Urna de Pólya e a Distribuição Beta

Carolina Fabiana Svetliza

DISSERTAÇÃO APRESENTADA
AO
INSTITUTO DE MATEMÁTICA E ESTATÍSTICA
DA
UNIVERSIDADE DE SÃO PAULO
PARA
OBTENÇÃO DO GRAU
DE
MESTRE EM ESTATÍSTICA

Área de Concentração: **Estatística**
Orientador: **Prof. Dr. Sergio Wechsler**

Durante este trabalho, o autor teve o apoio financeiro do CNPq

São Paulo, 16 de setembro de 1998

Urna de Pólya e a Distribuição Beta

Este exemplar corresponde à redação final da dissertação devidamente corrigida e defendida por Carolina Fabiana Svetliza e que foi aprovada pela comissão julgadora.

São Paulo, 16 de setembro de 1998.

Comissão Julgadora:

- Prof. Dr. Sergio Wechsler (orientador) - IME-USP
- Prof. Dr. Vladimir Belitsky - IME-USP
- Prof. Dr. Dani Gamerman - UFRJ

A meus pais e amigos.

*Las leyes del Cielo y del Infierno son versátiles
que vayas a un lugar o a otro depende
de un ínfimo detalle.*
Silvina Ocampo

AGRADECIMENTOS

Ao Professor Sergio Wechsler pela sua orientação.

Aos Professores Vladimir Belitsky e Nelson Tanaka pelos aportes dados.

A Veronica e Luis Gustavo pelas suas contribuições.

A Marcia, Juan e Laika por deixá-me formar parte de seu hogar.

A todos meus amigos que de uma forma ou outra colaborarão e me apoiaron durante este trabalho. Também a meus amigos que estão longe, mas sempre estiverão perto.

A minha mãe, a meu pai, a Javier, Irina e Leandro que de longe me apoiaron e me derão carinho em todo momento.

Ao CNPq pelo apoio financiero, que me permitiu desenvolver este trabalho.

Resumo

Este trabalho descreve o processo de urna de Pólya, que é uma particularização dos processos de urna permutáveis. Considerando o espaço amostral dado pelo conjunto $\{0, 1\}$, o Teorema da Representação de Bruno De Finetti para processos de Pólya surge em sua forma mais simples, e *a priori* é essencialmente membro da família Beta de distribuições. Dada a urna de Pólya com composição inicial de bolas (a_0, b_0) , sendo a_n e b_n quantidades de bolas azuis e brancas, respectivamente, no tempo n , as proporções de bolas $\{\frac{a_n}{a_n+b_n}\}_{n \geq 0}$ convergem a uma variável limite, que possui distribuição Beta com parâmetros dados pela composição inicial de bolas na urna.

Abstract

This work describes the Pólya's Urn Process, which is a particularization of exchangeable urn process. Considering the sample space given by the set $\{0, 1\}$, Bruno de Finetti's Theorem of the Representation for the Pólya's Process presents its simplest form, with *a priori* essentially belonging to the Beta's family. Given a Pólya's Urn with (a_0, b_0) as initial composition, i.e., a_n and b_n being the quantities of blue and white balls respectively, at time n , the proportions of balls $\{\frac{a_n}{a_n+b_n}\}_{n \geq 0}$ converge to a limit variable, which has a Beta distribution with parameters given by the initial composition in the urn.

ÍNDICE

1	INTRODUÇÃO	3
2	PERMUTABILIDADE	5
2.1	Permutabilidade	5
2.2	Permutabilidade Infinita	12
2.3	Estendibilidade	15
2.4	Teorema da Representação	18
2.5	Simulações	25
3	PROCESSOS DE URNA	28
3.1	Processo de Urna de Pólya	33
3.2	Permutabilidade	35
3.3	Caracterização da Priori para o Processo de Pólya	37
3.3.1	A urna de Bernard Friedman	38
3.4	Simulações	43
4	CONSIDERAÇÕES ASSINTÓTICAS	45

4.1	Processo de Pólya com parâmetro contínuo	46
4.2	Distribuições Assintóticas	49
5	GENERALIZAÇÕES-PROCESSOS PERMUTÁVEIS	53
5.1	Postulado de Suficiência de Johnson	56
5.1.1	Complementos e extensões	58
5.1.2	O Novo Postulado de Suficiência. O Postulado Z	59
6	CONCLUSÃO	63
A	DEFINIÇÕES	65
	REFERÊNCIAS BIBLIOGRÁFICAS	68

Capítulo 1

INTRODUÇÃO

O objetivo deste trabalho é descrever o processo de urna de Pólya, que é uma particularização dos processos de urna permutáveis. Considerando o espaço amostral dado pelo conjunto $\{0, 1\}$, o Teorema da Representação de Bruno De Finetti para processos de Pólya surge em sua forma mais simples, e a priori é essencialmente membro da família Beta de distribuições. A distribuição Beta constitui uma família conjugada de distribuições para verossimilhança gerada, por exemplo, por uma distribuição Binomial. Além disso, dada a urna de Pólya com composição inicial de bolas (a_0, b_0) , sendo a_n e b_n quantidades de bolas azuis e brancas, respectivamente, no tempo n , as proporções de bolas $\{\frac{a_n}{a_n+b_n}\}_{n \geq 0}$ convergem a uma variável limite, que possui distribuição Beta com parâmetros dados pela composição inicial de bolas na urna.

Como os processos estocásticos são uma ferramenta necessária no desenvolvimento deste trabalho, é preciso mencionar alguns dos conceitos utilizados, o que é feito no Apêndice. Introduzimos no Apêndice entre outros conceitos, a definição de Martingais que será usada na demonstração do Teorema de Bernard Friedman (Freedman, D. (1965) [11]) no capítulo 3, seção 3.3.1.

No capítulo 2, seção 2.1, apresentamos a definição e principais propriedades de

Permutabilidade. Na seção 2.2, introduzimos a definição de permutabilidade infinita. A definição de estendibilidade é dada na seção 2.3. Referimo-nos ao Teorema da Representação para variáveis aleatórias com valores no conjunto $\{0,1\}$, e são incluídos também comentários a respeito e corolários deste teorema na seção 2.4. Foram feitas simulações de processos aleatórios permutáveis na seção 2.5, utilizando o Teorema da Representação, obtendo matrizes de seqüências de variáveis aleatórias com valores no conjunto $\{0,1\}$.

Os processos de urna são considerados no capítulo 3. Na seção 3.1, apresentamos o processo de urna de Pólya; na seção 3.2, mostramos que este processo é permutável. Na subseção 3.3.1, baseados nos resultados de Freedman, D. (1965) [11], vamos mostrar que o processo $\{X_n\}_{n \geq 0}$, dado pela proporção de bolas de uma dada cor, converge a uma variável limite que possui distribuição Beta. Também neste capítulo são consideradas simulações de processos de urna, que são apresentadas na seção 3.4.

No capítulo 4 mostramos considerações assintóticas, onde temos que o processo de Pólya pertence à classe dos processos de Markov descontínuos, isto é, processos discretos no espaço, mas contínuos no tempo. Na seção 4.1, obtemos uma caracterização do Processo de Pólya como processo de parâmetro contínuo. Temos diferentes distribuições assintóticas e outros resultados na seção 4.2.

As generalizações dos processos permutáveis são apresentadas no capítulo 5, onde o Teorema central do artigo de Hill, Suddenth e Lane (1987) [14] é enunciado. O Postulado de Suficiência de Johnson (Zabell (1982) [27]) é dado na seção 5.1. O novo postulado de Suficiência (Zabell (1997) [28]) é mencionado na subseção 5.1.2.

No capítulo 6 apresentamos uma conclusão do trabalho.

Capítulo 2

PERMUTABILIDADE

2.1 Permutabilidade

Definição 2.1 : Dizemos que duas variáveis aleatórias X e Y em um espaço de probabilidade $(\Omega, \mathcal{F}, P)$ são permutáveis quando (X, Y) e (Y, X) possuem a mesma distribuição conjunta.

Proposição 2.1 : *Se X e Y são permutáveis, então X e Y são identicamente distribuídas.*

Prova: (Caso discreto). Para todo x no suporte da distribuição de X ,

$$P(X = x) = \sum_y P(X = x, Y = y) = \sum_y P(X = y, Y = x) = P(Y = x).$$

É claro que duas variáveis aleatórias independentes e identicamente distribuídas são permutáveis, mas variáveis aleatórias podem ser permutáveis sem que sejam independentes, como se exhibe nos seguintes exemplos:

Exemplo 2.1 : Consideremos uma urna contendo A bolas azuis e B bolas brancas, com $A+B = M$, da qual serão retiradas n bolas sem reposição, uma de cada vez, sempre com mesma probabilidade para as bolas restantes na urna. Sejam $X_1, X_2, \dots, X_n$

os indicadores de bola azul (isto é, $X_i = 1$, se a i^{a} bola retirada for azul, $X_i = 0$ em caso contrário, para $i = 1, 2, \dots, n$). As variáveis X_i e X_j não são independentes, mas são permutáveis, sempre que $i \neq j$, para $i, j = 1, 2, \dots, n$.

Exemplo 2.2 : Se o vetor aleatório (X, Y) possui distribuição normal bivariada com vetor de médias $\mu = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}$ e matriz de covariâncias $\Sigma = \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_2\sigma_1 & \sigma_2^2 \end{pmatrix}$, com $\rho \neq 0$, então as variáveis X e Y não são independentes, mas são permutáveis desde que $\mu_1 = \mu_2$ e $\sigma_1 = \sigma_2$.

Exemplo 2.3 : Consideremos um vetor aleatório (X, Y) , onde X possui distribuição Normal com média $\mu_X = 0$ e variância $\sigma_X^2 = 1$ e Y , dado $X = x$, é uma mistura de duas normais $N(\frac{\sqrt{2}}{2}x, \frac{1}{2})$ e $N(-\frac{\sqrt{2}}{2}x, \frac{1}{2})$, com pesos $1/2$. A função de densidade conjunta de (X, Y) apresenta $\text{Cov}(X, Y) = 0$ e, embora as variáveis não sejam independentes, X e Y são permutáveis, dada a simetria observada no seguinte gráfico.

Figura 2.1. Curvas de Nível para a densidade $f(x,y)$.

Definição 2.2 : Dizemos que as variáveis aleatórias $X_1, X_2, \dots, X_N$ são permutáveis quando o vetor $(X_{\pi(1)}, \dots, X_{\pi(N)})$ possui a mesma distribuição conjunta, qualquer que seja a permutação π do conjunto $\{1, \dots, N\}$.

Proposição 2.2 : *Subconjuntos de duas ou mais variáveis aleatórias permutáveis são permutáveis.*

Prova: (Caso absolutamente contínuo). Sejam $X_1, \dots, X_N$ variáveis aleatórias permutáveis. Para $r = 2, \dots, N - 1$ e $I = \{i_1, \dots, i_r\} \subset \{1, \dots, N\}$, a densidade conjunta

$$P(x_{i_1}, x_{i_2}, \dots, x_{i_r}) = \int P(x_{i_1}, x_{i_2}, \dots, x_{i_r}, x_{k_1}, \dots, x_{k_{N-r}}) dx_{k_1} \dots dx_{k_{N-r}};$$

Consideremos, contudo, qualquer permutação π de $\{1, \dots, N\}$ que satisfaça $\pi(k) = k$, sempre que $k \notin I$. Teremos, então, por hipótese,

$$\begin{aligned} \int P(x_{i_1}, x_{i_2}, \dots, x_{i_r}, x_{k_1}, \dots, x_{k_{N-r}}) dx_{k_1} \dots dx_{k_{N-r}} &= \\ \int P(x_{\pi(i_1)}, x_{\pi(i_2)}, \dots, x_{\pi(i_r)}, x_{k_1}, \dots, x_{k_{N-r}}) dx_{k_1} \dots dx_{k_{N-r}} &= P(x_{\pi(i_1)}, x_{\pi(i_2)}, \dots, x_{\pi(i_r)}). \end{aligned}$$

Corolário 2.1 : *As variáveis aleatórias $X_1, \dots, X_N$ são permutáveis se, e somente se, todas as marginais de mesmo comprimento forem identicamente distribuídas.*

Prova: (Caso absolutamente contínuo).

($\Leftarrow$) Basta considerar todas as marginais de comprimento N . Por definição, $(X_1, \dots, X_N)$ são permutáveis.

($\Rightarrow$) Seja $r \in \{1, 2, \dots, N\}$. Sejam $\mathbf{i} = (i_1, i_2, \dots, i_r)$ e $\mathbf{j} = (j_1, j_2, \dots, j_r)$ vetores de r índices distintos de $\{1, 2, \dots, N\}$ cada. (Mas um mesmo índice pode estar no vetor

i e no vetor j . Temos que provar que $(X_{i_1}, X_{i_2}, \dots, X_{i_r})$ possui a mesma distribuição que $(X_{j_1}, X_{j_2}, \dots, X_{j_r})$. Pela **Proposição 2.2**, estes dois vetores são, cada um, de variáveis permutáveis. Portanto, basta provar que $(X_{i_1}, X_{i_2}, \dots, X_{i_r})$ e $(X_{j_1}, X_{j_2}, \dots, X_{j_r})$ possuem a mesma distribuição. Portanto, sem perda de generalidade, podemos supor que $1 \leq i_1 < i_2 < \dots < i_r \leq N$. Seja π uma permutação de $(1, 2, \dots, N)$ que satisfaz

$$\pi(i_1) = j_1, \pi(i_2) = j_2, \dots, \pi(i_r) = j_r. \text{ Então}$$

$$\begin{aligned} P(x_{i_1}, \dots, x_{i_r}) &= \int P(x_{i_1}, x_{i_2}, \dots, x_{i_r}, x_{k_1}, \dots, x_{k_{N-r}}) dx_{k_1} \dots dx_{k_{N-r}} \\ &\text{Por hipótese} \\ &= \int P(x_{\pi(i_1)}, x_{\pi(i_2)}, \dots, x_{\pi(i_r)}, x_{\pi(k_1)}, \dots, x_{\pi(k_{N-r})}) dx_{\pi(k_1)} \dots dx_{\pi(k_{N-r})} \\ &= \int P(x_{j_1}, x_{j_2}, \dots, x_{j_r}, x_{\pi(k_1)}, \dots, x_{\pi(k_{N-r})}) dx_{\pi(k_1)} \dots dx_{\pi(k_{N-r})} \\ &= P(x_{j_1}, x_{j_2}, \dots, x_{j_r}). \end{aligned}$$

Proposição 2.3 : *Sejam $X_1, \dots, X_N$ variáveis aleatórias permutáveis. Sejam $I = \{i_1, \dots, i_r\}$ e $J = \{j_1, \dots, j_s\}$ subconjuntos de $\{1, \dots, N\}$, com $I \cap J = \emptyset$, para $2 \leq r \leq N-1$, $1 \leq s \leq N-2$. Então as variáveis aleatórias $X_{i_1}, \dots, X_{i_r}$, dado $X_{j_1} = x_{j_1}, \dots, X_{j_s} = x_{j_s}$, possuem distribuição condicional de marginais $X_{i_1}, X_{i_2}, \dots, X_{i_r}$ permutáveis.*

Prova:(Caso absolutamente contínuo). A densidade condicional

$$P(x_{i_1}, \dots, x_{i_r} | x_{j_1}, \dots, x_{j_s}) = \frac{P(x_{i_1}, \dots, x_{i_r}, x_{j_1}, \dots, x_{j_s})}{P(x_{j_1}, \dots, x_{j_s})};$$

seja π uma permutação qualquer do conjunto $\{1, \dots, N\}$, tal que $\pi(j) = j$, sempre que $j \notin I$. Considerando a permutação π na densidade conjunta $P(x_{i_1}, \dots, x_{i_r}, x_{j_1}, \dots, x_{j_s})$, vamos ter, por hipótese,

$$\begin{aligned}\frac{P(x_{i_1}, \dots, x_{i_r}, x_{j_1}, \dots, x_{j_s})}{P(x_{j_1}, \dots, x_{j_s})} &= \frac{P(x_{\pi(i_1)}, \dots, x_{\pi(i_r)}, x_{j_1}, \dots, x_{j_s})}{P(x_{j_1}, \dots, x_{j_s})} \\ &= P(x_{\pi(i_1)}, \dots, x_{\pi(i_r)} | x_{j_1}, \dots, x_{j_s}).\end{aligned}$$

Proposição 2.4 : Se $X_1, X_2, \dots, X_N$ são variáveis aleatórias permutáveis com $E(X_1^2) < \infty$, então a correlação $\rho(X_i, X_j) \geq -1/(N-1)$, $i \neq j$, para $i, j = 1, \dots, N$, com $N > 1$.

Prova: Seja $m = E(X_i)$, $\sigma^2 = V(X_i)$ e $\lambda = \rho(X_i, X_j)$. Do **Corolário 2.1**, temos que todas as distribuições marginais bivariadas coincidem, o mesmo com as univariadas, portanto o vetor de médias e a matriz de covariância são dados por:

$$\mathbf{m} = m \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix}$$

$$\Sigma = \sigma^2 \begin{pmatrix} 1 & \lambda & \dots & \dots & \lambda \\ \lambda & 1 & \lambda & \dots & \lambda \\ \vdots & & \vdots & & \vdots \\ \lambda & \lambda & \dots & \dots & 1 \end{pmatrix} = (1-\lambda)\sigma^2 I_N + \lambda\sigma^2 J_N.$$

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

$\mathbf{1}_N^t \Sigma \mathbf{1}_N = (1 + \lambda(N-1), \dots, 1 + \lambda(N-1)) \mathbf{1}_N = N + \lambda(N-1)N = (1-\lambda)N + \lambda N^2 > 0$, concluindo-se a partir desta desigualdade que $\lambda > -1/(N-1)$.

