

ELEMENTOS DE
INFERÊNCIA
BAYESIANA

CARLOS ALBERTO DE
BRAGANÇA PEREIRA
E
MARLOS AUGUSTO
GOMES VIANA

5° SIMPÓSIO
NACIONAL
DE PROBABILIDADE
E ESTATÍSTICA



ELEMENTOS DE INFERÊNCIA BAYESIANA

CARLOS ALBERTO DE BRAGANÇA PEREIRA
Instituto de Matemática e Estatística
UNIVERSIDADE DE SÃO PAULO

MARLOS AUGUSTO GOMES VIANA
Instituto de Matemática
UNIVERSIDADE FEDERAL DO RIO DE JANEIRO

"It is a fact that our degrees of confidence in a proposition habitually change when we make new observations or new evidence is communicated to us by somebody else, and this change constitutes the essential feature of all learning from experience. We must therefore be able to express it".

Harold Jeffreys

PREFÁCIO

Este trabalho pretende ser uma introdução ao Método Bayesiano e é destinado à comunidade de usuários da Estatística no Brasil. Essa introdução será feita através da discussão das soluções Bayesianas para alguns problemas clássicos da Estatística.

Como o objetivo é atingir uma platéia numerosa e também evitar que a compreensão das idéias seja afetada por dificuldades de natureza matemática, nos restringimos ao tratamento dos modelos estatísticos mais simples, isto é, aqueles que envolvem apenas famílias de distribuições discretas.

Apesar dessa restrição, o Capítulo 1 tem um caráter bastante geral e é nele que procuramos apresentar (e defender) o argumento Bayesiano na Inferência Estatística. Além disso, o Capítulo 1 contém também a caracterização de classes de distribuições importantes para a teoria e que contribuem também para a simplificação do método, do ponto de vista analítico.

O Capítulo 2 apresenta um sumário das propriedades das principais distribuições discretas (que constituem alguns modelos estatísticos) e das distribuições contínuas usadas como distribuições a priori (dos modelos discretos). Destaque especial é dado às distribuições que raramente são tratadas nos textos Clássicos tais como a de Dirichlet (ou Beta generalizada) e Dirichlet-Multinomial.

Os autores lamentam que a premência do tempo e as dificuldades impostas pela distância tenham impedido um maior número de reuniões entre si, que certamente teriam contribuído para uma maior unidade do conjunto. Esses fatores forçaram a uma divisão

de tarefas, ficando o primeiro autor responsável pela redação do Capítulo 1 e o segundo autor pela redação do Capítulo 3. Na redação do Capítulo 2 os autores participaram em conjunto, através de uma correspondência intensa entre eles.

Esperamos ter a oportunidade, numa edição posterior, de sanar algumas das falhas que certamente tenham resultado dessa divisão imposta pelas circunstâncias.

Os autores desejam agradecer a todos aqueles que direta ou indiretamente favoreceram a elaboração dessas notas. Deixamos consignada nossa gratidão aos nossos colegas incentivadores e à Comissão Organizadora do 5º SINAPE pela confiança em nos oferecer a grande responsabilidade de um minicurso.

Finalmente, dedicamos esse trabalho às nossas famílias que durante esse primeiro semestre de 1982 foram obrigadas a "suportar" com paciência e compreensão a falta de nossa companhia, inclusive nos fins de semana.

ÍNDICE

CAPÍTULO 1 - O ARGUMENTO BAYESIANO E A VEROSSIMILHANÇA	1
1.1 - Introdução	1
1.2 - O Conceito de Probabilidade	2
1.3 - Verossimilhança: O recipiente das informações experimentais	5
1.4 - Operação Bayesiana: Calibrando a incerteza através dos resultados experimentais	11
1.5 - Classes Conjugadas de Distribuições	16
CAPÍTULO 2 - DISTRIBUIÇÕES ESPECIAIS	19
2.1 - Introdução	19
2.2 - Distribuições Derivadas do Processo de Bernoulli	21
2.3 - A Distribuição Multinomial	25
2.4 - A Distribuição Hipergeométrica Multivariada	38
2.5 - A Distribuição Dirichlet	40
2.6 - A Distribuição Dirichlet-Multinomial	58
CAPÍTULO 3 - INFERÊNCIAS PARA MODELOS DISCRETOS	63
3.1 - Distribuições a priori, a posteriori	63
3.2 - Regiões de Credibilidade	69
3.3 - Estimadores de Bayes, Fator de Bayes	77
3.4 - Proporções, Tabelas de Contingências e Aplicações	81
3.5 - Predição Bayesiana	94

CAPÍTULO 1

O ARGUMENTO BAYESIANO E A VEROSSIMILHANÇA

"A new scientific truth does not triumph by convincing the opponents and making them see the light, but rather because its opponents eventually die, and a new generation grows up that is familiar with it"

Max Planck

1.1 INTRODUÇÃO

A palavra chave da Estatística é informação. Informação sobre o valor de uma quantidade ou estado da natureza de interesse, θ . É evidente que quando o verdadeiro valor de θ é conhecido, a informação é total e nesse caso a consulta a um probabilista ou a um estatístico seria desnecessária. O trabalho destes começa quando θ é desconhecido e a redução desse desconhecimento é de grande interesse. Ao estabelecermos que θ é desconhecido, queremos dizer que seu verdadeiro valor não é conhecido embora, a intensidade da incerteza sobre o valor de θ possa assumir diferentes graus.

Na concepção Bayesiana da Probabilidade e da Estatística, os diferentes graus de incerteza, sobre o verdadeiro valor de θ , são representados por diferentes afirmações ou modelos probabilísticos sobre θ . Consideramos, assim, que o trabalho do estatístico é o de construir modelos probabilísticos, de acordo com a Teoria das Probabilidades, e desenvolver (baseado naqueles modelos) métodos de inferência sobre o verdadeiro valor de θ . Esses métodos visam a redução da Sua incerteza sobre θ e para tanto, a observação do máximo possível (dentro de Suas restrições operacionais) de situações e dados relacionados (em Sua opinião) a θ , torna-se da maior relevância.

Na apresentação acima, queremos salientar o fato de que diferentes pesquisadores podem possuir diferentes graus de incerteza que correspondem a modelos distintos. Isto nos leva a afirmar que o único "modelo perfeito" é aquele que exclue qualquer in

certeza sobre o verdadeiro valor de θ e para isso esse valor deveria ser conhecido. Assim, falar em modelo perfeito seria absurdo, pois se θ fosse conhecido não seria necessária a consideração de modelos sobre θ .

1.2 O CONCEITO DE PROBABILIDADE

A seção anterior é baseada na visão subjetivista da teoria das probabilidades que considera a probabilidade de um evento A como a medida da incerteza sobre a ocorrência ou não de A . Como essa incerteza sobre A pode variar, de indivíduo para indivíduo, o valor da probabilidade de A , $P(A)$, também sofre o mesmo tipo de variação. Por esta razão, alguns autores consideram apenas probabilidades condicionais para salientar, por exemplo, que o valor que Você atribui à probabilidade de A é condicional ao Seu conhecimento (ou Sua incerteza) sobre a ocorrência de A . Bruno de Finetti, em seus trabalhos, escreve $P(A|H)$ em lugar de $P(A)$, para indicar que essa medida da incerteza é função do Seu "conhecimento" que por outro lado depende do Seu passado (ou história, representada por H).

Para fazer sentido a perspectiva subjetivista discutida acima, vamos considerar o mais familiar dos "experimentos" (uma definição precisa de experimento é apresentada na sequência); o lançamento de uma moeda objetivando a observação da face que ficará voltada para cima. Propositadamente escolhemos esse exemplo por ser êle muito utilizado na defesa da visão objetivista da teoria das probabilidades.

É razoável admitir-se que todos sabem que existem apenas dois resultados possíveis; cara e coroa. Acreditamos, contudo, que ao se afirmar que a probabilidade de qualquer dos dois resultados é $\frac{1}{2}$, deseja-se representar a falta total de informação, dos indivíduos envolvidos, sobre o resultado do experimento. Isto é, nenhum desses indivíduos possui controle sobre o lançamento e todos admitem que a moeda é "honesta". Para que o leitor admita alguma subjetividade na medida $p = P\{\text{cara}\} = \frac{1}{2}$, suponha que desse experimento participem Você e João. João vai lançar a moeda e pede a

Você para predizer o resultado. Como a Sua incerteza é total, Você pensaria em considerar o mesmo peso $p = \frac{1}{2}$ para os dois resultados possíveis e o mesmo se aplicaria a João. Aqui, tanto Você quanto João possuem a mesma incerteza, medida por $p = \frac{1}{2}$. Suponha que após o lançamento, João observa e esconde de Você o resultado, impedindo que Você ganhe alguma informação relevante. Como antes, João solicita que Você advinhe o resultado que ele já conhece. Agora, João tem certeza do resultado, enquanto Você, exatamente como no início, possui total incerteza e é obrigado a manter a mesma medida $p = \frac{1}{2}$. O importante é realçar que o valor $\frac{1}{2}$ para p não é consequência do resultado ser produzido pelo lançamento da moeda (da aleatorização, na linguagem clássica) e sim pela Sua incerteza total sobre o resultado. Note que mesmo após o lançamento Você manteve $p = \frac{1}{2}$ enquanto que João mudou a probabilidade do resultado obtido de $\frac{1}{2}$ para 1.

