

O exemplo acima não envolve parâmetros “nuisance”. A seguir estudamos esta situação quando a variância é desconhecida.

Exemplo 4: Sejam $X_1, X_2, \dots, X_N$ variáveis aleatórias com distribuição comum Normal de média μ e variância σ^2 , ambas desconhecidas. Segundo foi estabelecido no Exemplo 2 da seção anterior, a distribuição de $\mathbf{X}_n / \sum_{i=1}^N X_i = \tau$ é normal n -variada com média $(\frac{\tau}{N} \mathbf{1}_n)$ e matriz de covariâncias $\sigma^2 (I_n - \frac{1}{N} J_n)$. Nesta situação, o preditor de máxima verossimilhança é dado por $\hat{\tau} = \frac{N}{n} t_s$, que coincide com o preditor obtido no exemplo acima, quando σ^2 é conhecido. Um intervalo de predição para τ pode ser obtido notando (através de resultados de independência entre formas lineares e quadráticas) que sob o modelo modificado

$$W = \frac{T_s - \frac{n}{N} \tau}{S \sqrt{n \left(1 - \frac{n}{N}\right)}},$$

onde $S^2 = \frac{1}{n-1} \sum_{i=1}^n \left(X_i - \frac{T_s}{n}\right)^2$, tem distribuição t -student com $n-1$ graus de liberdade. Assim, um intervalo preditivo para τ é da forma

$$\hat{\tau} \pm NwS \sqrt{\frac{1}{n} - \frac{1}{N}},$$

onde w é tal que $\Pr(-w < W < w) = \gamma$.

Notemos que o intervalo obtido coincide com o usado na amostragem clássica com amostra casual simples.

Os dois exemplos a seguir consideram o problema de predição do total populacional sob modelo de superpopulação normal com uma covariável.

Exemplo 5: Sejam $X_1, X_2, \dots, X_N$ variáveis aleatórias estatisticamente independentes com distribuição Normal, variância comum c^2 e média de X_i , para $i = 1, \dots, N$ igual a $\beta_0 + \beta_1 W_i$, onde W_i é um valor fixado de uma covariável W e β_0, β_1 são parâmetros desconhecidos. Neste caso, o total populacional τ é suficiente específico para β_0 . Se a única quantidade de interesse é o total populacional τ , então podemos utilizar o modelo modificado obtido a partir da distribuição condicional de $\mathbf{X}_n / \sum_{i=1}^n X_i = \tau$, que é normal

n -variada com média igual a

$$\frac{\tau}{N} \mathbf{1}_n + \beta_1(\mathbf{W} - \bar{W}_N \mathbf{1}_n) ,$$

matriz de variâncias igual a

$$c^2 \left(I_n - \frac{1}{N} J_n \right) ,$$

onde $\mathbf{1}_n$, I_n e J_n foram definidos no Exemplo 2 e

$$\mathbf{W} = (W_1, \dots, W_n)^t , \quad \bar{W}_N = \frac{1}{N} \sum_{i=1}^N W_i ,$$

isto é, $\bar{W}_N$ é a medida populacional dos W_i 's. Da mesma maneira, $\bar{W}_n = \frac{1}{n} \sum_{i=1}^n W_i$ é a média amostral de $W_1, \dots, W_n$. Notemos que o modelo modificado pode ser reescrito na forma

$$X_i = \frac{\tau}{N} + \beta_1(W_i - \bar{W}_N) + \varepsilon_i , \quad i = 1, \dots, n ,$$

onde $\boldsymbol{\varepsilon} = (\varepsilon_1, \dots, \varepsilon_n) \sim N(o_n, c^2 (I_n - \frac{1}{N} J_n))$.

A partir disto e depois de algumas operações algébricas, obtemos

$$\hat{\beta}_1 = \frac{\sum_{i=1}^n (W_i - \bar{W}_N)(X_i - \frac{t_s}{n})}{\sum_{i=1}^n (W_i - \bar{W}_N)^2} \quad (9)$$

e

$$\hat{\tau} = \frac{N}{n} t_s - N \hat{\beta}_1 (\bar{W}_n - \bar{W}_N) . \quad (10)$$

Um intervalo de predição para τ pode ser obtido notando que sob o modelo modificado

$$Z = \frac{\hat{\tau} - \tau}{N c \sqrt{\frac{R(W)}{n} - \frac{1}{N}}} ,$$

tem distribuição normal padrão, onde

$$R(W) = \frac{\sum_{i=1}^n (W_i - \bar{W}_N)^2}{\sum_{i=1}^n (W_i - \bar{W}_n)^2} .$$

Exemplo 6: Suponhamos que no exemplo anterior, além de estarmos interessados na predição do total populacional, interessa também fazer predição para

$$R = \sum_{i=1}^N X_i W_i .$$

Denotaremos por ρ um valor de R .

O vetor $(\sum_{i=1}^N X_i, \sum_{i=1}^N X_i W_i)$ é uma estatística suficiente minimal para o modelo original em relação a (β_0, β_1) . Seja $R_s = \sum_{i=1}^n X_i W_i$, então, notando que (T_s, R_s) é uma redução suficiente minimal da amostra (sob o modelo original), temos pelo Lema 1 que a verossimilhança preditiva pode ser obtida a partir da distribuição condicional de $T_s, R_s | \sum_{i=1}^N X_i = \tau, \sum_{i=1}^N X_i W_i = \rho$, a qual também tem distribuição Normal. Para descrever a média e a matriz de covariâncias desta distribuição, introduzimos a seguinte notação:

$$S_n = \frac{1}{n} \sum_{i=1}^n W_i^2, \quad S_N = \frac{1}{N} \sum_{i=1}^N W_i^2, \quad \Sigma_{11} = \Sigma_{12} = n \begin{pmatrix} \frac{1}{\bar{W}_n} & \bar{W}_n \\ \bar{W}_n & S_n \end{pmatrix}$$

$$\Sigma_{22} = \begin{pmatrix} \frac{1}{\bar{W}_N} & \bar{W}_N \\ \bar{W}_N & S_N \end{pmatrix} \quad \text{e} \quad \Sigma = \Sigma_{11} - \Sigma_{12} \Sigma_{22}^{-1} \Sigma_{12} .$$