Proposição 2.5 : *O conjunto de funções de distribuição conjunta de variáveis aleatórias permutáveis em $\mathbb{R}^N$ é um conjunto convexo.*

Prova: Sejam F_1, F_2 duas funções de distribuição de variáveis aleatórias permutáveis em $\mathbb{R}^N$ e seja $F = \alpha F_1 + (1 - \alpha)F_2$, com $0 \leq \alpha \leq 1$. Vamos verificar que $F(x_1, \dots, x_N)$ é uma função de distribuição, para isto deve satisfazer as seguintes propriedades.

I) $F(x_1, \dots, x_N)$ é não decrescente em cada uma das variáveis. Por exemplo, é não decrescente em x_1 : se $w < z$, então

$$\begin{aligned} F(w, \dots, x_N) &= \alpha F_1(w, \dots, x_N) + (1 - \alpha)F_2(w, \dots, x_N) \\ &< \alpha F_1(z, \dots, x_N) + (1 - \alpha)F_2(z, \dots, x_N) \\ &= F(z, \dots, x_N) \end{aligned}$$

Portanto,

$$F(w, \dots, x_N) < F(z, \dots, x_N).$$

Analogamente para $x_2, x_3, \dots, x_N$.

II) F é contínua à direita em cada uma das variáveis, por exemplo, se $w_m \downarrow x_1$ quando $m \rightarrow \infty$, então

$$\begin{aligned} F(w_m, \dots, x_N) &= \alpha F_1(w_m, \dots, x_N) + (1 - \alpha)F_2(w_m, \dots, x_N) \\ &\downarrow \alpha F_1(x_1, \dots, x_N) + (1 - \alpha)F_2(x_1, \dots, x_N) \\ &= F(x_1, \dots, x_N), \end{aligned}$$

quando $m \rightarrow \infty$.

Analogamente quando $w_m \downarrow x_2, w_m \downarrow x_3, \dots$

III) Para todo i ,

$$\begin{aligned}\lim_{x_i \rightarrow -\infty} F(x_1, \dots, x_N) &= \alpha \lim_{x_i \rightarrow -\infty} F_1(x_1, \dots, x_N) + (1 - \alpha) \lim_{x_i \rightarrow -\infty} F_2(x_1, \dots, x_N) \\ &= 0\end{aligned}$$

Também,

$$\begin{aligned}\lim_{\forall i, x_i \rightarrow +\infty} F(x_1, \dots, x_N) &= \alpha \lim_{\forall i, x_i \rightarrow +\infty} F_1(x_1, \dots, x_N) + (1 - \alpha) \lim_{\forall i, x_i \rightarrow +\infty} F_2(x_1, \dots, x_N) \\ &= 1\end{aligned}$$

IV) Esta propriedade pode ser escrita por meio de operadores de diferença. Para $I = (a, b]$, com $a < b$ e $g : \mathbb{R}^k \rightarrow \mathbb{R}$ definimos

$$\Delta_{k,I} g(x_1, \dots, x_k) = g(x_1, \dots, x_{k-1}, b) - g(x_1, \dots, x_{k-1}, a).$$

De modo que aplicando esta propriedade a $F(x_1, \dots, x_N)$ temos

$$\begin{aligned}\Delta_{1,I_1}, \dots, \Delta_{N,I_N} F(x_1, \dots, x_N) &= \alpha \Delta_{1,I_1}, \dots, \Delta_{N,I_N} F_1(x_1, \dots, x_N) + \\ &+ (1 - \alpha) \Delta_{1,I_1}, \dots, \Delta_{N,I_N} F_2(x_1, \dots, x_N) \geq 0,\end{aligned}$$

$$\forall I_k = (a_k, b_k], a_k < b_k, k = 1, \dots, N.$$

Estas quatro propriedades são suficientes para caracterizar funções de distribuição, portanto temos que combinações convexas de funções de distribuição é uma função de distribuição. Então para $(x_1, x_2, \dots, x_N) \in \mathbb{R}^N$ e uma permutação π de $\{1, 2, \dots, N\}$, teremos

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

Um conceito importante relacionado com esta última Proposição é o de mistura de funções de distribuição.

Definição 2.3 : Seja $\mathcal{F} = \{F(\cdot|\theta) : \theta \in \Theta\}$ uma família de funções de distribuição sobre $\mathbb{R}^N$ e Q uma função de distribuição sobre Θ . A função de distribuição $F(\cdot)$ é denominada mistura sobre $\mathcal{F}$, definida pela função de distribuição Q , sempre que

$$F(\cdot) = \int_{\Theta} F(\cdot|\theta) dQ(\theta)$$

Desta definição, e da **Proposição 2.5**, podemos concluir que se $F(\cdot|\theta)$ é uma função de distribuição permutável, então, a função de distribuição mistura $F(\cdot)$ é permutável.

2.2 Permutabilidade Infinita

Os estatísticos clássicos usam o termo evento como um nome coletivo, enquanto as simples repetições desse evento são denominadas ensaios; cada ensaio possui probabilidade desconhecida de sucesso (espaço amostral dado pelo conjunto $\{0, 1\}$). O limite das frequências relativas de ocorrência de um evento em uma sequência de repetições de ensaios independentes, e sob as mesmas condições é a probabilidade de sucesso do evento, a partir do ponto de vista clássico. Os problemas nesta definição estão dados pelo limite das frequências, que na realidade não é possível atingir, e também na condição de ensaios independentes, dado que se usa o termo “independência” antes de sua própria definição, a qual como sabemos necessita o conceito de probabilidade. Na verdade, não existem ensaios sob as mesmas condições de experimento algum, já que as condições vão mudando de um instante a outro, isto é, os eventos são únicos e impossíveis de serem repetidos (De Finetti, 1937 [7]). A partir do ponto de vista

Bayesiano, a suposição de ensaios independentes e identicamente distribuídos é então substituída pela suposição de eventos permutáveis.

O conceito de permutabilidade é subjetivo, pois leva a considerar que a ordem da ocorrência dos eventos é irrelevante, na opinião do pesquisador. Dado que a probabilidade de ocorrência de eventos depende somente de quantos ocorreram e não quais. Neste contexto, a probabilidade da ocorrência de certos eventos não depende da sequência na qual eles aparecem. Isto é, se eles foram permutados de alguma maneira, as probabilidades não serão modificadas.

Para De Finetti, permutabilidade representa a forma correta de expressar a idéia de “ensaios independentes com probabilidade constante mas desconhecida”. Por exemplo, temos uma urna com bolas de duas cores, cuja composição é desconhecida, fazemos retiradas de bolas com reposição da mesma bola retirada. Interessa a quantidade de bolas de uma dada cor que são retiradas. A frase entre aspas significa que em relação a todas as possíveis composições da urna, as retiradas de bolas podem ser vistas como independentes com probabilidade de obter uma bola da cor de interesse, constante, mas desconhecida. De Finetti esclarece que o que é desconhecida é a composição da urna, não a probabilidade. Esta última é sempre conhecida e é dada pela opinião subjetiva sobre a composição da urna. Esta opinião modifica-se a medida que novas retiradas são feitas. Porque o conhecimento ao respeito da composição da urna vai mudando à medida que são realizadas novas retiradas de bolas, levando em consideração as frequências relativas observadas de bolas da cor de interesse.

Para os estatísticos Bayesianos, a frequência limite não é uma probabilidade, embora a probabilidade frequentista e a probabilidade (subjetiva) coincidam no caso hipotético em que a frequência limite seja conhecida. Mas, em geral, a probabilidade definida como limite de frequências relativas de sucessos é desconhecida.

Enquanto os estatísticos clássicos consideram somente ensaios independentes com probabilidade de sucesso desconhecida, os estatísticos Bayesianos consideram estes ensaios como permutáveis. Segundo o Teorema da Representação que será apresentado na próxima seção, os ensaios são condicionalmente independentes dada a frequência limite desconhecida de sucesso. A subjetividade entra na opinião pessoal sobre esta frequência limite. Este Teorema apresenta uma forte conexão entre a probabilidade (subjetiva) e a sequência de frequências relativas.

De Finetti manifesta que “probabilidade não existe”, isto é, ele refere-se ao aspecto físico, para ele a probabilidade (subjetiva) está dada por uma medida pessoal da incerteza sobre algum evento. O Teorema da Representação de Bruno De Finetti considera uma sequência infinita de variáveis aleatórias permutáveis, de modo que a seguir vamos definir permutabilidade infinita.

Definição 2.4 : (Permutabilidade infinita). A sequência de variáveis aleatórias $X_1, X_2, \dots$ é permutável se toda subsequência finita ordenada $(X_1, \dots, X_N)$ for permutável, no sentido da **Definição 2.2**.

Definição 2.5 : Um processo $X_1, X_2, \dots$ é permutável se sua distribuição for invariante sob permutações finitas de seus índices.

Teorema 2.1 : *As Definições 2.4 e 2.5 são equivalentes.*

Prova: Da **Definição 2.4** resulta a **Definição 2.5** de forma imediata.

Agora provamos que a partir da **Definição 2.5** podemos obter a **Definição 2.4**. Consideremos a função de distribuição do processo, dado um M arbitrário, com $M > N$. Seja π uma permutação de $\{1, 2, \dots, N\}$, então

$$\begin{aligned}
F(x_1, x_2, \dots, x_N) &= \lim_{\substack{x_{N+K} \rightarrow \infty \\ 1 \leq K \leq M-N}} F(x_1, x_2, \dots, x_N, x_{N+1}, \dots, x_M) \\
&= \lim_{\substack{x_{N+K} \rightarrow \infty \\ 1 \leq K \leq M-N}} F(x_{\pi(1)}, x_{\pi(2)}, \dots, x_{\pi(N)}, x_{N+1}, x_{N+2}, \dots, x_M) \\
&= F(x_{\pi(1)}, x_{\pi(2)}, \dots, x_{\pi(N)}),
\end{aligned}$$

para $N \geq 2$.

2.3 Estendibilidade

Definição 2.6 : Dizemos que uma coleção finita de variáveis aleatórias permutáveis $X_1, X_2, \dots, X_N$ é estendível se existe um processo $Z = (Z_1, Z_2, \dots)$ permutável tal que $(X_1, X_2, \dots, X_N)$ e $(Z_1, Z_2, \dots, Z_N)$ possuam a mesma distribuição para todo $N \geq 2$.

Nem todas as seqüências finitas de variáveis aleatórias permutáveis podem ser estendidas a seqüências infinitas permutáveis. Quando não é possível estender uma coleção finita de variáveis aleatórias permutáveis, o interesse centra-se na estimação da distância entre as distribuições da coleção finita e da mistura correspondente de uma seqüência infinita.

A seguir apresentamos um exemplo de seqüências finitas não estendíveis.

Exemplo 2.4 : Consideremos três variáveis aleatórias X_1, X_2, X_3 tais que $X_i = 0$ ou $X_i = 1$ para $i = 1, 2, 3$, com função de densidade conjunta dada por:

$$\begin{aligned} P(X_1 = 0, X_2 = 1, X_3 = 1) &= P(X_1 = 1, X_2 = 0, X_3 = 1) \\ &= P(X_1 = 1, X_2 = 1, X_3 = 0) \\ &= 1/3. \end{aligned}$$

Qualquer outra combinação das variáveis tem probabilidade zero. As variáveis aleatórias X_1, X_2, X_3 são permutáveis. Agora introduzimos X_4 , com $X_4 = 0$ ou $X_4 = 1$, tal que X_1, X_2, X_3, X_4 sejam permutáveis. Para isso, requeremos que

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

mas

$$\begin{aligned} P(X_1 = 0, X_2 = 1, X_3 = 1, X_4 = 0) &= \\ &= P(X_1 = 0, X_2 = 1, X_3 = 1) - P(X_1 = 0, X_2 = 1, X_3 = 1, X_4 = 1) \\ &= 1/3 - P(X_1 = 0, X_2 = 1, X_3 = 1, X_4 = 1) \\ &= 1/3 - P(X_1 = 1, X_2 = 1, X_3 = 1, X_4 = 0), \end{aligned}$$

mas

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

de modo que

$$P(X_1 = 0, X_2 = 1, X_3 = 1, X_4 = 0) = 1/3;$$

mas também temos que

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

portanto

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

Proposição 2.6 : *Seja $X_1, X_2, \dots$ um processo estocástico permutável, então*

$$Cov(X_i, X_j) \geq 0.$$

Prova: Consideremos variáveis aleatórias $X_1, X_2, \dots$ condicionalmente independentes e identicamente distribuídas dada $\mathcal{F}$, a σ -álgebra dos eventos permutáveis. Logo,

$$\begin{aligned} Cov(X_i, X_j) &= E(E(X_i X_j | \mathcal{F})) - (E(E(X_i | \mathcal{F})))^2 \\ &= E(E^2(X_i | \mathcal{F})) - (E(E(X_i | \mathcal{F})))^2 \\ &= Var(E(X_i)) \geq 0. \end{aligned}$$

As variáveis aleatórias em uma seqüência permutável infinita estão positivamente correlacionadas. Enquanto que em uma seqüência finita de variáveis aleatórias

permutáveis com correlação negativa, as variáveis não podem ser condicionalmente independentes e identicamente distribuídas, portanto não são estendíveis. De modo que a correlação não-negativa é uma condição necessária para estender uma coleção finita de variáveis aleatórias permutáveis.

Exemplo 2.5 : Voltemos ao **Exemplo 2.1** e calculemos a covariância entre as variáveis X_i e X_j ,

$$\begin{aligned}
 Cov(X_i, X_j) &= EX_i X_j - E^2 X_i \\
 &= P(X_i = 1, X_j = 1) - P^2(X_i = 1) \\
 &= \frac{A}{M} \frac{A-1}{M-1} - \left(\frac{A}{M}\right)^2 \\
 &= \frac{A}{M} \left(\frac{A-1}{M-1} - \frac{A}{M}\right) < 0.
 \end{aligned}$$

Portanto as variáveis aleatórias permutáveis $X_1, \dots, X_M$ não são estendíveis.

2.4 Teorema da Representação

Este Teorema foi apresentado pela primeira vez, junto com o conceito de permutabilidade por Bruno De Finetti no artigo “La Prévision ses lois logiques, ses sources subjectives”(1937, [7]), que foi determinante para o resurgimento da Estatística Bayesiana. De Finetti rejeitou a consideração de probabilidades objetivas e de modelos paramétricos arbitrários, onde os parâmetros são obtidos sem nenhum julgamento de invariância nos dados que o geram. (Wechsler,1993 [24]), (Barlow, 1991 [3]).

A força do Teorema da Representação surge do fato que somente é preciso para sua aplicação que as variáveis aleatórias sejam permutáveis. Em geral este Teorema fornece um resultado de existência, sem apresentar uma forma funcional para a verossimilhança na representação do processo. No entanto, quando o espaço amostral é o

conjunto $\{0,1\}$, o Teorema da Representação de Bruno de Finetti surge na sua forma mais simples. Neste caso, a verossimilhança na representação surge da estrutura do espaço amostral, e a priori, é essencialmente membro da família Beta de distribuições.

Consideramos o caso de uma seqüência infinita $X_1, X_2, \dots$ de variáveis aleatórias permutáveis com valores em $\{0,1\}^\infty$. A representação resultante para a distribuição conjunta $P(x_1, x_2, \dots, x_n)$ das n primeiras variáveis aleatórias $X_1, \dots, X_n$, é obtida da seguinte Proposição.

Proposição 2.7 : *(Teorema da Representação para variáveis aleatórias com valores em $\{0,1\}$)*

1) *Seja $X_1, X_2, \dots$, uma seqüência (infinita) de variáveis aleatórias permutáveis com valores em $\{0,1\}^\infty$, com medida de probabilidade P . Então, existe uma função de distribuição Q com suporte em $[0,1]$, tal que a densidade conjunta $P(X_1 = x_1, \dots, X_n = x_n)$, $\forall n \geq 1$ e $\forall x_i = 0,1$, apresenta a seguinte forma*

$$P(X_1 = x_1, \dots, X_n = x_n) = \int_0^1 \theta^{\sum_{i=1}^n x_i} (1 - \theta)^{(n - \sum_{i=1}^n x_i)} dQ(\theta).$$

2) *Se $X_1, X_2, \dots$ são permutáveis, então $P\{\sum_{i=1}^n X_i/n \rightarrow \}$ = 1*

3) **Observação:**

$$\lim_{n \rightarrow \infty} \left(\sum_{i=1}^n x_i/n \right) \stackrel{\text{notação}}{=} \theta$$

ou

$$\lim_{n \rightarrow \infty} \left(\sum_{i=1}^n X_i/n \right) \stackrel{\text{notação}}{=} \Theta$$

Q acima é a função de distribuição de Θ .