Neste trabalho, de uma maneira geral, as quantidades desconhecidas são representadas por letras maiúsculas, X, Y, \dots , e quando substituídas pelas minúsculas correspondentes, x, y, \dots , indicamos que passaram de desconhecidas para conhecidas. Isto é, o verdadeiro valor de X é x , de Y é y , e assim por diante. Para incluirmos os eventos como casos particulares, consideramos apenas as funções indicadoras da ocorrência destes. Por exemplo, $X=1$ se o evento A ocorre e $X=0$ em caso contrário. Para mantermos a linguagem pré-estabelecida, denominaremos as quantidades desconhecidas de quantidades (ou variáveis, quando seus verdadeiros valores são números reais) aleatórias. Para tornar clara a analogia com a Estatística "comum", letras gregas serão utilizadas para as quantidades desconhecidas de interesse; parâmetros na linguagem clássica da Estatística. Uma definição precisa de parâmetro será apresentada na próxima seção.

Em resumo, a posição Bayesiana é que a incerteza sobre o valor de qualquer quantidade X seja descrita probabilisticamente, ou seja por $P(X|H)$. Essa medida $P(\cdot|H)$ deve seguir as seguintes regras:

R1 - Regra da Convexidade: $0 \leq P(X|H) \leq 1$ e $P(H|H) = 1$.

R2 - Regra da Adição: $P(X_1 \text{ ou } X_2 | H) = P(X_1 | H) + P(X_2 | H) - P(X_1 \text{ e } X_2 | H)$.

R3 - Regra da Multiplicação: $P(X_1 \text{ e } X_2 | H) = P(X_1 | H)P(X_2 | X_1 \text{ e } H)$.

Em favor da simplicidade, não enunciaremos a regra da adição infinita. Uma discussão mais formal das regras da Probabilidade, é deixada a cargo de um curso básico de Teoria das Probabilidades. Para exemplificar a utilização dessas regras, consideremos $X = (X_1, X_2)$ e um experimento que observa o valor x_2 de X_2 . Deseja-se calcular $P(X_1 = x_1 | x_2 \text{ e } H)$ que por R3 pode ser escrita como

$$\frac{P(X_1 = x_1 \text{ e } X_2 = x_2 | H)}{P(X_2 = x_2 | H)}$$

onde

$$P(X_2 = x_2 | H) = \sum_1 P(X_1 = x_1 \text{ e } X_2 = x_2 | H)$$

e \sum_1 indica que a soma é sobre o conjunto dos pontos que Você considera impossíveis para X_1 . Finalmente, usando-se R3 mais uma vez, obtêm-se a fórmula de Bayes

$$P(X_1 = x_1 | x_2 \text{ e } H) = \frac{P(X_1 = x_1 | H)}{P(X_2 = x_2 | H)} P(X_2 = x_2 | x_1 \text{ e } H).$$

Essa fórmula mostra como "resultados experimentais" influenciam a opinião do pesquisador. As definições abaixo formalizam alguns conceitos usados anteriormente:

Definição 1.1 - Um conjunto de afirmações probabilísticas obedecendo as regras acima, é dito ser coerente. O conjunto coerente de uma quantidade aleatória X é denominado distribuição de probabilidades de X .

Definição 1.2 - A qualquer mecanismo que transforma uma quantidade desconhecida em conhecida damos o nome de Experimento. Os resultados de tais operações são denominados dados ou resultados

experimentais.

O objetivo principal deste trabalho é discutir o uso das regras acima, após a consideração de um conjunto coerente de afirmações probabilísticas, em presença de resultados experimentais. A forma de acesso a esse conjunto coerente é assunto para um trabalho futuro.

1.3 VEROSSIMILHANÇA: O RECIPIENTE DAS INFORMAÇÕES EXPERIMENTAIS

Nas situações mais comuns ao Estatístico, não existem, como no caso do lançamento da moeda, experimentos (realizáveis na prática) que observam o verdadeiro valor da quantidade de interesse, θ . Por exemplo, suponha que θ seja o número de fumantes da população brasileira. Intuitivamente, é impraticável a consideração de um experimento para observar o valor exato desse número. Outras vezes, embora, o experimento seja realizável, Você é solicitado a tomar decisões antes da realização deste. As qualidades dessas decisões dependem do resultado do experimento cuja realização (ou não), por sua vez, faz parte de Suas decisões. Por exemplo, suponha que Você é consultado no momento de decidir-se sobre a perfuração (ou não) de um poço petrolífero. Aqui, $\theta = 1$ indica que a extração de petróleo é viável e $\theta = 0$ em caso contrário. O exato valor de θ só será conhecido após a perfuração. Contudo, no caso de decidir-se pela não perfuração, esse valor nunca será conhecido e, por outro lado, não existirão despesas de perfuração. Completando o nosso raciocínio, suponha que um médico vive uma situação onde deve decidir qual das duas drogas, A_1 ou A_2 , deve administrar a um de seus pacientes. Após escolher uma das drogas, digamos A_1 , observa se o resultado é positivo ou negativo. Aqui, $\theta = (\theta_1, \theta_2)$ onde $(i = 1, 2)\theta_i = 1$ se o uso de A_i produzir resultado positivo e $\theta_i = 0$ em caso contrário. Como decidiu-se pela droga A_1 , apenas θ_1 será observado enquanto o valor de θ_2 nunca será conhecido.

Em situações análogas às descritas acima, Você é obrigado a considerar experimentos realizáveis que observam os valores x de quantidades X que estão, de certa forma, relacionados

com θ . O objetivo é reduzir a incerteza sobre θ . Essa redução é baseada na "informação relevante" (sobre θ) que esses resultados experimentais possam, em Sua opinião, possuir. A seguir, desejamos o tipo de relação, entre X e θ , que normalmente se considera.

Seja X uma quantidade aleatória e $P(X|H)$ a distribuição de probabilidade associada por Você a X . Da mesma forma, para qualquer quantidade aleatória θ , $P(\theta|H)$ representa Sua distribuição de probabilidade para θ e podemos escrever

$$P(X|H) = \sum_{\theta} P(X|H \text{ e } \theta)P(\theta|H)$$

onde \sum_{θ} indica soma sobre todos os valores possíveis de θ .

Definição 1.3 - A quantidade θ é chamada de parâmetro para X se, $P(X|H \text{ e } \theta) = P(X|\theta)$ e nesse caso $P(X|\theta)$ é denominado de (o Seu) modelo estatístico. O conjunto Θ de valores possíveis (em Sua opinião) de θ é denominado espaço paramétrico. O conjunto X dos valores (que Você considera antes da realização do experimento) possíveis de X é denominado espaço amostral.

Em outras palavras, θ é um parâmetro para X , se θ contém toda a informação relevante (representada por H) que Você possui sobre X . Isto é, no "momento" que θ fosse conhecido, Sua incerteza sobre X tornar-se-ia independente de qualquer conhecimento anterior. Em termos mais formais, X e H são condicionalmente independentes dado θ e escreve-se $X \perp\!\!\!\perp H | \theta$. Em resumo, se θ é um parâmetro para X , temos

$$P(X|H) = \sum_{\theta} P(X|\theta)P(\theta|H).$$

Ao definirmos um modelo estatístico $P(X|H)$, na verdade, estamos considerando a seguinte família de distribuições para X :

$$P = \{P(X|\theta); \theta \in \Theta\}.$$

Por outro lado, após a realização do experimento que observa o valor x de X , temos a função $v_x(\theta) = P(x|\theta)$ definida em θ . Note

que antes da realização do experimento, definimos o seguinte conjunto dessas funções de θ :

$$V = \{P(x|\theta); x \in X\} = \{v_x(\theta); x \in X\}.$$

Por definição, existe uma correspondência entre X e V que, a cada resultado possível x , associa a função $v_x(\theta)$.

É importante salientarmos o fato de a única ligação entre X e θ estar descrita apenas pelos conjuntos P e V . Após a realização do experimento e a obtenção do resultado x , a função $v_x(\theta)$ é o único elemento que permanece para descrever o relacionamento entre θ (desconhecido) e o resultado experimental x (conhecido). R.A. Fisher em seu último livro "Statistical Methods and Scientific Inference" publicado em 1956, descreve esse fato da seguinte forma: Toda a informação relevante da amostra x , sobre o parâmetro θ , está contida na função $v_x(\theta)$. Para uma análise mais profunda dessa afirmação, veja Basu (1975). A definição abaixo traduz esse pensamento.