A distribuição condicional de $T_s, R_s | \sum_{i=1}^N X_i = \tau, \sum_{i=1}^N X_i W_i = \rho$ é normal com média μ e matriz de covariâncias Σ , onde

$$\mu = \frac{n}{N(S_N - \bar{W}_N^2)} \begin{pmatrix} S_N - \bar{W}_N \bar{W}_n & \bar{W}_n - \bar{W}_N \bar{W}_n \\ \bar{W}_n S_N - \bar{W}_N S_n & S_n - \bar{W}_N \bar{W}_n \end{pmatrix} \begin{pmatrix} \tau \\ \rho \end{pmatrix} = B \begin{pmatrix} \tau \\ \rho \end{pmatrix} .$$

Claramente, o preditor de máxima verossimilhança de $(\tau, \rho)^t$ é dado por $B^{-1}(T_s, R_s)^t$. Não é difícil ver depois de algumas operações algébricas que a primeira componente deste vetor coincide com a expressão (8). Uma região de predição para (τ, ρ) pode ser deduzida da maneira usual, como será mostrado no próximo exemplo.

Nos exemplos anteriores consideramos modelos de superpopulação Normais associados a variáveis aleatórias univariadas. Em muitas situações práticas, associada à i -ésima unidade da população temos um vetor $\mathbf{X}_i = (X_{i1}, \dots, X_{ik})$ e o interesse pode ser obter um preditor para o vetor $\mathbf{T}_N = \sum_{i=1}^N \mathbf{X}_i$. Denotaremos por τ um valor particular de

$\mathbf{T}_N$. Notemos que τ é agora um vetor k -dimensional. Para uma amostra de tamanho n ($n < N$) o vetor $\mathbf{T}_s = \sum_{i=1}^n \mathbf{X}_i$ denotará o vetor de totais na amostra. Neste caso, o modelo modificado decorrente do modelo de superpopulação assumido consiste da família de distribuições condicionais de $\mathbf{X}_1, \dots, \mathbf{X}_n | \mathbf{T}_N = \tau$.

O exemplo a seguir ilustra uma dessas situações.

Exemplo 7: Sejam $\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_N$ vetores aleatórios estatisticamente independentes com distribuição normal k -variada com média comum μ desconhecida e matriz de covariâncias comum Σ conhecida. Notando que $\mathbf{T}_N = \sum_{i=1}^N \mathbf{X}_i$ é uma redução suficiente minimal de $(\mathbf{X}_1, \dots, \mathbf{X}_N)$ em relação a μ e que $\mathbf{T}_s = \sum_{i=1}^n \mathbf{X}_i$ é uma redução suficiente minimal de $(\mathbf{X}_1, \dots, \mathbf{X}_n)$ em relação a μ , temos a partir do Lema 1, que a verossimilhança preditiva pode ser obtida a partir da distribuição condicional de $\mathbf{T}_s | \mathbf{T}_N = \tau$. Esta distribuição é normal multivariada com vetor de médias $\left[\frac{n}{N}\right] \tau$ e matriz de covariâncias $\left[1 - \frac{n}{N}\right] n \Sigma$. A partir disto temos que o preditor de máxima verossimilhança é $\hat{\tau} = \frac{N}{n} \mathbf{T}_s$ e uma região preditiva para τ é dada por

$$\left\{ \tau : \frac{N}{(N-n)n} \left(\mathbf{T}_s - \frac{n}{N} \tau \right) \Sigma^{-1} \left(\mathbf{T}_s - \frac{n}{N} \tau \right) \leq \chi^2 \right\} ,$$

onde χ^2 é o valor da distribuição qui-quadrado com k graus de liberdade que produz $100\gamma\%$ de confiança.

Nos exemplos acima, com excessão dos casos que consideram covariáveis, o preditor de máxima verossimilhança é o usual estimador do total populacional utilizado na amostragem clássica.

O exemplo a seguir ilustra uma situação diferente.

Exemplo 8: Sejam $X_1, X_2, \dots, X_N$ variáveis aleatórias estatisticamente independentes com distribuição comum Gama (a, β) , onde a é conhecido e $\beta > 0$, desconhecido.

A verossimilhança preditiva para o total populacional é obtida a partir da distribuição

condicional de $T_s / \sum_{i=1}^N X_i = \tau$, produzindo

$$l(\tau/t_s) \propto \left(1 - \frac{t_s}{\tau}\right)^{(N-n)a-1} \left(\frac{t_s}{\tau}\right)^{na}.$$

O preditor de máxima verossimilhança é $\hat{\tau} = (N - a^{-1}) \frac{t_s}{na}$. Para obter um intervalo de predição para τ com $100\gamma\%$ de confiança, notamos que sob o modelo modificado $W = \frac{T_s}{\tau}$ tem distribuição Beta($na, (N - n)a$). A partir disto, um intervalo preditivo para τ é da forma

$$\left[\frac{t_s}{d}, \frac{t_s}{c}\right],$$

onde c e d podem ser escolhidos de tal forma que

$$\frac{1}{c} - \frac{1}{d} \text{ é mínimo sob a restrição } \int_c^d f(w) dw = \gamma.$$

Para finalizar a seção, apresentamos exemplos associados a modelos de superpopulação discretos.

Exemplo 9: Sejam $X_1, X_2, \dots, X_N$ variáveis aleatórias estatisticamente independentes com distribuição comum $B(1, \theta)$. Neste caso, a distribuição condicional de $X_1, X_2, \dots, X_n$ dado $\sum_{i=1}^N X_i = \tau$ coincide com a obtida na seção 3.4.3 para análise de dados binários. Assim, o preditor de máxima verossimilhança coincide com o obtido nesta seção, sob priori uniforme para o total populacional. Isto é, $\hat{\tau}$ é um inteiro no intervalo

$$\left[(N+1)\frac{t_s}{n} - 1, (N+1)\frac{t_s}{n}\right].$$

Notemos a partir da Proposição 4 do Capítulo 1, que o total $T = \sum_{i=1}^N X_i$ é uma estatística de resumo para a família de distribuições permutáveis sobre $\{0, 1\}^N$. Assim, o modelo modificado é comum a esta família. Notemos também a partir do Teorema 6 do Capítulo 1, que quando N é suficientemente grande e n pequeno em relação a N , o modelo modificado pode ser bem aproximado pelo modelo associado a n variáveis aleatórias independentes com distribuição comum Binomial (τ/N) .

