Coerência, Probabilidade

e Calibração

Rosângela Helena Loschi

Dissertação Apresentada
ao
Instituto de Matemática e Estatística
da
Universidade de São Paulo
para a Obtenção do Grau de
Mestre em
Estatística

Área de concentração: **Estatística**
Orientador: **Sergio Wechsler**

- São Paulo, Outubro 1992 -

ABSTRACT

Subjective probability is obtained starting from the Principle of Coherence. The particular process of construction based on bets introduced by de Finetti (1931) is described in detail. The problems related to group decisions are also discussed with several properties of consensus-distributions and criteria for evaluation of probabilities presented (in particular, the Calibration Criterion). Methods for generation of consensus-distributions are also discussed and a coherent method suggested by de Finetti for construction of a group's "opinion" is developed.

SUMÁRIO

A construção de probabilidade subjetiva é feita partindo-se do princípio fundamental de Coerência. O particular processo de construção baseado em apostas proposto por de Finetti(1931) é descrito detalhadamente. A problemática que envolve decisões em grupo também é abordada com a apresentação de propriedades desejáveis em distribuições-consenso e também de critérios para avaliação de probabilidades (particularmente, é visto o Critério de Calibração). Além disso, são exibidos alguns métodos de gerar distribuições-consenso e desenvolvido um método coerente sugerido por de Finetti para estabelecer a "opinião" de um grupo.

AGRADECIMENTOS

Ao CNPq e à FAPESP pelo apoio financeiro.

Aos professores Nelson e Júlio por todo apoio e incentivo durante este longo tempo de mestrado, ao Carlinhos pelo primeiro de Finetti (este a gente nunca esquece), ao Caio por ter provado que PT é a solução (e isso já em 89) e ao Chico Teoria pelo empurrãozinho inicial.

À Eunice por ter tirado do fundo do baú vários artigos fundamentais que já figuravam no rol dos esquecidos (acho mesmo que dos próprios autores).

À todos os meus amigos (e aqui serei bem Mineira), daqui e de além mar, por todos os bons papos, pelos cafezinhos, pelas noites de gandaia, pelos ombros que utilizei várias vezes para chorar as pitangas e por sentarem ao meu lado na sarjeta sempre que precisei. Vocês foram demais. Jamais esquecerei o carinho, incentivo e a amizade.

Ao meu pai que, tenho certeza, torce por mim de onde estiver.

À minha Família pela garra que ela tem. Vocês ensinaram-me que vida sem alegria é a maior baixaria. Vocês são sábios. Obrigada pela força, pelo bom exemplo e por tudo.

Ao Sergio pela excelente orientação e mais uma lista de coisas que segue em anexo. Sem a sua mão este trabalho jamais teria ficado como ficou.

Valeu, gente !!!!

*À Dona Helena e à Hilda
minhas eternas musas*

ÍNDICE

Introdução	1
Capítulo Um - Coerência e Probabilidade	
1. Introdução	7
2. Medida Numérica de Incerteza	7
Capítulo Dois - O Problema da Decisão em Grupo	
1. Introdução	23
2. Algumas Características e Propriedades	25
2.1. Marginalização	26
2.2. Bayesianidade Externa	29
2.3. Correlação entre Opiniões dos Membros do Grupo	33
2.4. Conhecimento Específico	33
2.5. Sinceridade	33
2.6. Probabilidades ou Dados?	36
2.7. Calibração	37
3. Formando Distribuições-Consenso	48
3.1. Agrupamento Linear de Probabilidades	49
3.2. Agrupamento Logarítmico de Probabilidades	53
3.3. A Distribuição-Consenso de Morris DeGroot	55
4. Campos de Coincidência de Opiniões	63
Conclusão	66
Referências	

INTRODUÇÃO

O objetivo deste trabalho é mostrar detalhadamente que medidas de incerteza coerentes são probabilidades e apresentar uma discussão ampla sobre o problema de decisão em grupo.

O capítulo um está endereçado à construção de probabilidade subjetiva. Considerando-se de um conjunto Θ de valores que um estado da natureza θ pode assumir, o objetivo deste capítulo é provar que medidas numéricas de incerteza que Você estabelece coerentemente sobre valores de θ são distribuições de probabilidade.

O particular processo de construção de probabilidade considerado é o processo baseado em apostas sugerido por de Finetti em (1931).

Um cenário constituído por um cassino e dois jogadores (o “Bookmaker” e o “Você”) é considerado com o intuito de tornar mais claro este processo de construção de probabilidade - além de tornar mais intuitivas as demonstrações dos resultados do capítulo.

No cassino são negociados todos os tipos de bilhetes de loteria formados a partir dos possíveis valores do estado da natureza θ a respeito dos quais Você está incerto. Os prêmios a serem possivelmente pagos para cada bilhete de loteria deste tipo são estabelecidos pelo Bookmaker e completamente desconhecidos por Você. A Você cabe a determinação das taxas de negociação de cada um destes bilhetes - o que, obviamente, será feito sem o conhecimento do valor de θ . A taxa p_i estabelecida para a negociação do bilhete i revela a sua indiferença entre a quantia $p_i S_i$ certa e o bilhete de loteria que paga um prêmio S_i caso $\theta = \theta_i$. As taxas estabelecidas por Você desta forma deter-

minam a sua medida numérica de incerteza sobre θ . Você deve ser coerente ao estabelecer suas taxas, isto é, Você deve estipular suas taxas de tal forma que esta escolha não o leve a perder com certeza do Bookmaker.

A força intuitiva deste cenário enquanto representação da noção de incerteza é usada na demonstração de dois resultados fundamentais. O primeiro deles estabelece a relação entre medidas de incerteza coerentes e os axiomas de Kolmogorov e o segundo a relação entre medidas de incerteza coerentes e a definição de probabilidade condicional (também sugerida por Kolmogorov). Em resumo, é provado que se Você é coerente, suas medidas de incerteza são probabilidades.

O capítulo dois destina-se a discutir o problema de decisão em grupo, com algumas tentativas de trazer o conceito de coerência para a nova situação.

Na seção 1, o problema de decisão em grupo é apresentado juntamente com algumas polêmicas que surgem nas tentativas de sua solução. O Teorema da Impossibilidade de Arrow(1951), que é apresentado em linhas gerais, mostra a inexistência de uma solução para este problema consideradas algumas condições mínimas de racionalidade para um grupo. É discutida também a idéia simplista de Lindley(1985) que “soluciona” este problema elegendo um líder para o grupo. Além disso, a subdivisão em tipos de problema de decisão em grupo sugerida por French(1985) é apresentada com uma descrição de cada um dos tipos estabelecidos, a saber, os problemas do expert, do livro-texto e o de decisão em grupo propriamente dito.

Na seção 2, define-se distribuição-consenso e apresenta-se propriedades e características importantes para a formação destas distribuições e também para a avaliação de probabilidades individuais.

As propriedades de Bayesianidade Externa e Marginalização são exibidas como propriedades à primeira vista desejáveis na formação de uma distribuição-consenso. Um processo de combinação de opiniões satisfará a propriedade de Marginalização se gerar distribuições-consenso marginais coincidentes em ambos os casos: ou através de uma combinação direta de opiniões marginais individuais, ou através da marginalização da distribuição-consenso formada pela combinação das opiniões individuais completas. A propriedade de Bayesianidade Externa, por sua vez, será satisfeita por processos de agrupamento de opiniões quando a distribuição-consenso obtida por uma combinação de distribuições a posteriori individuais coincidir com a distribuição-consenso a posteriori obtida por uma atualização, via teorema de Bayes, da distribuição-consenso formada pela combinação de distribuições a priori individuais. Alguns resultados importantes e pontos frágeis destas propriedades também são destacados.

Duas questões interessantes são abordadas quando é levada em conta a informação possuída pelos membros do grupo. O compartilhamento de informações pelos membros do grupo gera uma correlação entre as opiniões. E a especificidade no assunto, caracterizada pela quantidade de informação de cada membro do grupo, possibilitará uma atribuição de pesos distintos às suas opiniões.

Uma característica importante que se deve almejar nas probabilidades individuais anunciadas é que elas representem sinceramente a incerteza da pessoa sobre um dado evento. Na sub-seção 2.5, as regras de penalidades próprias são definidas e sugeridas como um mecanismo operacional usado para induzir a sinceridade de uma pessoa. Um outro bom mecanismo para

induzir a sinceridade são as loterias (vistas no capítulo um). Alguns resultados estabelecidos por Kadane e Winkler(1988) relacionando estes processos de expressão sincera de probabilidades com a função de utilidade da pessoa também são apresentados.

Outra questão amplamente discutida no trabalho é a interpretação de probabilidades alheias. A proposta razoável para esta questão é que sejam consideradas como dados e não como probabilidades, conforme sugere de Finetti(1972).

Guardando relação com esta questão, um critério para avaliar probabilidades subjetivas individuais é descrito - o Critério de Calibração. O objetivo deste critério é relacionar probabilidades anunciadas com frequências relativas. A maneira como este relacionamento ocorre é confrontada com a definição freqüentista de probabilidade, enfatizando-se que a relação entre freqüência e probabilidade sugerida pelo critério de calibração é bem mais tênue do que a que norteia a definição freqüentista. O critério é estudado em dois contextos distintos: situações com ou sem feedback. Para ambos os casos, são estabelecidos resultados que descrevem quando Você tem certeza (na sua opinião) de uma boa calibração assintótica. Uma ampla discussão sobre as relações entre coerência e calibração é apresentada. Fica estabelecido que o uso da informação de boa ou má-calibração não conduzirá a qualquer incoerência se usada na avaliação de probabilidades alheias. O uso desta informação, no entanto, para uma auto-recalibração é incoerente.

As seções 3 e 4 propõem-se à construção de uma “opinião” para o grupo. As distribuições-consenso que são normalmente interpretadas como uma opinião do grupo são abordadas na seção 3. Três métodos para gerar consen-

so, partindo de distribuições individuais, são apresentados: os agrupamentos Linear e Logarítmico de opiniões e a forma-consenso de DeGroot. As características de cada método são ressaltadas dentro do proposto na seção 2 e, também, suas particularidades.

Uma particularidade do método de agrupamento linear de opiniões é que neste método, se uma revisão nos pesos é permitida, a propriedade de Bayesianidade Externa passa a ser satisfeita.

Comparado com os outros dois métodos, o método de gerar consenso proposto por DeGroot(1974) mostra-se um pouco mais atraente por fornecer explicitamente uma maneira de obtenção dos pesos dados às opiniões dos membros do grupo. Outro ponto de comparação é a maneira como o consenso é gerado. Aqui, o consenso é gerado através de uma atualização constante de opiniões até que estas não mais se alterem com novos diálogos. Nos outros dois métodos o consenso é simplesmente uma média generalizada de probabilidades. Nesta seção, também são exibidas uma tentativa de definição para Coerência Conjunta e uma crítica à natureza das “distribuições” geradas pelos métodos de consenso. Infelizmente, o que se conclui é que grupos não têm opinião no sentido individual, ou seja, “opiniões” de grupos não expressam coerência.

Por fim, na seção 4, são mostrados os Campos de Coincidência de Opiniões propostos por de Finetti(1953). A idéia deste método é determinar uniões de intervalos onde a opinião de todos os membros do grupo coincidem. A probabilidade associada à união destes intervalos é então tomada como a opinião do grupo. Embora este método não gere uma distribuição de probabilidade completa, ele é apresentado como a única opinião genuína

e coerente que um grupo de pessoas pode expressar.

CAPÍTULO UM

COERÊNCIA E PROBABILIDADE

1.INTRODUÇÃO

Este capítulo foi baseado em alguns artigos de Bruno de Finetti sobre construções de probabilidade subjetiva. O objetivo do capítulo é estabelecer resultados que comprovem que se Você estabelece suas medidas numéricas de incerteza coerentemente, estas seguirão as leis do Cálculo de Probabilidades.

Com este objetivo, será considerado um conjunto Θ constituído de todos os valores possíveis θ_j que um “estado da natureza” θ pode assumir. A expressão *estado da natureza* denota uma condição cujos valores são desconhecidos por Você e que, por exemplo em um problema de decisão, afetam as conseqüências de cada decisão disponível para Você. Por simplicidade, Θ será considerado um conjunto finito.

2.MEDIDA NUMÉRICA DA INCERTEZA

Sob o ponto de vista Bayesiano, todas as medições de incerteza coerentes devem ser distribuições de probabilidade.

Aqui, deseja-se quantificar a incerteza que Você tem em relação a um estado da natureza θ . Esta quantificação, se coerente, será necessariamente uma distribuição de probabilidade sobre Θ . Tal distribuição de probabilidade é subjetiva e depende de toda a informação que Você possui sobre θ .

Existem vários “processos de construção” de probabilidade subjetiva. (A expressão *processo de construção* significa partir-se de descrições de propriedades naturais da noção de incerteza para obter-se necessariamente - isto é,

provar-se um teorema - que a medida numérica da incerteza deve obedecer às Leis do Cálculo de Probabilidades).

Processos de construção mais diretos, baseados em propriedades de uma relação subjetiva de “ordenação e comparação” dos θ_j 's, podem ser encontrados em de Finetti (1931), Fishburn (1986), dentre outros. Aqui, será descrito o particular processo de construção de probabilidades baseado em apostas proposto por de Finetti (1931). Apesar de tal construção envolver a noção de dinheiro que é até certo ponto exógena ao problema, “tendo sido provado que se pode superar a desconfiança decorrente da natureza, de certa forma muito concreta e talvez artificial da definição baseada em apostas, esta é preferível, antes de tudo, por sua clareza” (de Finetti (1937) - pg.61).

Para de Finetti (1931): a sua medida de incerteza a respeito de um determinado valor θ_j de θ é entendida como o número p_j que torna Você indiferente entre possuir a quantia $p_j S$ com certeza, ou possuir um bilhete que valerá S caso $\theta = \theta_j$ e 0 (zero) caso $\theta \neq \theta_j$. O processo de construção, então, baseia-se no fato óbvio que uma pessoa racional não participará de um jogo no qual ela perderá dinheiro certamente, qualquer que seja o seu resultado.

Será considerado que o estado da natureza θ , sobre o qual Você está incerto, pode assumir apenas um número finito de valores, $\theta_1, \theta_2, \dots, \theta_n$, isto é, $\Theta = \{\theta_1, \dots, \theta_n\}$. Um subconjunto E de Θ , qualquer que seja ele, será um *evento*. Assim, existem $r = 2^n$ eventos distintos de Θ .