Definição 1.4 - Após a observação do resultado experimental x , a função $v_x(\theta)$ é denominada Verossimilhança de θ , resultante da observação x .

Antes de prosseguir a discussão, recordemos alguns conceitos importantes: (1) A qualquer transformação T que associa pontos amostrais a elementos de um outro conjunto, dá-se o nome de estatística. (2) Se essa transformação T , que na verdade é uma redução dos dados experimentais, não acarretar perda de informação sobre θ , diz-se que T é suficiente. Isto é traduzido por $P(X|T \text{ e } \theta) = P(X|T)$ indicando que ao conhecer T , os dados já não possuirão nenhuma informação relevante sobre θ , pois $X \perp\!\!\!\perp \theta | T$. (3) Se T é suficiente e puder ser escrita como reduções de qualquer outra estatística suficiente, então T é denominada suficiente mínima; isto é, qualquer redução de T acarretaria em perda de informação sobre θ .

No estudo das estatísticas suficientes, a função de verossimilhança tem um papel bastante relevante. Esse papel fica

realçado quando entende-se que; (i) existe uma estatística T associando a cada ponto x do espaço amostral X , o elemento $v_x(\theta)$ de V , e (ii) essa transformação $T: X \rightarrow V$ é uma estatística suficiente mínima. Assim, para inferências sobre θ , no lugar de anotar-se o resultado x , é bastante que se registre o gráfico da função $v_x(\theta)$. Essas propriedades da função de verossimilhança deram origem ao seguinte princípio da Estatística.

Sejam duas quantidades aleatórias X e Y que possuem a mesma quantidade θ como parâmetro. Considere dois experimentos que observam, respectivamente, os valores x e y de X e Y . Denotemos as respectivas verossimilhanças por $v_x(\theta)$ e $v_y^*(\theta)$.

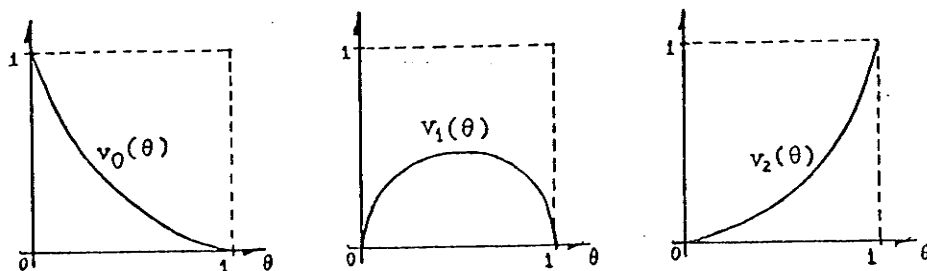
PRINCÍPIO DA VEROSSIMILHANÇA

Os dois resultados experimentais x e y possuem a mesma informação sobre θ se $v_x(\theta)$ e $v_y^*(\theta)$ são proporcionais (escreve-se $v_x(\theta) \propto v_y^*(\theta)$); isto é, se existir uma constante c tal que, $v_x(\theta) = cv_y^*(\theta)$ para qualquer $\theta \in \Theta$. Uma outra representação usada é $x \sim_{\theta} y$ para indicar equivalência segundo θ pois, x e y devem produzir a mesma inferência sobre θ .

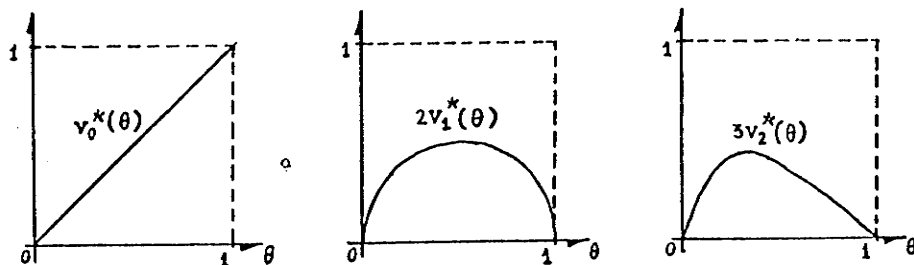
Acreditamos que, mesmo os leitores que aceitam o ponto de vista clássico da estatística, são forçados a admitir o fato da discussão acima ser razoavelmente lógica. Contudo, a maioria dos métodos clássicos de análise aceitos por esses leitores são contraditórios segundo o princípio da verossimilhança. Lembre-se que a qualidade de uma inferência clássica é definida por um estudo detalhado dos resultados que poderiam ter ocorrido (mas na verdade não ocorreram); isto é, um estudo detalhado do espaço amostral. Por outro lado, o princípio da verossimilhança indica que a qualidade da inferência depende da qualidade da informação contida no resultado específico do experimento. Para que o leitor possa se sentir mais a vontade nesse debate, apresentamos os seguintes exemplos. O objetivo principal é ilustrar os conceitos introduzidos, realçando, no último exemplo, a contradição que acabamos de sugerir.

Exemplo 1.1

Suponha que uma proporção θ é o parâmetro de interesse e nesse caso o espaço paramétrico é o intervalo $(0,1)$. Considere um experimento que observa o valor x de uma variável aleatória X , cujo modelo estatístico (denominado de binomial para duas observações) é representado por $P(X=x|\theta) = \binom{2}{x}\theta^x(1-\theta)^{2-x}$, onde $X = \{0,1,2\}$ é o espaço amostral. Se o resultado experimental for $x=1$, a função de verossimilhança será dada por $v_1(\theta) = 2\theta(1-\theta)$. O conjunto V é formado pelas três funções abaixo:

Exemplo 1.2

Suponha que no exemplo 1.1 no lugar de X , o experimento que se considera é o de observar o valor y de uma variável aleatória Y com espaço amostral $Y = \{0,1,2,\dots\}$ e cujo modelo estatístico (conhecido como modelo geométrico) é representado por $P(Y=y|\theta) = \theta(1-\theta)^y$. Se o resultado experimental for $y=1$, a função de verossimilhança será $v_1^*(\theta) = \theta(1-\theta)$. Aqui, $V = \{v_0^*(\theta), v_1^*(\theta), v_2^*(\theta), \dots\}$ é um conjunto infinito enumerável cujos três primeiros elementos estão ilustrados abaixo. Como verossimilhanças proporcionais se equivalem, decidimos apresentar, para uma ilustração mais clara, os gráficos de $v_0^*(\theta)$, $2v_1^*(\theta)$ e $3v_2^*(\theta)$.



Exemplo 1.3

Consideremos que dada uma população, θ representa a proporção de indivíduos, dessa população, que possuem a característica C . Com o objetivo de se fazer inferências sobre θ , Você seleciona, com reposição, dois indivíduos dessa população e registra o número x , dentre esses, que possuem a característica C . Dessa forma, Você considera o modelo descrito no Exemplo 1.1 João, não tendo conhecimento do Seu trabalho, decide selecionar, sequencialmente e com reposição, indivíduos até que encontre o primeiro com a característica C . Assim, João considera o modelo do Exemplo 1.2 onde $Y+1$ é o número de indivíduos que serão selecionados. Suponha agora que, após a realização dos dois experimentos, Você observa $x=1$ e João, por sua vez, observa $y=1$. Como as verossimilhanças são proporcionais, $v_1(\theta) = 2v_1^*(\theta)$, os dois resultados são, pelo princípio da verossimilhança, equivalentes e resultariam em inferências idênticas sobre θ . Note que este não é o caso dos métodos clássicos de inferência. Mesmo no caso do método da máxima verossimilhança, que embora produzisse estimativas pontuais idênticas, teríamos intervalos de confiança distintos.

Poderia parecer para muitos que os exemplos acima estariam sugerindo problemas com a lógica do princípio da verossimilhança. No entanto, os autores pensam que é justamente a lógica dos métodos clássicos que deveria estar sendo revisitada. O argumento que usamos é bastante simples, senão vejamos. Primeiramente note que Você e João possuem o mesmo objetivo; conhecer o valor de θ . Em segundo lugar, note que Você e João obtiveram a mesma informação sobre θ pois, Você observou dois indivíduos dos quais apenas um possuía a característica C e exatamente o mesmo foi observado por João. Assim, não podemos concordar que Você e João, baseados apenas nas mesmas informações experimentais, produzam diferentes inferências sobre θ , mesmo possuindo a mesma verossimilhança. Para uma análise mais profunda desses exemplos, veja Lindley and Phillips (1976).

Como veremos a seguir, no momento de incorporar as informações experimentais ao conhecimento prévio, os métodos Bayesianos

utilizam apenas as funções de verossimilhança. Isto nos leva a afirmar que o princípio da verossimilhança não será violado ao optar-se pelo ponto de vista Bayesiano da Estatística.