Prova (da parte 1): (De Finetti 1937 [7]; segue-se aqui a prova dada por Heath e Sudderth (1976) [13], Barlow, 1991 [3]). Suponha $X_1, \dots, X_N$, uma seqüência de

variáveis aleatórias permutáveis com valores em $\{0, 1\}^N$. Sejam $\sum_{i=1}^n X_i = y_n$, onde $0 \leq y_n \leq n \leq N$, e $\sum_{i=1}^N X_i = y_N$. Por hipótese temos que para todo $0 \leq y_n \leq n$ e $\forall n > 0$,

$$P\left(\sum_{i=1}^n X_i = y_n\right) = \binom{n}{y_n} P(X_1 = 1, \dots, X_{y_n} = 1, X_{y_n+1} = 0, \dots, X_n = 0),$$

se $y_n \leq y_N \leq N - (n - y_n)$ temos que,

$$\begin{aligned} P(\sum_{i=1}^n X_i = y_n) &= \sum_{y_N=y_n}^{N-(n-y_n)} P(\sum_{i=1}^n X_i = y_n | \sum_{i=1}^N X_i = y_N) P(\sum_{i=1}^N X_i = y_N) \\ &= \sum_{y_N=y_n}^{N-(n-y_n)} \binom{N}{n}^{-1} \binom{y_N}{y_n} \binom{N-y_N}{n-y_n} P(\sum_{i=1}^N X_i = y_N) \\ &= \binom{n}{y_n} \sum_{y_N=y_n}^{N-(n-y_n)} ((y_N)_{y_n} (N - y_N)_{n-y_n} / (N)_n) P(\sum_{i=1}^N X_i = y_N), \end{aligned}$$

onde $(X)_k = \prod_{j=0}^{k-1} (X - j)$.

Intuitivamente é possível imaginar que retiramos n bolas sem substituição de uma urna contendo N bolas, das quais y_N são uns e $N - y_N$ zeros, o que corresponde à distribuição Hipergeométrica.

Se agora se define $Q_N(\theta)$ sobre $\mathbb{R}$, sendo uma função escada com 0 para $\theta < 0$ e saltos de $P(\sum_{i=1}^N X_i = y_N)$ em $\theta = y_N/N$, com $y_N = 0, \dots, N$, observa-se que,

$$P(X_1 = x_1, \dots, X_n = x_n) = \int_0^1 ((\theta N)_{y_n} ((1 - \theta)N)_{n-y_n} / (N)_n) dQ_N(\theta),$$

onde

$$((\theta N)_{y_n} ((1 - \theta)N)_{n-y_n} / (N)_n) \xrightarrow{N \rightarrow \infty} \theta^{y_n} (1 - \theta)^{(n-y_n)}.$$

Pelo Teorema de Helly existe uma subsequência $Q_{N_1}, Q_{N_2}, \dots$ tal que,

$$\lim_{N_j \rightarrow \infty} Q_{N_j} = Q,$$

onde Q é uma função de distribuição.

Comentários:

1) A “Lei dos Grandes Números” (parte 2) do Teorema) estabelece que a sequência $\sum_{i=1}^n X_i/n$ vai se estabilizar em torno de um certo valor e que a probabilidade (subjativa) desse fato acontecer é um. Na parte 3) do Teorema, tem-se que

$$\lim_{n \rightarrow \infty} \left(\sum_{i=1}^n x_i/n \right) \stackrel{\text{notação}}{=} \theta,$$

determina a conexão entre a probabilidade (subjativa) e a sequência de frequências relativas, onde Q representa a sua opinião inicial sobre estas propensões (ou tendências) da ocorrência de um dado evento, em uma série infinita de ensaios.

2) Condiionadas em θ , as variáveis aleatórias $X_1, X_2, \dots$ se distribuem como Bernoulli com parâmetro θ , gerando a distribuição conjunta:

$$P(X_1 = x_1, \dots, X_n = x_n | \theta) = \prod_{i=1}^n \theta^{x_i} (1 - \theta)^{1-x_i} = \theta^{y_n} (1 - \theta)^{n-y_n}.$$

3) A função de densidade Binomial como “verossimilhança” surge então como uma consequência do fato de ser variáveis aleatórias permutáveis.

4) Na estatística clássica a “probabilidade desconhecida” de sucessos em um experimento se transforma em um parâmetro, designando-lhe o símbolo θ . O Teorema de De Finetti pode ser visto como justificando a imposição do modelo clássico de experimentos independentes condicionados no parâmetro desconhecido θ , e descreve a forma em que pode se incluir parâmetros no modelo.

O limite da seqüência de freqüências relativas, assim como os parâmetros nos modelos, em geral, não são observáveis, seus valores somente podem ser conhecidos após uma série de infinitas observações (o que é impossível de realizar). Os parâmetros nos modelos são arbitrários. Esta arbitrariedade surge de não considerar julgamentos de invariância sobre os dados observados que os geram, como permutabilidade. A eliminação desta arbitrariedade é um elemento básico do programa reducionista de De Finetti e esta atitude de restringir a consideração de parâmetros a julgamentos de invariância é denominada “preditivismo” (Wechsler,(1993) [24]).

A recíproca da parte 1 do Teorema da Representação.

Teorema 2.2 : *Seja Θ uma variável aleatória com valores em $[0, 1]$ e sejam $X_1, X_2, \dots$ variáveis condicionalmente independentes $Bernoulli(\theta)$ dada a variável $\Theta = \theta$. Então o processo $\{X_i\}_{i \geq 1}$ é permutável.*

Prova: Consideremos

$$P(X_1 = x_1, \dots, X_n = x_n) = \int_0^1 \theta^{\sum_{i=1}^n x_i} (1 - \theta)^{(n - \sum_{i=1}^n x_i)} dQ(\theta).$$

A soma é comutativa, de modo que qualquer que seja a ordem das variáveis, a probabilidade conjunta vai ser a mesma para cada $n \geq 1$, portanto pela definição de permutabilidade as variáveis $X_1, X_2, \dots$ são permutáveis.

Comentários

1) Se Q fosse uma função de distribuição degenerada, as variáveis além de serem permutáveis seriam independentes (e identicamente distribuídas).

2) O Teorema da Representação assume uma seqüência infinita de variáveis aleatórias permutáveis, e não é válido para seqüências finitas permutáveis, como vemos no seguinte exemplo:

Exemplo 2.6 : Sejam $n = 2$ e X_1, X_2 variáveis aleatórias finitamente permutáveis tais que

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

$$P(X_1 = 1, X_2 = 0) = P(X_1 = 0, X_2 = 1) = 1/2;$$

se fosse possível usar o Teorema da Representação:

$$\int_0^1 \theta^2 dQ(\theta) = \int_0^1 (1 - \theta)^2 dQ(\theta) = 0$$

para alguma $Q(\cdot)$. Logo não é possível, pois isto implicaria que $P(\theta = 0) = 1$ e $P(\theta = 1) = 1$, obtendo um absurdo.

Corolário 2.2 : Dadas as condições da **Proposição 2.7**, temos que

$$P(X_1 + \dots + X_n = y_n) = \binom{n}{y_n} \int_0^1 \theta^{y_n} (1 - \theta)^{(n-y_n)} dQ(\theta).$$

Prova: Segue da **Proposição 2.7** e do seguinte fato,

$$P(X_1 + \dots + X_n = y_n) = \binom{n}{y_n} P(X_1 = x_1, \dots, X_n = x_n),$$

para todo $x_1, x_2, \dots, x_n$ tal que $\sum_{i=1}^n x_i = y_n$.

Em muitas aplicações onde são consideradas variáveis aleatórias com valores no conjunto $\{0, 1\}$, pode ser de mais interesse trabalhar com a soma das variáveis, que com a seqüência individual das variáveis aleatórias. Isto providencia uma justificativa da nossa opinião sobre y_n , considerando a verossimilhança dada por uma Binomial($y_n|\theta, n$) e uma distribuição a priori $Q(\theta)$ para o parâmetro Θ da Binomial.

Corolário 2.3 : Se $X_1, X_2, \dots$ é uma seqüência permutável infinita de variáveis aleatórias com valores no $\{0, 1\}^\infty$ com medida de probabilidade P , a probabilidade condicional $P(X_{m+1} = x_{m+1}, \dots, X_n = x_n | X_1 = x_1, \dots, X_m = x_m)$ para $X_{m+1}, \dots, X_n$ dado $x_1, \dots, x_m$, apresenta a seguinte forma

$$P(X_{m+1} = x_{m+1}, \dots, X_n = x_n | X_1 = x_1, \dots, X_m = x_m) = \int_0^1 \prod_{i=m+1}^n \theta^{(x_i)} (1 - \theta)^{(1-x_i)} dQ(\theta | x_1, \dots, x_m),$$

com $1 \leq m < n$, onde

$$dQ(\theta | x_1, \dots, x_m) = \frac{\prod_{i=1}^m \theta^{(x_i)} (1 - \theta)^{(1-x_i)} dQ(\theta)}{\int_0^1 \prod_{i=1}^m \theta^{(x_i)} (1 - \theta)^{(1-x_i)} dQ(\theta)},$$

e

$$Q(\theta) = \lim_{n \rightarrow \infty} P(y_n/n \leq \theta).$$

Prova:

$$\begin{aligned} P(x_{m+1}, \dots, x_n | x_1, \dots, x_m) &= \frac{P(X_1 = x_1, \dots, X_n = x_n)}{P(X_1 = x_1, \dots, X_m = x_m)} \\ &= \frac{\int_0^1 \prod_{i=1}^n \theta^{(x_i)} (1 - \theta)^{(1-x_i)} dQ(\theta)}{\int_0^1 \prod_{i=1}^m \theta^{(x_i)} (1 - \theta)^{(1-x_i)} dQ(\theta)}, \text{ pela Proposição 2.7} \\ &= \frac{\int_0^1 \prod_{i=m+1}^n \theta^{(x_i)} (1 - \theta)^{(1-x_i)} \prod_{i=1}^m \theta^{(x_i)} (1 - \theta)^{(1-x_i)} dQ(\theta)}{\int_0^1 \prod_{i=1}^m \theta^{(x_i)} (1 - \theta)^{(1-x_i)} dQ(\theta)}, \end{aligned}$$

a distribuição a priori para Θ é atualizada por meio do Teorema de Bayes, obtendo desse modo a distribuição a posteriori dada por $Q(\theta | x_1, \dots, x_m)$.

Corolário 2.4 : Se $X_1, X_2, \dots$ são variáveis aleatórias permutáveis, então

$$P(X_{n+1} = x_{n+1} | X_1 = x_1, \dots, X_n = x_n) = \begin{cases} E(\theta | y_n), & \text{se } x_{n+1} = 1 \\ 1 - E(\theta | y_n), & \text{se } x_{n+1} = 0 \end{cases}$$

para toda função de distribuição a priori $Q(\cdot)$, com $y_n = \sum_{i=1}^n x_i$.

Prova:

$$\begin{aligned}
P(X_{n+1} = x_{n+1} | X_1 = x_1, \dots, X_n = x_n) &= \frac{P(X_1 = x_1, \dots, X_n = x_n, X_{n+1} = x_{n+1})}{P(X_1 = x_1, \dots, X_n = x_n)} \\
&= \frac{\int_0^1 \theta^{\sum_{i=1}^{n+1} x_i} (1 - \theta)^{n+1 - \sum_{i=1}^{n+1} x_i} dQ(\theta)}{\int_0^1 \theta^{y_n} (1 - \theta)^{n - y_n} dQ(\theta)} \\
&= \frac{\int_0^1 \theta^{y_n} (1 - \theta)^{n - y_n} \theta^{x_{n+1}} (1 - \theta)^{1 - x_{n+1}} dQ(\theta)}{\int_0^1 \theta^{y_n} (1 - \theta)^{n - y_n} dQ(\theta)} \\
&= \int_0^1 \theta^{x_{n+1}} (1 - \theta)^{1 - x_{n+1}} dQ(\theta | y_n) \\
&= \begin{cases} E(\theta | y_n), & \text{se } x_{n+1} = 1 \\ 1 - E(\theta | y_n), & \text{se } x_{n+1} = 0. \end{cases}
\end{aligned}$$

As igualdades surgem da hipótese de permutabilidade e da **Proposição 2.7**.

2.5 Simulações

A idéia que surge após introduzir os conceitos teóricos é fazer simulações de processos aleatórios permutáveis de variáveis aleatórias com valores no conjunto $\{0, 1\}^\infty$ a partir do Teorema da Representação, considerando diferentes casos, com o objetivo de observar que todas levam à geração de seqüências permutáveis equivalentes. Dadas as hipóteses da **Proposição 2.7**, a densidade conjunta é dada por,

$$P(X_1 = x_1, \dots, X_n = x_n) = \int_0^1 \theta^{\sum_{i=1}^n x_i} (1 - \theta)^{(n - \sum_{i=1}^n x_i)} dQ(\theta).$$

Para começar vamos simular um processo aleatório usando o fato que a probabilidade marginal pode ser escrita como o produto de probabilidades condicionais, isto

é,

$$P(X_1 = x_1, \dots, X_n = x_n) = P(X_1 = x_1)P(X_2 = x_2|X_1 = x_1) \dots P(X_n = x_n|X_1 = x_1, \dots, X_{n-1} = x_{n-1}).$$

Vamos gerar matrizes (10x20) de observações do processo $X_1, X_2, \dots$ de variáveis aleatórias com valores no conjunto $\{0, 1\}^\infty$, representando 10 seqüências de 20 observações cada uma. Serão considerados dois casos, a saber:

I) As seqüências de variáveis aleatórias com valores em $\{0, 1\}^{20}$ foram obtidas através do cálculo das probabilidades condicionais mediante as correspondentes esperanças condicionais. Neste caso a distribuição a priori Q utilizada é $\text{Beta}(a, b)$, escolhendo arbitrariamente os valores de $a = 2$ e $b = 3$. De modo que no início temos $P(X_1 = 1) = E_Q(\theta) = \frac{a}{a+b}$. Calculamos esta probabilidade e comparamos seu valor com o valor de uma uniforme, obtido de forma aleatória; se a probabilidade de $X_1 = 1$ é maior do que o valor gerado pela densidade uniforme, então damos o valor 1 para a variável aleatória X_1 , 0 em caso contrário. Logo para calcular as probabilidades condicionais seguintes se faz do mesmo jeito, só que usamos a esperança condicional da posteriori, que pertence também à família de densidades Beta, mas com os parâmetros atualizados. Deste modo, geramos 20 valores da primeira seqüência; para as outras seqüências, só repetimos este procedimento.

II) A distribuição a priori utilizada neste caso também é uma $\text{Beta}(2,3)$. Geramos um valor desta distribuição e comparamos com um valor gerado por uma distribuição uniforme. Assim, determinamos o valor da variável na seqüência. Com este valor atualizamos o valor da esperança, agora condicionada a esse valor. Logo, geramos uma Beta com os parâmetros também atualizados e continuamos com este procedimento até obter os 20 valores da seqüência. Repetimos estes passos para formar a matriz dada por 10 seqüências.

Observação: A diferença entre estes dois métodos consiste em que no primeiro,

utilizamos o valor da esperança condicional atualizada para comparar com um valor gerado por uma distribuição uniforme, enquanto, no segundo método, geramos um valor de uma distribuição Beta que tem a esperança condicional atualizada.

Agora vamos gerar a matriz(10x20) a partir do segundo membro do Teorema da Representação.

III) A distribuição a priori que vamos utilizar é a mesma dos casos anteriores, isto é, uma distribuição Beta(2,3). Para o cálculo, consideramos todos os valores possíveis de somas (que representam a quantidade de variáveis aleatórias com valor um). Cada soma possui uma dada probabilidade de sair. Dividimos o intervalo $[0,1]$ em subintervalos, de modo que cada um deles corresponda à probabilidade relacionada com o valor da soma. Para obter a seqüência de $\{0,1\}$, sorteamos uma uniforme; vemos a qual subintervalo pertence, e portanto, sabemos qual é o valor da soma correspondente. Deste modo conhecemos a quantidade de variáveis aleatórias com valor um que a seqüência terá. Logo se procede a colocar ao acaso na seqüência as variáveis aleatórias com valores uns, preenchendo os restantes lugares com variáveis aleatórias com valores zeros.

Todas estas matrizes foram geradas usando-se a mesma expressão, isto é, o Teorema da Representação, só que utilizando diferentes métodos de obtenção das seqüências. Nos três casos considerados foi utilizada a mesma distribuição Beta(2,3) como distribuição a priori. As seqüências de zeros e uns foram obtidas ao acaso usando um método ou outro. O interessante é ver que a partir do Teorema da Representação é possível gerar seqüências de variáveis aleatórias permutáveis. É claro que aqui foram consideradas só três formas de construir matrizes de zeros e uns, existindo talvez outras formas possíveis de gerá-las.

Capítulo 3

PROCESSOS DE URNA

Suponhamos que uma urna contém inicialmente b bolas brancas e v bolas vermelhas e, a cada passo, uma bola é escolhida ao acaso e retornada à urna acompanhada de uma nova bola da mesma cor. O processo de Pólya é um caso especial de uma família de processos estocásticos, cuja característica distinguível é dada porque a sequência de observações pode ser representada por retiradas sucessivas de urnas com composições que vão mudando.

Definição 3.1 : Um processo estocástico $(Y_1, Y_2, \dots)$ definido em um espaço de probabilidade $(\Omega, \mathcal{F}, P)$, com valores em $\{0, 1\}^\infty$, denomina-se “processo de Urna” com composição inicial (b, v) , onde $b, v \in \mathbb{N}$ e uma função $f: [0, 1] \rightarrow [0, 1]$, sempre que,

1) X_0 é uma variável aleatória com distribuição degenerada em $b/(b + v)$.