O modelo deduzido no exemplo acima pode ser visto como a versão não bayesiana, do modelo estudado na seção 3.4.3 para análise de dados binários. Uma situação similar ocorre quando consideramos $X_1, X_2, \dots, X_N$ estatisticamente independentes com distribuição comum Geométrica (θ) com média $(1 - \theta)/\theta$. A verossimilhança preditiva a partir da distribuição condicional de $T_s / \sum_{i=1}^N X_i = \tau$ é a deduzida na seção 3.2.4 para análise de dados inteiros, sob a avaliação de que

$$X_1, X_2, \dots, X_N | \sum_{i=1}^N X_i = t \sim U((x_1, \dots, x_N) \in \mathbb{N}_0^N : \sum_{i=1}^N x_i = t), t \in \mathbb{N}_0.$$

Então, o preditor de máxima verossimilhança $\hat{\tau}$, para este caso, coincide com aquele que maximiza a densidade a posteriori na seção 3.4.3, sob priori imprópria para o total populacional. Isto é, $\hat{\tau}$ é um inteiro no intervalo

$$\left[(N-1) \frac{t_s}{N} - 1, (N-1) \frac{t_s}{N} \right].$$

Também neste caso o modelo modificado pode ser visto como a versão não bayesiana do modelo considerado na seção 3.2.4 para análise de dados inteiros. Segue também, a partir do Teorema 8 do Capítulo 1, que, para valores de N grande e n pequeno em relação a N , o modelo modificado pode ser bem aproximado pelo modelo associado a n variáveis aleatórias independentes com distribuição comum Geométrica $\left(\frac{t}{N+t}\right)$.

Exemplo 10: Sejam $X_1, X_2, \dots, X_N$ variáveis aleatórias estatisticamente independentes com distribuição comum Poisson (θ). A verossimilhança preditiva obtida da distribuição condicional de $T_s | \sum_{i=1}^N X_i = \tau$ é binomial com parâmetros τ e $\frac{n}{N}$. O preditor de máxima verossimilhança é um inteiro no intervalo

$$\left[N \frac{t_s}{n} - 1, N \frac{t_s}{n} \right].$$

A verossimilhança coincide com aquela obtida a partir da avaliação preditivista que $X_1, X_2, \dots, X_N | \sum_{i=1}^N X_i = t \sim \text{Multinomial}(t, \frac{1}{N} \mathbf{1}_N)$, onde $\mathbf{1}_N$ é o vetor de 1's N -

dimensional. No caso de priori imprópria para o total populacional, o preditor que maximiza a densidade a posteriori coincide com $\hat{\tau}$.

4.4 Comentários

Na seção anterior foi ilustrado através de vários exemplos, um procedimento para fazer inferências em populações finitas utilizando-nos de métodos de inferência estatística clássica. Como foi ilustrado no Exemplo 2, a quantidade populacional de interesse nem sempre é uma redução suficiente para $X_1, X_2, \dots, X_N$ em relação ao superparâmetro. Desta maneira, o problema de inferência envolve parâmetros “nuisance”. Alternativamente, pode ser considerada uma redução suficiente a qual não é necessariamente a quantidade de interesse, porém esta última pode ser expressa como função dessa redução dos dados. Neste caso as inferências sob o modelo modificado estão relacionadas a funções do parâmetro. O exemplo a seguir ilustra esta situação.

Exemplo 11: Sejam $X_1, X_2, \dots, X_N$ variáveis aleatórias com distribuição comum $F \in \mathcal{F}$, onde $\mathcal{F}$ é a família de funções de distribuição sobre a reta. Neste caso o superparâmetro F é infinito dimensional, contudo o vetor de estatísticas de ordem $\mathbf{Y}_N = (X_{11}, \dots, X_{NN})$ é uma redução suficiente de $X_1, X_2, \dots, X_N$ em relação a F . Seja $\boldsymbol{\theta} = (\theta_1, \dots, \theta_N)$ um valor particular de $\mathbf{Y}_N$ e M o número de componentes diferentes em $\boldsymbol{\theta}$ com multiplicidades $N_1, N_2, \dots, N_M$, respectivamente. Então o modelo modificado é

$$P\{X_1 = x_1, \dots, X_n = x_n / \mathbf{Y}_N = \boldsymbol{\theta}\} = \frac{\prod_{i=1}^m n_i!}{n!} \frac{\prod_{i=1}^m \binom{N_i}{n_i}}{\binom{N}{n}} \prod_{i=1}^m I_{\{\theta_1, \dots, \theta_N\}}(x_i), \quad (11)$$

onde m ($m \leq M$) é o número de x_i 's diferentes com multiplicidades $n_1, n_2, \dots, n_m$, respectivamente, e $\sum_{i=1}^M N_i = N$, $\sum_{i=1}^m n_i = n$.

Não é difícil ver a partir da fórmula (11) que

$$E(T_s|\theta) = \sum_{i=1}^N \theta_i \quad \text{e} \quad V(T_s|\theta) = \left(1 - \frac{n}{N}\right) \sum_{i=1}^N \frac{(\theta_i - \bar{\theta})^2}{(N-1)n}, \quad (12)$$

onde $T_s = \sum_{i=1}^N X_i$ e $\bar{\theta} = \frac{1}{N} \sum_{i=1}^N \theta_i$.

As expressões em (12) são de fato os resultados deduzidos do modelo de população fixada (seção 1.5, fórmula 22), com amostragem casual simples. Também, a verossimilhança em (11) é proporcional à deduzida por Royall (1968) a partir do modelo de população fixada quando os rótulos são ignorados e a amostragem é casual simples.