Admite-se que sobre cada um destes r eventos ($E_1, \dots, E_r$) existam bilhetes de loteria i , $i = 1, \dots, r$. Um bilhete do tipo i é um bilhete que premiará seu dono com uma quantia S_i (positiva ou negativa) se o evento E_i ocorrer, ao passo que se o evento E_i não ocorrer, o dono do bilhete não receberá nada.

Diz-se que um evento E “ocorre” quando verifica-se que o valor θ_j de θ é um elemento de E . O fato do prêmio S_i , a ser possivelmente pago ao portador do bilhete i , poder ser uma quantia negativa ou positiva é interpretado como resultado de uma venda ou uma compra do bilhete i , respectivamente.

O “preço” a ser pago por um bilhete i deste tipo é a quantia $p_i S_i$, onde p_i é uma taxa estabelecida por Você de forma que lhe seja indiferente possuir a quantia $p_i S_i$ com certeza ou o bilhete i descrito acima. Saber que o preço $p_i S_i$ pode ser tanto um preço de compra quanto um preço de venda faz com que Você evite dizer um valor de p_i menor (maior) do que aquele que Você realmente julga ser justo para negociar o bilhete i , uma vez que Você pode ser obrigado a vender (comprar) o bilhete i por aquele preço.

Um “bookmaker” convida Você a participar de um jogo. Ele negocia todos os bilhetes de loteria formados a partir dos r eventos de Θ , alguns dos quais, em particular, ele deseja negociar com Você. Ele permite que Você determine as taxas $p_1, \dots, p_r$, de modo que, na sua opinião, o preço justo a ser pago pelo bilhete de loteria i , é a quantia $p_i S_i$, $i = 1, \dots, r$. Os prêmios $S_1, \dots, S_r$ a serem eventualmente pagos pelos bilhetes $1, \dots, r$, respectivamente, são estabelecidos pelo bookmaker e seus valores somente são revelados após Você declarar os valores das taxas $p_1, \dots, p_r$. Assim, Você só saberá que se trata de uma compra ou de uma venda do bilhete i , quando S_i for anunciado após todos os p_i 's terem sido estabelecidos.

A única característica dos S_i 's conhecida por Você, antes de Você declarar o valor dos p_i 's, é que eles são quantias pequenas. Esta exigência é feita com o intuito de eliminar do problema a utilidade do dinheiro estimulando Você a dizer sinceramente os seus valores para as taxas p_i . Você deve dizer

exatamente o menor (maior) - se tratar-se de uma venda (compra) - valor de p_i que o torna indiferente entre a quantia $p_i S_i$ com certeza e o bilhete de loteria i . As taxas $p_1, \dots, p_r$, definidas assim, revelam a sua incerteza sobre a ocorrência dos eventos $E_1, \dots, E_r$. Em Kadane e Winkler(1988) pode-se encontrar mais detalhes sobre a relação entre processos de elicitação de probabilidades e utilidade do dinheiro.

Até agora, fez-se apenas a descrição de um cenário que torna mais intuitiva a demonstração dos teoremas que relacionam coerência com as Leis do Cálculo de Probabilidades. Sabe-se que coerência é uma exigência mínima que se faz sobre a opinião das pessoas. Mas, o que é uma pessoa coerente? Dentro do contexto apresentado, onde $E_1, \dots, E_r$ são todos os eventos de Θ , coerência pode ser assim definida.

DEFINIÇÃO 1: (*Coerência*) - Você é coerente se o conjunto $\{p_1, \dots, p_r\}$ das taxas que Você estabelece para a compra ou venda dos r bilhetes de loteria do "bookmaker" é tal que, qualquer que seja a escolha dos S_i 's feita por ele, seu ganho será não-negativo para pelo menos um valor θ_j .

DEFINIÇÃO 2: (*Incoerência*) - Você é incoerente se não é coerente.

Da definição 1 pode-se concluir, por exemplo, que uma escolha de taxas onde, na melhor das hipóteses, Você tem um ganho igual a zero é coerente. Ou seja, esta definição confirma que coerência é realmente uma exigência muito fraca sobre a opinião de uma pessoa, é o mínimo que se exige na tentativa de definir racionalidade. E, ao mesmo tempo, fortalece ainda mais a idéia que opiniões de pessoas incoerentes não devem ser levadas a sério, a

menos que se queira perder sempre.¹

Antes de apresentar os principais resultados deste capítulo, outros resultados úteis (e que também surgem naturalmente da condição de coerência) são exibidos.

LEMA 1 - Se Você é coerente, para toda partição $\pi = \{E_1, \dots, E_k\}$ de Θ ,

$$p_1 + p_2 + \dots + p_k = 1,$$

onde p_i representa a sua medida numérica de incerteza sobre a ocorrência de E_i .

DEMONSTRAÇÃO: Suponha, por absurdo, que Você estabeleça suas taxas de compra ou venda $p_1, \dots, p_k$ de tal forma que $\sum_{i=1}^k p_i \neq 1$. Considere, agora, que um “bookmaker” convida Você a jogar. Ele pretende apenas comprar ou vender bilhetes referentes aos eventos E_i da partição π .

Para todo i ($i = 1, \dots, k$), caso E_i ocorra, o seu ganho líquido será dado pela seguinte expressão:

$$G_i = S_i - \sum_{i=1}^k p_i S_i \quad (I)$$

onde S_i é o prêmio a ser possivelmente pago pelo bilhete i e é desconhecido por Você.

Como π é uma partição de Θ , a ocorrência de qualquer um dos θ_j , $j = 1, \dots, n$, implicará a ocorrência de exatamente um dos eventos E_i de

¹Esta idéia é chamada de “The Dutch Book Argument” - o argumento da pule holandesa. Veja Hill(1991) sobre a origem do termo.

π . E, então, o sistema de equações (I) para $i = 1, \dots, k$ realmente apresenta todos os seus possíveis ganhos.

O determinante do sistema gerado por (I) é dado pela seguinte expressão:

$$\begin{vmatrix} 1 - p_1 & -p_2 & \dots & -p_k \\ -p_1 & 1 - p_2 & \dots & -p_k \\ \vdots & \vdots & \ddots & \vdots \\ -p_1 & -p_2 & \dots & 1 - p_k \end{vmatrix} = 1 - p_1 - \dots - p_k,$$

que, por hipótese de absurdo, é diferente de zero.

Assim, da teoria de sistemas lineares decorre que o sistema dos possíveis ganhos possui uma única e bem definida solução. Logo, se Você determinou os seus p_i 's de forma que $\sum_{i=1}^k p_i \neq 1$, Você deu uma oportunidade ao “book-maker” de escolher convenientemente valores de $S_i, i = 1, \dots, k$, de maneira que os seus ganhos sejam sempre negativos. Absurdo, pois Você é coerente por hipótese. $\square$

Um segundo resultado que é também de grande interesse antes da apresentação dos teoremas principais é a aditividade finita, apresentada no lema a seguir.

LEMA 2 - Sejam $E_1, \dots, E_m$ eventos de Θ , dois-a-dois disjuntos, e sejam $p_1, \dots, p_m$ as suas medidas numéricas de incerteza sobre a ocorrência de $E_1, \dots, E_m$, respectivamente. Se Você é coerente, então

$$p_{\cup_{i=1}^m E_i} = \sum_{i=1}^m p_i.$$

DEMONSTRAÇÃO :

1) Por indução, considere primeiramente $m = 2$. E tome duas partições distintas de Θ , $\pi = \{E_1, E_2, \overline{E_1 \cup E_2}\}$ e $\tau = \{E_1 \cup E_2, \overline{E_1 \cup E_2}\}$. Para os bilhetes de loteria formados a partir dos eventos de π e τ Você estabelece, respectivamente, as taxas $\{p_1, p_2, p\}$ e $\{p_{12}, p\}$.

Do lema 1, sabe-se que

$$\begin{cases} p_1 + p_2 + p = 1 \\ p_{12} + p = 1 \end{cases}$$

Este fato permite concluir que $p_{12} = p_1 + p_2$.

2) Suponha a hipótese de indução: para $E_1, \dots, E_{k-1}$, eventos dois-a-dois disjuntos de Θ ,

$$p_{\cup_{i=1}^{k-1} E_i} = \sum_{i=1}^{k-1} p_i.$$

Para $E_1, \dots, E_k$ eventos dois-a-dois disjuntos de Θ , B e E_k são eventos disjuntos, onde $B = \cup_{i=1}^{k-1} E_i$.

Então, pela parte (1) da demonstração,

$$p_{E_k \cup B} = p_k + p_B.$$

Por hipótese de indução,

$$p_B = \sum_{i=1}^{k-1} p_i.$$

Daí,

$$p_{E_k \cup B} = p_k + \sum_{i=1}^{k-1} p_i = \sum_{i=1}^k p_i.$$

□

French (1988) questiona esta demonstração argumentando que não existe garantia de que Você (mesmo sendo coerente) atribuirá a mesma taxa a um bilhete quando este refere-se a um mesmo evento que, no entanto, faça parte de partições distintas. Isto é, para ele, declarar que a taxa do bilhete $\overline{E_1} \cap \overline{E_2}$ é p tanto na partição π quanto na partição τ “é uma falácia”². Segundo a visão de de Finetti (1931) sobre coerência, esta questão de French não possui qualquer sentido. A taxa de compra (ou venda) do bilhete i , p_i , é estabelecida independentemente da particular partição à qual o evento E_i possa pertencer. Esta taxa deve revelar a incerteza que Você possui sobre E_i , e dentro desta visão, não faz sentido algum falar em taxas distintas para um mesmo bilhete de loteria i .

Pretende-se mostrar agora a ligação entre as Leis do Cálculo de Probabilidades (Axiomas de Kolmogorov) e medidas numéricas coerentes de incerteza.

Sejam $A_1 = \{\theta_1\}$, ..., $A_n = \{\theta_n\}$ todos os eventos unitários de Θ e seja $p_{\{\theta_j\}}$ a probabilidade do evento A_j , $j = 1, \dots, n$. Sendo Θ um conjunto finito, os axiomas de Kolmogorov para probabilidades estabelecidas sobre os

²French utiliza uma definição mais fraca de coerência - isto é, ele usa uma definição de coerência em conjunto de taxas por partições de Θ , apenas. (Este fato faz com que a demonstração do lema 2 fique mais complicada.) Isto pode ter sido sugerido pela construção apresentada por de Finetti (1931), que, no entanto, parece ter sido feita apenas por simplicidade.

eventos de Θ tornam-se equivalentes aos seguintes:

$$\text{Ax.1)} \quad p_{\{\theta_j\}} \geq 0, \forall j = 1, \dots, n$$

$$\text{Ax.2)} \quad \sum_{j=1}^n p_{\{\theta_j\}} = 1.$$

Ax.3) Seja A um evento qualquer de Θ . Então,

$$p_A = \sum_{j: \theta_j \in A} p_{\{\theta_j\}}.$$

Sejam novamente $p_1, \dots, p_r$ as medidas numéricas de sua incerteza sobre cada um dos r eventos de Θ . Em particular, sejam $p_{\{\theta_1\}}, \dots, p_{\{\theta_n\}}$ as descrições numéricas de sua incerteza sobre cada um dos n eventos unitários de Θ , $A_1, \dots, A_n$, respectivamente.

TEOREMA 1 - Você é coerente se, e somente se, o conjunto das suas medidas numéricas de incerteza satisfaz os três axiomas de Kolmogorov acima.

DEMONSTRAÇÃO:

1) Suponha, por absurdo, que exista pelo menos um j tal que, para Você, $p_{\{\theta_j\}} < 0$. Considere ainda que conhecendo o conjunto de p 's que Você estabeleceu, o "bookmaker" estabeleça um prêmio $S_j < 0$ para o bilhete formado sobre o evento A_j , enquanto que para os demais bilhetes ele estabeleça um prêmio $S_i = 0, i \neq j$. Os seus possíveis ganhos com a compra (ou venda) dos r tipos de bilhetes serão, então:

$$\begin{cases} G_1 = (1 - p_{\{\theta_j\}})S_j & \text{se } A_j \text{ ocorrer} \\ G_2 = -p_{\{\theta_j\}}S_j & \text{se } A_j \text{ não ocorrer.} \end{cases}$$

Como $p_{\{\theta_j\}} < 0$, $(1 - p_{\{\theta_j\}})$ e $-p_{\{\theta_j\}}$ serão números positivos e sendo S_j negativo, $G_1 < 0$ e $G_2 < 0$. Absurdo!

2) Os eventos $A_1, \dots, A_n$ formam uma partição de Θ . Assim, pelo lema 1,

$$\sum_{j=1}^n p_{\{\theta_j\}} = 1.$$

3) Considere A um evento qualquer de Θ . A é obviamente uma união finita de eventos unitários distintos, isto é, $A = \cup_{j:\theta_j \in A} \{\theta_j\}$.

Pelo lema 2, então,

$$p_A = \sum_{j:\theta_j \in A} p_{\{\theta_j\}}$$

Reciprocamente, sejam $a_{ij} = \mathcal{I}(\theta_j \in E_i)$, para $i = 1, \dots, r$ e $j = 1, \dots, n$.

O sistema de seus possíveis ganhos é

$$G_j = \sum_{i=1}^r S_i(a_{ij} - p_i), \quad j = 1, \dots, n.$$

Deste sistema, a seguinte relação é obtida:

$$\begin{aligned} \sum_{j=1}^n p_{\{\theta_j\}} G_j &= \sum_{j=1}^n p_{\{\theta_j\}} [\sum_{i=1}^r S_i(a_{ij} - p_i)] \\ &= \sum_{i=1}^r S_i [\sum_{j=1}^n p_{\{\theta_j\}} (a_{ij} - p_i)] \\ &= \sum_{i=1}^r S_i [\sum_{j=1}^n p_{\{\theta_j\}} a_{ij} - \sum_{j=1}^n p_{\{\theta_j\}} p_i] \\ &= (\text{pelo Ax.2}) = \sum_{i=1}^r S_i [\sum_{j=1}^n p_{\{\theta_j\}} a_{ij} - p_i]. \end{aligned}$$

Já do Ax.3, conclui-se que

$$\sum_{j=1}^n p_{\{\theta_j\}} a_{ij} = \sum_{j=1}^n p_{\{\theta_j\}} \mathcal{I}(\theta_j \in E_i) = p_i,$$

de modo que para todos os valores de $S_1, \dots, S_r$,

$$\sum_{j=1}^n p_{\{\theta_j\}} G_j = \sum_{i=1}^r S_i (p_i - p_i) = 0. \quad (II)$$

Suponha no entanto, por absurdo, que exista pelo menos um valor de $\{S_1, \dots, S_r\}$ tal que $G_j < 0, \forall j = 1, \dots, n$. Então, para que (II) seja satisfeita, pelo menos um dos $p_{\{\theta_j\}}$'s será necessariamente negativo. (A não ser que, para todo $j, p_{\{\theta_j\}} = 0$ e, neste caso, o Ax.2 será contradito). Absurdo, pois valem, por hipótese, os axiomas de Kolmogorov. $\square$

Portanto, o conjunto de suas medidas numéricas de incerteza sobre os eventos de Θ é coerente se, e somente se, os seus elementos são estabelecidos satisfazendo os Axiomas de Kolmogorov.