2) X_1 é uma variável aleatória com distribuição condicional,

$P(X_1 = \frac{b+y_1}{b+v+1} | Y_1 = y_1) = 1$, para $y_1 \in \{0, 1\}$, onde Y_1 é variável aleatória com distribuição condicional,

$$P(Y_1 = y_1 | X_0 = x_0) = f(x_0);$$

$Y_1 | X_0 = x_0$ possui distribuição Bernoulli($f(x_0)$).

3) X_2 é uma variável aleatória com distribuição condicional,

$P(X_2 = \frac{b+y_1+y_2}{b+v+2} | Y_1 = y_1, Y_2 = y_2) = 1$, para $(y_1, y_2) \in \{0, 1\}^2$, onde Y_2 é variável aleatória com distribuição condicional,

$$P(Y_2 = y_2 | X_1 = x_1) = f(x_1);$$

$Y_2 | X_1 = x_1$ possui distribuição Bernoulli($f(x_1)$).

4) Em geral, X_n é uma variável aleatória com distribuição condicional,

$P(X_n = \frac{b+\sum_{i=1}^n y_i}{b+v+n} | Y_1 = y_1, \dots, Y_n = y_n) = 1$, para $(y_1, \dots, y_n) \in \{0, 1\}^n$, onde Y_n é variável aleatória com distribuição condicional,

$$P(Y_n = y_n | X_{n-1} = x_{n-1}) = f(x_{n-1});$$

$Y_n | X_{n-1} = x_{n-1}$ possui distribuição Bernoulli($f(x_{n-1})$), para $n \geq 1$.

Observação: A função $f: [0, 1] \rightarrow [0, 1]$ é denominada função de urna.

Comentário: Seja $Y_1, Y_2, \dots$ uma seqüência de variáveis aleatórias permutáveis, a conexão que existe entre $P(Y_n = y_n | X_{n-1} = x_{n-1})$ e $P(Y_n = y_n | Y_1 = y_1, \dots, Y_{n-1} = y_{n-1})$ é dada considerando que $P(Y_1 + \dots + Y_{n-1} = k) = \binom{n-1}{k} P(Y_1 = y_1, \dots, Y_{n-1} = y_{n-1})$, com $x_{n-1} = \frac{b+\sum_{k=1}^{n-1} y_k}{b+v+n-1}$, onde $b, v \in \mathbb{N}$.

Definição 3.2 : O processo de Urna de Pólya é um processo de Urna com função de urna dada por $f(x) = x$.

As perguntas que surgem agora são as seguintes: Será que todos os processos de urna são permutáveis? Todos os processos estocásticos permutáveis podem ser representados como processos de urna? Existem seqüências $(Y_1, Y_2, \dots)$ de variáveis aleatórias com valores em $\{0, 1\}^\infty$ que não são permutáveis?

Para responder estas questões, vejamos os seguintes exemplos.

Exemplo 3.1 : Nem todos os processos de urna são permutáveis. Por exemplo, consideremos o processo $\{Y_n\}_{n \geq 1}$ de variáveis aleatórias permutáveis com valores em $\{0, 1\}^\infty$, definido em um espaço de probabilidade $(\Omega, \mathcal{F}, P)$, e com composição inicial dada por $(1, 2)$, isto é, uma bola vermelha e duas brancas. Seja a seguinte função de urna: $f(X)$ onde para cada x valor possível de X , $f(x) = 1 - x$. Vamos demonstrar que $P(Y_1 = 1, Y_2 = 0)$ é diferente de $P(Y_1 = 0, Y_2 = 1)$, assim, considerando primeiro $Y_1 = 1, Y_2 = 0$, temos o seguinte processo de urna,

$$x_0 = \frac{v}{v+b} = 1/3, f(x_0) = 1 - x_0 = 2/3,$$

$$x_1 = \frac{v+1}{v+b+1} = 2/4, f(x_1) = 1 - x_1 = 1/2.$$

$$\text{daqui que } P(Y_1 = 1, Y_2 = 0) = P(Y_2 = 0 | Y_1 = 1)P(Y_1 = 1) = (1/2)(2/3) = 1/3.$$

Enquanto que se o processo fosse o processo de urna de Pólya,

$$P(Y_1 = 1, Y_2 = 0) = P(Y_2 = 0 | Y_1 = 1)P(Y_1 = 1) = (1/2)(1/3) = 1/6.$$

Considerando agora que $Y_1 = 0$ e $Y_2 = 1$, vamos ter que,

$$x_0 = \frac{v}{v+b} = 1/3, f(x_0) = 1 - x_0 = 2/3,$$

$$x_1 = \frac{v}{v+b+1} = 1/4, f(x_1) = 1 - x_1 = 3/4,$$

$$\text{então } P(Y_1 = 0, Y_2 = 1) = P(Y_2 = 1 | Y_1 = 0)P(Y_1 = 0) = (3/4)(1/3) = 1/4.$$

Se o processo de urna fosse o processo de urna de Pólya,

$$P(Y_1 = 0, Y_2 = 1) = P(Y_2 = 1 | Y_1 = 0)P(Y_1 = 0) = (1/4)(2/3) = 1/6.$$

Observamos que o processo de urna, fixado $n = 2$, com função de urna dada por $f(x) = 1 - x$, não é permutável com essa composição inicial, contudo, conferimos que o processo de Pólya nessas condições é permutável.

Também nem todos os processos permutáveis podem ser representados como processos de urna, como vemos no próximo exemplo.

Exemplo 3.2 : Seja $(Y_1, Y_2, \dots)$ uma seqüência de variáveis aleatórias permutáveis com valores em $\{0, 1\}^\infty$, e seja Θ uma variável aleatória tal que dado $\Theta = \theta$, com $\theta \in [0, 1]$, as variáveis aleatórias $Y_1, Y_2, \dots$ são Bernoulli(θ) independentes. Considerando as seguintes probabilidades,

$$P(\theta = 1/3) = 1/2, P(\theta = 2/3) = 1/2.$$

é simples ver que o processo $\{Y_n\}_{n \geq 1}$ não é um processo de urna dado pela **Definição 3.1**, já que este processo pode consistir em duas urnas com diferentes composições iniciais de bolas, isto é, uma bola vermelha e duas brancas na primeira urna e duas bolas vermelhas e uma branca na segunda urna, e para retirar uma bola vamos escolher uma urna, por exemplo, jogando uma moeda honesta.

Comentários:

i) Observando a definição do processo de Pólya, nota-se que basta conhecer $S_n = \sum_{i=1}^n y_i$, com $n \geq 1$, para que o processo esteja determinado, porque é preciso saber a quantidade de variáveis aleatórias iguais a um para conhecer a probabilidade de Y_{n+1} , assumir valor um.

ii) O processo de urna de Pólya $\{Y_n\}_{n \geq 1}$, com composição inicial (b, v) , $b, v \in \mathbb{N}$, não apresenta a propriedade markoviana, porque pelo comentário i) nota-se que a probabilidade condicional de $Y_n = y_n$ dado $Y_1 = y_1, \dots, Y_{n-2} = y_{n-2}, Y_{n-1} = y_{n-1}$, não depende somente de y_{n-1} , mas também de $y_1, \dots, y_{n-2}$, isto é,

$$P(Y_n = y_n | Y_1 = y_1, \dots, Y_{n-1} = y_{n-1}) = \begin{cases} \frac{b+S_{n-1}}{v+b+(n-1)}, & \text{se } y_n = 1 \\ \frac{v+n-1-S_{n-1}}{v+b+(n-1)}, & \text{se } y_n = 0, \end{cases}$$

com $S_{n-1} = \sum_{i=1}^{n-1} y_i$.

Observações:

1) Se no processo de urna $\{Y_n\}_{n \geq 1}$ a função de urna é constante, isto é, $f(x) = p$, então o processo de urna é denominado processo Bernoulli, e as variáveis aleatórias $Y_1, Y_2, \dots$ são independentes com distribuição Bernoulli(p). Em consequência o processo Bernoulli é markoviano.

2) A distribuição condicional $P(Y_n = y_n | Y_1 = y_1, \dots, Y_{n-1} = y_{n-1})$ é função de $S_n = \sum_{i=1}^n y_i$ (vide **Definição 3.1**), e como a soma é invariante sob permutações de $(Y_1, \dots, Y_n)$, então o processo estocástico $\{Y_n\}_{n \geq 1}$ é permutável.

3) É interessante notar que o processo $\{S_n\}_{n \geq 1}$ apresenta a propriedade markoviana,

$$P(S_n = s_n | S_1 = s_1, \dots, S_{n-1} = s_{n-1}) = \begin{cases} \frac{b+s_{n-1}}{v+b+(n-1)}, & s_n = 1 + s_{n-1} \\ \frac{v+n-s_{n-1}-1}{v+b+(n-1)}, & s_n = s_{n-1}. \end{cases}$$

Observa-se que a lei de transição depende do tempo n , em outras palavras, o processo $\{S_n\}_{n \geq 1}$ é um processo de Markov com lei de transição não-estacionária, isto é, a distribuição condicional depende de n e não do estado ao tempo n .

iii) **Teorema:** *Seja $\{Y_n\}_{n \geq 1}$ um processo de urna definido no espaço de probabilidade $(\Omega, \mathcal{F}, P)$. Conhecendo*

$$P(Y_1 = y_1), P(Y_2 = y_2 | Y_1 = y_1), \dots, P(Y_n = y_n | Y_1 = y_1, \dots, Y_{n-1} = y_{n-1}),$$

para $n \geq 2$, podemos obter a distribuição conjunta do processo de urna para um n arbitrário.

Prova: (Teorema da Probabilidade Composta), para $n \geq 2$,

$$P(Y_1 = y_1, \dots, Y_n = y_n) = P(Y_1 = y_1)P(Y_2 = y_2 | Y_1 = y_1) \dots P(Y_n = y_n | Y_1 = y_1, \dots, Y_{n-1} = y_{n-1}),$$

iv) Pelo comentário iii), a definição do processo de Pólya de fato define a distribuição do próprio processo, fixado o valor n .

3.1 Processo de Urna de Pólya

O modelo de Pólya (1923) representa uma forma interessante para gerar seqüências de variáveis aleatórias com valores em $\{0, 1\}^\infty$. Neste esquema uma única urna contém b bolas brancas e v bolas vermelhas. Uma bola é retirada ao acaso e substituída por $s + 1$ bolas da mesma cor, este procedimento é repetido n vezes.

Denominemos U à variável aleatória que representa o número de vezes que uma bola vermelha é obtida nos n ensaios, então a distribuição de U está dada por,

$$P(U = k) = \binom{n}{k} \frac{v(v+s)\dots(v+(k-1)s)b(b+s)\dots(b+(n-k-1)s)}{(b+v)(b+v+s)\dots(b+v+(n-1)s)} \quad (3.1)$$

com $k = 0, 1, \dots, n$, sendo s o número de bolas adicionadas à urna. O fator $\binom{n}{k}$ representa o número de possíveis seqüências de “ k -sucessos”. Quando $s = 1$, podemos escrever (3.1) da seguinte forma,

$$P(U = k) = \binom{n}{k} \frac{[v]_k [b]_{n-k}}{[b+v]_n} = \binom{n}{k} \frac{\prod_{j=0}^{k-1} (v+j) \prod_{j=0}^{n-k-1} (b+j)}{\prod_{j=0}^{n-1} (b+v+j)}, \quad (3.2)$$

com $[x]_y = \prod_{j=0}^{y-1} (x+j)$.

Observação: No caso especial de ter $s = 0$, a distribuição da variável U é a distribuição Binomial com parâmetros n e $v/(b+v)$. Agora se $b = v = s$, a distribuição de U se transforma em uma distribuição Retangular, e por último quando $s = -1$, ela possui distribuição Hipergeométrica.

A probabilidade $P(U = k)$ pode ser escrita de muitas outras maneiras, assim particularmente, podemos considerar a seguinte seqüência de k -sucessos, onde denominando $\alpha = v/s$ e $\beta = b/s$, vamos ter

$$P(U = k) = \binom{n}{k} \frac{\alpha(\alpha+1)\dots(\alpha+(k-1))\beta(\beta+1)\dots(\beta+n-k-1)}{(\alpha+\beta)(\alpha+\beta+1)\dots(\alpha+\beta+(n-1))} \quad (3.3)$$

$$= \binom{n}{k} \frac{\Gamma(\alpha+\beta)\Gamma(\alpha+k)\Gamma(\beta+n-k)}{\Gamma(\alpha)\Gamma(\beta)\Gamma(\alpha+\beta+n)}. \quad (3.4)$$

Sendo $Beta(\alpha+k, \beta+n-k) = \frac{\Gamma(\alpha+k)\Gamma(\beta+n-k)}{\Gamma(\alpha+\beta+n)}$, e $Beta(\alpha+\beta) = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}$,

podemos escrever (3.4) da seguinte forma,

$$P(U = k) = \binom{n}{k} \frac{Beta(\alpha+k, \beta+n-k)}{Beta(\alpha, \beta)}. \quad (3.5)$$

A distribuição da variável aleatória U também pode ser escrita desta maneira

$$P(U = k) = \binom{n}{k} \frac{[\alpha]_k [\beta]_{n-k}}{[\alpha+\beta]_n} = \binom{n}{k} \frac{\prod_{j=0}^{k-1} (\alpha+j) \prod_{j=0}^{n-k-1} (\beta+j)}{\prod_{j=0}^{n-1} (\alpha+\beta+j)}, \quad (3.6)$$

com $[x]_y = \prod_{j=0}^{y-1} (x+j)$.

Observamos que existe uma dependência entre n e os quocientes $\alpha = v/s$ e $\beta = b/s$. Além disso sugere-se alguma conexão com as distribuições da família Beta.

Observação: (Johnson, N., Kotz, S. (1977) [17]). A característica comum destes modelos consiste em que a amostragem é feita junto com alguma forma de substituição da bola, que não é simplesmente a devolução da bola que foi retirada. A maneira na qual a substituição é feita determina a particular subclasse de distribuições. Em particular, a forma de reposição depende da cor da bola retirada.

Existem várias formas em que estes modelos podem ser generalizados. Mencionamos algumas:

1) A cada etapa, s' bolas da cor oposta daquela retirada pode ser adicionada junto com s bolas da mesma cor.

- 2) As quantidades s e s' podem depender da cor da bola retirada.
- 3) s (ou s') pode ser negativa (neste caso se $s = -1$ e $s' = 0$ temos amostragem sem reposição).
- 4) Os valores de s (e s') podem variar com o número de ensaios, isto é, podemos especificar previamente o valor s_t de s (s'_t para s') no t -ésimo ensaio.
- 5) s_t (e s'_t) podem ser variáveis aleatórias.
- 6) Na urna poderiam existir mais de duas cores de bolas, o que conduz uma distribuição multivariada.

Observação: modificações de 1) até o 5) podem ser combinadas com 6).