Gostaríamos que, ao terminar esta Seção, o leitor tenha percebido que a inferência Bayesiana não pode ser encarada como um capítulo da teoria clássica da inferência estatística. Essa idéia, foi uma distorção que os livros clássicos introduziram aos nossos estudantes. Note que, a teoria Bayesiana e a teoria clássica são filosoficamente opostas, assim como o subjetivismo e o objetivismo. É importante realçar que todo problema estatístico (teste de hipótese, intervalos de confiança, inferência não paramétrica, etc) pode ser atacado tanto da forma clássica quanto da forma Bayesiana. A escolha da linha filosófica, depende da coerência lógica de cada um.

1.4 OPERAÇÃO BAYESIANA: CALIBRANDO A INCERTEZA ATRAVÉS DOS RESULTADOS EXPERIMENTAIS

Como foi discutido nas seções anteriores, a "novidade" da posição Bayesiana é que, a incerteza é descrita probabilisticamente. Dessa forma, a receita Bayesiana é simples: Para uma quantidade desconhecida de interesse, θ , conhecendo-se H , calcula-se as probabilidades condicionais de θ dado H usando-se as regras do cálculo de probabilidades, R1, R2 e R3. Note que essa receita está, na verdade, encobrendo um processo sequencial. Se H representa o Seu conhecimento até um determinado momento, no instante seguinte deverá conter também as informações adicionais que Você poderia estar adquirindo naquele momento. O objetivo da inferência Bayesiana é justamente estudar as formas com as quais os resultados experimentais, adicionados a H , alteram $P(\theta|H)$.

Seja X uma quantidade aleatória que possui θ , a quantidade de interesse, como parâmetro; isto é, $X||H|\theta$. Considere um experimento E que observa o valor de X . Com as regras do cálculo das probabilidades, podemos deduzir que

$$P(\theta|H \text{ e } x) = \frac{P(\theta|H)}{P(X=x|H)} P(X=x|\theta) = \frac{P(\theta|H)}{C_x} v_x(\theta)$$

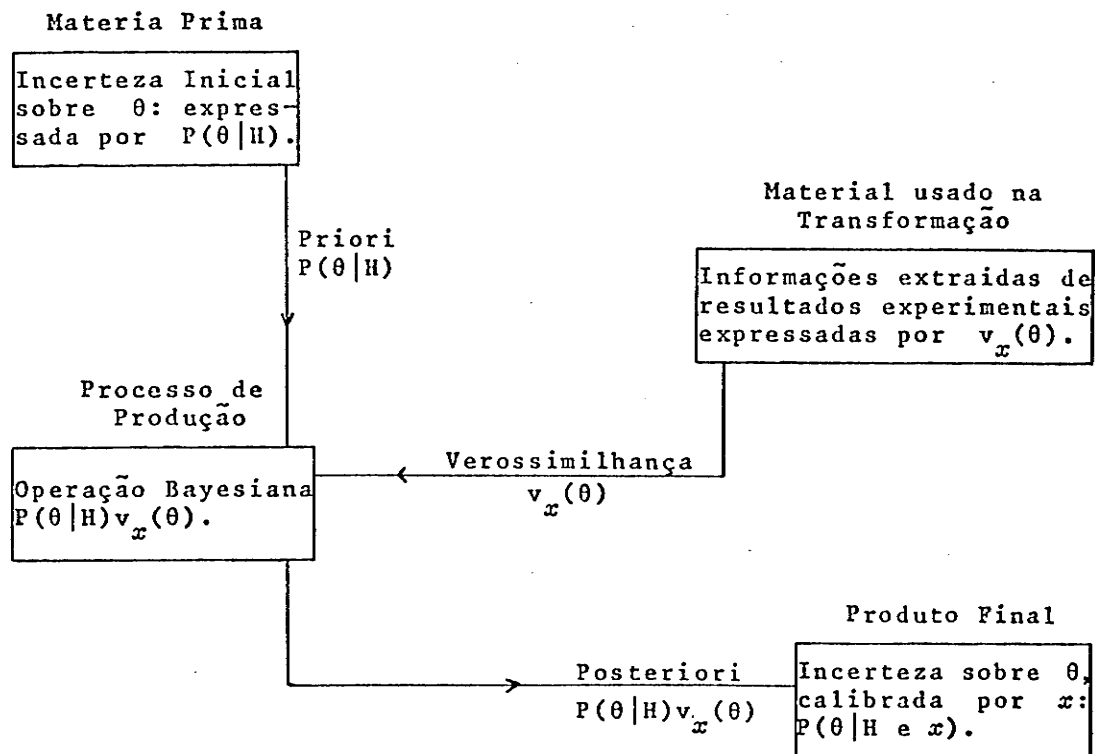
onde $C_x = \sum_{\theta} v_x(\theta) P(\theta|H)$ é uma constante conhecida, após observar-se o resultado x de X , pois \sum_{θ} é uma soma (ou integral) sobre θ . A expressão acima, visto que representa uma distribuição de probabilidades, pode ser escrita como

$$P(\theta|H \text{ e } x) \propto P(\theta|H) v_x(\theta).$$

Note que aqui $P(\theta|H)$ e $P(\theta|H \text{ e } x)$ representam, respectivamente, a distribuição de θ antes de observar-se o valor x de X e a distribuição de θ após essa observação.

Esse mecanismo de passar de $P(\theta|H)$ para $P(\theta|H \text{ e } x)$ é conhecido como operação Bayesiana e nesse caso $P(\theta|H)$ e $P(\theta|H \text{ e } x)$ são chamadas de distribuição a priori e distribuição a posteriori, respectivamente. Essas denominações, a priori e a posteriori, referem-se ao conhecimento dos resultados experimentais. No relacionamento com um experimento sobre X , anterior a um experimento sobre Y , o que está sendo considerado como distribuição a priori para θ , em relação a Y , pode ter sido uma distribuição a posteriori para θ , em relação a X . Por outro lado, no relacionamento com um experimento sobre Z , posterior a Y , a distribuição a posteriori de θ , em relação a (X,Y) , pode ser considerada como a a priori para θ , em relação a Z . Este é, exatamente, o espírito dinâmico ou sequencial da inferência Bayesiana. Um outro fato de realce é que a distribuição a posteriori de θ , em relação aos dois experimentos sobre X e Y , não sofre nenhuma influência da ordem com que X e Y foram observadas.

O seguinte diagrama é uma tentativa de ilustrar a operação Bayesiana.



A simplicidade do seguinte exemplo tem o objetivo de realçar o processo descrito.

Exemplo 1.4

Dr. André é um especialista em medicina e considera que, dentre os indivíduos que estão nas mesmas condições clínicas do Marcel, um entre dez é portador de um determinado tipo de Virus. Acredita-se que, das pessoas portadoras, 80% reagiriam positivamente a um teste de laboratório, enquanto que apenas 30% das não portadoras, produziram uma reação positiva. Dessa forma Dr. André solicitou esse teste para Marcel e este reagiu positivamente. Usando os conceitos descritos, como Dr. André poderia, coerentemente, descrever sua incerteza com respeito ao fato de Marcel ser

($\theta = 1$) ou não ($\theta = 0$) portador do Vírus?

Com o que foi considerado, uma medida razoável da incerteza do Dr. André seria $P(\theta = 1 | H) = 1/10$. Como os resultados possíveis do experimento (verificação da reação ao teste) são $X = 1$ para reação positiva e $X = 0$ para negativa, torna-se razoável considerarmos o seguinte como função de verossimilhança:

$$v_1(\theta) = \begin{cases} 8/10 & \text{se } \theta = 1 \\ 3/10 & \text{se } \theta = 0. \end{cases}$$

Assim, após o resultado (positivo) do teste, a distribuição (a posteriori) de θ é dada por

$$P(\theta | H \text{ e } X = 1) = \begin{cases} 8/35 & \text{para } \theta = 1 \\ 27/35 & \text{para } \theta = 0. \end{cases}$$

Note que o conhecimento do Dr. André, representado por H , é usado no momento que ele identifica as pessoas (entre seus pacientes anteriores) que estão nas mesmas condições clínicas de Marcel.

No exemplo acima, descrevemos um problema de predição; isto é, um problema cujo objetivo é adivinhar o valor de θ . É claro que após terminado o tratamento ao caso Marcel, Dr. André terá mais um caso estudado que deverá enriquecer os seus conhecimentos que proporcionará, em casos semelhantes no futuro, diagnósticos mais "precisos".

Enquanto nesse exemplo, o objetivo era um elemento de uma população, na maioria das análises que serão discutidas nessas notas, estaremos voltados aos problemas de decisão que dizem respeito a população como um todo. Especificamente, apresentaremos versões Bayesianas para a trindade da Estatística Clássica: Estimação (pontual), Intervalos de Confiança e Testes de Hipóteses.