O modelo em (11) (a menos de uma constante de proporcionalidade que depende do plano amostral) foi também deduzido por Sudgen (1982) a partir do modelo de população fixada. Dentro dessa perspectiva, Sudgen considerou uma reparametrização do vetor de quantidades populacionais, observando que existe uma correspondência biunívoca entre este vetor e as seguintes três componentes: o conjunto de rótulos, os valores ordenados do vetor de quantidades populacionais e uma função que associa a cada indivíduo na população um rótulo. Utilizando-se da suposição que os rótulos são alocados aleatoriamente, elimina a terceira componente por integração deduzindo a partir disto o modelo em (11). Sudgen fornece uma ampla discussão em relação a suficiência e completividade do vetor de estatísticas de ordem associada à amostra $\mathbf{X}_n$, sob planejamentos amostrais não informativos, mostrando por exemplo que a média amostral é um estimador ótimo (no sentido de variância uniformemente mínima) da média populacional.

O modelo em (11) é também denominado modelo de permutação aleatória, o qual é deduzido a partir da suposição que os valores associados às unidades na população são uma permutação aleatória de um conjunto de números desconhecidos. Note também que este modelo pode ser deduzido a partir da avaliação de permutabilidade sobre $X_1, \dots, X_N$. De fato, em todos os enfoques acima mencionados está implícita a noção de permutabilidade.

Através deste exemplo e os apresentados nas seções anteriores, resgatamos alguns resultados da teoria de amostragem clássica. Contudo, o artifício de tratar a quan-

tidade populacional como constante fixa (desconhecida) depois do condicionamento, é questionável mesmo dentro de um contexto não bayesiano, quando o modelo original é de superpopulação. Com efeito, sob o modelo de superpopulação a quantidade populacional é de fato aleatória. Através do processo de condicionamento perdemos a informação em relação a esta, contida no modelo de superpopulação. Suponhamos, por exemplo, que $X_1, X_2, \dots, X_N$ são variáveis aleatórias independentes com distribuição comum normal de média zero e variância desconhecida. O modelo modificado derivado da distribuição condicional de $X_1, X_2, \dots, X_n | \sum_{i=1}^N X_i = \tau$, neste caso, não se altera se no lugar de média zero tivéssemos assumido média um ou qualquer outro valor.

Poderíamos argumentar que a quantidade populacional neste caso está estreitamente relacionada com a média do modelo de superpopulação e portanto a informação em relação à média do modelo de superpopulação está contida na amostra. Porém, isto não está explicitado no modelo modificado. Assim, este argumento poderia estar sujeito às mesmas críticas feitas sobre os métodos baseados em aleatorização, pois o modelo modificado não reflete inteiramente as relações entre as uniades da população induzidas pelo modelo de superpopulação.

Neste capítulo, modelos indexados pelas quantidades de interesse foram gerados a partir dos modelos de superpopulação. Alternativamente, modelos indexados pela quantidade de interesse podem ser construídos sem partir de um modelo de superpopulação. Esta perspectiva foi considerada recentemente, para alguns modelos discretos, em Irony e Pereira (1992).

CONCLUSÕES E COMENTÁRIOS

Ao longo do Capítulo 1, vimos como alguns modelos paramétricos podem ser interpretados a partir julgamentos de invariância que expressam opiniões preditivas sobre seqüências de variáveis aleatórias.

O Teorema de de Finetti (Teorema 2, Capítulo 1) garante que a lei de uma seqüência infinita de variáveis aleatórias permutáveis pode ser representada como mistura de medidas produtos. O teorema é essencialmente um resultado de existência, não fornecendo necessariamente a forma funcional da medida na representação, a menos que a estrutura do espaço de estados do processo seja muito simples, isto é, finita. Quando a estrutura do espaço de estados não é simples, são necessárias condições mais fortes que permutabilidade para obter a forma funcional do modelo na representação. No caso de seqüências finitas de variáveis aleatórias permutáveis $\{X_k : 1 \leq k \leq N\}$, a lei da seqüência admite representação como mistura de distribuições de permutação aleatória (Proposição 1, Capítulo 1), ficando como único elemento de incerteza inicial a medida que expressa incerteza a respeito do vetor de estatísticas de ordem da seqüência. Embora a forma fun-

cional do modelo na representação seja explícita, o espaço paramétrico é N -dimensional. Contudo, como no caso infinito, restrições sobre o espaço de estados ou condições mais fortes que permutabilidade reduzem a dimensão do espaço paramétrico e simplificam a forma funcional do modelo na mistura. O modelo na representação não é necessariamente uma medida produto. No entanto, nas situações examinadas no Capítulo 1, a lei de subcoleções $\{X_k : 1 \leq k \leq n\}$ da sequência finita, ($n < N$) foi representada como quase mistura (formas finitas de teoremas do tipo de Finetti) de medidas produto cujo parâmetro é uma função de $\{X_k : 1 \leq k \leq N\}$.

A representação finita, juntamente com a lei forte dos grandes números e a convergência uniforme do modelo na representação finita, foram os elementos básicos para se obter versões infinitas na seção 3 do Capítulo 1. Contudo, como foi notado nesse capítulo, existem outras maneiras de se obter tais versões. A vantagem do procedimento utilizado na seção 3 desse capítulo (devido a Heath e Suderth, (1977)) é que além de utilizar-se ferramentas básicas, evidencia o sentido dos parâmetros que aparecem na representação. Isto é notável no caso 0-1, onde a relação entre propensões e probabilidade – que de Finetti entende como rigorosamente subjetiva – é evidenciada.

O procedimento acima descrito tem sido utilizado recentemente por Barlow e Spizzichino (1992), Hayakawa (1992) e Wechsler (1993), para obter versões finitas e infinitas de teoremas do tipo de Finetti envolvendo outros membros da família exponencial. Formas finitas dos teoremas não foram consideradas por esses autores.

As formas finitas também podem ser utilizadas para se obter a versão infinita, como foi ilustrado na seção 1.4. Contudo, formas finitas possuem interesse próprio quando a sequência finita não é estendível.

No Capítulo 2, mostramos como modelos paramétricos uniformes sobre a reta (dependendo de um ou dois parâmetros) podem ser compreendidos dentro da perspectiva preditivista.