3.2 Permutabilidade

Vamos mostrar por indução que o processo de Pólya é permutável. Para isto consideramos o processo $(Y_1, Y_2, \dots)$ definido em um espaço de probabilidade $(\Omega, \mathcal{F}, P)$, com composição de urna dada por (a, v) , onde $a, v \in \mathbb{N}$, e seja π qualquer permutação do conjunto $\{1, \dots, n\}$. Provamos que a propriedade é válida para $n = 2$,

$$\begin{aligned}
 P(Y_1 = y_1, Y_2 = y_2) &= \left(\frac{a + y_1}{a + v + 1}\right)^{y_2} \left(1 - \frac{a + y_1}{a + v + 1}\right)^{1-y_2} \left(\frac{a}{a + v}\right)^{y_1} \left(1 - \frac{a}{a + v}\right)^{1-y_1} I_{\{0,1\}^2}(y_1, y_2) \\
 &= \left(\frac{a + y_2}{a + v + 1}\right)^{y_1} \left(1 - \frac{a + y_2}{a + v + 1}\right)^{1-y_1} \left(\frac{a}{a + v}\right)^{y_2} \left(1 - \frac{a}{a + v}\right)^{1-y_2} I_{\{0,1\}^2}(y_1, y_2) \\
 &= P(Y_1 = y_2, Y_2 = y_1),
 \end{aligned}$$

vamos supor que a propriedade é válida para $n - 1$ variáveis aleatórias,

$$P(Y_1 = y_1, \dots, Y_{n-1} = y_{n-1}) = \prod_{j=1}^{n-1} \left(\frac{a + \sum_{k=0}^{j-1} y_k}{a + v + (j-1)}\right)^{y_j} \left(1 - \frac{a + \sum_{k=0}^{j-1} y_k}{a + v + (j-1)}\right)^{(1-y_j)} I_{\{0,1\}^n}(y_1, \dots, y_j)$$

$$= \prod_{j=1}^{n-1} \left(\frac{a + \sum_{k=0}^{j-1} y_{\pi(k)}}{a + v + (j-1)} \right)^{y_{\pi(j)}} \left(1 - \frac{a + \sum_{k=0}^{j-1} y_{\pi(k)}}{a + v + (j-1)} \right)^{(1-y_{\pi(j)})}$$

$$I_{\{0,1\}^J}(y_{\pi(1)}, \dots, y_{\pi(j)})$$

$$= P(Y_1 = y_{\pi(1)}, \dots, Y_{n-1} = y_{\pi(n-1)}).$$

Verificamos a propriedade para n variáveis aleatórias,

$$\begin{aligned} P(Y_1 = y_1, \dots, Y_n = y_n) &= \left(\frac{a + \sum_{k=0}^{n-2} y_k + y_{n-1}}{a + b + n - 1} \right)^{y_n} \left(1 - \frac{a + \sum_{k=0}^{n-2} y_k + y_{n-1}}{a + b + n - 1} \right)^{(1-y_n)} I_{\{0,1\}}(y_n) \\ &\quad \prod_{j=1}^{n-1} \left(\frac{a + \sum_{k=0}^{j-1} y_k}{a + b + (j-1)} \right)^{y_j} \left(1 - \frac{a + \sum_{k=0}^{j-1} y_k}{a + b + (j-1)} \right)^{(1-y_j)} I_{\{0,1\}^J}(y_1, \dots, y_j) \quad (3.7) \\ &= \left(\frac{a + \sum_{k=0}^{n-2} y_{\pi(k)} + y_{\pi(n-1)}}{a + b + n - 1} \right)^{y_{\pi(n)}} \left(1 - \frac{a + \sum_{k=0}^{n-2} y_{\pi(k)} + y_{\pi(n-1)}}{a + b + n - 1} \right)^{(1-y_{\pi(n)})} \\ &\quad I_{\{0,1\}}(y_{\pi(n)}) \prod_{j=1}^{n-1} \left(\frac{a + \sum_{k=0}^{j-1} y_{\pi(k)}}{a + b + (j-1)} \right)^{y_{\pi(j)}} \left(1 - \frac{a + \sum_{k=0}^{j-1} y_{\pi(k)}}{a + b + (j-1)} \right)^{(1-y_{\pi(j)})} \\ &\quad I_{\{0,1\}^J}(y_{\pi(1)}, \dots, y_{\pi(j)}). \quad (3.8) \end{aligned}$$

Das equações (3.7) e (3.8) temos que

$$\begin{aligned} P(Y_1 = y_1, \dots, Y_n = y_n) &= \prod_{j=1}^n \left(\frac{a + \sum_{k=0}^{j-1} y_k}{a + b + (j-1)} \right)^{y_j} \left(1 - \frac{a + \sum_{k=0}^{j-1} y_k}{a + b + (j-1)} \right)^{(1-y_j)} I_{\{0,1\}^J}(y_1, \dots, y_j) \\ &= \prod_{j=1}^n \left(\frac{a + \sum_{k=0}^{j-1} y_{\pi(k)}}{a + b + (j-1)} \right)^{y_{\pi(j)}} \left(1 - \frac{a + \sum_{k=0}^{j-1} y_{\pi(k)}}{a + b + (j-1)} \right)^{(1-y_{\pi(j)})} I_{\{0,1\}^J}(y_{\pi(1)}, \dots, y_{\pi(j)}) \\ &= P(Y_1 = y_{\pi(1)}, \dots, Y_n = y_{\pi(n)}). \end{aligned}$$

Portanto,

$$P(Y_1 = y_1, \dots, Y_n = y_n) = P(Y_1 = y_{\pi(1)}, \dots, Y_n = y_{\pi(n)}).$$

3.3 Caracterização da Priori para o Processo de Pólya

Sabendo que o processo de Pólya $\{Y_n\}_{n \geq 1}$ é permutável, podemos aplicar o Teorema da Representação de De Finetti para obter a probabilidade da variável aleatória resumo U (que representa maior interesse que a distribuição conjunta do processo $\{Y_n\}_{n \geq 1}$). A representação pode ser obtida como uma mistura de distribuições de uma seqüência de variáveis aleatórias independentes e identicamente distribuídas Bernoulli(θ), condicionadas em $\Theta = \theta$, com $Q(\theta)$ a distribuição a priori na mistura,

$$P(U = k) = \binom{N}{k} \int_0^1 \theta^k (1 - \theta)^{N-k} dQ(\theta), \quad (3.9)$$

$\forall N \in \mathbb{N}$ e $0 \leq k \leq N$. Vimos em (3.4) que podemos escrever a $P(U = k)$ da seguinte forma,

$$P(U = k) = \binom{N}{k} \frac{\Gamma(\alpha + \beta) \Gamma(\alpha + k) \Gamma(\beta + N - k)}{\Gamma(\alpha) \Gamma(\beta) \Gamma(\alpha + \beta + N)},$$

então de (3.9) e (3.4), igualando vamos ter que,

$$\int_0^1 \theta^k (1 - \theta)^{N-k} dQ(\theta) = \frac{(\alpha + \beta - 1)! (\alpha + k - 1)! (\beta + N - k - 1)!}{(\alpha - 1)! (\beta - 1)! (\alpha + \beta + N - 1)!} \quad (3.10)$$

Podemos atribuir uma função de distribuição a priori para Θ que satisfaça a igualdade (3.10), a candidata natural é a distribuição Beta cuja função de densidade está dada por

$$f(\theta) = \frac{1}{\beta(a, b)} \theta^{(a-1)} (1 - \theta)^{(b-1)}, \quad (0 \leq \theta \leq 1), \quad a, b > 0, \quad (3.11)$$

$$\text{com } \beta(a, b) = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}.$$

$dQ(\theta) = c\theta^x(1-\theta)^y d\theta$, onde c é uma constante positiva que vamos determinar, assim substituindo e integrando obtemos

$$\int_0^1 \theta^k(1-\theta)^{N-k} f(\theta) d\theta = c \int_0^1 \theta^{k+x}(1-\theta)^{N-k+y} d\theta = c \frac{(k+x)!(N-k+y)!}{(N+x+y+1)!} \quad (3.12)$$

para que a igualdade entre as equações (3.10) e (3.12) seja verificada é preciso tomar

$$c = \frac{(\alpha+\beta-1)!}{(\alpha-1)!(\beta-1)!}, \text{ com } x = \alpha - 1 \text{ e } y = \beta - 1,$$

sendo então

$$f(\theta) = \frac{(\alpha+\beta-1)!}{(\alpha-1)!(\beta-1)!} \theta^{(\alpha-1)} (1-\theta)^{(\beta-1)} I_{[0,1]}(\theta).$$

de modo que $f(\theta)$ representa a função de densidade de uma Beta com parâmetros $\alpha = v/s$ e $\beta = b/s$. Portanto o processo de Pólya além de ser permutável, possui a distribuição Beta($v/s, b/s$) como medida de De Finetti.

3.3.1 A urna de Bernard Friedman

(Freedman, D. (1965) [11]) Bernard Friedman em 1949 propôs um modelo que consiste em uma urna que contém a_n bolas azuis e b_n bolas brancas no tempo n . Uma bola é retirada ao acaso e logo substituída por α bolas da mesma cor da bola retirada e β bolas da cor oposta, a bola retirada também é colocada na urna. Quando $\beta = 0$ e $\alpha = 1$ temos a Urna de Pólya; as proporções $X_n = \{\frac{a_n}{b_n+a_n}\}_{n \geq 0}$ convergem com probabilidade 1 a uma variável aleatória limite Θ , que possui distribuição Beta com parâmetros a_0 e b_0 . Dado $\Theta = \theta$, as diferenças sucessivas $Y_n = a_{n+1} - a_n$, para $n \geq 0$, são condicionalmente independentes e identicamente distribuídas, sendo $Y_n = 1$ com probabilidade θ ou $Y_n = 0$ com probabilidade $1-\theta$.

Lema 3.1 : *Seja $\{a_n, b_n\}_{n \geq 0}$, um processo de urna com parâmetros $\alpha = 1$ e $\beta = 0$,*

definido em um espaço de probabilidade $(\Omega, \mathcal{F}, P)$. O processo $X_n = \{\frac{a_n}{b_n+a_n}\}_{n \geq 0}$ é uma martingala, e converge com probabilidade 1 a uma variável aleatória limite Θ .

Prova: A primeira afirmação é provada de forma simples, assim, verificamos que $\{\frac{a_n}{b_n+a_n}\}_{n \geq 0}$ é martingala com respeito à σ -álgebra $\mathcal{F}_n$, onde $\mathcal{F}_n = \sigma(X_0, X_1, \dots, X_n)$. X_n é $\mathcal{F}_n$ - mensurável e $E(\frac{a_n}{a_n+b_n}) < \infty$ para $n \geq 0$ então,

$$\begin{aligned} E(X_{n+1}|\mathcal{F}_n) &= E(\frac{a_n + Y_{n+1}}{a_n + b_n + 1}|\mathcal{F}_n) \\ &= E(\frac{a_n}{a_n + b_n + 1}|\mathcal{F}_n) + E(\frac{Y_{n+1}}{a_n + b_n + 1}|\mathcal{F}_n) \\ &= \frac{a_n}{a_n + b_n + 1} + E(\frac{Y_{n+1}}{a_n + b_n + 1}|\mathcal{F}_n) \\ &= \frac{a_n}{a_n + b_n + 1} + \frac{P(Y_{n+1} = 1|\mathcal{F}_n)}{(a_n + b_n + 1)} \\ &= \frac{a_n}{a_n + b_n + 1} + \frac{a_n}{(a_n + b_n)(a_n + b_n + 1)} \\ &= \frac{a_n(b_n + a_n + 1)}{(a_n + b_n + 1)(a_n + b_n)} = \frac{a_n}{(a_n + b_n)} = X_n. \end{aligned}$$

Sendo $E(X_{n+1}|\mathcal{F}_n) = X_n$, o processo $\{\frac{a_n}{a_n+b_n}\}_{n \geq 1}$ é uma martingala.

A segunda questão surge diretamente de considerar o Teorema de Convergência de Martingais (Doob, 1953 Teorema 4.1(i): Se o $\lim_{n \rightarrow \infty} E\{|X_n|\} = K < \infty$, então $\lim_{n \rightarrow \infty} X_n = X_\infty$ existe com probabilidade 1 e $E\{|X_\infty|\} \leq K$).

Teorema 3.1 : A σ -álgebra caudal de $\{a_n, b_n\}_{n \geq 0}$ é equivalente a σ -álgebra gerada por Θ . Dada esta σ -álgebra, as diferenças sucessivas Y_n , $n \geq 0$ são independentes e identicamente distribuídas, sendo $Y_i = 1$ com probabilidade θ e $Y_i = 0$ com probabilidade $(1-\theta)$, para $i=0,1,2,\dots,n$.

Prova: Seja $\mathcal{F}_n$ a σ -álgebra gerada por $\{a_j, b_j\}_{0 \leq j \leq n}$, e $\mathcal{F}^{(n)}$ a σ -álgebra gerada por $\{a_j, b_j\}_{j \geq n+1}$. A σ -álgebra caudal de $\{a_n, b_n\}_{n \geq 0}$ é $\bigcup_{n=0}^{\infty} \mathcal{F}^{(n)}$. Esta coincide com

σ -álgebra de eventos mensuráveis em $\{a_{n+1} - a_n : n \geq 0\}$ e são invariantes sob permutações finitas.

Vimos que o processo $\{Y_n\}_{n \geq 0}$ é permutável. Seja $\epsilon_i = 0$ ou $\epsilon_i = 1$, para $i \geq 0$, segue que

$$P(Y_j = \epsilon_j, 0 \leq j \leq k | \mathcal{F}^{(n)}) = \frac{\binom{n-k+1}{(a_{n+1} - a_0 - \sum_{j=0}^k \epsilon_j)}}{\binom{n+1}{(a_{n+1} - a_0)}}, \quad (3.13)$$

para $n \geq k - 1$.

À medida que n tende ao infinito, o lado esquerdo de (3.13) converge com probabilidade 1 para $P(Y_j = \epsilon_j, 0 \leq j \leq k | \mathcal{F}^{(\infty)})$ pelo Teorema de Convergência de Martingais (Doob, 1943, Teorema 4.2). Observando o **Lema 3.1** e a **Definição 3.1**, o lado direito da expressão (3.13) converge com probabilidade 1 para $\theta^{\sum_{j=0}^k \epsilon_j} (1 - \theta)^{k - \sum_{j=0}^k \epsilon_j}$, como vamos provar a continuação, para isto consideramos uma particular seqüência de $\sum_{j=0}^k \epsilon_j$ sucessos. Denominamos $W = \sum_{j=1}^k \epsilon_j$ por simplicidade.

$$\begin{aligned} P(Y_j = \epsilon_j, 0 \leq j \leq k | \mathcal{F}^{(n)}) &= \frac{\binom{n-k+1}{(a_{n+1} - a_0 - \sum_{j=0}^k \epsilon_j)}}{\binom{n+1}{(a_{n+1} - a_0)}}, \quad n \geq k - 1 \\ &= \frac{(a_{n+1} - a_0)!(n+1 - (a_{n+1} - a_0))!(n-k+1)!}{(a_{n+1} - a_0 - W)!(n+1 - k - (a_{n+1} - a_0) - W)!(n+1)!} \\ &= \frac{(a_{n+1} - a_0) \dots (a_{n+1} - a_0 - W + 1)(n+1 - (a_{n+1} - a_0)) \dots (n+1 - k - (a_{n+1} - a_0) - W + 1)}{(n+1)(n) \dots (n+1 - (k-1))} \\ &= P(Y_1 = \epsilon_1 | \mathcal{F}^{(n)}) \dots P(Y_k = \epsilon_k | \mathcal{F}^{(n)}). \end{aligned} \quad (3.14)$$

De (3.14) e considerando a seguinte notação,

$$\begin{aligned} P(Y_j = \epsilon_j, 0 \leq j \leq k | \mathcal{F}^{(n)}) &= \frac{\prod_{j=0}^{W-1} (a_{n+1} - a_0 - j) \prod_{j=0}^{k+W-1} (n+1 - (a_{n+1} - a_0) - j)}{\prod_{j=0}^{k-1} (n+1 - j)} \\ &= \frac{(a_{n+1} - a_0)_W (n+1 - (a_{n+1} - a_0))_{(k-W)}}{(n+1)_k}. \end{aligned}$$

Agora aplicando o **Lema 3.1** a (3.14), temos que

$$\lim_{n \rightarrow \infty} P(Y_j = \epsilon_j, 0 \leq j \leq k | \mathcal{F}^{(n)}) = \lim_{n \rightarrow \infty} \frac{(a_{n+1} - a_0)_W (n+1 - (a_{n+1} - a_0))_{(k-W)}}{(n+1)_k},$$

$$P(Y_j = \epsilon_j, 0 \leq j \leq k | \mathcal{F}^{(\infty)}) = \theta^{\sum_{j=1}^k \epsilon_j} (1 - \theta)^{k - \sum_{j=1}^k \epsilon_j}.$$

Pólya (1931) provou que $\{\frac{a_n}{b_n + a_n}\}_{n \geq 0}$ converge a uma distribuição Beta com parâmetros a_0, b_0 . Como veremos no próximo Teorema, utilizando o **Lema 3.1**.

Teorema 3.2 : *A variável aleatória Θ possui distribuição Beta com parâmetros a_0, b_0 .*

Prova: $P(Y_j = 1, 0 < j \leq k+1) = E(\theta^{k+1})$, do **Teorema 3.1**, sabemos que dada uma σ -álgebra, as variáveis $Y_n : n > 0$ são independentes e identicamente distribuídas, assumindo o valor 1 com probabilidade θ e 0 com probabilidade $1 - \theta$, então

$$P(Y_j = 1, 0 < j \leq k+1 | \mathcal{F}) = P(Y_1 = 1 | \mathcal{F}) \dots P(Y_k = 1 | \mathcal{F}) P(Y_{k+1} = 1 | \mathcal{F}) = \theta^{k+1},$$

$$E(P(Y_j = 1, 0 < j \leq k+1 | \mathcal{F})) = E(\theta^{k+1}) = P(Y_j = 1, 0 < j \leq k+1),$$

a última igualdade segue como consequência do Teorema da Representação. Considerando a **Definição 3.1** resulta que,

$$\begin{aligned} P(Y_j = 1, 0 < j \leq k+1) &= P(Y_1 = 1)P(Y_2 = 1|Y_1 = 1)\dots P(Y_{k+1} = 1|Y_1 = 1, \dots, Y_k = 1) \\ &= \frac{a_0}{a_0 + b_0} \frac{a_0 + 1}{a_0 + b_0 + 1} \dots \frac{a_0 + k}{a_0 + b_0 + k}. \end{aligned} \quad (3.15)$$

A $E(\theta^{k+1}) = \prod_{j=0}^k (\frac{a_0+j}{a_0+b_0+j})$ representa o $k+1$ -ésimo momento da distribuição Beta com parâmetros a_0 e b_0 . Conhecidos todos os momentos de uma distribuição, quando existir, é equivalente ao conhecimento da função de distribuição.

Corolário 3.1 : *Seja $Y_1, Y_2, \dots$ uma seqüência de variáveis aleatórias permutáveis com valores em $\{0, 1\}^\infty$. Considerando a **Proposição 2.7**, e sendo (a, b) a composição inicial da urna, isto é, a bolas azuis e b bolas brancas, a probabilidade conjunta do processo de urna de Pólya,*

$$P(Y_1 = y_1, \dots, Y_n = y_n) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \int_0^1 \theta^{\sum_{i=1}^n y_i + a - 1} (1 - \theta)^{(n - \sum_{i=1}^n y_i) + b - 1} d\theta.$$

Prova: É imediata.