Suponha que a quantidade desconhecida, θ , é um número real. O problema da estimação é o de encontrar, baseado na incerteza atual sobre θ , um número $\hat{\theta}$ que você aceitaria usar como

predição de θ . Como a distribuição a posteriori de θ é a expressão atualizada da Sua incerteza, o valor esperado de θ , baseado nessa distribuição, $[E(\theta|H e x) = \int \theta P(\theta|H e x)]$ pode ser considerado como um Estimador razoável de θ e é conhecido como Estimador de Bayes. É evidente que outras características, da distribuição a posteriori, poderiam ser usadas no lugar do valor esperado. Os exemplos mais naturais seriam a moda e a mediana dessa distribuição. A escolha de uma dessas características depende das suas qualidades, que por sua vez dependem do problema que se deseja resolver. A nossa opção pelo valor esperado deve-se, apenas, ao fato de esta ter recebido maior divulgação na comunidade.

Um outro aspecto, que pode surpreender o estatístico clássico, é o fato de na estimação Bayesiana pontual desejar-se estimadores $\hat{\theta}$ com máxima variância possível. Esse aspecto fica claro ao aceitar-se que a variância de θ seja um bom indicador do grau de incerteza sobre o valor de θ . Assim, a diferença entre as variâncias "a priori" e "a posteriori" $[Var(\theta|H) e Var(\theta|H e x)]$ poderia ser considerada como um indicador razoável da informação proporcionada pelo resultado experimental. O valor esperado dessa diferença indica a informação que se espera do experimento em questão. Isto é, $I = Var(\theta|H) - E[Var(\theta|H e X)]$ pode ser considerada como uma medida razoável da informação proporcionada pelo experimento que observa o valor de X . Um bom treino para o leitor seria verificar que

$$I = Var[E(\theta|H e X)] = Var(\hat{\theta})$$

e assim poderemos entender que maximizar a variância de $\hat{\theta}$ é o objetivo da inferência Bayesiana pontual.

Intervalos (ou conjuntos) de Credibilidade são os correspondentes Bayesianos aos Intervalos de Confiança. A idéia é bastante simples, pois o Bayesiano, por princípio, já admite afirmações probabilísticas sobre θ . Assim, ao descrever-se as incertezas pela distribuição a posteriori, é natural que se procure o menor conjunto C_x de θ , tal que, para $\alpha \in (0,1)$ fixado, $P(\theta \in C_x|H e x) = \alpha$. O conjunto C_x , assim construído, é conhecido

como Intervalo ou conjunto de credibilidade de nível α .

Ao construirmos os conjuntos C_x , podemos definir um método de decisão que corresponderia, de certa forma, aos testes de hipótese da teoria clássica. Fixado um nível α , decide-se em favor de $\theta = \theta_0$ se $\theta_0 \in C_x$ e contra se $\theta_0 \notin C_x$. Um outro procedimento, devido a Harold Jeffreys e mais razoável no caso de hipóteses compostas, baseia-se na consideração de probabilidades, além das distribuições a priori, para as hipóteses em questão. Seja ξ a probabilidade a priori para o evento $H_0: \theta \in C$ e $\xi(x)$ a probabilidade a posteriori. Para um valor k fixado, se $\frac{\xi(x)}{1 - \xi(x)} > k$ decide-se em favor de H_0 e em caso contrário, decide-se contra H_0 . A generalização para o caso das decisões múltiplas é intuitiva. O capítulo 3 será dedicado as aplicações dos métodos acima.

1.5 CLASSES CONJUGADAS DE DISTRIBUIÇÕES

Como foi discutido nas seções anteriores, a operação Bayesiana envolve os seguintes entes: distribuição a priori, função de verossimilhança e distribuição a posteriori. Na maioria das vezes, considera-se uma sequência de n experimentos para se observar os valores de X_1, X_2, \dots e X_n que possuem a propriedade de serem, condicionalmente a θ , mutuamente independentes e identicamente distribuídas. Dessa forma, existe um consenso que admite $\prod_{i=1}^n v_{x_i}(\theta)$ como a verossimilhança de θ no resultado experimental (x_1, \dots, x_n) de (X_1, \dots, X_n) . Para obter-se a distribuição a posteriori, objetivo da operação Bayesiana, necessita-se especificar a distribuição a priori. A seguir, definiremos uma classe de distribuições que, além de ser "rica" em opiniões, permite a obtenção de expressões analíticas gerais.

A propriedade sequencial da operação Bayesiana nos permite escrever

$$P(\theta | H \text{ e } (x_1, \dots, x_n)) \propto P(\theta | H \text{ e } (x_1, \dots, x_{m-1})) \prod_{i=1}^n v_{x_i}(\theta),$$

onde $m \in (1, n]$ é um inteiro positivo. Com essa propriedade sequencial, fica clara a conveniência de se especificar uma distri-

buição a priori que "conviva intimamente" com o modelo. A classe C de distribuições desse tipo, deve ser bastante numerosa para acomodar vários tipos de opiniões acerca de θ . Se, além da priori pertencer a C , a distribuição a posteriori for um elemento de C , qualquer que seja o resultado possível, então C é denominada classe conjugada (com o modelo) de distribuições.

Muitas vezes ocorre que a família C coesiste naturalmente com o modelo estatístico. Suponha que para este, existe $m > 0$ tal que para qualquer resultado (x_1, \dots, x_n) , $n > m$, a verossimilhança é integrável (segundo uma medida μ) sobre o espaço paramétrico θ . Considere assim a família C_0 de distribuições de probabilidade sobre θ que satisfaz a seguinte propriedade:

Se $P(\theta|H)$ é um membro de C_0 , então existe um vetor $y = (y_1, \dots, y_k)$, onde k é um inteiro positivo, tal que a função de densidade (ou probabilidade) associada a $P(\theta|H)$ é dada por

$$p(\theta|H) = \left[\int_{\theta} v_y(\theta) d\theta \right]^{-1} v_y(\theta).$$

[No caso geral a integral do denominador é calculada em relação a medida μ]. Se y é um vetor de k resultados experimentais já obtidos, a distribuição a priori que está sendo considerada é produto exclusivo dos dados experimentais y pois é obtida da verossimilhança de θ no resultado y . Assim, o conhecimento passado, representado por H , depende apenas de y , não incluindo conhecimentos não experimentais. Neste caso, pode-se considerar que a distribuição de θ , antes de observar-se y , representa a ignorância completa sobre θ . Essa distribuição é conhecida como não-informativa. Outra característica da classe C_0 é que após observar-se (x_1, \dots, x_n) , a distribuição (a posteriori) de θ pode ser obtida da verossimilhança de θ no vetor $(y_1, y_2, \dots, y_k, x_1, \dots, x_n)$ que contém todos os resultados experimentais antigos e novos. Como C_0 também contém a distribuição a posteriori, será chamada de classe conjugada natural de distribuições.

Nos próximos capítulos, a escolha de distribuições a priori será restrita as classes C_0 e assim o evento H passa a

ser irrelevante e pode ser substituído pelo particular y .

Se existir uma estatística suficiente $T = (t_1, \dots, t_d)$ com dimensão (independente do "tamanho da amostra a priori, k ") constante $d > 1$, a família C_0 toma uma forma mais simples. Neste caso, $v_y(\theta) \propto V(\theta|k, T)$, onde T é função de y ; isto é, $T(y_1, \dots, y_k) = (t_1, \dots, t_d)$ e d é uma constante (k sendo variável). Finalmente, a família C_0 de distribuições conjugadas naturais é caracterizada pelos seguintes $d+1$ "super-parâmetros"; t_1, t_2, \dots, t_d e k . Por esta razão, na sequência deste trabalho, H será substituído propriamente pelos super-parâmetros específicos.

Terminamos esse capítulo com o seguinte exemplo que ilustra a discussão acima.

Exemplo 1.5

Considere uma sequência X_1, \dots, X_n de variáveis aleatórias (de Bernoulli) tais que, para θ conhecido, X_1, \dots, X_n são independentes com mesma distribuição $P\{X_i = 1|\theta\} = 1 - P\{X_i = 0|\theta\} = \theta$. Assim, a verossimilhança no resultado (x_1, \dots, x_n) é dada por

$$\prod_{i=1}^n v_{x_i}(\theta) = \theta^t (1 - \theta)^{n-t}, \text{ onde } t = \sum_{i=1}^n x_i.$$

Se a distribuição a priori é Beta com parâmetros t_1 e $k - t_1$, então a distribuição a posteriori será Beta com parâmetros $t_1 + t$ e $k + n - (t_1 + t)$ e assim C_0 é a classe de distribuições Beta. Normalmente os parâmetros da Beta são representados por α e β e assim, os super-parâmetros de C_0 serão $t_1 = \alpha$ e $k = \alpha + \beta$.