Na seção 2.2, a discussão gira em torno de seqüências de variáveis aleatórias $\{X_k : 1 \leq k \leq N\}$ assumindo valores em $\mathbb{N}_0$ e satisfazendo a propriedade de que a distribuição condicional de $X_1, \dots, X_N | \max\{X_1, \dots, X_N\} = m$ seja uniforme sobre o conjunto $\{(x_1, \dots, x_N) \in \mathbb{N}_0^N : \max\{x_1, \dots, x_N\} = m\}$, $m \in \mathbb{N}_0$. Consequentemente, a classe de medidas de probabilidade consideradas nessa seção são aquelas representadas como misturas dessas distribuições condicionais. A medida na mistura expressa incerteza a respeito da variável aleatória $M = \max\{X_1, \dots, X_N\}$. A partir desta avaliação concluímos que qualquer distribuição n -dimensional ($n \leq N$) dessa classe pode ser representada como

$$P_n(\cdot) = \int Q_{Nmn}(\cdot) d\mu(m) ,$$

onde Q_{Nmn} é a distribuição definida pela expressão (4) da seção 2.2 e μ a P_N -lei de M . A medida na representação não é uma medida produto. No Teorema 1, mostramos que aquela pode ser aproximada uniformemente pela lei de n variáveis aleatórias independentes com distribuição comum $U\{0, 1, \dots, m\}$. Este foi o resultado chave para obtermos as formas finita e infinita de teoremas do tipo de Finetti envolvendo a família de distribuições $U\{0, 1, \dots, \theta\}$, $\theta \in \mathbb{N}_0$. Para obter a versão infinita (Teorema 3), utilizamos a forma finita e verificamos a condição de tightness da seqüência de medidas de probabilidade $\{\mu_N | N \geq 1\}$, onde μ_N é a P_N -lei de $\max\{X_1, \dots, X_N\}$, diferentemente dos resultados examinados no Capítulo 1, onde a lei forte dos grandes números auxilia este processo.

O caso discreto considerado na seção 2.2 descreve o procedimento geral utilizado ao longo do Capítulo 2. Porém, a construção das distribuições condicionais que aparecem nas seções 2.3 até 2.5 é menos simples que no caso discreto. As construções e definições utilizadas nos casos Normal e Exponencial da seção 1.4 nos serviram de base para obtermos os resultados estabelecidos nessas seções. Isto pode ser notado na seção 2.4, onde, ao invés de N -esferas, temos bordos de hipercubos N -dimensionais. A distribuição uniforme sobre o intervalo $(-1, 1)$ e a distância do máximo na Proposição 5 do Capítulo 2 desempenharam um papel similar ao da distribuição $N(0, 1)$ e ao da distância euclidiana na Proposição 11

do Capítulo 1, respectivamente.

No Capítulo 2, as classes de medidas de probabilidade permutáveis consideradas foram definidas a partir da distribuição condicional ou sistema de distribuições condicionais, no caso de seqüências infinitas. Note-se que, ao contrário do Capítulo 1, julgamentos em termos de invariância não determinam o sistema de distribuições condicionais. Uma caracterização em termos de invariância poderia ainda ser feita no Capítulo 2. Também os resultados estabelecidos na seção 2.2 podem ser estendidos seguindo o mesmo procedimento para obter formas finitas de teoremas do tipo de Finetti envolvendo distribuições discretas dependendo de dois parâmetros ou sobre conjuntos da forma $\{-\theta, -\theta+1, \dots, 0, \dots, \theta-1, \theta\}$, $\theta \in \mathbb{N}_0$.

O teorema da representação de de Finetti e seus fortalecimentos nos mostram, entre outras coisas, o sentido dos parâmetros na representação. Estes últimos, como funções da seqüência infinita das variáveis aleatórias sob consideração, nos sugeriram que a abordagem bayesiana usual de superpopulação não é apropriada em termos práticos no problema de predição em populações finitas. Isto vem do fato que o conhecimento a priori que o pesquisador possui está mais associado a quantidades da própria população do que a quantidades de população infinita. Tendo isto em vista, no Capítulo 3 foi formulada uma abordagem completamente preditivista para fazer inferências em populações finitas. Nesse capítulo, foram formulados modelos probabilísticos que descrevem o conhecimento a priori a respeito de quantidades de população finita envolvidas no estudo. A idéia de suficiência é colocada em termos do problema de predição que interessa resolver.

Na seção 3.4, a discussão girou em torno de modelos discretos localmente uniformes. Dentro dessa classe de modelos, a especificação do modelo preditivo depende apenas do número de elementos em cada parte da partição do espaço amostral induzida pela quantidade de interesse e da distribuição a priori desta última. Nessa seção, foram identificados, através da Proposição 4, modelos de predição que admitem como estatística preditiva

suficiente a restrição da quantidade populacional à amostra.

Conceitos estatísticos clássicos são de utilidade para a descrição de modelos. Isto é ilustrado na subseção 3.4.1, onde a noção de estatística de resumo permite deduzir os modelos sob consideração sem a necessidade de recorrer-se a técnicas de contagem.

A questão de estendibilidade associada a seqüências finitas de variáveis aleatórias permutáveis permite estabelecer conexões com a abordagem bayesiana usual de superpopulação. De uma maneira geral, na subseção 3.4.2, o problema consistiu em determinar condições sobre a medida a priori T_N de forma que uma particular seqüência finita de variáveis aleatórias permutáveis possa ser mergulhada numa seqüência infinita de variáveis aleatórias permutáveis. Condições de estendibilidade nos casos Geométrico e Bernoulli são traduzidas em termos de problemas clássicos de momentos. A relação entre estendibilidade de seqüências finitas de variáveis aleatórias permutáveis 0-1 com o problema de momentos reduzido foi sugerida originalmente por de Finetti (1937). Diferentemente do caso uniforme abordado na subseção 3.4.2, nos casos Bernoulli e Geométrico as condições sobre T_N que garantem estendibilidade não permitem especificar a lei do processo onde a seqüência finita possa ser mergulhada.