Corolário 3.2 : *Utilizando o **Corolário 3.1**, temos que a probabilidade conjunta do processo de Pólya, com distribuição Beta(a, b) como distribuição a priori é dada por,*

$$P(Y_1 = y_1, \dots, Y_n = y_n) = \frac{a \binom{a+b-1}{b-1}}{(n+b - \sum_{i=1}^n y_i) \binom{n+a+b-1}{\sum_{i=1}^n y_i + a - 1}}.$$

Prova:

$$\begin{aligned}
P(Y_1 = y_1, \dots, Y_n = y_n) &= \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \int_0^1 \theta^{\sum_{i=1}^n y_i + a - 1} (1-\theta)^{(n - \sum_{i=1}^n y_i) + b - 1} d\theta. \\
&= \frac{a(a + \sum_{i=1}^n y_i - 1)!(b + n - \sum_{i=1}^n y_i)!(b + a - 1)!}{a!(b-1)!(b + a + n - 1)!(b + n - \sum_{i=1}^n y_i)} \\
&= \frac{a \binom{a+b-1}{b-1}}{(n+b - \sum_{i=1}^n y_i) \binom{n+a+b-1}{\sum_{i=1}^n y_i + a - 1}}.
\end{aligned}$$

3.4 Simulações

Vamos gerar matrizes de 10 linhas por 20 colunas que representam 10 seqüências de 20 variáveis aleatórias com valores em $\{0,1\}$ cada uma. Estas seqüências vão simular o processo de Pólya. Consideramos diferentes formas para gerar estas matrizes.

I) Geração a partir da definição do processo de Pólya: partindo de uma urna de Pólya com composição inicial (2,3), isto é, 2 bolas brancas e 3 bolas pretas, esta escolha é arbitrária. Calcula-se a proporção de bolas brancas x_0 ; a função de urna no caso do processo de Pólya está dada por $f(x_0) = x_0$ e esta é a probabilidade de sair uma bola branca no próximo passo. Com esta probabilidade vamos determinar aleatoriamente o valor da variável aleatória Y_1 na seqüência, comparando esta probabilidade com o valor obtido gerando uma distribuição uniforme. Repete-se o procedimento, atualizando as proporções de bolas brancas, para gerar a primeira seqüência de 20 observações. Isto é feito para as 10 seqüências que formam as linhas da matriz, começando todas com a mesma composição inicial.

II) Outra maneira de simular esta matriz é usando o Teorema da Representação. Determinando o valor da integral, considerando uma Beta(2,3) como distribuição

a priori e, obtendo o valor na seqüência de forma aleatória, comparando este com um valor gerado por uma distribuição uniforme, logo, atualizando os parâmetros da posteriori obtida, e assim sucessivamente até obter os 20 valores da seqüência, e repetindo para as 10 seqüências.

III) Finalmente o cálculo da matriz foi feito por meio do **Corolário 3.2**.

A priori é $\text{Beta}(a, b)$, com parâmetros $a = 2$ e $b = 3$. Fez-se o cálculo para todas as somas possíveis, repetindo o procedimento utilizado para gerar a matriz de zeros e uns na simulação do item III) da seção 2.5. Nesse caso foram calculadas as probabilidades conjuntas a partir do segundo membro do Teorema da Representação. E também calculamos essas probabilidades conjuntas, só que utilizando o **Corolário 3.2**, onde estas probabilidades são obtidas usando o Teorema da Representação.

Estas matrizes foram obtidas considerando o mesmo processo estocástico, isto é, o processo de Pólya. Embora estas seqüências de $\{0, 1\}^\infty$ sejam geradas a partir de diferentes métodos, elas são equivalentes dado que surgem através do Teorema da Representação, utilizando uma distribuição a priori Beta, com parâmetros escolhidos arbitrariamente.

Capítulo 4

CONSIDERAÇÕES ASSINTÓTICAS

Mostramos no capítulo 3 que $\{S_n\}_{n \geq 0}$ é uma cadeia de Markov não-homogênea discreta no tempo, com probabilidade de transição dada por,

$$P_{i,i}(n, (n+1)) = 1 - P_{i,i+1}(n, (n+1)),$$

$$P_{i,i+1}(n, (n+1)) = P(S_{n+1} = i+1 | S_n = i) = \frac{a+i}{a+b+n}, i = 0, 1, \dots, n,$$

onde $P_{i,i+1}(n, (n+1))$ é a probabilidade de estar no estado $i+1$ no tempo $(n+1)$, dado que no tempo n está no estado i . Sendo a e b as quantidades iniciais de bolas azuis e brancas na urna.

O processo de Pólya também é uma cadeia de Markov com tempo contínuo e espaço de estados finito. Os processos desta classe são denominados Processos de Markov Descontínuos. No processo de Pólya, a probabilidade de uma mudança no intervalo $(t, t + \Delta t)$ é uma função do estado do sistema ao tempo t , mas também do tempo t .

4.1 Processo de Pólya com parâmetro contínuo

Caracterizações do Processo de Pólya como processo de parâmetro contínuo e Equações de Kolmogorov

Suposições para o processo de Pólya (Bharucha-Reid, A.T. (1960) [4]):

I) Se no tempo t o sistema está no estado x , com probabilidade $P_x(t)$, $x = 0, 1, 2, \dots$, a probabilidade de transição de x a $x+1$ no intervalo $(t, t+\Delta t)$, $P_{x,x+1}(t, t+\Delta t)$ está dada por: $\lambda \frac{1+\alpha x}{1+\alpha \lambda t} \Delta t + o(\Delta t)$, onde α e λ são constantes não negativas.

II) A Probabilidade de uma transição do estado x a um estado diferente de $x+1$ é $o(\Delta t)$.

III) A probabilidade de não ter nenhuma mudança está dada por: $1 - \lambda \frac{1+\alpha x}{1+\alpha \lambda t} \Delta t + o(\Delta t)$.

Temos a seguinte relação considerando as suposições dadas:

$$P_x(t + \Delta t) = (1 - \lambda \frac{1+\alpha x}{1+\alpha \lambda t} \Delta t) P_x(t) + (\lambda \frac{1+\alpha x}{1+\alpha \lambda t} \Delta t) P_{x-1}(t) + o(\Delta t),$$

de onde,

$$P_x(t + \Delta t) - P_x(t) = -\lambda \frac{1+\alpha x}{1+\alpha \lambda t} \Delta t P_x(t) + (\lambda \frac{1+\alpha x}{1+\alpha \lambda t} \Delta t) P_{x-1}(t) + o(\Delta t),$$

$$\lim_{\Delta t \rightarrow 0} \frac{P_x(t+\Delta t) - P_x(t)}{\Delta t} = -\lambda \frac{1+\alpha x}{1+\alpha \lambda t} P_x(t) + \lambda \frac{1+\alpha x}{1+\alpha \lambda t} P_{x-1}(t), \text{ para } x = 1, 2, \dots$$

de modo que vamos ter as seguintes equações,

$$dP_x(t)/dt = -\lambda \frac{1+\alpha x}{1+\alpha \lambda t} P_x(t) + \lambda \frac{1+\alpha x}{1+\alpha \lambda t} P_{x-1}(t), \quad (4.1)$$

para $x = 1, 2, \dots$

definimos $P_{-1}(t) = 0$, assim

$$dP_0(t)/dt = -\lambda\left(\frac{1}{1+\alpha\lambda t}\right)P_0(t). \quad (4.2)$$

As equações (4.1) e (4.2) caracterizam o processo de Pólya com parâmetro contínuo. Para resolver estas equações consideramos as seguintes condições iniciais,

$$\begin{aligned} P_x(0) &= 1, \text{ para } x = 0, \\ P_x(0) &= 0, \text{ para } x = 1, 2, \dots \end{aligned}$$

de (4.2) temos que

$$P_0(t) = (1 + \alpha\lambda t)^{-1/\alpha}, \quad (4.3)$$

e a solução de (4.1) é dada pela seguinte expressão,

$$P_x(t) = \left(\frac{(\lambda t)^x}{x!}\right)(1 + \alpha\lambda t)^{(-x)-(1/\alpha)} \prod_{i=1}^{x-1} (1 + \alpha i), \quad (4.4)$$

com $x = 1, 2, \dots$

Para começar com o estudo das propriedades do Processo de Pólya, vamos mostrar primeiro que $\sum_{x=0}^{\infty} P_x(t) = 1$, então

$$\begin{aligned} \sum_{x=0}^{\infty} P_x(t) &= (1 + \alpha\lambda t)^{-1/\alpha} \sum_{x=0}^{\infty} \frac{(\lambda t)^x}{x!} (1 + \alpha\lambda t)^{-x} \prod_{i=1}^{x-1} (1 + \alpha i) \\ &= (1 + \alpha\lambda t)^{-1/\alpha} \sum_{x=0}^{\infty} (-\alpha)^x \binom{-1/\alpha}{x} \left(\frac{\lambda t}{1 + \alpha\lambda t}\right)^x \\ &= (1 + \alpha\lambda t)^{-1/\alpha} \sum_{x=0}^{\infty} \binom{-1/\alpha}{x} \left(\frac{-\alpha\lambda t}{1 + \alpha\lambda t}\right)^x \\ &= (1 + \alpha\lambda t)^{-1/\alpha} \left(1 - \frac{\alpha\lambda t}{1 + \alpha\lambda t}\right)^{-1/\alpha} = 1, \end{aligned}$$

para esta prova se usa o fato que: $1/x! \prod_{i=1}^{x-1} (1 + \alpha i) = (-\alpha)^x \binom{-1/\alpha}{x}$.

Agora vamos determinar a média do processo de Pólya,

$$\begin{aligned}
 m(t) = E(X_t) &= \sum_{x=0}^{\infty} x P_x(t) \\
 &= (1 + \alpha \lambda t)^{-1/\alpha} \sum_{x=0}^{\infty} x \binom{-1/\alpha}{x} \left(\frac{-\alpha \lambda t}{1 + \alpha \lambda t} \right)^x \\
 &= (1 + \alpha \lambda t)^{-1/\alpha} (-1/\alpha) \left(\frac{-\alpha \lambda t}{1 + \alpha \lambda t} \right) \sum_{x=0}^{\infty} \binom{-(1 + 1/\alpha)}{x-1} \left(\frac{-\alpha \lambda t}{1 + \alpha \lambda t} \right)^{x-1} \\
 &= \lambda t (1 + \alpha \lambda t)^{-1-1/\alpha} \left(1 - \frac{\alpha \lambda t}{1 + \alpha \lambda t} \right)^{-1-1/\alpha} = \lambda t,
 \end{aligned}$$

Usamos o fato que $x \binom{-1/\alpha}{x} = -(1/\alpha) \binom{-(1+1/\alpha)}{x-1}$.

De forma similar se faz o cálculo do segundo momento:

$$\begin{aligned}
 (1 + \alpha) m^2(t) + m(t) &= (1 + \alpha) ((\lambda t)^2 + \lambda t, \\
 \text{Var}(X_t) &= \lambda t (1 + \lambda \alpha t).
 \end{aligned}$$

Observe-se que a média é função linear do tempo e a variância é função quadrática do tempo t .

As funções de intensidade dependem do tempo t e do estado i , assim

$$\begin{aligned}
 q_i(t) &= \lambda \left(\frac{1 + \alpha i}{1 + \alpha \lambda t} \right), \text{ para } i = 0, 1, \dots \\
 Q_{ij} &= 1, \quad j = i + 1, \quad i = 0, 1, 2, \dots \\
 Q_{ij} &= 0, \text{ caso contrário.}
 \end{aligned}$$

Vide Apêndice, Definição A.6.

Portanto a matriz de transição fica da seguinte forma:

$$A(t) = \begin{pmatrix} -\lambda \frac{1}{1 + \alpha \lambda t} & \lambda \frac{1}{1 + \alpha \lambda t} & 0 & \dots & 0 \\ 0 & -\lambda \frac{1 + \alpha}{1 + \alpha \lambda t} & \lambda \frac{1 + \alpha}{1 + \alpha \lambda t} & \dots & 0 \\ \vdots & & \vdots & & \vdots \\ \dots & \dots & \dots & \dots & \dots \end{pmatrix}$$

As equações de Kolmogorov para o processo de Pólya estão dadas por,

$$dP_{ij}(t)/dt = -\lambda \frac{1+\alpha j}{1+\alpha \lambda t} P_{i,j}(t) + \frac{1+\alpha(j-1)}{1+\alpha \lambda t} P_{i,j-1}(t),$$

$$dP_{ij}(t)/dt = -\lambda \frac{1+\alpha i}{1+\alpha \lambda t} P_{i,j}(t) + \frac{1+\alpha(i+1)}{1+\alpha \lambda t} P_{i+1,j}(t).$$

Estas equações também caracterizam o processo de Pólya.

Comentário: Considerando a composição inicial da urna dada por (a, b) , onde $a, b \in \mathbb{N}$, as constantes α e λ estão dadas por: $\alpha = \frac{1}{a}$ e $\lambda = \frac{a}{a+b}$, e $i = x = \sum_{k=1}^t y_k$, onde y_k é variável indicadora de bolas azuis, isto é, $y_k = 1$ se a k -ésima bola retirada for azul e $y_k = 0$ se a bola retirada for branca, com $k = 1, 2, \dots, t$.

4.2 Distribuições Assintóticas

Caracterização Assintótica do Processo de Pólya

Seja $\{Y_n\}_{n \geq 1}$ o processo de Pólya, consideremos uma urna com composição inicial dada por a bolas azuis e b bolas brancas, com a e b fixos, para determinar a probabilidade de retirar uma bola azul da urna no instante $t = n$, quando $n \rightarrow \infty$, isto é,

$$\lim_{n \rightarrow \infty} P(Y_n = 1),$$

usamos o fato de que as variáveis são permutáveis, e a **Definição 3.1**, então,

$$P(Y_1 = 1) = P(Y_2 = 1) = \dots = P(Y_n = 1) = \frac{a}{a+b},$$

portanto,

$$\lim_{n \rightarrow \infty} p(Y_n = 1) = \frac{a}{a+b}.$$

A continuação apresentamos diversos enfoques da abordagem de distribuição assintótica para uma urna de Pólya.

Vamos lembrar que a distribuição de Pólya possui a seguinte função de probabilidade,

$$P(X = j) = \binom{n}{j} \frac{[a]_j [b]_{n-j}}{[b+a]_n} = \frac{\prod_{k=0}^{j-1} (a+k) \prod_{k=0}^{n-j-1} (b+k)}{\prod_{k=0}^{n-1} (b+a+k)}, \quad (4.5)$$

para $j = 0, 1, 2, \dots$ e com $[x]_y = \prod_{j=0}^{y-1} (x+j)$.

Esta distribuição foi obtida supondo uma urna contendo b bolas brancas e a bolas azuis iniciais. X é uma variável aleatória que representa o número de vezes que uma bola azul é retirada da urna. As seguintes distribuições limites podem ser estabelecidas mediante argumentos elementares e diretos:

1) Se $b \rightarrow \infty$, $a \rightarrow \infty$, com $a/(b+a) \rightarrow p$ ($0 < p < 1$), com n fixo. Então,

$$P(X = j) \rightarrow \binom{n}{j} p^j (1-p)^{n-j}, j = 0, 1, \dots, n. \quad (4.6)$$

a distribuição limite de X é Binomial com parâmetros n , p . A prova da distribuição limite é feita de forma similar à usada na demonstração do Teorema da Representação (seção 2.4) para provar a passagem da distribuição Hipergeométrica à distribuição Binomial.

2) Se

$$\begin{cases} a \rightarrow \infty & b \rightarrow \infty, & \text{com } \frac{a}{a+b} \rightarrow 0 \\ n \rightarrow \infty, & \text{com } \frac{na}{a+b} \rightarrow \lambda, & \text{e } \frac{b+a}{n} \rightarrow \gamma, \end{cases}$$

sendo que $0 < \lambda < \infty$ e $0 < \gamma < \infty$, então

$$P(X = j) \rightarrow \binom{\lambda\gamma + j - 1}{j} (\gamma/(1+\gamma))^{\lambda} (1/(1+\gamma))^j, j = 0, 1, \dots, n. \quad (4.7)$$

a distribuição limite de X é a Binomial Negativa com parâmetros $\lambda\gamma$ e $(1 + \gamma)^{-1}$.

Para obter (4.7) vamos considerar uma particularização de $P_x(t)$ dada por (4.4), considerando $\lambda = \frac{a}{a+b}$, $\alpha = \frac{1}{a}$ e $t = n$,

$$P_x(n) = \left(\frac{(an/a + b)^x}{x!}\right) \left(\frac{a + b + n}{a + b}\right)^{(-a-x)} \prod_{i=1}^{x-1} \frac{a + i}{a},$$

com $a \rightarrow \lambda\gamma$, trocando o índice x por j , podemos escrever a expressão da seguinte forma

$$P_j(n) = \binom{a + j - 1}{j} \left(\frac{(a + b)/n}{(a + b + n)/n}\right)^a \left(\frac{n}{a + b + n}\right)^j a^j a^{1-j} a,$$

então quando $n \rightarrow \infty$ e fazendo as substituições correspondentes, vamos obter uma distribuição Binomial Negativa com parâmetros $\lambda\gamma$ e $(1 + \gamma)^{-1}$.