Todos os resultados importantes do Cálculo de Probabilidades surgem como consequência deste teorema. Como exemplo, pode-se citar:

COROLÁRIO 1- Se Você é coerente, para todo $i = 1, \dots, r, p_i \geq 0$.

COROLÁRIO 2- Se Você é coerente, a sua medida de incerteza sobre o evento vazio ϕ é zero.

Este fato que surge como consequência direta do Teorema 1 pode ser justificado de maneira intuitiva. Considere um bilhete de loteria formado sobre o evento ϕ . Este bilhete, dentro da estrutura estabelecida anteriormente, é um bilhete que premiará o seu portador com uma quantia S se ϕ ocorrer,

enquanto que se ϕ não ocorrer, o prêmio será zero. Como Θ contém todos os possíveis valores de θ , certamente um dos θ_j 's irá ocorrer. Isto indica que o evento vazio jamais ocorrerá. E, portanto, se Você for coerente, a taxa de compra ou venda de um bilhete deste tipo será $p_\phi = 0$, para que Você não dê oportunidade ao “bookmaker” de vencê-lo sempre.

COROLÁRIO 3 - Se Você é coerente, a sua medida numérica de incerteza sobre o evento Θ é um.

Este também é um resultado que pode ser intuitivamente explicado usando-se bilhetes de loteria. Considere um bilhete formado sobre o evento Θ . Suponha que, para Você, a taxa de compra ou venda justa para este bilhete seja $p \neq 1$. Considere que para esta sua escolha, o “bookmaker” escolha o seguinte conjunto de prêmios para os r bilhetes de loteria de Θ , $\{0, \dots, 0, S\}$, onde $S \neq 0$ é o prêmio a ser pago pelo bilhete formado sobre o evento Θ . Neste caso, como Θ é um evento que certamente ocorre, o seu possível ganho é $G_j = (1 - p)S$, $\forall j = 1, \dots, n$. Se $p < 1$ ($p > 1$) basta que o “bookmaker” escolha S negativo (positivo) para que Você perca certamente.

COROLÁRIO 4 - Se Você é coerente, então a sua medida numérica de incerteza sobre qualquer evento A é tal que $0 \leq p_A \leq 1$

COROLÁRIO 5 - Se Você é coerente, as suas medidas numéricas de incerteza para eventos complementares, A e $\bar{A} = \Theta - A$, são relacionadas por:

$$p_{\bar{A}} = 1 - p_A.$$

Para que coerência possa ser vista como o alicerce de toda a teoria de probabilidades é necessário ainda estabelecer a relação entre coerência e a

“definição” de probabilidade condicional estabelecida por Kolmogorov. Kolmogorov postula que a probabilidade de um evento A condicional a um evento B é o número $P(A|B)$ dado pela seguinte expressão:

$$P(A|B) = \frac{P(A \cap B)}{P(B)},$$

desde que $P(B) \neq 0$. Ou seja, ele toma a *lei do produto* como a “definição” de probabilidade condicional. Esta definição estabelecida por Kolmogorov é desprovida de um significado substantivo e é restrita aos casos onde $P(B) \neq 0$. Embora isto aconteça, como toda definição é arbitrária, não existe qualquer erro em fazer-se este tipo de consideração.

Não é difícil perceber, no entanto, que o nome *probabilidade condicional* tem por si só um significado intuitivo. Parece mais natural, então, construir a noção de probabilidade condicional partindo deste significado, ao invés de defini-la diretamente, como fez Kolmogorov. Com a construção intuitiva, a lei do produto - sugerida por Kolmogorov como a definição de probabilidade condicional - é obtida inevitavelmente como uma consequência da condição de coerência, e também, ela é válida sem qualquer restrição sobre $P(B)$.

No contexto de apostas, a noção de probabilidade condicional pode ser intuitivamente construída de forma análoga ao que foi feito para probabilidades “não condicionais”, usando bilhetes de loteria. Probabilidade condicional, como lá, também deve ser entendida como uma taxa de compra ou venda de bilhetes de loteria. Só que aqui, estes bilhetes de loteria são formados sobre *eventos condicionais*.

Sejam A e B dois eventos de Θ . De Finetti(1937) define um evento

condicional³ $A|B$ de Θ como um evento que descreve a ocorrência de A , sob a premissa da ocorrência de B . Esse é um tipo de evento que assume três valores lógicos - verdadeiro, falso ou vazio - de acordo com

$$A|B = \begin{cases} \text{verdadeiro} & \text{se } A \text{ e } B \text{ ocorrerem} \\ \text{falso} & \text{se } \overline{A} \text{ e } B \text{ ocorrerem} \\ \text{vazio} & \text{se } \overline{B} \text{ ocorrer.} \end{cases}$$

Um bilhete de loteria baseado em um evento condicional premiará seu dono com uma quantia S , se $A|B$ for verdadeiro; se $A|B$ for falso, o dono do bilhete não receberá nada; e, no caso em que $A|B$ for vazio, a aposta será cancelada e o dinheiro pago pelo bilhete será devolvido ao comprador.

Representando por $p_{E_1|E_2}$ as suas medidas numéricas de incerteza sobre os eventos $E_1|E_2$ e considerando que estes *bilhetes condicionais* também são negociados, obtemos o seguinte resultado:

TEOREMA 2 - Se as suas medidas numéricas de incerteza sobre os eventos (condicionais ou não)⁴ de Θ formam um conjunto coerente, então para quaisquer dois eventos E_1 e E_2 de Θ ,

³Apesar deste tipo de evento não ter sido mencionado antes, ele não deve representar qualquer novidade, uma vez que toda medida numérica de incerteza que Você estabelece depende de todo o conhecimento acumulado por Você até o momento de expressá-la, isto é, está condicionada à sua história H (de Finetti(1974)). Como em um instante fixo de tempo, a sua história não sofre alterações, ao mencionar-se, aqui, p_E , pretende-se que seja entendido $p_{E|H}$. O H foi apenas subentendido, não desconsiderado. Mais sobre isso pode ser encontrado em Coerência Temporal de M. Goldstein (1985).

⁴A extensão da definição de coerência para bilhetes condicionais é óbvia.

$$p_{E_1 \cap E_2} = p_{E_1|E_2} p_{E_2}.$$

DEMONSTRAÇÃO : Admita que o “bookmaker” permite que Você adquira também bilhetes condicionais. Considere que para os bilhetes formados sobre os eventos E_2 , $E_1 \cap E_2$ e $E_1|E_2$, Você atribua, respectivamente, as seguintes taxas, p_{E_2} , $p_{E_1 \cap E_2}$ e $p_{E_1|E_2}$.

Suponha, por absurdo, que na sua opinião $p_{E_1 \cap E_2} \neq p_{E_1|E_2} p_{E_2}$.

Considere, então, que para estes bilhetes o “bookmaker” determine os prêmios S_2 , S_{12} e S , respectivamente, enquanto que para os demais bilhetes ele estabeleça um prêmio nulo.

O sistema de todos os seus possíveis ganhos será então:

$$\begin{cases} G_1 = (1 - p_{E_1 \cap E_2})S_{12} + (1 - p_{E_2})S_2 + (1 - p_{E_1|E_2})S & \text{se } E_1 \text{ e } E_2 \text{ ocorrerem} \\ G_2 = -p_{E_1 \cap E_2}S_{12} + (1 - p_{E_2})S_2 - p_{E_1|E_2}S & \text{se } \overline{E_1} \text{ e } E_2 \text{ ocorrerem} \\ G_3 = -p_{E_1 \cap E_2}S_{12} - p_{E_2}S_2 & \text{se } \overline{E_2} \text{ ocorrer.} \end{cases}$$

E o determinante gerado por este sistema ,

$$\begin{vmatrix} 1 - p_{E_1 \cap E_2} & 1 - p_{E_2} & 1 - p_{E_1|E_2} \\ -p_{E_1 \cap E_2} & 1 - p_{E_2} & -p_{E_1|E_2} \\ -p_{E_1 \cap E_2} & -p_{E_2} & 0 \end{vmatrix} = p_{E_1 \cap E_2} - p_{E_1|E_2} p_{E_2},$$

é diferente de zero, por hipótese de absurdo.

Assim, o sistema dos possíveis ganhos apresenta uma única e bem definida solução. E, com isso, Você estará dando ao “bookmaker” a oportunidade de vencê-lo sempre. Isto contradiz a hipótese de coerência! $\square$

Perceba que a relação

$$p_A = p_{A|\Theta}, \text{ para todo evento } A,$$

surge como corolário do teorema 2. Se esta relação, que de resto é intuitiva, fosse definida desde o início, não seria necessária a consideração de dois “tipos” de bilhetes, como foi feito.

De uma combinação direta dos teoremas 1 e 2 surge, como consequência, o *Teorema de Bayes* que é um resultado fundamental para a Inferência Estatística. Se Você é coerente, Você necessariamente usa o teorema de Bayes para obter suas probabilidades posteriores, pois ele é o único mecanismo de atualização de probabilidades que gera distribuições posteriores coerentes. Este é de fato um dos primeiros motivos para usar-se modelos probabilísticos em Inferência. Além disso, consegue-se perceber a partir dos teoremas apresentados que os Axiomas de Kolmogorov são condições que caracterizam opiniões coerentes e, em resumo, permitem concluir que suas medidas numéricas de incerteza, desde que sejam estabelecidas coerentemente, são probabilidades.

CAPÍTULO DOIS

O PROBLEMA DA DECISÃO EM GRUPO

1. INTRODUÇÃO

O que se nota no dia-a-dia é que na maior parte das vezes existem não uma, mas várias pessoas responsáveis por decisões. Em experimentos clínicos, por exemplo, é comum uma junta médica opinar sobre os tratamentos a serem aplicados a cada paciente.

Como se sabe, quando existe apenas uma pessoa envolvida, o problema de decisão admite uma solução “racional”. Tal solução racional está fundamentada em maximização de utilidade esperada. Agora, quando há um grupo de pessoas envolvidas no problema, ainda não está estabelecida uma teoria normativa que permita ao grupo tomar decisões racionalmente. Mais que isso, em 1951, Arrow provou a impossibilidade de um grupo de pessoas tomar decisões racionalmente.

Considerando-se um grupo constituído por n Bayesianos e sendo $\mathcal{D}$ e $\mathcal{C}$, respectivamente, os conjuntos de decisões e conseqüências disponíveis comuns aos membros do grupo, Arrow(1951) demonstra que ainda que cada membro k do grupo possua uma relação de ordem coerente R_k sobre $\mathcal{D}$, é impossível determinar a partir de $\{R_k : k = 1, 2, \dots, n\}$ uma relação de ordem R_G sobre $\mathcal{D}$ para o grupo que satisfaça as seguintes propriedades (julgadas desejáveis para tais relações racionais):

1. Racionalidade Coletiva - todo conjunto de n relações de ordem individuais $R_1, \dots, R_n$ deve ter uma relação de ordem R_G (do grupo) associada a ele.
2. Princípio de Pareto - toda relação de ordem do grupo R_G deve implementar

a unanimidade, ou seja, se todas as pessoas do grupo preferem x a y , o grupo (visto como uma única pessoa cuja relação de ordem seria R_G) também deve preferir x a y .

3. Não-Ditadura - a relação de ordem do grupo R_G não deve ser sempre a ordenação de alguma pessoa. Não devem ser adotadas invariavelmente pelo grupo as preferências de uma única pessoa do grupo.

4. Transitividade - a relação de ordem para o grupo R_G bem como as relações de ordem individuais $R_1, \dots, R_n$, devem ser transitivas.

O resultado acima é conhecido como o *Teorema da Impossibilidade de Arrow*.

Mesmo tendo sido provado que não, Lindley(1985) acha que todo problema de decisão em grupo tem solução uma vez que se pode escolher um líder para o grupo responsável pela decisão conjunta. Fazendo isto, o problema ficaria reduzido a um problema de decisão uni-pessoal para o qual existe um procedimento normativo de solução. Mas, como escolher este líder? A menos de uma ditadura, tal escolha também será um problema de decisão em grupo.

A ausência de um paradigma coerente para tomada de decisões em grupo faz com que surjam várias questões. O que é racionalidade para um grupo? Como deve ser feita a quantificação numérica das opiniões e preferências de um grupo? Faz sentido combinar as opiniões (preferências) de cada membro do grupo em única distribuição de probabilidade (função de utilidade)? O que é otimalidade para um grupo? O que é a opinião de um grupo?

Neste capítulo, tenta-se mostrar algumas discussões realizadas com o intuito de responder tais questões. Para isso, na seção 2 serão apresentadas

algumas características importantes para a avaliação de probabilidades individuais e para a formação de uma distribuição-consenso para o grupo. Na seção 3, serão apresentadas as maneiras usuais de se formar a distribuição-consenso a partir das opiniões de cada pessoa do grupo. E, por fim, na seção 4 será apresentada a proposta de de Finetti(1953) para a obtenção de uma “opinião” consenso para o grupo onde as opiniões iniciais dos membros do grupo são mantidas.

Será mencionada várias vezes a divisão (proposta por French (1985)) para os problemas de decisão em grupo. Por isso, antes de iniciar a próxima seção cabe descrever tal divisão . Para French, os problemas de decisão em grupo devem ser abordados segundo a tarefa que o grupo deve executar:

1. *Problema do Expert* - o grupo de pessoas possui um líder (que pode ou não pertencer ao grupo) que será responsável pela decisão do grupo.

2. *Problema de Decisão em Grupo* (propriamente dito) - todo o grupo é responsável pela decisão , não havendo apenas uma pessoa responsável pela decisão do grupo.

3. *Problema do Livro-Texto* - aqui não existe um problema de decisão pré-definido. As opiniões são solicitadas para que outras pessoas as usem em algum problema futuro.