CAPÍTULO 2

DISTRIBUIÇÕES ESPECIAIS

2.1 INTRODUÇÃO

O objetivo deste capítulo é apresentar as propriedades das distribuições usadas na sequência. As distribuições contínuas, aqui discutidas, serão utilizadas como distribuições a priori e, com exceção da Dirichlet-Multinomial, as distribuições discretas constituirão os modelos estatísticos $P = \{P(\cdot | \theta); \theta \in \Theta\}$.

Neste trabalho, um experimento será idealizado por uma variável (ou vetor) aleatória, X , associada com um espaço amostral X (que é subconjunto de \mathbb{R}^n , para algum $n = 1, 2, \dots$) e uma função de probabilidade, $p(\cdot)$, sobre X . Lembre-se que, como consideramos apenas experimentos discretos, X é um conjunto enumerável e assim não é absurdo considerar-se apenas funções de probabilidade $p: X \rightarrow [0, 1]$ tal que, $\forall A \subset \mathbb{R}^n$, $P(A) = \sum_A p(x)$ onde \sum_A indica soma sobre o conjunto $A \cap X$. No caso de $A \cap X = \emptyset$ (o conjunto vazio), definimos $P(A) = 0$. O uso das funções de densidade será restrito a alguns casos onde temos uma variável (ou vetor) aleatória não observável, θ . Esta representação de incerteza ocorrerá sempre que o conjunto Θ (que é um subconjunto de \mathbb{R}^n , para algum $n = 1, 2, \dots$), de possíveis valores de θ , for não enumerável e a qualquer subconjunto enumerável de elementos de Θ estiver associado probabilidade zero. Assim, na maioria das vezes, a incerteza sobre θ será representada por uma função (de densidade de probabilidade) $f: \mathbb{R}^n \rightarrow (0, \infty)$, tal que $f(y) \neq 0 \forall y \in \Theta$ e a integral (de Riemann) de f sobre qualquer conjunto (integrável) $B \subset \mathbb{R}^n$ é a probabilidade do evento $\theta \in B$.

O valor esperado de uma função real $g(h)$ de X (de θ) é definido como

$$E\{g(X)\} = \sum_x g(x)p(x)$$

e

$$E\{h(\theta)\} = \int_{\theta} h(\theta) f(\theta) d\theta.$$

Note que, com a notação do capítulo anterior e as considerações feitas acima, poderemos representar P como a família $\{p(\cdot|\theta); \theta \in \Theta\}$ de funções de probabilidade sobre X e a família das possíveis densidades a posteriori por $F = \{f(\cdot|x); x \in X\}$. Assim, os valores esperados condicionais

$$E\{g(X) | \theta\} = \sum_x g(x) p(x|\theta)$$

e

$$E\{h(\theta) | X = x\} = \int_{\theta} h(\theta) f(\theta|x) d\theta$$

são de importância e representam os valores esperados das quantidades $g(X)$ e $h(\theta)$ para cada membro das famílias P e F .

O leitor deve-se lembrar a todo momento que o único ente objetivo da análise é o valor x , da quantidade X , o qual será observado. Normalmente, a quantidade de interesse θ é não observável. Contudo, antes do experimento X ser realizado, pode haver interesse em descrever-se a distribuição marginal de X . Isto é, a distribuição conjunta de θ e X pode ser calculada pelo produto $f(\theta)p(x|\theta)$ e assim a função de probabilidade marginal de X é dada por

$$p(x) = \int_{\theta} p(x|\theta) f(\theta) d\theta \quad \forall x \in X$$

que na verdade, para cada ponto x de X , é o valor esperado de $p(x|\theta)$. Esta função de probabilidade é denominada de "função de probabilidade preditiva" ou "função de probabilidade de predição". É necessário ressaltar o fato de que, muitas vezes, a distribuição a priori de θ , no experimento X , pode ser a distribuição a posteriori de θ no experimento Y . Ou seja, no lugar de $f(\theta)$ poderíamos escrever $f(\theta|y)$ e no caso de θ ser um parâmetro associado a X e Y teríamos $X || Y | \theta$ o que implicaria em substituímos $p(x)$ por $p(x|y)$. Isto é, se θ é um parâmetro associado a X e Y , a função de probabilidade preditiva de X , posterior

a Y , é calculada como

$$p(x|y) = \int_{\Theta} p(x|\theta) f(\theta|y) d\theta.$$

2.2 DISTRIBUIÇÕES DERIVADAS DO PROCESSO DE BERNOULLI

Consideremos uma sequência U_1, U_2, \dots de experimentos onde, $\forall i=1, 2, \dots, U_i$ pode assumir apenas dois valores 1 ou zero. Suponha também que, na Sua opinião, essa sequência seja permutável. Ou seja, $\forall n=1, 2, \dots$, se (i_1, i_2, \dots, i_n) é uma permutação de $(1, 2, \dots, n)$, então $P(U_1 = u_1, \dots, U_n = u_n) = P(U_1 = u_{i_1}, \dots, U_n = u_{i_n})$ onde (u_1, \dots, u_n) é um vetor de zeros e uns. Com apenas essa restrição de permutabilidade, que corresponde a dizer que os rótulos são não informativos, pode-se provar que existe uma quantidade desconhecida $\theta \in (0, 1)$ tal que, condicionalmente a θ , U_1, U_2, \dots são mutuamente independentes (escreve-se $(U_1 \perp\!\!\!\perp U_2 \perp\!\!\!\perp \dots) | \theta$ ou $\prod_{i=1}^{\infty} U_i | \theta$) e $P(U_i = 1) = \theta \quad \forall i=1, 2, \dots$. Este resultado é conhecido como Teorema de De Finetti e tem consequências importantes para a inferência Bayesiana.

Condicionalmente a θ , a sequência U_1, U_2, \dots recebe o nome de processo de Bernoulli e muitas das distribuições discretas conhecidas são obtidas de transformações desse processo. Inicialmente, consideremos uma regra de parada δ que, para cada $i=1, 2, \dots$, após observar (U_1, \dots, U_n) decide se observa ou não U_{n+1} .

A primeira distribuição que discutiremos aqui é a Binomial. Se δ_0 é a regra que sempre decide parar na observação de U_n onde n é um inteiro fixado, temos que para qualquer vetor (u_1, \dots, u_n) de zeros e uns,

$$P(U_1 = u_1, \dots, U_n = u_n | \theta) = \theta^x (1 - \theta)^{n-x},$$

onde $x = u_1 + \dots + u_n$ é o número de uns da sequência. A variável $X = U_1 + \dots + U_n$ é dita ter distribuição Binomial com parâmetros n e θ e escreve-se $X|(n, \theta) \sim b(n, \theta)$. Quando n (ou δ_0) estiver implícito no contexto, escrevemos $X|\theta$ no lugar de $X|(n, \theta)$.

A distribuição de probabilidade associada a X tem a seguinte expressão:

$$p(x|\theta) = \binom{n}{x} \theta^x (1-\theta)^{n-x}$$

onde $\binom{n}{x} = \frac{n!}{x!(n-x)!}$ e $x \in \{0, 1, \dots, n\}$. As propriedades da distribuição Binomial serão discutidas de forma mais geral na seção 2.3.

Consideremos uma regra δ_α que decide parar na observação de $U_{y+\alpha}$ se o vetor $(u_1, u_2, \dots, u_{y+\alpha})$ possuir um número α (fixado) de uns; isto é, se $u_1 + \dots + u_{y+\alpha} = \alpha$. Assim, podemos definir um experimento Y tal que $Y+k$ é o número de observações (de experimentos U_i) necessárias para obter-se um número fixado α de sucessos (de uns). Dessa forma Y é o número de fracassos (de zeros) antes de se obter o α -ésimo sucesso. A variável Y assim definida é dita ter distribuição Binomial negativa com parâmetros α e θ e escreve-se $Y|(\alpha, \theta) \sim \text{bn}(\alpha, \theta)$. Não é difícil provar-se que a distribuição de probabilidade associada a Y tem a seguinte expressão:

$$p(y|\theta, \alpha) = \binom{\alpha+y-1}{y} \theta^\alpha (1-\theta)^y = \frac{\Gamma(\alpha+y)}{\Gamma(\alpha)y!} \theta^\alpha (1-\theta)^y$$

onde $y \in \{0, 1, 2, \dots\}$ e $\Gamma(\cdot)$ é a função Gama.