3) Se quando $n \rightarrow \infty$, $n^{-2}(a+b) \rightarrow \theta$ ($0 < \theta < \infty$) e $na/(b+a) \rightarrow \lambda$ ($0 < \lambda < \infty$), então

$$P(X = j) \rightarrow \lambda^j / j! e^{-\lambda}, j = 0, 1, \dots, n. \quad (4.8)$$

Para obter este limite, utilizamos uma particularização de $P_x(t)$ dada por (4.4), considerando $t = n$ e $x = j$,

$$P_j(n) = \left(\frac{(an/a + b)^j}{j!}\right) \left(\frac{a + b + n}{a + b}\right)^{(-a-j)} \prod_{i=1}^{j-1} \frac{a + i}{a},$$

considerando neste caso que $na/(b + a) \rightarrow \lambda$, com $0 < \lambda < \infty$, vamos ter que,

$$\lim_{a \rightarrow \infty} \left(1 + \left(\frac{1}{a}\right) \left(\frac{an}{a + b}\right)\right)^{j-a} = e^{-an/a+b},$$

portanto,

$$P(X = j) \rightarrow \lambda^j/j!e^{-\lambda}, j = 0, 1, \dots, n. \quad (4.9)$$

a distribuição limite é poisson com parâmetro λ . A mesma distribuição poisson é obtida quando $n^{-2}(a + b) \rightarrow 0$ e $nb/(b + a) \rightarrow \lambda$ ($0 < \lambda < \infty$), quando $n \rightarrow \infty$, ou também podemos obter esta distribuição se $n^{-2}(a + b) \rightarrow \infty$, $n^{-3}(a + b) \rightarrow 0$, e $nb/(b + a) \rightarrow \lambda$ ($0 < \lambda < \infty$), com $n \rightarrow \infty$.

Outros resultados assintóticos foram obtidos na seção 3.3.1, quando demonstramos os Teoremas do artigo de Freedman(1965) [11].

Capítulo 5

GENERALIZAÇÕES-PROCESSOS PERMUTÁVEIS

No capítulo 3, vimos que o processo de Pólya é um caso especial de uma família de processos estocásticos, cuja característica distinguível consiste em que a seqüência de observações é dada por sucessivas retiradas de bolas de urnas com composições que vão mudando.

Para definir esta família, consideramos uma urna com composição inicial (v, b) , onde v é o número de bolas vermelhas e b o número de bolas brancas. Com $X_0 = v/(v + b)$, a proporção inicial de bolas vermelhas e uma função $f : [0, 1] \rightarrow [0, 1]$, denominada função de urna. Suponhamos que uma bola vermelha é adicionada à urna com probabilidade $f(x_0)$, portanto uma bola branca é adicionada com probabilidade $1 - f(x_0)$.

Seja Y_n o indicador do evento que a n -ésima bola é vermelha. O processo $Y_1, Y_2, \dots$ é um processo de urna com composição inicial (v, b) e função de urna f . A distribuição de um processo de urna $\{Y_n\}_{n \geq 0}$ está completamente determinada pela composição inicial e pelos valores da função de urna f para as sucessivas proporções $X_0, X_1, \dots$.

As únicas proporções possíveis são da forma $(v + k)/(v + b + n)$, onde k e n são inteiros não negativos, $k \leq n$.

O esquema de urna de Pólya é um processo que possui a função de urna dada por $f(x) = x$. Vimos, que o processo $\{Y_n\}_{n \geq 0}$ é permutável com a distribuição Beta como medida de De Finetti.

Quando a função de urna é constante, isto é, $f(x) = \theta_0$, o processo $\{Y_n\}_{n \geq 1}$ é gerado por $Y_1, Y_2, \dots$, variáveis aleatórias independentes Bernoulli(θ_0). Este processo é claramente permutável e a medida de De Finetti está concentrada no ponto θ_0 . Para verificar isto, lembremos que U foi definida na seção 3.1 como a variável aleatória que conta o número k de bolas de uma determinada cor (vermelhas neste caso), para $k = 0, 1, \dots, n$. Como as variáveis aleatórias Y_n , $n \geq 1$, são Bernoulli(θ_0), a variável aleatória U possui distribuição Binomial(n, θ_0), então

$$P(U = k) = \binom{n}{k} \theta_0^k (1 - \theta_0)^{n-k}. \quad (5.1)$$

Pelo Teorema da Representação de Bruno De Finetti, obtemos,

$$P(U = k) = \binom{n}{k} \int_0^1 \theta^k (1 - \theta)^{n-k} dF(\theta), \quad (5.2)$$

como consequência,

$$F(\theta) = \begin{cases} 1 & \text{se } \theta \geq \theta_0 \\ 0 & \text{se } \theta < \theta_0. \end{cases}$$

Resolvendo a integral de Riemann-Stieltjes temos que,

$$P(U = k) = \binom{n}{k} \theta_0^k (1 - \theta_0)^{n-k}, \quad (5.3)$$

portanto, verificamos que a medida de De Finetti está concentrada no ponto θ_0 .

Também existe uma coleção trivial de processos de urna permutáveis que são os processos determinísticos. Suponhamos que retiramos a primeira bola da urna. A cor

da bola é vermelha com probabilidade p e é branca com probabilidade $1 - p$. Todas as bolas retiradas em continuação são da mesma cor da primeira bola retirada.

Neste caso, a medida de De Finetti atribui probabilidade p a $\{1\}$ e $1 - p$ a $\{0\}$. Utilizando a mesma metodologia que aplicamos nos casos anteriores temos que,

O Teorema da Representação estabelece que,

$$P(U = k) = \binom{n}{k} \int_0^1 \theta^k (1 - \theta)^{n-k} dF(\theta) \quad (5.4)$$

de modo que, se $k = n$

$$\begin{aligned} P(U = n) &= \binom{n}{n} \int_0^1 \theta^n dF(\theta) \\ &= p, \end{aligned} \quad (5.5)$$

deste modo temos que $P(U = n) = p$.

Se $k = 0$

$$\begin{aligned} P(U = 0) &= \binom{n}{0} \int_0^1 (1 - \theta)^n dF(\theta) \\ &= 1 - p, \end{aligned} \quad (5.6)$$

deste modo temos que $P(U = 0) = 1 - p$.

No caso de conhecer a cor das bolas na urna,

$$P(U = n) = \begin{cases} 1 & \text{se são todas vermelhas} \\ 0 & \text{caso contrário,} \end{cases}$$

sendo $k = n$, pelo Teorema da Representação temos que

$$\begin{aligned} P(U = n) &= \binom{n}{n} \int_0^1 \theta^n dF(\theta) \\ &= 1, \end{aligned}$$

portanto $P(\theta = 1) = 1$, em conseqüência

$$F(\theta) = \begin{cases} 1 & \text{se } \theta \geq 1 \\ 0 & \text{se } \theta < 1. \end{cases}$$

Sendo $k = 0$, o Teorema da Representação estabelece que

$$\begin{aligned} P(U = 0) &= \binom{n}{0} \int_0^1 (1 - \theta)^n dF(\theta) \\ &= 0, \end{aligned}$$

De modo que $P(\theta = 0) = 1$, portanto

$$F(\theta) = \begin{cases} 1 & \text{se } \theta \geq 0 \\ 0 & \text{se } \theta < 0. \end{cases}$$

De modo que a medida de De Finetti está dada por uma distribuição Bernoulli(p), quando não conhecemos a cor das bolas inicialmente, ou por uma distribuição degenerada, quando a cor é conhecida.

Enunciamos o seguinte Teorema que nos diz quais são os processos de urna permutáveis.

Teorema 5.1 : *Supondo que o processo $Y_1, Y_2, \dots$ de variáveis aleatórias com valores em $\{0, 1\}$ é um processo de urna permutável. Então Y é Pólya ou Bernoulli ou Determinístico.*

O resultado deste Teorema pode se expressar de forma semelhante ao Postulado de Suficiência de Johnson que veremos a seguir.

5.1 Postulado de Suficiência de Johnson

(Zabell, S. (1982) [26]) W.E. Johnson (1932) mostrou através de seu Postulado de Suficiência que seqüências multinomiais infinitamente permutáveis podem ser caracterizadas por distribuições a priori pertencentes à família de distribuições Dirichlet

simétricas. Johnson provou este resultado prévio à aparição do trabalho de Bruno De Finetti sobre permutabilidade e o Teorema da Representação.

Seja $Y_1, Y_2, \dots$ uma seqüência de variáveis aleatórias permutáveis com valores em $\{0, 1\}^\infty$, $S_N = Y_1 + \dots + Y_N$, pelo Teorema da Representação vamos ter que

$$P(S_N = k) = \binom{N}{k} \int_0^1 p^k (1-p)^{N-k} dF(p),$$

para $N \geq 1$, $0 \leq k \leq N$, com $F(p) = P(Z \leq p)$ e $Z = \lim_{N \rightarrow \infty} \frac{S_N}{N}$.

Nessa época usar distribuições a priori uniforme $dF(p) = dp$, eram interpretadas como completa ignorância sobre o valor p . Laplace defendeu seu uso invocando o princípio da razão insuficiente. Outros defensores deste princípio argumentavam que era justificada sua utilização na base da experiência e da informação obtida, mas esta posição sugeria que qualquer distribuição a priori poderia ser usada.

Bayes e Laplace generalizaram o resultado para o caso de seqüências multinomiais com t categorias, assumindo uma priori uniforme. Johnson inicialmente concordou com aquela priori, mas logo não satisfeito com seu postulado de equi-probabilidade propôs outro postulado mais geral, que forçava a $dF(p)$ a pertencer à família Dirichlet.

$$Dir(k_1, \dots, k_t) = p_1^{k_1-1} \dots p_t^{k_t-1} dp_1 \dots dp_{t-1},$$

para $k_i > 0$, $1 \leq i \leq t$, com $t > 1$.

Vamos fazer menção a alguns resultados e definições a respeito deste Postulado.

Seqüências finitas permutáveis

Seja $X_1, X_2, \dots, X_{N+1}$ uma seqüência de variáveis aleatórias com valores no conjunto $t = \{1, 2, \dots, t\}$, para $N \geq 1$ e $t \leq \infty$, tal que,

$$P\{X_1 = i_1, \dots, X_N = i_N\} > 0, \quad (5.7)$$

para todo $(i_1, \dots, i_N) \in \mathbf{t}^N$.

Seja $\mathbf{n} = \mathbf{n}(X_1, \dots, X_N)$ o vetor das frequências, isto é, $\mathbf{n} = (n_1, \dots, n_t)$, onde $n_i = n_i(X_1, \dots, X_N) = \#\{X_j = i\}$, para $j = 1, \dots, N$, com $\#\{X_j = i\} = 0, \dots, N$.

O postulado de Suficiência de Johnson assume que:

$$P\{X_{N+1} = i | X_1, \dots, X_N\} = f_i(n_i), \quad (5.8)$$

a probabilidade condicional da saída na i -ésima cela dado $X_1, \dots, X_N$ depende somente de n_i , o número de saídas prévias nessa cela. Se $X_1, \dots, X_{N+1}$ são permutáveis, então, $f_i(n_i) = P\{X_{N+1} = i | \mathbf{n}\} = P\{X_{N+1} = i | n_i\}$.

Lema 5.1 : Se $t > 2$, (5.7) e (5.8) verificam-se, então existem constantes $a_i \geq 0$ e b , tais que para todo i ,

$$f_i(n_i) = a_i + bn_i, \quad (5.9)$$

com $a_i = f_i(0) > 0$ e $b = \Delta f_i(n_i)$, onde b não depende de i .

Observações: Se $t = 2$ o postulado é vazio e (5.9) não precisa se manter. Assim, no caso Binomial é necessário fazer a suposição adicional de linearidade. Em ambos os casos ($t = 2$ ou $t > 2$) é requerido $a_i > 0$.

5.1.1 Complementos e extensões

A Dirichlet simétrica. Johnson considerou o caso especial onde

(i) f_i é independente de i , isto é, para todo N , existe uma função f tal que,

$$P\{X_{N+1} = i | \mathbf{n}\} = f(n_i, N), \text{ para todo } i,$$

(ii) b é positivo (este é o principal artifício na prova de Johnson. Se $\{X_1, X_2, \dots\}$ é infinitamente permutável, mas não independentes, a suposição que b é positiva é trivial).

Sob estas condições $t < \infty$, $a_i \equiv a$, $k_i \equiv k > 0$, $P\{X_1 = i\} = \frac{1}{t}$,

$$P\{X_{N+1} = i|n\} = \frac{n_i+k}{N+kt},$$

e $X_1, \dots, X_N$ pode ser estendida como uma seqüência permutável infinita, cuja medida na mistura na representação de De Finetti é uma distribuição Dirichlet simétrica com parâmetro k .

Diaconis e Ylvisaker (1979) [9] provaram que a distribuição Beta é a única distribuição que permite esperança linear a posteriori em amostragem Binomial permutável, isto é, $t = 2$ e $\{X_n\}_{n \geq 0}$ infinitamente permutável. O Teorema é enunciado a seguir.

Teorema 5.2 : *Seja Q uma distribuição a priori no conjunto de Borel $[0, 1]$ e assumindo que Q não se concentra em um único ponto. Se para todo $n = 1, 2, \dots$ existem números a_n e b_n para os quais:*

$$\int_0^1 p^{k+1}(1-p)^{n-k} dQ(p) = (a_n k + b_n) \int_0^1 p^k(1-p)^{n-k} dQ(p) \quad (5.10)$$

para $k = 0, 1, \dots, n$. Então $a_n = \frac{a}{1+(n-1)a}$, $b_n = \frac{b}{1+(n-1)a}$,

com $a > 0$ e $b > 0$, $a + b < 1$ e Q possui a distribuição Beta.

5.1.2 O Novo Postulado de Suficiência. O Postulado Z

Zabell (1997) [28] propôs um novo Postulado de Suficiência, tratando de solucionar um problema que se apresenta quando, por exemplo, consideramos o caso em que há

três categorias c_1, c_2 e c_3 , sendo $X_1, \dots, X_{10}$ as saídas, isto é, $t = 3$, $n = 10$, e dadas as seguintes amostras:

caso 1: $\mathbf{n}_1 = (0, 0, 10)$

caso 2: $\mathbf{n}_2 = (0, 5, 5)$

é natural esperar que $P[X_{11} = c_1 | \mathbf{n}_1] < P[X_{11} = c_1 | \mathbf{n}_2]$,

isto é, porque o primeiro caso sugere a possibilidade que todas as saídas continuem na categoria 3, não sendo assim no segundo caso. O Postulado de Suficiência diz com respeito à c_1 , que as duas probabilidades são as mesmas, isto porque as probabilidades preditivas para a primeira categoria depende somente de n_1 , que em ambos os casos está dada por $n_1 = 0$.

Como veremos na continuação, existe uma conexão entre o Postulado de Johnson e a Regra da Sucessão de Laplace, esta última pode ser interpretada da seguinte maneira: Consideremos $N + 1$ urnas numeradas desde $k = 0, \dots, N$, todas as urnas possuem N bolas, onde k bolas são vermelhas e $N - k$ azuis. Sendo todas as urnas igualmente prováveis, escolhemos aleatoriamente uma urna. Retiram-se n bolas da urna, se todas as bolas retiradas são vermelhas, a probabilidade de que a bola seja vermelha na próxima retirada é dada por $\frac{n+1}{n+2}$.

Seja $X_1, X_2, \dots$ uma seqüência de variáveis aleatórias permutáveis, com valores no conjunto $\{c_1, c_2, \dots, c_t\}$, $2 \leq t \leq \infty$, para $n \geq 1$, seja $n_i = n_i(X_1, \dots, X_n)$ o número de vezes que $X_k = c_i$ na seqüência finita, $k = 1, \dots, n$ e seja $\mathbf{n} = (n_1, \dots, n_t)$. O Princípio Johnson-Carnap “continuum of inductive methods” (Zabell (1982)[27]) considera que as probabilidades preditivas (ou regra da sucessão) tomam a seguinte forma: Se $\alpha_i > 0$ e $\alpha = \sum \alpha_i$, então $\forall n \geq 0$, e $\mathbf{n} = (n_1, n_2, \dots, n_t)$, vamos ter,

$$P[X_{n+1} = c_i | \mathbf{n}] = \frac{n_i + \alpha_i}{n + \alpha}. \quad (5.11)$$

Esta forma não é arbitrária. Johnson provou que se todas as seqüências são possíveis, existindo pelo menos três categorias ($t \geq 3$), sendo a regra da sucessão função somente da categoria i , da freqüência n_i , do tamanho de amostra n (isto é, $P[X_{n+1} = c_i | \mathbf{n}] = f(n_i, n)$), os elementos da seqüência são independentes ou as probabilidades preditivas possuem a forma indicada por (5.11) para alguns valores de $\alpha_i > 0$.

Uma crítica a este “continuum” está dada porque não confirma as generalizações universais, isto é, se todas as observações são da classe i ($n_i = n$, e $n_j = 0$ para $i \neq j$), então nunca vamos ter que

$$P(X_{n+1} = X_{n+2} = \dots = c_i | n_i = n) > 0, \quad (5.12)$$

dado que o Postulado de Suficiência atribui probabilidade positiva para qualquer seqüência, e (5.12) estaria contradizendo esse fato.

Então, fazendo pequenas modificações no Postulado de Suficiência, surge o novo Postulado, denominado Postulado Z.