2. ALGUMAS CARACTERÍSTICAS E PROPRIEDADES

Nesta seção serão mostradas algumas características e propriedades que podem ser satisfeitas por probabilidades individuais e distribuições-consenso.

Por uma *distribuição-consenso*, Φ , deve-se entender uma distribuição formada por uma “combinação ” de distribuições de probabilidade. Em outras palavras, $\Phi = \Phi(F_1, \dots, F_n)$ é a distribuição-consenso formada a partir das

funções de distribuição individuais $F_1, \dots, F_n$ das pessoas $1, \dots, n$ do grupo.

Por exemplo, considere $w_1, \dots, w_n$ como pesos atribuídos às opiniões das pessoas $1, \dots, n$, respectivamente, e considere um estado da natureza θ .

EXEMPLO 1- Uma distribuição-consenso formada a partir de uma combinação linear das distribuições individuais é a função de distribuição Φ dada por:

$$\Phi_{(F_1, \dots, F_n)}(\theta) = \sum_{i=1}^n w_i F_i(\theta)$$

com $\sum_{i=1}^n w_i = 1$ e $w_i \geq 0, \forall i = 1, \dots, n$.

EXEMPLO 2 - Φ é uma distribuição-consenso formada a partir de uma combinação logarítmica de densidades individuais se

$$\Phi_{(f_1, \dots, f_n)}(\theta) = \frac{\prod_{i=1}^n [f_i(\theta)]^{w_i}}{\int_{\Theta} \prod_{i=1}^n [f_i(\theta)]^{w_i} d\theta}$$

com $w_i \geq 0, \forall i = 1, \dots, n$.

As formas de agrupamento de probabilidades acima serão apresentadas com mais detalhes na seção 3.

2.1. MARGINALIZAÇÃO

Considere que Você está interessado em uma particular sub- σ -álgebra $\mathcal{F}$ de eventos de Θ sobre a qual cada um (de um grupo de pessoas) deverá expressar sua opinião . E considere que há interesse em formar um consenso destas opiniões . Este consenso poderá ser obtido de duas formas:

1. As opiniões são combinadas, primeiramente, em uma distribuição-consenso sobre toda a σ -álgebra original $\mathcal{A}$ e então a distribuição marginal sobre

a sub- σ -álgebra $\mathcal{F}$ é tomada.

2. Cada pessoa do grupo expressa sua opinião sobre os eventos da sub- σ -álgebra $\mathcal{F}$ (ou seja, fornece sua distribuição marginal sobre esta sub- σ -álgebra) e uma distribuição-consenso marginal é formada a partir destas.

Talvez seja intuitivo que as distribuições-consenso obtidas pelas duas formas descritas acima coincidam. Quando tal coincidência ocorre para toda sub- σ -álgebra $\mathcal{F}$ de $\mathcal{A}$, dizemos que a distribuição-consenso satisfaz a propriedade de Marginalização .

EXEMPLO - Considere um conjunto de estados da natureza $\Theta = \{\theta_1, \theta_2, \theta_3\}$ e considere $\mathcal{A} = \mathcal{P}(\Theta)$ como a σ -álgebra original de sub-conjuntos de Θ .

Seja $\mathcal{F} = \{\phi, \{\theta_1\}, \{\theta_2, \theta_3\}, \Theta\}$ uma sub- σ -álgebra de $\mathcal{A}$. Duas pessoas coerentes são ouvidas e suas opiniões sobre Θ são as seguintes:

	θ_1	θ_2	θ_3
<i>pessoa 1</i>	p_{11}	p_{12}	p_{13}
<i>pessoa 2</i>	p_{21}	p_{22}	p_{23}

Se a forma de agrupamento linear de probabilidades for admitida e se o peso w_i é atribuído à opinião da pessoa i ($i = 1, 2$), obtém-se:

- 1) Marginalizando primeiro e depois agrupando

	ϕ	$\{\theta_1\}$	$\{\theta_2, \theta_3\}$	Θ
<i>pessoa 1</i>	0	p_{11}	$p_{12} + p_{13}$	1
<i>pessoa 2</i>	0	p_{21}	$p_{22} + p_{23}$	1

Neste caso, a distribuição-consenso do grupo é caracterizada por:

$$p_G(\{\theta_2, \theta_3\}) = w_1(p_{12} + p_{13}) + w_2(p_{22} + p_{23}) \quad (I)$$

2) Por outro lado, primeiro agrupando e depois marginalizando

	θ_1	θ_2	θ_3
grupo	$w_1p_{11} + w_2p_{21}$	$w_1p_{12} + w_2p_{22}$	$w_1p_{13} + w_2p_{23}$

Neste caso, obtém-se para o grupo

$$p_G(\{\theta_2, \theta_3\}) = (w_1p_{12} + w_2p_{22}) + (w_1p_{13} + w_2p_{23}). \quad (II)$$

Assim, como (I) e (II) são iguais, verifica-se que a propriedade de Marginalização é válida para o agrupamento linear.

Em 1981, McConway demonstrou um teorema estabelecendo que a coincidência entre os dois métodos vistos só ocorre se a distribuição-consenso for função apenas das probabilidades individuais atribuídas a um determinado evento e do próprio evento em questão . Formalmente,

TEOREMA : Seja Ψ uma distribuição-consenso. Ψ satisfaz a propriedade de Marginalização se, e somente se, existe uma função $G : \mathcal{A} \times [0, 1]^n \mapsto [0, 1]$ tal que, para qualquer evento A pertencente à σ -álgebra $\mathcal{A}$,

$$\{\Psi(\pi_1, \dots, \pi_n)\}(A) = G(A, \pi_1(A), \dots, \pi_n(A))$$

onde π_i , $i = 1, \dots, n$ é a probabilidade atribuída ao evento A pelo i -ésimo indivíduo do grupo.

A demonstração deste teorema pode ser encontrada em McConway (1981-pg 411).

Um argumento intuitivo para que tal coincidência não seja tão desejável é dado por Lindley (1985). Para ele, a distribuição-consenso obtida pelo primeiro método será feita com mais informação, uma vez que esta distribuição é estabelecida através das opiniões individuais sobre a σ -álgebra completa e, no segundo método, a distribuição-consenso é formada a partir das opiniões individuais apenas sobre a sub- σ -álgebra $\mathcal{F}$ - o que representa menos informação.

2.2. BAYESIANIDADE EXTERNA

Bayesianidade Externa é também uma propriedade à primeira vista razoável no processo de formação de uma opinião-consenso (ou melhor, de uma distribuição-consenso) a partir das opiniões emitidas por cada membro de um grupo. Para se falar em Bayesianidade Externa, é necessário que todas as pessoas do grupo tenham acesso à mesma informação X que guarda conexão com o estado da natureza θ através de um modelo estatístico e também concordem sobre a *verossimilhança* dos dados, isto é, o grupo deve ter uma função de verossimilhança comum. E partindo disto, a propriedade de Bayesianidade Externa diz que para uma pessoa externa ao grupo, este parece agir como um único ente bayesiano ao atualizar suas opiniões. Isto significa dizer que as distribuições-consenso posteriores obtidas pelos métodos que seguem devem coincidir.

1. Cada membro do grupo atualiza sua distribuição a priori (coerentemente) através do Teorema de Bayes e, então, a distribuição-consenso

posterior é formada a partir das distribuições posteriores individuais;

2. Forma-se uma distribuição-consenso a priori a partir das distribuições a priori individuais e então esta é atualizada através do Teorema de Bayes, originando a distribuição-consenso posterior.

EXEMPLO 1 - Considere que duas pessoas coerentes são convidadas a opinar sobre um estado da natureza $\theta \in \Theta$. Elas tiveram acesso à mesma informação x sobre θ e, mais ainda, compartilham de uma mesma função de verossimilhança $L_x(\theta) = f(x|\theta)$ para θ , isto é, $f(x|\cdot)$ é unânime para o grupo.

Suponha que a distribuição a priori da pessoa i para θ é a densidade $p_i(\theta)$, $i = 1, 2$. Se for adotado o método de agrupamento logarítmico para a obtenção da distribuição-consenso, tem-se :

- 1) Fazendo-se primeiro o agrupamento, a seguinte distribuição a priori é obtida:

$$p_G(\theta) = \frac{p_1^{w_1}(\theta)p_2^{w_2}(\theta)}{\int_{\Theta} p_1^{w_1}(\theta)p_2^{w_2}(\theta)d\theta}.$$

Usando a verossimilhança unânime como válida para o grupo e atualizando $p_G(\cdot)$ através da fórmula de Bayes, obtem-se a seguinte distribuição-consenso posterior:

$$p_G(\theta|x) = \frac{p_1^{w_1}(\theta)p_2^{w_2}(\theta)f(x|\theta)}{\int_{\Theta} p_1^{w_1}(\theta)p_2^{w_2}(\theta)f(x|\theta)d\theta}. \quad (I)$$

- 2) Por outro lado, fazendo-se primeiro a atualização e depois o agrupamento logarítmico, $p_G(\theta|x)$ é assim obtida:

$$p_G(\theta|x) = \frac{p_1^{w_1}(\theta)p_2^{w_2}(\theta)f^{w_1+w_2}(x|\theta)}{\int_{\Theta} p_1^{w_1}(\theta)p_2^{w_2}(\theta)f^{w_1+w_2}(x|\theta)d\theta}. \quad (II)$$

E desde que $w_1 + w_2 = 1$, a propriedade de Bayesianidade Externa será satisfeita pelo processo de agrupamento logarítmico.

EXEMPLO 2 - Com as mesmas considerações do exemplo 1 e assumindo agora o processo de agrupamento linear, verifica-se que:

1) Pelo primeiro método de obter-se distribuições-consenso posteriores a seguinte distribuição-consenso é encontrada:

$$p_G(\theta|x) = w_1p_1(\theta|x) + w_2p_2(\theta|x), \quad (I)$$

onde $p_i(\cdot|x)$ é a distribuição posterior de θ para a pessoa i , $i = 1, 2$.

2) Pelo método 2, partindo-se da distribuição-consenso do grupo

$$p_G(\theta) = w_1p_1(\theta) + w_2p_2(\theta),$$

a seguinte distribuição-consenso posterior é obtida:

$$p_G(\theta|x) = \frac{w_1p_1(\theta)f(x|\theta) + w_2p_2(\theta)f(x|\theta)}{w_1 \int_{\Theta} p_1(\theta)f(x|\theta)d\theta + w_2 \int_{\Theta} p_2(\theta)f(x|\theta)d\theta}. \quad (II)$$

Como as expressões (I) e (II) não coincidem em Θ , a propriedade de Bayesianidade Externa não é satisfeita pelo processo de agrupamento linear.

Para Genest e Zidek (1986), a propriedade de Bayesianidade Externa é bastante razoável quando a distribuição-consenso é formada pelo grupo. Se,

no entanto, existe uma pessoa externa ao grupo responsável pela formação da distribuição-consenso, a propriedade de Bayesianidade Externa faz menos sentido, uma vez que a obtenção de novos dados (informações) poderia e deveria mudar a avaliação que esta pessoa faz sobre o conhecimento de cada membro do grupo, alterando inclusive o peso dado a cada opinião. E com isso, a distribuição-consenso formada através do método 1 seria alterada. No método 2, como a distribuição-consenso é formada **antes** de se receber as novas informações, estas serviriam apenas para a atualização da distribuição-consenso, sem proporcionar qualquer alteração no peso das opiniões dos membros do grupo. McConnway (French (1985) - discussão) e Lindley (1985) também compartilham desta mesma opinião. Lindley vai ainda mais longe, afirmando que a propriedade de Bayesianidade Externa não passa de uma falácia, uma vez que ele julga essencial a presença de um líder no grupo, responsável pela combinação das probabilidades individuais.

No exemplo 2, foi visto que a forma de agrupamento linear não possui a propriedade de Bayesianidade Externa. Mais adiante, neste mesmo capítulo, será provado que se for permitida uma mudança (feita pelo líder) nos pesos individuais com a aquisição de novos dados tal propriedade será satisfeita por esta nova forma de agrupamento linear de opiniões. Este fato confirma a afirmação feita por Genest e Zidek(1986).

Um outro ponto vulnerável da propriedade de Bayesianidade Externa é a exigência de uma mesma verossimilhança para o grupo. Esta exigência é bastante forte e faz com que tal propriedade perca um pouco de seu apelo.

2.3. CORRELAÇÃO ENTRE OPINIÕES DOS MEMBROS DO GRUPO

Este é um outro fator de importância na discussão sobre a formação de uma distribuição-consenso. Quando uma pessoa é solicitada a expressar sua opinião sobre um determinado evento, ela o fará baseada nas informações que possui sobre tal evento. Normalmente, o que ocorre quando se considera um grupo de pessoas, é que as pessoas envolvidas no grupo poderão ter acesso às mesmas informações e, portanto, raramente suas opiniões não estarão correlacionadas.

2.4. CONHECIMENTO ESPECÍFICO

Deve-se levar em conta que as pessoas não têm o mesmo grau (quantidade) de conhecimento sobre todos os fatos. Elas podem ter (e, provavelmente, terão) formas diferentes de assimilar informações sobre os diversos fatos, de acordo com o seu grau de especificidade no assunto em questão. Esta diversidade deve ser considerada por distribuições-consenso razoáveis, ou seja, toda distribuição-consenso razoável deve incorporar e ponderar o conhecimento específico dos membros do grupo. Saber o grau de conhecimento de uma pessoa sobre um determinado assunto possibilita atribuir um peso à sua opinião. Supõe-se que quanto mais alguém entenda de um assunto, mais informativa será sua opinião e portanto mais peso esta opinião deverá ter ao se formar o consenso de opiniões.

Mais sobre correlação entre opiniões e conhecimento específico pode ser encontrado em French(1985) e Genest e Zidek(1986).

2.5. SINCERIDADE

Quando a opinião de uma pessoa sobre um determinado evento é solicitada, espera-se que a pessoa expresse, com base em suas informações, uma distribuição de probabilidade de forma sincera.