Na seção 3.5, foram considerados modelos não discretos localmente uniformes, com ênfase nas inferências que podem ser feitas a partir desses modelos. A distribuição de Pareto foi estudada como distribuição a priori para as quantidades populacionais de interesse que aparecem nessa seção. Para inferências sobre o valor máximo na população, a distribuição a posteriori (no sentido de distância de variação total) é próxima de uma distribuição de Pareto. A distribuição de Pareto bivariada poderia ainda ser examinada como distribuição a priori para o valor mínimo e máximo na população sob o modelo apresentado na seção 2.5. A questão de estendibilidade e a derivação de prioris conjugadas são tópicos que também poderiam ser abordados para a classe de modelos estudados na seção 3.5. Finalmente, modelos de população finita gerados como na seção 3.3, estão sujeitos

ao mesmo tipo de análise feita sobre modelos bayesianos usuais. Algumas destas questões foram abordadas em Hayakawa (1992) para modelos de população finita no contexto de análise de sobrevivência.

Por último, no Capítulo 4, foi considerada uma solução não bayesiana para inferências em populações finitas. Modelos são gerados a partir de modelos de superpopulação por consideração da distribuição condicional dos dados, dada a quantidade populacional de interesse. Resultados comumente usados na teoria de amostragem clássica (Exemplos 4 e 11) são explicados dentro desta perspectiva.

Sob condições de suficiência da quantidade populacional de interesse relativa ao modelo de superpopulação, o modelo modificado é indexado unicamente por esta quantidade. Nestes casos, os modelos deduzidos poderiam ser interpretados como a versão clássica da abordagem descrita na seção 3.3 e as comparações usuais podem ser feitas. Por exemplo, quando o modelo deduzido desta maneira corresponde ao modelo especificado no primeiro estágio da seção 3.2, o preditor de máxima verossimilhança é equivalente ao valor que maximiza a distribuição a posteriori com distribuição a priori imprópria uniforme. Tais comparações são, evidentemente, de natureza teórica, pois o raciocínio utilizado no Capítulo 3 não está presente no Capítulo 4.

Os resultados obtidos no Capítulo 4 poderiam ainda ser comparados aos obtidos através da abordagem clássica de superpopulação e àqueles deduzidos através da utilização de outras funções de verossimilhança (existem pelo menos 14 versões segundo Bjørnstand (1990)).

Ao longo desta discussão pode-se notar que existem vários tópicos que ainda podem ser desenvolvidos, em especial no que se refere à abordagem preditivista para fazer inferências em populações finitas. O teorema de de Finetti e seus fortalecimentos são fundamentais para a compreensão de como tais inferências devem ser feitas dentro da visão subjetivista.

REFERÊNCIAS

- Ash, R.B. (1972). Real analysis and probability. New York, Academic Press.
- Aldous, D. (1985): Exchangeability and related topics. Lecture Notes in Mathematics 1117. Springer, Verlag.
- Barndorff-Nielsen, O. (1973): On M -ancillarity. *Biometrika*, **60**, 447-455.
- Barlow, R.E. (1991): De Finetti's "La prévision: ses lois logiques, ses sources subjectives". In Breakthroughs in Statistics. Johnson, N.; Kotz, S. (editores). Springer.
- Barlow, R.E. and Mendel M.B. (1992): De Finetti-type representation for life distribution. *Journal of the American Statistical Association*, to appear.
- Barlow, R.E. and Spizzichino, F. (1992): Schur-concave survival functions and survival analysis. *The Journal of Computational and Applied Mathematics*, **40**. To appear.
- Basu, D. (1958): On statistics independent of sufficient statistics. *Sankhyā*, **20**, 223-226.
- Basu, D. (1959): The family of ancillary statistics. *Sankhyā*, A, **26**, 3-16.
- Basu, D. and Ghosh, J.K. (1967): Sufficient statistic in sampling from a finite universe. *Bull. Int. Inst.*, **42**, 850-859.
- Basu, D. (1969): Role of the sufficiency and likelihood principles in sample survey theory. *Sankhyā*, A, **31**, 441-454.
- Basu, D. (1971): An essay on the logical foundations of survey sampling. Foundations of Statistical Inference. Edited by V.P. Godambe and D.A. Sprott. Hoet, Rinehart and Winston, Toronto: pp. 203-242.
- Basu, D. (1975): Statistical information and likelihood. *Sankhyā* A, **37**, 1-71.
- Basu, D. (1977): On the elimination of nuisance parameters. *Journal of American Statistical Association*, **72**, 355-66.

- Berman, S. (1962): Limiting distribution of maximum term in sequences of dependent variables. *Ann. Math. Statist*, **33**, 894-908.
- Billingsley, P. (1968): Convergence of probability measures. Wiley, New York.
- Billingsley, P. (1986): Probability and measures. Wiley, New York.
- BJORnstad, J.F. (1990): Predictive likelihood: a review (with discussion). *Statistical Science*, **5**, 242-65.
- Blum, J.R. Chernoff, H. Rosenblatt, M. and Teicher, H. (1958): Central limit theorem for interchangeable processes. *Canad. J. Math*, **10**, 222-229.
- Branco, M. (1991): Um estudo sobre suficiência, ancilaridade, independência estatística e os teoremas de Basu. Dissertação de Mestrado. Instituto de Matemática e Estatística. Universidade de São Paulo. Brasil.
- Bolfarine, H.; Zacks, S. (1991): Prediction theory for finite population quantities under superpopulation models. Springer Series in Statistics. Em fase de publicação.
- Bühlman, H. (1987): Seminar on Exchangeable Random Variables and Conditional Independence. University of California at Berkeley.
- Butler, R.W. (1986): Predictive likelihood inference with applications (with discussion). *J. Roy. Statist. Soc., B*, **47**, 1-38.
- Campanino, M.; Spizzichino, F. (1981): Prediction sufficiency and representation of infinite sequence of exchangeable random variables. Technical report, Institute of Mathematics <<G. Castel Nuovo>> University of Rome. (Roma).
- Cassel, C.M., Särndal, C.E. and Wretman, J.H. (1977): Foundations of inference in survey sampling. Wiley, New York.
- Cifarelli, D. and Regazzini, E. (1982): Some consideration about mathematical statistics teaching methodology suggested by the concept of exchangeability. In exchangeability in Probability and Statistics edited by G. Kosch and F. Spizzichino (North Holland, Amsterdam) p. 185.
- Cochran, W.G. (1963): Sampling techniques, 2.ed. Wiley, New York.