Postulado Z: Para todo $n \geq 1$, e todos os vetores de freqüência $\mathbf{n}$, as probabilidades preditivas $p[X_{n+1} = c_i | \mathbf{n}]$ são funções $f_i(n_i, n)$ de n_i e de n somente, exceto quando $n_i = 0$ e $t = 1$, onde t é o número de espécies observadas na amostra de tamanho n .

Este postulado é criticado, pois por um lado existem casos onde ele não é válido, além de supor que o número de espécies t tem que ser conhecido previamente, o que

não é real e também, a restrição que existe para o número de categorias, que deve ser pelo menos três, o que não é natural. Existe uma solução pragmática Bayesiana, baseada no Teorema da Representação de Bruno De Finetti (Jeffreys, Wrinch, 1919).

Voltando para o Princípio de Suficiência de Johnson, que assume uma seqüência finita de variáveis aleatórias $Y_1, \dots, Y_{n+1}$ com valores em $\{0, 1, \dots, t-1\}$, sendo independentes dado o vetor de probabilidade $\mathbf{p} = (p_0, \dots, p_{t-1})$. Suponhamos que $t = 2$, e que a seqüência pode ser infinitamente estendida, assumimos que para todo n , a seguinte probabilidade condicional é função de $X_n = x_n$, isto é, da proporção de bolas vermelhas na urna.

$$P(Y_{n+1} = 1 | Y_1 = y_1, \dots, Y_n = y_n) = f(x_n). \quad (5.13)$$

O **Teorema 5.1** enunciado nos mostra que as únicas distribuições para $\mathbf{p}$ consistentes com esta suposição são: as distribuições Beta ou ponto massa ou aquelas distribuições concentradas em um ponto do $\{0, 1\}$.

Seja $S_n = Y_1 + \dots + Y_n$. A proporção $X_n = x_n$ de bolas vermelhas no instante n pode ser escrita da seguinte maneira:

$$X_n = \frac{v + S_n}{v + b + n},$$

onde (v, b) é a composição inicial da urna, isto é, v bolas vermelhas e b bolas brancas. Para processos de urna, a probabilidade preditiva (5.13) é função de x_n . Uma pergunta que você pode se fazer é se existem processos permutáveis $\{Y_n\}_{n \geq 0}$ para os quais a probabilidade (5.13) é função fixa de $\frac{S_n}{n}$, isto é, não dependem da composição inicial da urna. A resposta para isto é que os únicos processos possíveis são Bernoulli ou Determinísticos. Para ver isto observemos que dado $Y_1 = 1(0)$, $Y_2, Y_3, \dots$ é um processo de urna permutável com composição inicial $(1, 0)((0, 1))$. Aplicando o **Teorema 5.1** e a **Definição 3.1**, notamos que o processo não pode ser Pólya.

Capítulo 6

CONCLUSÃO

O Teorema da Representação de Bruno De Finetti apresenta interesse para estatísticos, matemáticos e filósofos. Vimos que entre suas muitas conseqüências, destacam-se a possível justificativa de modelagens paramétricas, principalmente, estabelece uma conexão entre probabilidade e frequências relativas, tornando clara, deste modo, a versão subjetivista da Lei dos Grandes Números.

O Teorema da Representação possui a forma mais simples quando as variáveis aleatórias permutáveis tomam valores no conjunto $\{0, 1\}$. Neste caso a priori é dada por uma distribuição que pertence à família Beta de distribuições. Sendo a verossimilhança uma distribuição binomial, a distribuição a posteriori também pertence à família Beta de distribuições. O Processo de urna de Pólya $\{Y_n\}_{n \geq 1}$, gerado por variáveis aleatórias permutáveis $Y_1, Y_2, \dots$ com valores em $\{0, 1\}^\infty$, com função de urna dada por $f(x) = x$, possui uma distribuição a priori dada por uma distribuição Beta, com parâmetros dados pela composição inicial da urna. As proporções de bolas de uma determinada cor convergem a uma distribuição limite, dada por uma distribuição Beta.

A importância do estudo de urna de Pólya, além de seu estudo teórico, está em que ela possui várias aplicações, entre elas podemos mencionar o uso da urna de Pólya no

caso de ensaios médicos seqüenciais, assim supondo que a resposta de um paciente a um tratamento é dicotômica, se trata de analisar esquemas para aplicar tratamentos, com tendência a pôr mais pacientes nos melhores tratamentos (Wei,L.(1979)[25]). Outro caso de sua utilização é dado quando um grande número de gerentes deve adotar uma nova tecnologia, a qual existe em 2 classes, A e B. Cada gerente analisa que tecnologia é escolhida por r gerentes selecionados aleatoriamente, e se não menos que um dado número m de gerentes usa a tecnologia A, com $r > m$, então ele escolhe a tecnologia A, no outro caso, escolhe a B. (Arthur,B.,Ermol'ev,M.Yu., Kaniovskii,M.Yu.(1983)[3]).

Quanto à simulações, mostramos que é possível gerar seqüências de variáveis aleatórias permutáveis, através do Teorema da Representação de Bruno De Finetti, utilizando diferentes métodos. Uma tarefa que fica para o futuro é determinar critérios de comparação entre as matrizes geradas por estos métodos.

Uma questão interessante é que ainda não há solução quando temos uma seqüência de variáveis aleatórias, com valores em $\{0, 1\}^\infty$, e sem nenhuma outra informação a respeito; devemos dizer se as variáveis nesta seqüência são permutáveis.

Apêndice A

DEFINIÇÕES

Definição A.1 : Seja $(\Omega, \mathcal{F}, P)$ um espaço de probabilidades. Um processo estocástico $\{X_t\}_{t \in T}$ em $(\Omega, \mathcal{F}, P)$ é uma coleção de variáveis aleatórias indexadas em T , isto é, para todo t de um conjunto de índices T , $X(t)$ é uma variável aleatória em $(\Omega, \mathcal{F}, P)$. Às vezes, interpreta-se t como o instante do tempo e $X(t)$ é denominada o estado do processo no instante t . Se o conjunto de índices T é enumerável, então dizemos que $\{X_t\}_{t \in T}$ é um processo estocástico discreto no tempo, e se T for contínuo, dizemos que o processo estocástico é contínuo no tempo.

Definição A.2 : O processo $\{X_n\}_{n \geq 0}$ é uma cadeia de Markov se satisfaz a condição de Markov,

$$P(X_n = s | X_0 = x_0, X_1 = x_1, \dots, X_{n-1} = x_{n-1}) = P(X_n = s | X_{n-1} = x_{n-1})$$

para todo $n \geq 0$ e para todo $s, x_0, x_1, \dots, x_{n-1} \in S$, onde S é um conjunto enumerável e designa-se espaço de estados.

Exemplo A.1 : Modelo de Difusão de Ehrenfest. Considera-se o movimento de M moléculas, distribuídas entre duas urnas. Uma molécula é escolhida ao acaso entre

as M moléculas, retirada da urna e colocada na outra urna. Se denominamos X_n o número de moléculas na primeira urna imediatamente após a n -ésima retirada, então, $\{X_n\}_{n \geq 0}$ é uma cadeia de Markov com as seguintes probabilidades de transição,

$$\begin{aligned} P_{i,i+1} &= \frac{M-i}{M}, & 0 \leq i \leq, M \\ P_{i,i-1} &= \frac{i}{M}, & 0 \leq i \leq, M \\ P_{ij} &= 0, & |i-j| > 1. \end{aligned}$$

Definição A.3 : A cadeia de Markov $\{X_n\}_{n \geq 0}$ é denominada homogênea se

$$P(X_{n+1} = j | X_n = i) = P(X_1 = j | X_0 = i),$$

para todo n, i, j .

Definição A.4 : Seja $\{X(t)\}_{t \geq 0}$ uma família de variáveis aleatórias com valores em um espaço de estados enumerável S e indexadas em $[0, \infty)$. $\{X(t)\}_{t \geq 0}$ é denominada cadeia de Markov a tempo-contínuo quando satisfaz o seguinte,

$$P(X(t_n) = j | X(t_1) = i_1, \dots, X(t_{n-1}) = i_{n-1}) = P(X(t_n) = j | X(t_{n-1}) = i_{n-1}),$$

para todo $j, i_1, \dots, i_{n-1} \in S$ e $t_1 < t_2 < \dots < t_n$.

Definição A.5 : A cadeia de Markov $\{X(t)\}_{t \geq 0}$ é denominada homogênea se

$$P(X_t = j | X_s = i) = P(X_{t-s} = j | X_0 = i),$$

para todo i, j, s, t , com $s \leq t$.

Teorema A.1 : *Equações de Chapman-Kolmogorov.*

$$P(X_{m+n} = j | X_0 = i) = \sum_k P(X_{m+n} = j | X_m = k) P(X_m = k | X_0 = i).$$

Definição A.6 : Seja $\{X(t)\}_{t \geq 0}$ uma cadeia de Markov descontínua, isto é, uma cadeia de Markov discreta no espaço, mas contínua no tempo, a função de intensidade $q_i(t)$ possui a seguinte interpretação: $q_i(t)\Delta t + o(\Delta t)$ é a probabilidade de $X(t)$ mudar no intervalo $(t, t + \Delta t)$ quando $X(t) = i$; em consequência, $1 - q_i(t)\Delta t + o(\Delta t)$ é a probabilidade de não ter nenhuma mudança no intervalo $(t, t + \Delta t)$. A função de probabilidade de transição relativa $Q_{ij}(t)$ se define como a probabilidade condicional que $X(t + \Delta t) = j$, dado que $X(t) = i$.

Definição A.7 : Seja $X: \Omega \rightarrow \mathbb{R}$ uma variável aleatória em um espaço de probabilidade $(\Omega, \mathcal{F}, P)$. Dada $\mathcal{G} \subseteq \mathcal{F}$, dizemos que X é $\mathcal{G}$ -mensurável se

$$X^{-1}(B) \in \mathcal{G} \text{ para } B \in \mathcal{B},$$

onde $X^{-1}(B) \in \sigma(X)$, esta é a σ -álgebra gerada pela variável X , e $\mathcal{B}$ é a σ -álgebra de Borel em $\mathbb{R}$.

Definição A.8 : Sejam $X_1, X_2, \dots$ variáveis aleatórias permutáveis definidas em um espaço de probabilidade $(\Omega, \mathcal{F}, P)$, denomina-se σ -álgebra dos eventos permutáveis à σ -álgebra gerada pelas variáveis aleatórias permutáveis.

Definição A.9 : Sejam $(\Omega, \mathcal{F}, P)$ um espaço de probabilidade, $\mathcal{G}$ uma σ -álgebra tal que $\mathcal{G} \subseteq \mathcal{F}$ e X uma variável aleatória no $(\Omega, \mathcal{F}, P)$. Dizemos que a esperança condicional de X (v.a. não negativa) com respeito à σ -álgebra $\mathcal{G}$, é uma variável aleatória Y , isto é, $Y = E(X|\mathcal{G})$, tal que

a) Y é $\mathcal{G}$ mensurável,

b) $\int_A X dP = \int_A Y dp$, para todo $A \in \mathcal{G}$.

Definição A.10 : Seja $(\Omega, \mathcal{F}, P)$ um espaço de probabilidade e $\{\mathcal{F}_n\}$ uma família de σ -álgebras, tal que $\mathcal{F}_0 \subseteq \mathcal{F}_1 \subseteq \dots \subseteq \mathcal{F}$. Seja $X_0, X_1, \dots$ uma seqüência de variáveis aleatórias em $(\Omega, \mathcal{F}, P)$. Dizemos que $X = (X_n, \mathcal{F}_n)$, $n \geq 0$ é uma seqüência estocástica se X_n é $\mathcal{F}_n$ -mensurável para cada $n \geq 0$.

Definição A.11 : Dizemos que uma seqüência estocástica $X = (X_n, \mathcal{F}_n)$ é uma Martingala se $\forall n \geq 0$ se verificam,

a) $E|X_n| < \infty$,

b) $E(X_{n+1}|\mathcal{F}_n) = X_n$ (P q.c.).

Exemplo A.2 Seja $X_1, X_2, \dots$ uma seqüência de variáveis aleatórias independentes com $E(|X_k|) < \infty$, $\forall k$ e $E(X_k) = 0$, $\forall k$.

Definimos $S_n = X_1 + X_2 + \dots + X_n$, com $S_0 = 0$;

e $\mathcal{F}_n = \sigma(X_1, X_2, \dots, X_n)$, com $\mathcal{F}_0 = \{\emptyset, \Omega\}$,

para $n \geq 1$ logo

$$E(S_n|\mathcal{F}_{n-1}) = E(S_{n-1}|\mathcal{F}_{n-1}) + E(X_n|\mathcal{F}_{n-1}) = S_{n-1} + E(X_n) = S_{n-1}.$$

a primeira igualdade surge da propriedade de linearidade das esperanças condicionais. A segunda igualdade provém do fato de S_{n-1} ser $\mathcal{F}_{n-1}$ mensurável, então $E(S_{n-1}|\mathcal{F}_{n-1}) = S_{n-1}$ (P q.c.), como X_n é independente de $\mathcal{F}_{n-1}$, temos que

$$E(X_n|\mathcal{F}_{n-1}) = EX_n \text{ (P q.c.)}.$$

REFERÊNCIAS BIBLIOGRÁFICAS

- [1] Aldous, D.J. (1983). Exchangeability and related topics, in *École d'Été de Probabilités de Saint-Flour XIII* - edited by a. Dold and B. Eckman; Springer-Verlag, Lecture Notes in Mathematics, New York, pp. 1-198.
- [2] Arthur B., Ermol'ev, M. Yu., Kaniovskii, M. Yu (1983). A Generalized Urn Problem and its applications. *Kibernetika*, 1, pp. 49-56.
- [3] 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.
- [4] Bharucha-Reid, A.T. (1960). *Elements of the theory of Markov Process and their applications*. McGraw-Hill.
- [5] Bernardo, J.M., Smith, A.F.M. (1994). *Bayesian Theory*. Chichester, John Wiley, (Wiley Series in Probability and Mathematical Statistics).
- [6] Cox, D.R., Snell, E.J. (1970). *Analysis of Binary Data*. Chapman and Hall. Segunda edição.
- [7] De Finetti, B. (1937). La prévision ses lois logiques, ses sources subjectives. *Annales de l'Institut Henri Poincaré*, 7, 1-68, in H.E. Kyburg, J. and H.E. Smokler (trans.), *Studies in Subjective Probability*. John Wiley, New York.

- [8] De Finetti, B. (1969). Sulla pregeuibilità di processo aleatori scanbialbili. Rend. Inst. Mat. Univ. Trieste, 1.
- [9] Diaconis, P., Ylvisaker, D. (1979). Conjugate Priors for Exponential Families. Ann. Statist. 7, 269-281.
- [10] Feller, W. An introduction to probability theory and its applications. Vol. I e II.
- [11] Freedman, D. A. (1965). Bernard Friedman's Urn. Ann. Math. Statist. 36, 956-970.
- [12] Grimmett, G., Stirzaker, D. Probability and Random Process. (1993). Oxford University Press Inc., New York.
- [13] Heath, D. and Sudderth, W. (1976). De Finetti's theorem on exchangeables variables. American Statistician, vol. 30, 4, pp. 188-9.
- [14] Hill, B. M., Lane, D., Sudderth, W. (1987). Exchangeable Urn Processes. The Annals of Probability 15, pp. 1586-1592.
- [15] Iglesias-Zuazola, P., Pereira, C. A., Wechsler, S. (1994). Predictivistic Statistical Inference in Finite Populations. IME, USP. São Paulo.
- [16] Iglesias-Zuazola, P., Silva, A. J. da, Wechsler, S. (1995). Partial Exchangeability and Extendibility. IME, USP. São Paulo.
- [17] Johnson, N., Kotz, S. (1977). Urn Models and Their Application. John Wiley Sons, Inc.
- [18] Koch, G. and Spizzichino, F. eds (1982). Exchangeability in Probability and Statistics, 9, pp. 45-58.

- [19] O'Hagan, A., Kendall, M.G.(1994). The advanced Theory of Statistics, Bayesian Theory, vol. 2B, London, Edward Arnold (Kendall's Advanced Theory of Statistics).
- [20] Pólya, G.(1931). Sur quelques points de la theorie des probabilites. Ann.Inst. H.Poincaré 1, pp. 117-161.
- [21] Ross, S. Stochastic Processes. (1983). John Wiley Sons, Inc.
- [22] Silva A.J.,(1997). Tese de Doutorado: "Sobre Condições de Estendibilidade, Simetria Esférica e Inferência Bayesiana em Modelos Elípticos com erros nas Variáveis". IME-USP.
- [23] Spizzichino, F.(1982). Extendibility of symmetric probability distributions related bounds. In Exchangeability in Probability and Statistics, edited by G. Kosh and F. Spizzichino (North-Holland, Amsterdam) p.313.
- [24] Wechsler, S.(1993). Exchangeability and Predictivism. Erkenntnis 38, pp. 343-350.
- [25] Wei, L.J.(1979). The Generalized Pólya's Urn design for sequential medical trials. Annals of Statist. Vol.7, 2, 291-296.
- [26] Williams, D.(1991). Probability with martingales, Cambridge, University Press (Cambridge Mathematical Textbooks).
- [27] Zabell, S. (1982). W.E. Johnson's "Sufficientness" Postulate. The Annals of Statistics, vol. 10 No 4. pp. 1091-1099.
- [28] Zabell, S. (1997). Confirming Universal Generalizations. Erkenntnis 45, 267-283.