Note que, da forma como definimos Y , α é um inteiro positivo. Contudo, essa restrição não é tão importante pois, para qualquer valor real positivo de α , a função acima continua sendo uma função de probabilidade. Esta afirmação pode ser provada usando-se o seguinte fato:

$$\sum_{y=0}^{\infty} \frac{\Gamma(\alpha+y)}{\Gamma(\alpha)y!} (1-\theta)^y = \theta^{-\alpha}$$

para todo $\alpha \in (0, \infty)$ e qualquer $\theta \in (0, 1)$. Além dessa igualdade, um bom exercício para o leitor seria verificar se a média e a variância de Y possuem as seguintes expressões:

$$E\{Y | (\alpha, \theta)\} = \frac{\alpha(1 - \theta)}{\theta}$$

e.

$$\text{Var}\{Y | (\alpha, \theta)\} = \frac{\alpha(1 - \theta)}{\theta^2}.$$

No caso particular de $\alpha = 1$, a distribuição de Y toma o nome de distribuição geométrica com parâmetro θ . Se Y_1, Y_2, \dots, Y_a são a variáveis aleatórias independentes e identicamente distribuídas (i.i.d.) segundo uma geométrica de parâmetro θ , então, $Y = Y_1 + \dots + Y_a$ tem distribuição binomial negativa com parâmetros a e θ .

Voltando à sequência inicial, U_1, U_2, \dots , e fixando um inteiro positivo N , vamos supor que o parâmetro de interesse seja $\psi = U_1 + U_2 + \dots + U_N$. Contudo, suponha que apenas n ($< N$) dentre os N primeiros experimentos possam ser observados. Como a sequência U_1, U_2, \dots é permutável, pode-se considerar, sem perda de generalidade, a observação das n primeiras variáveis, U_1, U_2, \dots, U_n . Usando-se o resultado de De Finetti, tem-se que $X = U_1 + \dots + U_n$ e $\psi - X = U_{n+1} + \dots + U_N$ são condicionalmente independentes, dado θ . Isto é, $X \perp\!\!\!\perp (\psi - X) | \theta$; além disso, podemos escrever $X | \theta \sim b(n, \theta)$, $\psi | \theta \sim b(N, \theta)$ e $(\psi - X) | \theta \sim b(N - n, \theta)$. Note que, como ψ é o parâmetro de interesse e X , sendo suficiente para U_1, \dots, U_n com relação a θ , é o experimento observável, o nosso objetivo principal é encontrar a função de probabilidade de $X | \psi$. Por outro lado, ψ é suficiente para (U_1, U_2, \dots, U_N) em relação a θ e assim, como X é função desse vetor, $X | \psi$ tem a mesma distribuição de $X | (\theta, \psi)$. Usando-se as regras do cálculo de probabilidades e as propriedades acima, temos que:

$$\begin{aligned} P(X | \psi) &= P(X | \psi, \theta) = \frac{P(X, \psi | \theta)}{P(\psi | \theta)} = \\ &= \frac{P(X, \psi - X | \theta)}{P(\psi | \theta)} = \frac{P(X | \theta) P(\psi - X | \theta)}{P(\psi | \theta)}. \end{aligned}$$

Assim, $X|\psi$ tem função de probabilidade

$$p(x|\psi) = \frac{\binom{N}{x} \binom{N-n}{\psi-x}}{\binom{N}{\psi}} = \frac{\binom{\psi}{x} \binom{N-\psi}{n-x}}{\binom{N}{n}}$$

onde, $\max\{0, (n-N+\psi)\} \leq x \leq \min\{n, \psi\}$. A variável X é dita ter distribuição hipergeométrica com parâmetros N , n e ψ . A notação, quando N e n estão implícitos, será

$$X|\psi \sim h(N, n, \psi).$$

Um bom exercício para o leitor seria verificar as seguintes expressões para a média e a variância de X e a covariância entre U_i e U_j , $i, j \leq n$:

$$E\{X|\psi\} = np,$$

$$\text{Var}\{X|\psi\} = np(1-p) \frac{N-n}{N-1};$$

$$\text{Cov}\{U_i, U_j|\psi\} = p(1-p) \frac{1}{N-1}, \quad i \neq j = 1, 2, \dots, n,$$

onde $p = \frac{\psi}{N}$.

O elemento U_i da sequência U_1, U_2, \dots é na verdade uma variável que serve para indicar a categoria, dentre duas alternativas, a que o elemento i pertence. A generalização para $k (> 2)$ categorias é natural, bastando para isso que usemos a base ortogonal de \mathbb{R}^k no lugar de 0 e 1 ou, equivalentemente, no lugar de $\begin{pmatrix} 0 \\ 1 \end{pmatrix}$ e $\begin{pmatrix} 1 \\ 0 \end{pmatrix}$. As distribuições multinomial e hipergeométrica multivariada, discutidas na sequência, são distribuições derivadas de tais generalizações.

Para complementar esta seção, apresentaremos a derivação de uma distribuição bastante útil na inferência Bayesiana. Vamos supor que o parâmetro de interesse seja o número de fracassos λ antes de observar-se o α -ésimo sucesso. Contudo, apenas podemos observar o número Y de fracassos antes de obter-se o a -ésimo ($a < \alpha$) sucesso. Como α e a são fixados, podemos escrever,

$$\lambda | \theta \sim \text{bn}(\alpha, \theta), \quad Y | \theta \sim \text{bn}(a, \theta)$$

$$(\lambda - Y) | \theta \sim \text{bn}(\alpha - a, \theta) \quad \text{e} \quad Y \perp\!\!\!\perp (\lambda - Y) | \theta.$$

Como antes, o nosso objetivo é descrever a distribuição de $Y | \lambda$. Assim que o leitor perceber que (λ, α) é suficiente para Y e $(\lambda - Y)$, com respeito a θ , o problema torna-se semelhante ao caso da hipergeométrica e obtemos, analogamente,

$$P(Y | \lambda) = P(Y | \lambda, \theta) = \frac{P(Y | \theta) P(\lambda - Y | \theta)}{P(\lambda | \theta)}.$$

A função de probabilidade de $Y | \lambda$ tem a seguinte expressão:

$$p(y | \lambda) = \frac{\lambda! \Gamma(\alpha)}{\Gamma(\lambda + \alpha)} \frac{\Gamma(y + a)}{y! \Gamma(a)} \frac{\Gamma(\lambda - y + \alpha - a)}{(\lambda - y)! \Gamma(\alpha - a)},$$

$$y \in \{0, 1, \dots, \lambda\},$$

Para futuras referências, escrevemos a_0 no lugar de $\alpha - a$ e assim temos que Y tem distribuição Beta-Binomial com parâmetros, λ , a e a_0 e escrevemos

$$Y | (\lambda, a, a_0) \sim \text{Bb}(\lambda; a, a_0).$$

As propriedades dessa distribuição serão estudadas no caso geral da Dirichlet-Multinomial, que aparece na sequência.

2.3 A DISTRIBUIÇÃO MULTINOMIAL

Sejam Y_α , $\alpha = 1, 2, \dots$ variáveis aleatórias cujos possíveis resultados são os vetores

$$Y_\alpha = (e_{0\alpha}, e_{1\alpha}, \dots, e_{k\alpha})$$

com exatamente uma componente igual a 1 e as demais componentes iguais a 0. Sejam $\theta_0, \theta_1, \dots, \theta_k$ as probabilidades de $Y_\alpha = (1, 0, \dots, 0)$, $Y_\alpha = (0, 1, \dots, 0)$, \dots , $Y_\alpha = (0, 0, \dots, 1)$, respectivamente, com $\theta_i > 0$, $i = 0, 1, \dots, k$; $\alpha = 1, 2, \dots$. Dado um número inteiro $n > 0$, defina

$$X_i = \sum_{\alpha=1}^n e_{i\alpha} \quad , \quad i = 0, 1, \dots, k,$$

de modo que

$$\theta_0 = 1 - \sum_1^k \theta_i \quad ,$$

$$X_0 = n - \sum_1^k X_i \quad .$$

Seja ainda

$$X = (X_1, \dots, X_k),$$

e

$$\theta = (\theta_1, \dots, \theta_k).$$

Definição 2.1 - Se o conhecimento de θ torna as variáveis Y_1, \dots, Y_n independentes, então a distribuição condicional de X dado θ chama-se Distribuição Multinomial com parâmetros θ e n , e é dada por

$$P(X = x | \theta) = n! \prod_{i=0}^k \frac{\theta_i^{x_i}}{x_i!} \quad ,$$

definida para $x_i \geq 0$, inteiro, $i = 1, \dots, k$, $\sum_1^k x_i \leq n$, $\theta_i > 0$,
 $i = 1, \dots, k$, $\sum_1^k \theta_i < 1$. Para representar a distribuição multinomial escreveremos

$$X | \theta \sim M_k(n; \theta).$$

Quando $k = 1$, a distribuição de $X | \theta$ coincide com a distribuição binomial estudada anteriormente.

Exemplo 2.1 - Sejam A, B dois atributos, cada um deles presente ou ausente numa população. Então, cada elemento desta população pode ser classificado em exatamente uma das $2^2 = 4$ categorias

A	B	α	y_α
presente	presente	0	(1,0,0,0)
presente	ausente	1	(0,1,0,0)
ausente	presente	2	(0,0,1,0)
ausente	ausente	3	(0,0,0,1)

De acordo com a notação acima, podemos escrever

$$X|\theta \sim M_3(n;\theta).$$

Se $\theta = (.20, .30, .15)$ e $n = 10$, então

$$P(X = x|\theta) = 10! \prod_{i=0}^3 \frac{\theta_i^{x_i}}{x_i!},$$

com $x_0 = 10 - \sum_1^3 x_i$, $\theta_0 = 1 - \sum_1^3 \theta_i = .35$.