- Cramer, H. (1946): Mathematical methods of statistics. Princeton University Press.
- Chernoff, H. and Teicher, H. (1958): A central limit theorem for sums of interchangeable random variables. *Ann. Math. Statist.*, **29**, 118-130.
- Chow, Y.; Teicher, H. (1978): Probability theory. Springer, New York.

- Dawid, A.P. (1977): Invariant distribution and Analysis of Variance models, *Biometrika*, **64**, 291-7.
- Dawid, A.P. (1977): Spherical matrix distributions and a multivariate model, *J. Roy. Statist. Soc.*, B, **39**, 254-61.
- Dawid, A.P. (1978): Extendibility of spherical matrix distribution, *J. Mult. Anal.*, **8**, 559-66.
- Dawid, A.P. (1979): Conditional independence in statistical theory (with discussion). *J. Roy. Statist. Soc.*, B, **41**, 1-31.
- Dawid, A.P. (1982): Intersubjective statistical models. In Exchangeability in Probability and Statistics edited by C. Kosch and F. Spizzichino (North-Holland, Amsterdam) pp. 217-230.
- de Finetti, B. (1931): Funzione caratteristica di un fenomeno aleatorio. *Atti della R. Accademia Nazionale dei Lincei, Ser. 6, Memorie, classe di Scienze Fisiche, Matematiche e Naturale*, **4**, 251-299.
- de Finetti, B. (1937): Le prévision: ses lois logiques, ses sources subjectives. *Ann. Inst. H. Poincaré. Probab. Statist.*, **7**, 1-68.
- de Finetti, B. (1938): Sur la condition d'équivalence partielle. *Actualities Scientifiques et Industrielles*, no. 739. Hemon and Cii: Paris. Translated in *Studies in Inductive Logic and Probability*, II, R. Jeffrey Ed. University of California Press, Berkeley.
- de Finetti, B. (1949): Le vroi et le probable-dialectica, vol. 3, pp. 78-92.
- de Finetti, B. (1959): La probabilita e la statistica nei rapporti con induzione; secondo i diversi punti di vista. In corso C.I.M.E. su Induzione e Statistica. Cremonese, Rome.

- de Finetti, B. (1969): Sulla proseguibilità di processo aleatori scambiabili. *Rend. Inst. Mat. Univ. Trieste*, 1.
 - de Finetti, B. (1972): Probability, induction and statistics: Wiley , New York.
 - de Finetti, B. (1974): Theory of probability, vol. 1. Wiley, New York.
-
- De Groot, M.H. (1970): Optimal statistical decisions. Mc Graw-Hill, New York.
 - Diaconis, P. (1977): Finite forms of de Finetti's theorem on exchangeability. *Synthese*, vol. 3b: 271-281.
 - Diaconis, P. and Freedman, D. (1980): Finite exchangeable sequences. *Ann. Prob.*, 8, 745-764.
 - Diaconis, P. and Freedman, D. (1981): Partial exchangeability and sufficiency. In: Proceedings of the Indian Statistical Institute Golden Jubilee. International Conference on Statistics Applications and New Directions: 205-236.
 - Diaconis, P. and Ylvisaker, D. (1985): Quantifying prior opinion. In: Bayesian Statistics, 2, Elsevier Science Publishers B.V. and Valencia University Press.
 - Diaconis, P. and Freedman, D. (1987): A dozen de Finetti - style results in search of theory. *Ann. Inst. H. Poincaré. Probab. Statist.*, 23, 397-423,
 - Diaconis, P. and Freedman, D. (1988a): Conditional limit theorem for exponential families and finite versions of de Finetti's theorem on exchangeability. *Theoretical Prob.*, 1-50.
 - Diaconis, P. (1988): Recent progress on de Finetti's notions of exchangeability. Technical Report no. 297. Stanford University. Department of Statistics.
 - Diaconis, P.; Eaton, M. Lauritzen, S. (1988): Finite de Finetti theorems in linear models and multivariate analysis. Technical Report. Harvard University and University of Minnesota. 20 pp.
 - Diaconis, P. and Freedman, D. (1990): Cauchy's equation and de Finetti's theorem. *Scand. J. Statist.*, 17, 235-250.

- Dubins, L. and Freedman, D. (1986): Exchangeable processes need not mixtures of independent identically distributed random variables. *Z. Wahrscheinlichkeitstheorie und verw. Gebiet*, vol. 48, 1979: 115-132.
- Dynkin, E. (1978): Sufficient statistics and extreme points. *Ann. Prob.*, 6, 705-730.
- Eaton, M. (1981): On the projections of isotropic distributions. *Ann. Stat.*, 9, 391-400.
- Eaton, M. (1983): Multivariate statistics: a vector space approach. Wiley, New York.
- Eaton, M. (1989): Group invariance applications in statistics. CBMS Regional Conference Series in Probability and Statistics, vol. 1, IMS. Hayward, California.
- Ericson, W.A. (1969a): Subjective bayesian models in sampling finite populations (with discussion). *J. Roy. Statist. Soc.*, B, 31, 195-224.
- Ericson, W.A. (1969b): A note on the posterior mean of a population mean. *J. Roy. Statist. Soc.*, B, 31, 332-324.
- Feller, W. (1968): An introduction to probability theory and its applications, vol. I, Wiley, New York.
- Feller, W. (1966): An introduction to probability theory and its applications, vol. II. Wiley, New York.
- Freedman, D. (1962a): Mixtures of Markov processes. *Ann. Math. Statist.*, 33, 114-118.
- Freedman, D. (1962b): Invariant under mixing which generalize de Finetti's theorem. *Ann. Math. Statist.*, 33, 916-923.
- Godambe, V.P. (1955): A unified theory of sampling from finite populations. *J. of Statist. Soc.*, B, 17, 269-278.
- Godambe, V.P. (1966): A new approach to sampling from finite populations, *J. R. Statist. Soc.*, B, 28, 310-328.