Existem várias maneiras de fazer com que uma pessoa expresse suas opiniões sinceramente. Uma forma operacional freqüentemente usada com este objetivo baseia-se em *Regras de Penalidade Próprias*. Estas regras são assim denominadas por garantirem que a perda esperada para uma pessoa será mínima se esta anunciar sua verdadeira opinião. Dentre tais regras, a mais comum é a *Regra de Brier*. Esta regra determina que uma pessoa deve ser penalizada com a perda quadrática. Isto é, considerando um evento A , com probabilidade P de ocorrer para certa pessoa e considerando que a probabilidade anunciada para a ocorrência de A pela mesma pessoa é p , a perda quadrática é dada por:

$$L(p) = \begin{cases} (1 - p)^2 & \text{se } A \text{ ocorrer} \\ p^2 & \text{se } A \text{ não ocorrer} \end{cases}$$

O objetivo de uma pessoa coerente é ter perda esperada mínima. Quando a Regra de Brier é considerada, a perda esperada é dada pela seguinte expressão:

$$E[L(p)] = (1 - p)^2 P + p^2 (1 - P),$$

que será minimizada por $p = P$, isto é, a perda esperada será mínima quando a probabilidade anunciada por uma pessoa for igual à sua opinião sincera.

Na mesma situação anterior, admita agora que a pessoa será penalizada segundo uma *Regra de Penalidade Imprópria*. Por exemplo, considere a seguinte regra de penalidade

$$L(p) = \begin{cases} |1 - p| & \text{se } A \text{ ocorrer} \\ |p| & \text{se } A \text{ não ocorrer} \end{cases}$$

Neste caso, a perda esperada é dada por:

$$E[L(p)] = P + (1 - 2P)p$$

Aqui, nota-se que se $P \geq 0.5$, a perda esperada será mínima quando $p = 1$ e se $P \leq 0.5$, a perda mínima ocorrerá quando $p = 0$, ou seja, qualquer que seja o valor anunciado P em $(0, 1)$, a pessoa será estimulada a mentir, quando for escolhida uma regra de penalidade imprópria para penalizá-la.

Kadane e Winkler (1988) mostraram que as declarações de probabilidade - obtidas por processos de elicitación como regras de penalidade e loterias - estão relacionadas com a função de utilidade da pessoa não só para os ganhos inerentes ao processo de elicitación mas também para outros interesses que a pessoa possa ter no evento em questão. Com isso eles mostraram que as regras de penalidades próprias são bons mecanismos para induzir uma pessoa a ser sincera apenas nos casos onde esta possuir uma função de utilidade linear.

Um outro bom processo para elicitare probabilidades que induz sinceridade foi utilizado, aqui, no capítulo um - as loterias. Para este processo, Kadane e Winkler estabeleceram que a pessoa coerente é sincera não só nos casos em que a sua função de utilidade é linear mas também naqueles em que a pessoa

não possui outros interesses sobre o evento (além do possível lucro a ser pago pelo bilhete de loteria formado a partir deste) e sua função de utilidade é não-linear.

Além destes, outro resultado interessante foi mostrado por estes autores. Tanto para regras de penalidades quanto para loterias, se não existem outros interesses sobre o evento além dos lucros inerentes ao processo de elicitación e se estes lucros são quantias pequenas (que se aproximam de zero) a melhor estratégia é, ainda, anunciar probabilidades sinceras.

Em problemas de decisão em grupo, a sinceridade também é fortemente desejada, pelo simples fato de que a decisão do grupo pode ser radicalmente alterada se apenas um dos membros do grupo não for sincero. Por exemplo, no problema do expert, se um dos membros do grupo sabe, a princípio, o problema a ser resolvido com o uso de suas opiniões, ele poderá expressá-las de forma mais conveniente, de modo que na formação da opinião-consenso, o líder dê maior ou menor peso à sua opinião, comprometendo a decisão a ser tomada pelo grupo.

Este problema é estudado de forma mais abrangente por Bayarri e DeGroot (1988). Neste trabalho eles exibem estratégias ótimas para um membro do grupo que deseja ter seu peso a posteriori maximizado, considerando diversas situações e funções de utilidade. E além disto, exibem a única função de utilidade para a qual a melhor estratégia é sempre a sinceridade da pessoa. Para isto, eles admitem que os pesos a posteriori de cada membro do grupo são obtidos através do método proposto por Roberts (1965) - (ver expressão (II), sub-seção 3.1, neste capítulo).

2.6. PROBABILIDADES OU DADOS?

Segundo de Finetti (1972), a probabilidade só existe como tal para a pessoa que a estabelece, ou melhor, probabilidade só existe na mente da pessoa (**PROBABILIDADE NÃO EXISTE**) - é a descrição (pessoal) de sua incerteza sobre um determinado acontecimento. Para outra pessoa, tal probabilidade poderá ser também a sua probabilidade, apenas no caso muito especial em que ambas possuam a mesma opinião. Se não, qualquer probabilidade alheia deve ser considerada como um dado a mais. Assim é, por exemplo, no problema do expert onde o líder é responsável pela obtenção de uma distribuição-consenso que incorpore todo o conhecimento dos membros do grupo. Para o líder, as probabilidades dos membros do grupo são, mais do que nunca, dados para os quais ele deve ter uma verossimilhança. Na prática, a determinação desta verossimilhança é uma tarefa muito complicada de ser realizada (é a opinião sobre a opinião de uma outra pessoa). Para os casos onde as distribuições de probabilidades individuais são discretas, Lindley(1985) mostra em detalhes uma sugestão de verossimilhança.

2.7. CALIBRAÇÃO

Calibração é um critério usado para avaliar probabilidades subjetivas. Baseia-se no relacionamento de probabilidades anunciadas com a frequência relativa observada de ocorrência de eventos.

Esta comparação da probabilidade subjetiva anunciada com frequência relativa observada parece, à primeira vista, contradizer a visão bayesiana sobre probabilidade, uma vez que para seguidores desta corrente, a definição de probabilidade freqüentista é absurda. No entanto, o Critério de Calibração

não faz a associação fundamental entre probabilidade e frequência sobre a qual repousa a corrente freqüentista. Para os seguidores desta corrente, a probabilidade de um evento é interpretada como o limite da frequência relativa de ocorrências do evento em uma seqüência (infinita) de realizações independentes e em condições idênticas de um mesmo experimento. Diga-se de passagem, esta “definição” é confusa sobre vários aspectos. Além do fato óbvio de uma pessoa não ter a eternidade para ficar, por exemplo, lançando moedinha, ela também não terá condições de realizar todos os lançamentos em condições perfeitamente idênticas uma vez que pelo menos o tempo estará variando de lançamento para lançamento. Outro ponto vulnerável desta definição é a consideração “experimentos independentes”. Esta consideração gera um ciclo vicioso, pois independência é definida a partir de probabilidades. Apesar de todos estes aspectos confusos, o mais grave é, que para freqüentistas, probabilidade é, por exemplo, um atributo físico da moedinha.

No Critério de Calibração, o relacionamento de probabilidade com frequência relativa se dá como segue. Considere eventos $E_1, E_2, \dots, E_m$ aos quais Você atribui probabilidades de ocorrência, coerentemente. Considere a subcoleção C_p de eventos aos quais Você atribuiu uma probabilidade comum p . O Critério de Calibração compara p à frequência relativa f dos eventos de C_p que realmente ocorrem. E, assim, define-se uma pessoa como *bem-calibrada* se p e f coincidem, isto é, se $p = f$, para todo valor que p assume. (Por motivos práticos, é claro que C_p deve incluir todos os eventos cuja probabilidade anunciada é convenientemente próxima de p).

Analogamente, se $X_1, X_2, \dots, X_r$ são quantidades aleatórias contínuas, o Critério da Calibração compara a proporção f destas quantidades aleatórias

que tiveram seus resultados observados à esquerda do p -ésimo quantil de suas respectivas distribuições de probabilidades, com p . E, da mesma forma, se $f = p$, a pessoa que fez tais atribuições de probabilidade é chamada bem-calibrada.

Na prática, o que se espera é que de um conjunto de eventos, tomando-se apenas aqueles cuja probabilidade anunciada é, por exemplo, 0,3, aproximadamente 30% destes eventos efetivamente ocorram. E, quando se trata de quantidades aleatórias contínuas, espera-se, por exemplo, que aproximadamente 50% das quantidades aleatórias assumam valores à esquerda da mediana de suas respectivas densidades anunciadas.

Considere, agora, todos os valores de p para os quais $C_p \neq \phi$. A cada um destes p 's associe o valor correspondente da frequência relativa f . Isto é, a cada valor p_i anunciado por Você tal que $C_{p_i} \neq \phi$ associe o valor f_i correspondente. Plotando-se todos os pontos (p_i, f_i) em um gráfico, a curva assim obtida é definida como *curva de calibração*. Quanto mais esta curva aproxima-se da diagonal, mais bem calibrada será a pessoa.

Note-se que o conjunto de eventos (ou quantidades aleatórias) considerado na definição de Critério de Calibração não é necessariamente formado por eventos “realizados em condições idênticas”. Por exemplo, podemos ter E_1 = ocorrência de chuva em São Paulo, E_2 = o time do Botafogo ganhar o próximo campeonato nacional e E_3 = o sistema monárquico ser apontado como o melhor no plebiscito do ano que vem. Cabe lembrar, por exemplo, que os eventos E_1 = ocorrência de chuva amanhã e E_2 = ocorrência de chuva depois-de-amanhã, embora tratem do mesmo “assunto”, não são eventos realizados sobre condições idênticas (“eventos são únicos” - Bruno de Fi-

netti. Veja também de Finetti(1937), pag.80). Qualquer conexão percebida entre os eventos - por exemplo, permutabilidade - é rigorosamente subjetiva (Wechsler(1993)).

Kadane e Lichtenstein (1982) estabeleceram duas situações críticas para Calibração . A primeira delas é a situação com feedback - isto é, cada vez que Você faz uma atribuição de probabilidade, Você sabe o resultado de todos os eventos para os quais Você tinha feito uma atribuição de probabilidade anteriormente. E a segunda situação é a sem-feedback - isto é, Você faz um grande número de atribuições de probabilidade antes de saber o resultado de qualquer evento para o qual Você atribuiu uma probabilidade.

Dentro de cada um destes contextos existem vários aspectos importantes sobre calibração a serem explorados. O principal deles é o estabelecimento de uma relação entre Coerência e Calibração . Para prosseguir neste estudo, será lembrado que as pessoas ao atribuírem suas probabilidades são sinceras (e coerentes).

1) Situação Com Feedback

Em situações envolvendo atribuições de probabilidade com feedback, dois teoremas foram estabelecidos - um por Pratt(1962) e outro por Dawid(1982) - mostrando uma relação entre Coerência e Calibração.

Pratt considerou uma seqüência de quantidades aleatórias contínuas $X_1, X_2, X_3, \dots$ e, definindo

$$S_n = \#\{1 \leq i \leq n : X_i < F_{X_i}^{-1}(p)\},$$

demonstrou que $\forall p \in [0, 1]$ vale o seguinte:

TEOREMA 1- Supondo as considerações acima e feedback, se Você conhece $x_1, \dots, x_n$ quando Você fizer uma atribuição de probabilidade sobre X_{n+1} ,

a) Para Você S_n terá distribuição Binomial com parâmetros n e p .

b) Você está quase certo que $\frac{S_n}{n} \rightarrow p$, quando $n \rightarrow \infty$.

Em sua parte (a), o teorema obtém a distribuição do número de sucessos em n ensaios relativos às n quantidades aleatórias supostas, apenas, contínuas. (Uma pessoa bem-calibrada deve ter, para todo p , $S_n \approx np$). Na parte (b), o teorema diz que Você é *assintoticamente* bem-calibrado com (sua) probabilidade 1. Ou seja, Você tem certeza que é assintoticamente bem-calibrado.

Perceba-se que a virtude da boa calibração assintótica requer, dentre outras coisas, uma seqüência infinita de quantidades aleatórias, enquanto que o número de quantidades aleatórias usadas para a construção de uma curva de calibração real é sempre finito. (Para uma pessoa real esta sua curva de calibração poderá ou não estar próxima da diagonal em $[0, 1] \times [0, 1]$).

Um resultado análogo e um pouco mais forte foi estabelecido e demonstrado por Dawid (1982) considerando uma seqüência infinita de eventos, $E_1, E_2, \dots$. A estes eventos Você atribuiu probabilidades $p_1, p_2, \dots$, com feedback. Considerando Y_i como sendo o indicador do evento E_i , Dawid considera um conjunto-teste que é uma sub-seqüência de eventos escolhidos por Você antes de Você conhecer seus respectivos resultados. Os elementos do seu conjunto-teste podem ser determinados ou todo de uma vez, ou seqüencialmente - Você pode decidir pela inclusão ou não de E_i no conjunto-teste no mais tardar após conhecer o resultado de E_{i-1} .

Definindo ξ_i como o indicador de inclusão ou não do evento E_i no conjunto-

teste, $\eta_k = \sum_{i=1}^k \xi_i$ como sendo o número de eventos dentre os k primeiros que pertencem ao conjunto-teste, $\lambda_k = \eta_k^{-1} \sum_{i=1}^k \xi_i Y_i$ como a proporção destes que realmente ocorreram, e $\pi_k = (\sum_{i=1}^k \xi_i p_i) \eta_k^{-1}$ a média das probabilidades atribuídas aos eventos do conjunto-teste, Dawid demonstra o seguinte teorema.

TEOREMA 2- Supondo as considerações acima, se $\eta_k \rightarrow \infty$ Você está quase certo que $\lambda_k - \pi_k \rightarrow 0$, qualquer que seja o conjunto-teste.

O teorema 2 diz que a diferença entre a frequência relativa de ocorrências de eventos do conjunto-teste e a média amostral das probabilidades a eles atribuídas é, para Você, assintoticamente nula, para qualquer escolha de um conjunto-teste. Ou seja, Você tem certeza que a longo prazo as frequências relativas estarão próximas das suas probabilidades (médias).

Para perceber melhor a notação usada por Dawid, observe a seguinte tabela-exemplo:

E_i	E_1	E_2	E_3	E_4	E_5	
p_i	p_1	p_2	p_3	p_4	p_5	
Y_i	0	1	1	0	0	...
ξ_i	0	1	0	1	1	...
η_k	0	1	1	2	3	...
λ_k	-	1	1	1/2	1/3	...
π_k	-	p_2	p_2	$\frac{p_2+p_4}{2}$	$\frac{p_2+p_4+p_5}{3}$	...

As demonstrações dos teoremas estão omitidas porque o interesse maior aqui, está em suas interpretações. Em ambos os teoremas, a conclusão é que

em situações com feedback, coerência implica a certeza de boa calibração assintótica.

Estes teoremas são bastante fortes uma vez que a única exigência que fazem sobre a estrutura de probabilidade é feedback (e coerência).