- Hayakawa, Y. (1992): Bayesian parametric models for lifetimes from a subjectivistic view point: model construction and characterizations of aging. Ph. D. Theses, University of California at Berkeley.
 - Hewitt, E. and Savage, L.J. (1955): Symetric measures on Cartesian products. *Trans. Amer. Math. Soc.*, **80**, 470-501.
-
- Heath, D. and Sudderth, W. (1976): De Finetti's theorem on exchangeable variables, *American Statistician*, **30**, 188-9.
 - Hill, B.M.; Lane, D. and Sudderth, W. (1987): Exchangeable urn processes. *Ann. Prob.* vol. 15, number 4: 1586-1592.
 - Hinkley, D.V. (1979): Predictive Likelihood. *Ann. Statist.*, **7**, 718-728.
 - Iglesias, P.; Sandoval, M.; Pereira, C.A.B. (1991): Predictive likelihood in finite population. RT-MAE-9106. IME-USP. São Paulo.
 - Irony, T.Z. e Pereira, C.A.B. (1992): Motivation for the correct use of discrete models in quality assurance. RT-MAE-9215. IME-USP. São Paulo.
 - Kagan, A.M.; Linnik, Y.V. and Rao, C.R. (1972): Characterization problems in mathematical statistics. Wiley, New York.
 - Kingman, J.F.C. (1972): On random sequences with spherical simmetry. *Biometrika*, **59**, 492-4.
 - Kingman, J.F.C. (1978): Uses of exchangeability. *Ann. Prob.*, Vol 6, no. 2, 183-197.
 - Küchler, V. and Lauritzen, S. (1989): Exponential families, extreme point models and minimal space-time. Invariant Functions for Stochastic Processes with Stationary and Independent Increments. *Scand J. Statist.*, **16**, 237-261.
 - Lauritzen, S.L. (1974): Sufficiency, prediction and extreme model. *Scand. J. Statist.*, **1**, 128-134.
 - Lauritzen, S.L. (1974): On the interrelationships among sufficiency and some related concepts. Univ. Copenhagen, Inst. Math. Stat., Preprint No. 8.
 - Lauritzen, S.L. (1984): Extreme Point model in statitics. *Scand J. Statist.*, **11**, 65-91.

- Lauritzen, S.L. (1988): Extremal families and system of sufficient statistics. Springer Lecture Notes in Statistics, vol. 49. Springer, Berlin, New York.
 - Lindley, D.V. (1981): The role of exchangeability in inference. *Ann. Statist.*, Vol. 9, no. 1, 45-58.
 - Loève, M. (1963): Probability theory. Van Nostrand, New York.
-
- Nachbin, L. (1965): The Haar integral. Van Nostrand, Princeton, N.J.
 - Olshen, R. (1974): A note on exchangeable sequences. *Z. Wahrscheinlichkeitstheorie und verw. Gebiet*, **28**, 317-321.
 - Paulino, C.D. (1988): Análise de dados categorizados incompletos: fundamentos, métodos e aplicações. Tese de Doutorado, Instituto de Matemática e Estatística. Universidade de São Paulo. Brasil.
 - Pereira, C.A. (1980): Bayesian solutions to some classical problems of statistics. Ph.D. Thesis. The Florida State University, Tallahassee, Florida.
 - Pereira, C.A. (1985): Teste de hipótese definida em espaços de diferentes dimensões: visão bayesiana e interpretação clássica. Tese Livre Docência. Instituto de Matemática e Estatística. Universidade de São Paulo. Brasil.
 - Royall, R.M. (1970a): Finite population on labels in estimation. *Ann. Math. Statist.* **41**, 1774-1779.
 - Ressel, P. (1985): de Finetti - type theorems: an analytical approach. *Ann. Prob.*, **13**, 898-922.
 - Royall, R.M. (1971): Linear regression models infinite population sampling theory. In: V.P. Godambe & D.A. Sprott, eds., *Foundations of Statistical Inference*. Toronto: Holt, Rinehart & Winston: 259-274.
 - Rubin, D. (1978): Bayesian inference for casual effects: the role of randomization. *Ann. Statist.*, vol. 6, no. 1: 34-58.
 - Savage, L.J. (1954): The foundations fo statistics. Wiley , New York.
 - Scott, A.J. and Smith, T.M.F. (1973): Survey design, symmetry and posterior distributions. *J.Roy. Statist.Soc.*, B,**35**, 57-60.

- Shohat, J.A. and Tamarkin, J.D. (1943): The problem of moments. Amer. Math. Soc., New York.
- Smith, T.M.F. (1976): The foundations of survey sampling: a review (with discussion). *J. Roy. Statist. Soc. A*, **39**, 183-204.
- Smith, A.M.F. (1981): On Random sequences with centered spherical symmetry. *J. Roy. Statist. Soc. B*, **43**, 208-209.
- Sudgen, R.A. (1978): Exchangeability and the foundation of survey sampling. Ph.D. Thesis, University of Southampton.
- Sudgen, R.A. (1979): Inference on symmetric functions of exchangeable populations. *J.R. Statistics. Soc. B*, **2**, 269-73.
- Sudgen, R.A. (1982): Exchangeability and survey sampling inference. In Exchangeability in Probability and Statistics edited by G. Kosch and F. Spizzichino (North-Holland, Amsterdam) p. 321.
- Spizzichino, F. (1982): Extendibility of symmetric probability distributions and related bounds. In Exchangeability in Probability and Statistics edited by G. Kosch and F. Spizzichino (North-Holland, Amsterdam) p. 313.
- Spizzichino, F. (1988): Symmetry conditions on opinion assessment leading to time-transformed exponential models. Reprinted from accelerated life testing and expert's opinions in reliability. Soc. Italian di Fisica - Bologna - Italy. p. 83.
- Wechsler, S. (1993): Exchangeability and predictivism. *Erkenntnis – International Journal of Analytic Philosophy*. To appear.