Quando há feedback, Você conhece a sua curva de calibração atual antes de atribuir probabilidade para o próximo evento. Talvez seja natural que Você se deixe influenciar por uma má curva de calibração no momento da atribuição - Você poderia até alterar sua próxima probabilidade, ou seja, Você poderia *recalibrar-se*. Por exemplo, se conhecidos $e_1, e_2, \dots, e_{m-1}$ e a sua curva de calibração distar muito da diagonal, Você poderá usar este fato para encontrar uma transformação $\gamma(E_m; p)$ significando que Você pensava em atribuir p para E_m mas irá anunciar a “probabilidade” recalibrada γ

Se sua opinião inicial sobre o evento E_m é coerente, para que a coerência prevaleça, a transformação γ encontrada por Você deverá ser, necessariamente a identidade. Senão, Você estará atribuindo duas probabilidades distintas ao mesmo evento, o que é absurdo. Tal mudança seria incoerente.

Numa tentativa de entender esta situação, alguns autores apresentam as seguintes explicações.

Kadane (comentário; Dawid (1982)) argumenta que os teoremas apresentam resultados assintóticos, o que na prática significa apenas que uma pessoa real deve se recusar a recalibrar-se por achar que, com certeza, no **limite** ela é bem-calibrada. Isto, por sinal, lança a questão sobre o número de eventos que devem ser considerados (qual número é suficientemente grande) para que má-calibração real possa ser diagnosticada. Lindley (comentário; Dawid (1982)) sugere que a informação de má-calibração induz uma mudança brusca

na história H da pessoa (H como definido por de Finetti(1974)) e assim, nenhum princípio de coerência estaria sendo infringido. Para Goldstein (1985), toda atribuição de probabilidade é feita em algum instante de “tempo” e a exigência de coerência é feita em cada um destes instantes isoladamente. Por exemplo, a fórmula de Bayes valeria apenas temporalmente, não descrevendo mudanças bruscas de opinião. E, realmente, não parece razoável o uso da fórmula de Bayes para recalibração - seria uma mudança de opinião provocada pela (má calibração da) própria opinião. E, evidentemente, probabilidades anunciadas como sintomas de probabilidades (subjettivas) desconhecidas são inaceitáveis sob o ponto de vista operacionalista (de Finetti(1949)).

É claro que enquanto método de interpretação de probabilidades alheias, recalibração é perfeitamente razoável para Você, no entanto, auto-recalibração é incoerente. Apesar disto, se em um experimento clínico um médico descobre que está mal-calibrado na atribuição de seus diagnósticos, a recalibração (mesmo incoerente) parece bastante razoável.

O exemplo abaixo sugere que até mesmo na situação em que probabilidades alheias são consumidas por um chefe, uma curva bem calibrada pode não ser atrativa, sendo preferível para o chefe recalibrar probabilidades que geram uma curva má-calibrada.

EXEMPLO- Duas pessoas são indagadas sobre suas probabilidades para o evento E = “ocorrência de chuva em São Paulo”, seqüencialmente, por um período de 100 dias. Tais atribuições serão feitas com feedback, coerentemente, por ambos. Suponha que para a pessoa 1, a probabilidade do evento E ocorrer seja 0,5 em todos os 100 dias e, que para a pessoa 2, essa probabilidade seja, alternadamente, 0,6 e 0,4. Depois de transcorridos os 100 dias,

verifica-se que a proporção de dias chuvosos foi 50%. E, mais ainda, durante os 100 dias choveu dia sim, dia não. Portanto, a pessoa 1 é perfeitamente calibrada e a pessoa 2 é mal-calibrada. No entanto, se fossem ambas penalizadas pela Regra de Brier, a perda sofrida pela pessoa 1 a cada dia seria 0.25; enquanto para a pessoa 2 a perda diária seria 0,16 ($< 0,25$).

Obviamente, o chefe preferirá contratar a pessoa 2 e recalibrar os valores 0,4 e 0,6, a contratar a pessoa 1, apesar desta possuir calibração perfeita.

Alguns autores como Kadane e Lichtenstein (1982), Dawid (1982) e DeGroot (1980) são de opinião que nas situações onde os eventos são permutáveis é mais fácil ser bem calibrado ou melhorar-se a calibração. No entanto, esta questão também é ambígua, na medida que permutabilidade é tão subjetiva quanto as probabilidades em questão.

2) Situações Sem Feedback

Aqui também, estabeleceu-se uma relação entre coerência e calibração. Só que, como será visto mais adiante, neste caso, algumas suposições mais fortes que as exigidas nos teoremas 1 e 2 serão necessárias. A ausência de feedback dificilmente será encontrada em situações práticas (reais).

Antes de enunciar o teorema 3 que relaciona coerência e calibração na ausência de feedback, será apresentado o

TEOREMA FUNDAMENTAL DE BRUNO DE FINETTI

Seja $A_1, A_2, \dots$ uma seqüência de eventos permutáveis. Então, existe uma distribuição F em $[0, 1]$, tal que a probabilidade de que exatamente k dentre n destes eventos ocorram possui a seguinte representação:

$$\binom{n}{k} \int_0^1 p^k (1-p)^{n-k} dF(p), \quad (I)$$

quaisquer que sejam k e n .

E, mais ainda, se $A_1, A_2, \dots$ são eventos permutáveis para Você, a proporção $\frac{k}{n}$ dos eventos que efetivamente ocorrem converge com probabilidade (subjativa) 1. E a opinião que Você tem sobre o limite é a distribuição F da representação acima.

Uma das conseqüências do teorema acima é que se os eventos $A_1, A_2, \dots$ são na sua opinião permutáveis, isto significa, em particular, que para Você eles têm a mesma probabilidade de ocorrência igual a

$$P(A_1) = P(A_2) = \dots = \int_0^1 p dF(p),$$

que também é o valor esperado da distribuição F .

Este teorema garante que Você está certo que $\frac{k}{n}$ convergirá. O limite da convergência é desconhecido para Você e, portanto, aleatório. (Veja Wechsler(1993) e Barlow(1991) para interpretação e comentários sobre o Teorema da Representação de Bruno de Finetti.)

Se $f(p)$ concentra-se em torno de sua média, então Você espera estar bem-calibrado, por que daí

$$\lim_{n \rightarrow \infty} \frac{k}{n} \approx \int_0^1 p dF(p)$$

que é a probabilidade de ocorrência do evento, atribuída por Você. Caso não haja concentração em torno da média, a calibração não deve ser esperada

por Você.

Assim, o Teorema da Representação permite concluir que permutabilidade e coerência apenas não são suficientes para a certeza de boa-calibração assintótica. Um exemplo tornará este fato mais claro.

EXEMPLO: Considere uma cidade onde as pizzarias estarão ou todas abertas, ou todas fechadas domingo à noite. Seja E_i o indicador de que a i -ésima pizzeria estará aberta ($i = 1, 2, \dots$). Para Você, a probabilidade de estarem todas abertas é s . É claro que os eventos E_i são permutáveis e que $P(E_i) = s$, $i = 1, 2, \dots$. No entanto, Você tem certeza que será mal-calibrado, pois certamente a sua curva de calibração será ou o ponto $(s, 1)$ ou o ponto $(s, 0)$.

Portanto, permutabilidade, apenas, não basta para a certeza de boa-calibração assintótica.

Kadane e Lichtenstein (1982) enunciaram o seguinte teorema que estabelece a condição para a certeza de boa-calibração assintótica em situações sem feedback.

TEOREMA 3- Seja $A_1, A_2, \dots$ uma seqüência de eventos permutáveis para Você. Seja S_n o número de tais eventos que ocorrerão dentre os primeiros n e seja $p = P(A_i)$. Então, Você está quase certo que

$$\frac{S_n}{n} \rightarrow p$$

se, e somente se, a seqüência é independente (para Você).

O que se pode concluir deste teorema é que se na sua opinião uma seqüência de eventos é permutável e independente, então Você tem certeza de

ser uma pessoa assintoticamente bem-calibrada, em condições sem feedback.

A questão da independência, portanto, é fundamental para a calibração em situações sem feedback. E basta que para Você exista um elo entre os eventos para que não haja independência, na sua opinião. Por exemplo, se um médico sabe da presença de um vírus altamente contagioso em uma certa região, certamente o seu diagnóstico para um paciente será dependente da proporção de pacientes da mesma região que acusaram a presença do vírus. (O simples fato de uma mesma pessoa fazer atribuições de probabilidade para eventos diversos provoca uma dependência entre estas atribuições, na opinião de uma outra pessoa.)

Note-se que o julgamento de independência é equivalente a Você possuir uma priori F degenerada. E, assim, raramente Você espera ser bem-calibrado assintoticamente em situações sem feedback.

Para Kadane e Lichtenstein (1982), pesquisa sobre a calibração pode auxiliar na compreensão dos importantes processos conjugados de avaliação e expressão de opiniões, mais do que ser uma meta a ser perseguida por uma pessoa coerente. Calibração não oferece qualquer evidência sobre a coerência de uma pessoa e vimos que auto-recalibração é incoerente. Por outro lado, recalibração pode ser útil para uma pessoa que atribui pesos às opiniões alheias.

3. FORMANDO DISTRIBUIÇÕES-CONSENSO

Considere que um grupo composto de n pessoas é convidado a expressar suas opiniões sobre um determinado estado de natureza $\theta \in \Theta$. Considere,

ainda, que cada pessoa do grupo expressa suas opiniões sobre θ coerentemente através de distribuições de probabilidades e que a opinião da i -ésima pessoa do grupo é a distribuição p_i ($i = 1, \dots, n$). Nesta seção, pretende-se discutir maneiras de combinar-se essas probabilidades, p_i , (que representam as opiniões de cada membro do grupo) originando uma “distribuição de probabilidade” - p_G - que poderia ser interpretada como a opinião-consenso do grupo. (Veja também a seção 2)

3.1. AGRUPAMENTO LINEAR DE PROBABILIDADES

Uma das formas mais simples de agrupar-se probabilidades é através de uma combinação linear destas probabilidades. Considere que à opinião de cada pessoa do grupo seja atribuído um determinado peso w_i . É suposto que para cada pessoa i , w_i seja não-negativo e também que $\sum_{i=1}^n w_i = 1$. Um agrupamento linear de probabilidades é um operador p_G da seguinte forma:

$$p_G(\theta) = \sum_{i=1}^n w_i p_i(\theta) \quad (I).$$

Operadores determinados desta forma apresentam algumas características que devem ser comentadas. Por exemplo, este tipo de operador preserva a unanimidade, isto é, se para todo membro i do grupo, $p_i(\theta) = \alpha$, o valor do operador p_G no ponto θ também será α . É lógico pensar que se todos os membros do grupo compartilham da mesma opinião sobre a chance de ocorrência de θ , no consenso, o grupo deverá manter esta opinião. Em particular quando $p_i(\theta) = 0$ para todo $i = 1, \dots, n$, $p_G(\theta) = 0$.

Outra característica deste tipo de agrupamento é que ele possui a propriedade de Marginalização (ver seção 2.1). Isto é equivalente a dizer que o valor

do operador p_G no ponto θ em (I) é função apenas das opiniões dos membros do grupo sobre θ e do próprio θ (fato provado por McConway(1981)).

Os pesos, mesmo quando atribuídos por um líder externo, são meros reflexos do conhecimento específico de cada membro do grupo, não existindo para sua determinação qualquer forma normativa. Além disso, são elementos fixados a priori não sendo permitidas mudanças após o recebimento de novas informações. Esta característica dos pesos faz com que o operador linear de opiniões não possua a propriedade de Bayesianidade Externa, como foi exemplificado na seção 2.2.

Alguns autores (French (1980), Wechsler(1989), Roberts (1965)) têm estudado a mudança de opinião sobre a opinião dos outros. Eles mostraram que, recebidas novas informações x , o peso dado à i -ésima pessoa do grupo, w_i , pode ser atualizado da seguinte forma:

$$w_{i|x} = \frac{w_i f_i(x)}{\sum_{i=1}^n w_i f_i(x)}, \quad (II)$$

onde $f_i(\cdot)$ é a função preditiva para X anunciada pela i -ésima pessoa do grupo. Esta forma de atualização nos pesos também foi considerada no trabalho de Bayarri e DeGroot(1988) citado na seção 2.5.

O que se pode perceber na expressão (II) é que a pessoa i será premiada (ou penalizada) em seu peso $w_{i|x}$ de acordo com o valor de sua função preditiva no ponto x efetivamente observado. Se, por acaso, a pessoa i atribuiu um valor baixo à chance de ocorrência de x observado, $f_i(x)$, o peso inicial dado à sua opinião sofrerá uma redução por esta fórmula - o que é razoável de esperar-se.

Tal forma de atualização dos pesos se assemelha à fórmula de Bayes, mas não são equivalentes. Aqui, os w_i 's não são probabilidades, são simplesmente pesos.

Um fato a mais que deve ser observado nesta fórmula é que não é necessária a presença de um líder para proceder à atualização dos pesos. O próprio grupo pode responsabilizar-se por isso.

Admitindo-se que esta mudança nos pesos w_i possa ser feita, o seguinte resultado pode ser demonstrado.

TEOREMA: Admitindo-se a expressão (II), o agrupamento linear de opiniões torna-se Externamente Bayesiano.

DEMONSTRAÇÃO: Admita que o valor da opinião preditiva da pessoa i para x seja $p_i(x)$ e que $f(x|\theta)$ seja a função de verossimilhança para θ , sobre a qual todos os membros do grupo concordam (veja seção 2.2). A função preditiva da pessoa i sobre x pode ser escrita da seguinte forma:

$$p_i(x) = \int_{\Theta} f(x|\theta)p_i(\theta)d\theta \quad (III).$$

Para provar-se que o agrupamento linear de opiniões com pesos revisados é Externamente Bayesiano, tem-se que mostrar que

$$\sum_{i=1}^n w_{i|x} p_i(\theta|x) \propto [\sum_{i=1}^n w_i p_i(\theta)] f(\theta|x), \quad (IV)$$

ou seja, que a razão entre os termos de (IV) não depende de θ .

Desenvolvendo-se o lado esquerdo da expressão (IV), tem-se

$$\sum_{i=1}^n w_i p_i(\theta|x) = \sum_{i=1}^n \frac{w_i p_i(x)}{[\sum_{i=1}^n w_i p_i(x)]} p_i(\theta|x),$$

que, pelo o teorema de Bayes, torna-se igual a

$$\frac{1}{[\sum_{i=1}^n w_i p_i(x)]} \sum_{i=1}^n w_i p_i(x) \frac{p_i(\theta) f(x|\theta)}{\int_{\Theta} p_i(\theta) f(x|\theta) d\theta}.$$

Assim,

$$\sum_{i=1}^n w_i p_i(\theta|x) = \frac{f(x|\theta)}{[\sum_{i=1}^n w_i p_i(x)]} \sum_{i=1}^n w_i p_i(\theta)$$

é proporcional ao lado direito da expressão $[\sum_{i=1}^n w_i p_i(\theta)] f(\theta|x)$.

Portanto, verifica-se que a igualdade de densidades proposta na equação (IV) é satisfeita e conclui-se que com a mudança Bayesiana nos pesos, o agrupamento linear de opiniões torna-se Externamente Bayesiano. $\square$

Uma outra característica das formas de agrupamento linear é que elas não preservam a independência. Considere, por exemplo, dois eventos E_1 e E_2 que são considerados independentes por todos os membros do grupo - $p_i(E_1 \cap E_2) = p_i(E_1)p_i(E_2)$ para todo $i = 1, \dots, n$. Para o consenso, tem-se então

$$p_G(E_1 \cap E_2) = \sum_{i=1}^n w_i p_i(E_1 \cap E_2) = \sum_{i=1}^n w_i p_i(E_1) p_i(E_2),$$

que é igual a $p_G(E_1)p_G(E_2)$ apenas quando $w_i = 1$ para algum i . Ou seja, somente quando se tem uma ditadura - que é um método insatisfatório para agrupamentos de opiniões.

3.2. AGRUPAMENTO LOGARÍTMICO DE PROBABILIDADE

Uma outra forma bastante comum de agrupar-se probabilidades dando origem a uma distribuição-consenso é fazer uma combinação logarítmica de tais probabilidades. Um operador p_G é um agrupamento logarítmico de opiniões (probabilidades) quando p_G é da seguinte forma:

$$p_G(\theta) = \frac{\prod_{i=1}^n [p_i(\theta)]^{w_i}}{\int_{\Theta} \prod_{i=1}^n [p_i(\theta)]^{w_i} d\theta}$$

onde os pesos w_i ($i = 1, \dots, n$) são tais que a integral do denominador seja finita.

Nesse tipo de agrupamento de opiniões basta que um único membro do grupo considere nula a chance de ocorrência de θ para que o operador-consenso se anule. Embora a preservação de zero seja uma propriedade desejável em um operador-consenso, a forma drástica como ela ocorre aqui enfraquece este tipo de agrupamento. O fato de apenas uma opinião nula anular a opinião-consenso pode fazer com que p_G não seja uma boa representação da opinião do grupo.

Uma outra característica que desfavorece a forma de agrupamento logarítmico é a maneira como os pesos w_i são determinados. Assim como no agrupamento linear, os pesos são novamente fruto do conhecimento específico de cada membro do grupo e não possuem uma forma normativa de determinação. Além disso, tanto aqui quanto no agrupamento linear, as probabilidades alheias são arbitrariamente consideradas como probabilidades e não como dados, contrariando a visão de de Finetti(1972). (Ver também seção 2.6). Em particular, isto representa um problema quando existe um

líder responsável pelo agrupamento das opiniões.

Apesar destas, várias outras características favorecem tal forma de agrupamento. Ao contrário do que ocorria com o operador linear, o operador logarítmico preserva a independência de E_1 e E_2 , quando todos os membros do grupo acham que E_1 e E_2 são independentes. E, além disso, quando $\sum_{i=1}^n w_{ij} = 1$ tal agrupamento é Externamente Bayesiano (ver exemplo 1, seção 2.2) sem que haja uma atualização dos pesos (via fórmula de Bayes).

Genest (1984) mostrou que o operador logarítmico é o único operador Externamente Bayesiano que é função apenas das opiniões de cada membro do grupo. Isto é, Genest mostrou que o operador logarítmico, p_G , é o único operador Externamente Bayesiano que satisfaz a seguinte identidade em θ :

$$p_G(\theta) = kF(p_1(\theta), \dots, p_n(\theta)) \quad (I)$$

onde k é uma constante independente de θ . Esta condição foi interpretada por ele como uma versão do *Princípio da Verossimilhança* para fórmulas de agrupamento de probabilidades no sentido que “ exceto por uma constante de normalização independente de θ , é requerido da densidade da distribuição-consenso em θ depender apenas das densidades individuais no valor real da quantidade não observada, e não das densidades de valores que poderiam ter sido obtidos mas não foram” (Genest(1984) - pg1101), isto é, resultados não observados não devem desempenhar qualquer papel na determinação de uma distribuição-consenso num ponto particular θ . .

3.3. A DISTRIBUIÇÃO-CONSENSO DE MORRIS DEGROOT

A proposta feita por Morris DeGroot(1974) para a determinação de uma distribuição-consenso para o grupo está baseada no diálogo (que poderia estender-se infinitamente) entre os n membros de um grupo até que surja concordância. É claro que tal concordância pode não ser atingida. O método sugerido por DeGroot possui a forma de um agrupamento linear de opiniões, mas nele é descrita, explicitamente, a forma de atingir o consenso e de determinar os pesos usados.

Suponha que a opinião de cada membro i ($i = 1, \dots, n$) do grupo sobre uma quantidade desconhecida θ é a distribuição de probabilidade F_i . $F_1, \dots, F_n$ descrevem a incerteza que os indivíduos $1, \dots, n$, respectivamente, possuem sobre θ com base em toda a informação que possuem. Esta informação pode ser diferente para cada membro do grupo.

Uma pessoa i pode mudar sua opinião sobre θ , ou através de novas informações adquiridas sobre θ , ou então por saber a opinião sobre θ dos outros $(n - 1)$ membros do grupo. Nesse último caso, a distribuição atualizada do i -ésimo membro do grupo, F_{i1} , é formada do seguinte modo:

$$F_{i1} = \sum_{j=1}^n w_{ij} F_j, \quad (I)$$

onde w_{ij} representa o peso atribuído à pessoa j pela pessoa i , antes que esta tome conhecimento da distribuição F_j da pessoa j . É suposto, como antes, que $w_{ij} \geq 0$ para todo $i, j = 1, \dots, n$ e que $\sum_{j=1}^n w_{ij} = 1$, para todo i .

Essa pessoa i atribui peso w_{ij} à sua própria opinião. É claro que se ela atribui à sua opinião um peso muito grande, F_{i1} será muito próxima de F_i .

Suponha, agora, que o indivíduo i , já sabendo que sua distribuição de probabilidade sobre θ mudou de F_i para F_{i1} , foi informado que todos os outros $(n - 1)$ membros do grupo também mudaram suas distribuições de probabilidade. Ele também é informado sobre quais são essas distribuições. Considere ainda que não existe motivo para ele mudar sua opinião sobre os demais membros do grupo, de forma que os pesos que ele atribui a cada colega permanecem os mesmos. Com isso, a nova distribuição para θ da i -ésima pessoa do grupo é:

$$F_{i2} = \sum_{j=1}^n w_{ij} F_{j1}, \quad i = 1, \dots, n. \quad (II)$$

O processo continua desse modo até que a mudança nas distribuições não altere mais as distribuições obtidas anteriormente, isto é, até que para algum k , $F_{ik} = F_{i(k-1)}$, $\forall i = 1, \dots, n$.

Para escrever o processo em forma matricial, considere a matriz $\mathcal{W}$ formada pelos pesos w_{ij} ($i, j = 1, \dots, n$), onde w_{ij} representa o elemento da i -ésima linha e j -ésima coluna da matriz $\mathcal{W}$. $\mathcal{W}$, definida assim, é uma matriz estocástica. Considere, ainda, que $F^{(k)} = (F_{1k}, \dots, F_{nk})^t$ é o vetor das distribuições dos n membros do grupo, após k modificações e que $F_{i0} = F_i$. Com isso, as expressões (I) e (II) acima podem ser escritas, respectivamente, como:

$$F^{(1)} = \mathcal{W}F^{(0)}$$

$$F^{(2)} = \mathcal{W}F^{(1)} = \mathcal{W}^2 F^{(0)}.$$

Generalizando,

$$F^{(k)} = \mathcal{W}^k F^{(0)}. \quad (III)$$

As revisões nas distribuições são feitas com o intuito de chegar-se a uma distribuição que seja comum a todos os membros do grupo, ou seja, até que F_{ik} independa de i . Neste caso, $F_{i(k+1)} = F_{ik} = F^*$. Pode-se dizer que a distribuição-consenso é conseguida se, e somente se, todos os componentes do vetor $F^{(k)}$ convergirem para um limite comum, quando $k \rightarrow \infty$. Em outras palavras, o consenso é atingido se, e somente se, para todo $i = 1, \dots, n$,

$$\lim_{k \rightarrow \infty} F_{ik} = F^*, \quad (IV)$$

onde F^* é também uma função de distribuição.

Supondo verdadeira a expressão (III), o seguinte resultado pode ser demonstrado.

TEOREMA 1: Quando (III) é verdadeira, o consenso é atingido se, e somente se, existe um vetor $\Pi = (\pi_1, \dots, \pi_n)$ tal que, para todo $i, j = 1, \dots, n$,

$$\lim_{k \rightarrow \infty} w_{ij}^{(k)} = \pi_j,$$

onde $w_{ij}^{(k)}$ representa o elemento da i -ésima linha e j -ésima coluna da matriz $\mathcal{W}^k$.

DEMONSTRAÇÃO: Considere que $F^{(k)} = \mathcal{W}^k F^{(0)}$, isto é, $F_{ik} = \sum_{j=1}^n w_{ij}^{(k)} F_j$, para $i = 1, \dots, n$.

Suponha que exista um vetor $\Pi = (\pi_1, \dots, \pi_n)$ tal que, para todo $i, j = 1, \dots, n$,

$$\lim_{k \rightarrow \infty} w_{ij}^{(k)} = \pi_j.$$

Como $F_{ik} = \sum_{j=1}^n w_{ij}^{(k)} F_j$, temos para todo $i = 1, \dots, n$,

$$\begin{aligned} \lim_{k \rightarrow \infty} F_{ik} &= \lim_{k \rightarrow \infty} \sum_{j=1}^n w_{ij}^{(k)} F_j \\ &= \sum_{j=1}^n F_j \lim_{k \rightarrow \infty} w_{ij}^{(k)} \\ &= (\text{por hipótese}) = \sum_{j=1}^n F_j \pi_j, \text{ que não depende de } i. \end{aligned}$$

Portanto, podemos definir F^* por

$$F^* = \sum_{j=1}^n F_j \pi_j,$$

que é, efetivamente, uma função de distribuição, pois $\pi_j = \lim_{k \rightarrow \infty} w_{ij}^{(k)} \geq 0$ e $\sum_{j=1}^n \pi_j = \lim_{k \rightarrow \infty} \sum_{j=1}^n w_{ji}^{(k)} = \lim_{k \rightarrow \infty} 1 = 1$ (o produto de matrizes estocásticas é estocástico)

Logo, o consenso é alcançado.

Por outro lado, se para todo $i = 1, \dots, n$, $\lim_{k \rightarrow \infty} F_{ik} = F^*$, tem-se:

$$\begin{aligned} F^* &= \lim_{k \rightarrow \infty} \sum_{j=1}^n w_{ij}^{(k)} F_j \\ &= \sum_{j=1}^n F_j \lim_{k \rightarrow \infty} w_{ij}^{(k)}, \text{ para } i = 1, 2, \dots, n. \end{aligned}$$

O sistema acima implica

$$\lim_{k \rightarrow \infty} w_{ij}^{(k)} = \pi_j, \quad j = 1, \dots, n.$$

Logo, se o consenso for atingido, existe um vetor $\Pi = (\pi_1, \dots, \pi_n)$ tal que

$$\lim_{k \rightarrow \infty} w_{ij}^{(k)} = \pi_j, \quad \forall i, j = 1, \dots, n.$$

□

Como a matriz $\mathcal{W}$ é uma matriz estocástica, ela poderia ser interpretada como a matriz de transição em um passo de uma Cadeia de Markov (CM) com n estados de transição e com probabilidades de transição estacionárias. Por esse motivo, os teoremas limites da teoria de CM podem ser utilizados aqui, fornecendo condições para a convergência (e, conseqüentemente, o consenso) e uma forma para determinar-se os pesos.

As condições para a convergência ficam estabelecidas pelo seguinte teorema da teoria de CM.

TEOREMA 2: Quando todos os estados recorrentes da matriz finita de Markov $\mathcal{W}$ comunicam-se e são todos aperiódicos, existe um vetor Π tal que

$$\lim_{k \rightarrow \infty} w_{ij}^{(k)} = \pi_j, \quad i, j = 1, \dots, n.$$

A demonstração desse teorema pode ser encontrada, por exemplo, em Çinlar (1975).

Por esse teorema pode-se concluir que se todos os estados correspondentes à matriz $\mathcal{W}$ são recorrentes e aperiódicos, o consenso é atingido com

distribuição-consenso dada por

$$F^* = \sum_{j=1}^n \pi_j F_j. \quad (V)$$

Se o consenso é atingido e uma vez que $F^* = \sum_{j=1}^n \pi_j F_j$ será a distribuição comum a todos os membros do grupo, o vetor Π pode ser determinado prontamente, como mostra o teorema abaixo (também demonstrado em Çinlar(1975)).

TEOREMA 3: Suponha que um consenso é atingido e que $F^* = \sum_{j=1}^n \pi_j F_j$ é a distribuição comum aos membros do grupo. Então, $\Pi = (\pi_1, \dots, \pi_n)$ é o único vetor tal que

$$\Pi W = \Pi$$

$$\sum_{j=1}^n \pi_j = 1$$

$$\pi_j \geq 0, \quad \forall j = 1, \dots, n.$$

Com isso, DeGroot conseguiu uma distribuição-consenso para o grupo que, embora seja um agrupamento linear de opiniões, possibilita determinar o peso das opiniões dos membros do grupo (aqui os pesos representam a influência, a longo prazo, da pessoa j em um contexto de “trocas permanentes de opinião”).

Algumas falhas podem ser notadas no método proposto por DeGroot para a obtenção de uma distribuição-consenso. Assim como nos métodos apresentados anteriormente, novamente as probabilidades alheias são vistas

como probabilidades e não como dados, contrariando a posição de de Finetti (ver seção 2.6). Se as probabilidades alheias fossem vistas como dados, cada pessoa do grupo deveria descrevê-las através de uma função de verossimilhança e, se coerentes, deveriam atualizar suas opiniões através da Fórmula de Bayes (ver cap.1) e não através da expressão (I) acima proposta por DeGroot(1974). As propriedades de Calibração e Sinceridade também não são levadas em conta e, por possuir a forma de um agrupamento linear, a propriedade de Bayesianidade Externa não pode ser verificada.

Nenhuma das formas para agrupar probabilidades exibidas aqui exigiam a presença de uma pessoa responsável pela formação do consenso. Essa pessoa, é claro, poderia existir. Basta que os pesos atribuídos às opiniões fossem dados por uma pessoa (pertencente ou não ao grupo) de forma que a sua opinião surgisse daí.

Existem muitas outras maneiras de formar-se uma “distribuição-consenso de probabilidade”. Aqui apenas algumas foram apresentadas. Referências importantes sobre formas de agrupar-se probabilidades são Genest e Zidek(1986), Morris(1983), Bordley (1982).

Apesar de todos estes métodos produzirem uma distribuição-consenso, um “ultimato” pode ser lançado a respeito da natureza deste tipo de “distribuição”. Pode-se facilmente provar que as “distribuições de probabilidade” p_G geradas pelos métodos consenso são efetivamente distribuições de probabilidade, isto é, satisfazem todas as propriedades exigidas de uma função de distribuição. No entanto, não são distribuições de probabilidade no sentido apresentado no capítulo um. Somente em casos particulares, como em ditaduras e em unanimidade, estas p_G 's não contrariam a visão de de Finetti -

probabilidade só existe enquanto opinião de alguma pessoa. Nos demais casos, as p_G 's são desprovidas de sentido uma vez que são apenas combinações de opiniões. (O grupo não tem opinião.)

O principal fator que depõe contra a existência de uma distribuição de probabilidade para caracterizar a “opinião” de um grupo é a ausência de um conceito de racionalidade coletiva (Arrow(1951)).

Uma tentativa interessante de definição de racionalidade para grupo é apresentada por Nau e McCardle (1990). Em linhas gerais, a proposta de Nau e McCardle consiste em estender a idéia de coerência individual de de Finetti (ver cap.1) para o contexto de jogos não-cooperativos. Jogos não-cooperativos são problemas de decisão conjunta em que as estratégias de cada jogador são loterias cujos prêmios dependem da ação incerta de seus oponentes. Além disso, considera-se a presença de um “bookmaker” que é adversário de todos os jogadores. Neste tipo de jogo, Nau e McCardle estabelecem alguns comportamentos que julgam serem desejáveis para os jogadores nas escolhas de suas estratégias:

i) Cada jogador rejeita estratégias que o exponham individualmente à compra de um “Dutch book” (coerência individual (ver cap.1)).

ii) Cada jogador rejeita estratégias que o exponham individualmente à compra de um “Dutch Book”, consideradas todas as estratégias não rejeitadas por seus oponentes no jogo (quando estes também seguem (i) e (ii)).

iii) Os jogadores rejeitam todas as estratégias que exponham o grupo à compra de um “Dutch Book”.

Uma estratégia escolhida de acordo com a situação (iii) é definida como *conjuntamente coerente*. Ou seja, uma estratégia escolhida pelo grupo é

conjuntamente coerente se nenhum “Dutch Book” pode ser realizado contra ela.

Uma axiomatização deste conceito - baseada na extensão dos axiomas de racionalidade individual de Savage(1954) para o cenário de multi-jogadores - pode ser encontrada em Nau e McCardle(1990).

Embora esta definição apresentada por Nau e McCardle seja matematicamente razoável por basear-se em axiomas simples que usam o apelo intuitivo de apostas, ela não consegue realizar a mesma função que coerência exerceu no capítulo um: garantir que uma “opinião” conjuntamente coerente seja descrita numericamente por uma distribuição de probabilidade.

Sem a pretensão de estabelecer uma distribuição de probabilidade (consenso), a seguinte seção exhibe um método para efetivamente determinar a “opinião” do grupo.

4. CAMPOS DE COINCIDÊNCIA DE OPINIÕES

Em 1953, Bruno de Finetti sugeriu um método diferente para “combinar” distribuições de probabilidades individuais. Contrária aos métodos de consenso, a proposta de de Finetti consiste não em encontrar uma nova distribuição de probabilidade (consenso), mas em construir um grupo de intervalos onde as opiniões individuais coincidam - os *campos de coincidência de opiniões*. Para isso de Finetti provou os seguintes fatos:

TEOREMA 1 - Sejam F e G duas funções de distribuição contínuas. Então, existe um intervalo (a, b) tal que

$$F(b) - F(a) = G(b) - G(a) = 1/2 \quad (I)$$

DEMONSTRAÇÃO: Seja m uma mediana de F e considere o conjunto de todos os intervalos (r, s) satisfazendo a relação $F(s) - F(r) = 1/2$. Este conjunto pode ser completamente percorrido começando-se com $r = -\infty$, $s = m$ e fazendo-se r e s crescerem continuamente até $r = m$ e $s = \infty$. Para cada um destes intervalos (r, s) , estará associado um valor para $G(s) - G(r)$ que variará, também continuamente, de $G(m)$ até $1 - G(m)$. Obviamente, ou $G(m) \leq 1/2 \leq 1 - G(m)$, ou $1 - G(m) \leq 1/2 \leq G(m)$, garantindo-se assim a existência de algum intervalo (a, b) tal que (I) seja satisfeita. $\square$

Repetindo-se a argumentação acima no conjunto de intervalos que formam o complementar de (a, b) , segue o seguinte:

TEOREMA 2 - Sejam F e G duas funções de distribuição contínuas. Para todo inteiro $n \geq 1$, existe uma união finita A_n de no máximo $2^n - 1$ intervalos tais que

$$\int_{A_n} dF = \int_{A_n} dG = 1 - 1/2^n.$$

Uma demonstração deste fato pode ser encontrada em de Finetti(1953).

Se F e G representam a opinião de duas pessoas sobre uma mesma quantidade aleatória contínua θ , estes teoremas garantem que é sempre possível encontrar campos onde estas opiniões coincidam. Isto é, para qualquer “nível” α em $(0, 1)$, é sempre possível encontrar uma união finita A_n (onde $n = n(\alpha)$) de intervalos tal que

$$p(\theta \in A_n) \geq 1 - \alpha$$

na opinião de **todos** os membros do grupo. Ou seja, este método tem a vantagem de permitir aos membros do grupo manter a sua opinião individual na atribuição de “probabilidade” do grupo.

Embora não seja uma distribuição de probabilidade completa, “um campo de coincidência de opiniões (uma união finita de intervalos de probabilidade) é a **única** atribuição de probabilidade genuína que um grupo pode fazer. E esta atribuição de probabilidade também pode ser vista como “objetiva” no sentido que **todos** os membros do grupo concordam sobre ela”(Wechsler(1989)).

CONCLUSÃO

Coerência é fundamental. Este fato pode ser percebido ao longo de todo este trabalho.

No capítulo um, fica provado através dos teoremas 1 e 2 que coerência é o alicerce de todo o Cálculo de Probabilidades. Estes teoremas garantem que a única forma de Você expressar opiniões coerentes é expressá-las como probabilidades. Por ser uma exigência mínima, fica claro também que a falta de coerência na expressão de opiniões faz com que Você aja ingenuamente, participando de jogos em que Você é o perdedor com certeza, sempre.

No capítulo dois, a inexistência de um conceito de coerência para grupos levou à impossibilidade de determinação de opinião de grupo (no mesmo sentido do capítulo um), justificando a inexistência de um procedimento normativo racional para solucionar o problema de decisão em grupo. Sob a condição de coerência, a única “opinião” genuína do grupo conseguida - os campos de coincidência de opiniões - não é uma distribuição de probabilidade completa. No entanto, este método de gerar “opinião” de grupo se comparado aos métodos de consenso (discutidos seção 3) é melhor por manter inalterada as opiniões individuais (coerentes).

Ainda no capítulo dois, alguns resultados aparecem relacionando coerência e calibração. O critério de calibração, em si, não viola a condição de coerência. Tanto em situações com feedback como em situações sem feedback fica provado que coerência (aliada a algumas outras condições) implica a certeza de boa-calibração assintótica. A violação da condição de coerência ocorre quando é considerada a questão de recalibração. Você recalibrar pro-

habilidades alheias é razoável e coerente uma vez que para Você estas probabilidades nada mais são que dados. Neste caso, o princípio de coerência é seguido pois a informação de má-calibração dos outros seria mais um dado que Você poderia usar na sua verossimilhança para opiniões alheias. A incoerência surge em casos de auto-recalibração. Neste caso, Você seria incoerente atribuindo duas probabilidades distintas a um mesmo evento. Talvez este problema deva ser focado sob a definição de Coerência Temporal de Goldstein(1985). No contexto apresentado, pessoas coerentes não se auto-recalibram por estarem certas de boa-calibração assintótica. A constatação de (sua própria) má-calibração pode ser vista como uma mudança brusca na “história” H de Você.

REFERÊNCIAS

Arrow, K.J.(1951)

Social Choice and Individual Values.

2ª edição, Yale University Press.

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.

Bayarri, M.J. e DeGroot, M.H.(1988)

Gaining Weight: A Bayesian Approach.

In Bayesian Statistics 3. Bernardo J.M., DeGroot M.H., Lindley D.V.,

Smith A.F.M. (editores). Oxford University Press. Pg. 25-44.

Bordley, R.F.(1982)

A multiplicative Formula for Aggregating Probability Assessments.

Management Science, Vol 28, No 10, Pg 1137-1148.

Çinlar, E.(1975)

Introduction to Stochastic Processes.

Prentice-Hall, Inc.

Dawid, A.P.(1982)

The Well-Calibrated Bayesian.

Journal of the American Statistical Association, Vol 77, No 379,

Pg 605-613.

de Finetti, B.(1931)

Sul Significato Soggettivo della Probabilità.

Fundamenta Mathematicae, Vol. 17,Pg. 298 - 329.

de Finetti, B.(1937)

Foresight: Its Logical Laws, Its Subjetive Sources.

Traduzido e reimpresso em “Studies in Subjetive Probability”.

Kyburg, H.E. e Smokler, H.E. (editores), 2ª edição, Pg. 55 - 118.

Krieger, New York, 1980.

de Finetti, B.(1949)

Le Vrai et le Probable

Dialectica, Vol 3, Pg 78-92.

de Finetti, B.(1953)

*Trasformazioni di Numeri Aleatori Atte a Far Coincidere Distribuzioni
Diverse*

Giornale Dell'Istituto Italiano Degli Attuari, XVI, Pg 51-57

de Finetti, B.(1972)

Probability, Induction and Statisics.

John Wiley & Sons. New york.

de Finetti, B.(1974)

Theory of Probability, Vol 1.

John Wiley & Sons. New York.

DeGroot, M.H.(1974)

Reaching a Consensus.

Journal of the American Statistical Association, Vol 69, No 345,
Pg 118-121.

DeGroot, M.H.(1980)

Improving Predictive Distributions.

In Bayesian Statistics 1. Bernardo J.M., DeGroot M.H., Lindley D.V.,
Smith A.F.M. (editores). Valencia: University Press. Pg 385-395.

Fishburn, P.C.(1986)

The Axioms of Subjective Probability.

Statistical Science, Vol. 1, No. 3, Pg 335-358.

French, S.(1980)

Updating of Belief in the Light of Someone Else's Opinion.

Journal of the Royal Statistical Society, Vol 143, Parte 1, Pg 43-48.

French, S.(1985)

Group Consensus Probability Distributions: A Critical Survey.

In Bayesian Statistics 2. Bernardo J.M., DeGroot M.H., Lindley D.V.,
Smith A.F.M. (editores). Elsevier Science, North Holland. Pg 183-202.

French, S.(1988)

Decision Theory: An introduction to the Mathematics of rationality.

Jonh Wiley & Sons, New York.

Genest, C.(1984)

A Characterization Theorem for Externally Bayesian Groups.

The Annals of Statistics, Vol 12, Pg 1100-1105.

Genest, C. e Zidek, J.V.(1986)

Combining Probability Distributions: A Critique and an Annotated bibliography.

Statistical Science, Vol 1, No 1, Pg 114-148.

Goldstein, M.(1985)

Temporal Coherence.

In Bayesian Statistics 2. Bernardo J.M., DeGroot M.H., Lindley D.V., Smith A.F.M. (editores). Elsevier Science, North Holland. Pg 231-248.

Hill, B.M.(1991)

Dutch Books, the Jeffreys-Savage Theory of Hypothesis Testing and Bayesian Reliability.

Trabalho a ser publicado nas atas do Congresso sobre "Reliability and Decision Making". Siena. Itália. Outubro 1990.

Kadane, J.B. e Lichtenstein, S.(1982)

A subjectivist View of Calibration.

Technical Report, No 233. Carnegie-Mellow University.
Department of Statistics.

Kadane, J.B. e Winkler, R.L.(1988)

Separating Probability Elicitation From Utilities.

Journal of the American Statistical Association, Vol 83, No 402,
Pg 357-363.

Lindley, D. V.(1985)

Reconciliation of Discrete Probability Distributions.

In Bayesian Statistics 2. Bernardo J.M., DeGroot M.H., Lindley D.V.,

Smith A.F.M. (editores). Elsevier Science, North Holland. Pg 375-390.

McConway, K.J.(1981)

Marginalization and Linear Opinion Pools.

Journal of the American Statistical Association, Vol 76, No 374,
Pg 410-414.

Morris, P.A.(1983)

An Axiomatic Approach to Expert Resolution.

Management Science, Vol 29, No 1, Pg 24-32.

Nau, R.F. e McCardle, K.F.(1990)

Coherent Behavior in Noncooperative Games

Journal of Economic Theory, Vol 50, No 2, Pg 424-444.

Pratt, J.M.(1962)

Must Subjective Probabilities be Realized as Relative Frequencies?.

Artigo não-publicado. Harvard University Graduate School of Business
Education.

Roberts, H.V.(1965)

Probabilistic Prediction.

Journal of the American Statistical Association, Vol 60, Pg 50-62.

Savage, L.J.(1954)

The Foundations of Statistics

Wiley, New York.

Wechsler, S.(1989)

On de Finettian Decision-Making.

Departamento de Estatística. Universidade da Califórnia - Berkeley.

Wechsler, S.(1993)

Exchangeability and Predictivism.

Erkenntnis - International Journal of Analytic Philosophy.

(A ser publicado).