Assim, para calcularmos a probabilidade de $x = (3,2,3)$ dado θ , utilizamos a expressão acima e obtemos

$$P(X = x|\theta) = 10! \frac{(.35)^2 (.20)^3 (.30)^2 (.15)^3}{2!3!2!3!} = 0.75\% .$$

A seguir, verificaremos algumas propriedades da distribuição multinomial.

Resultado 2.1 - Se $X|\theta \sim M_k(n;\theta)$ então

1. $E(X_i) = n\theta_i$, $i = 1, \dots, k$,
2. $\text{Var}(X_i) = n\theta_i(1-\theta_i)$, $i = 1, \dots, k$,
3. $\text{Cov}(X_i, X_j) = -n\theta_i\theta_j$, $i \neq j$; $i, j = 1, \dots, k$.

Demonstração: Sem perda de generalidade, tomemos $i = 1$ e $j = 2$

$$E(X_1) = \sum_{s_k} x_1 n! \prod_{i=0}^k \frac{\theta_i^{x_i}}{x_i!},$$

onde o somatório é sobre o conjunto s_k dos pontos x_1, \dots, x_k in

teiros não negativos tais que $\sum_1^k x_i \leq n$. Notemos, de passagem, que

$$1 = (\theta_0 + \theta_1 + \dots + \theta_k)^n = \sum_{s_k} n! \prod_{i=0}^k \frac{\theta_i^{x_i}}{x_i!}.$$

Por outro lado, podemos reescrever $E(X_1)$ como

$$E(X_1) = n\theta_1 \sum_{s_k} (n-1)! \prod_{i=0}^k \frac{\theta_i^{x'_i}}{x'_i!},$$

onde

$$x' = (x_1 - 1, x_2, \dots, x_k)$$

e portanto

$$\sum_{i=1}^k x'_i = \sum_{i=1}^k x_i - 1 \leq n - 1.$$

Como as parcelas nas quais $x_1 = 0$ são nulas, podemos incluir apenas aquelas com $x_1 \geq 1$. Neste caso, x'_1, \dots, x'_k são inteiros não negativos e portanto, como observamos acima, tal soma na expressão de $E(X_1)$ é igual a 1. Logo $E(X_1) = n\theta_1$.

Para obtermos a $\text{Var}(X_1)$ basta calcularmos $E(X_1^2)$, já que $\text{Var}(X_1) = E(X_1^2) - E^2(X_1)$. Para isso, note que, usando a notação $x' = (x_1 - 1, x_2, \dots, x_k)$,

$$\begin{aligned} E(X_1^2) &= \sum_{s_k} x_1^2 n! \prod_{i=0}^k \frac{\theta_i^{x_i}}{x_i!} \\ &= n\theta_1 \sum_{s_k} x_1 (n-1)! \prod_{i=0}^k \frac{\theta_i^{x'_i}}{x'_i!} \\ &= n\theta_1 \sum_{s_k} [(x_1 - 1) + 1] (n-1)! \prod_{i=0}^k \frac{\theta_i^{x'_i}}{x'_i!} \\ &= n\theta_1 [(n-1)\theta_1 + 1]. \end{aligned}$$

Segue-se então o resultado proposto.

Analogamente, para obtermos a $\text{Cov}(X_1, X_2)$, basta calcularmos $E(X_1 X_2)$, já que $\text{Cov}(X_1, X_2) = E(X_1 X_2) - E(X_1)E(X_2)$.

Para isso, note que

$$\begin{aligned} E(X_1 X_2) &= \sum_{s_k} x_1 x_2 n! \prod_{i=0}^k \frac{\theta_i^{x_i}}{x_i!} \\ &= n(n-1) \theta_1 \theta_2 \sum_{s_k} (n-2)! \prod_{i=0}^k \frac{\tilde{x}_i}{\tilde{x}_i!}, \end{aligned}$$

onde

$$\tilde{x} = (x_1 - 1, x_2 - 1, \dots, x_k).$$

Como antes, o somatório pode ser restrito ao conjunto \tilde{s}_k dos inteiros \tilde{x}_i não negativos, de modo que $E(X_1 X_2) = n(n-1) \theta_1 \theta_2$. Como consequência, $\text{Cov}(X_1, X_2) = -n \theta_1 \theta_2$.

Exemplo 2.2 - Se $X|\theta \sim M_3(10; \theta)$, como no Exemplo 2.1, então, conclui-se a partir do resultado acima que

$$E(X) = (2, 3, 1.5),$$

enquanto que a matriz Σ de variâncias e covariâncias resulta em

$$\Sigma = \begin{bmatrix} 1.6 & -.6 & -.3 \\ -.6 & 2.1 & -.45 \\ -.3 & -.45 & 1.23 \end{bmatrix}$$

O leitor pode verificar que os resultados acima, bem como quaisquer momentos $E(X_1^{r_1} \dots X_k^{r_k})$ podem ser obtidos facilmente a partir da função característica para X .

O resultado seguinte apresenta uma caracterização da distribuição multinomial em termos da distribuição Poisson.

Resultado 2.2 - Suponha, conhecido $\lambda = (\lambda_0, \lambda_1, \dots, \lambda_k)$, que X_0, X_1, \dots, X_k sejam independentes com distribuição Poisson com pa

râmetros $\lambda_0, \lambda_1, \dots, \lambda_k$, respectivamente. Seja n um inteiro positivo. Então, dado λ ,

$$(X_0, X_1, \dots, X_k) \mid \sum_{i=0}^k X_i = n \sim M_k(n; \theta),$$

onde

$$\theta_i = \frac{\lambda_i}{\sum_{j=0}^k \lambda_j}, \quad i = 1, \dots, k.$$

Reciprocamente, conhecido θ , se $X \sim M_k(n; \theta)$ e $\tilde{X}_0, \tilde{X}_1, \dots, \tilde{X}_k$ são independentes com distribuição Poisson com parâmetros $\theta_0, \theta_1, \dots, \theta_k$ respectivamente, então

$$X \sim (\tilde{X}_0, \tilde{X}_1, \dots, \tilde{X}_k \mid \sum_{i=0}^k \tilde{X}_i = n).$$

Demonstração: A distribuição conjunta de X_0, X_1, \dots, X_k , dados $\lambda_0, \lambda_1, \dots, \lambda_k$ é

$$P(x_0, x_1, \dots, x_k) = \prod_{i=0}^k \frac{e^{-\lambda_i} \lambda_i^{x_i}}{x_i!}.$$

Além disso, como $\sum_{i=0}^k X_i \mid \lambda_0, \lambda_1, \dots, \lambda_k \sim \text{Poisson} \left(\sum_{i=0}^k \lambda_i \right)$, temos

$$P\left(\sum_{i=0}^k X_i = n\right) = e^{-\ell} \ell^n / n!$$

onde

$$\ell = \sum_{i=0}^k \lambda_i.$$

Portanto,

$$\begin{aligned}
 P(x_0, x_1, \dots, x_k \mid \sum_{i=0}^k X_i = n) &= \frac{P(\sum_{i=0}^k X_i = n \mid x_0, x_1, \dots, x_k) P(x_0, x_1, \dots, x_k)}{P(\sum_{i=0}^k X_i = n)} \\
 &= \frac{P(x_0, x_1, \dots, x_k)}{P(\sum_{i=0}^k X_i = n)}, \text{ caso } \sum_{i=0}^k x_i = n, \\
 &= 0, \text{ caso contrário.}
 \end{aligned}$$

Substituindo-se as expressões para $P(x_0, x_1, \dots, x_k)$ e $P(\sum_{i=0}^k X_i = n)$ no quociente acima obtém-se o resultado desejado, ou seja

$$X_0, X_1, \dots, X_k \mid \sum_{i=0}^k X_i = n \sim M_k(n; \theta),$$

com

$$\theta_i = \lambda_i / \sum_{j=0}^k \lambda_j, \quad i = 1, \dots, k.$$

Reciprocamente, se $X \mid \theta \sim M_k(n; \theta)$ podemos reproduzir a distribuição de $X \mid \theta$ a partir da distribuição de $\tilde{X}_0, \tilde{X}_1, \dots, \tilde{X}_k$ condicional em $\sum_{i=0}^k \tilde{X}_i = n$, tomando-se para isso

$$(\tilde{X}_0, \tilde{X}_1, \dots, \tilde{X}_k \mid \theta_0, \theta_1, \dots, \theta_k) \sim \prod_{i=0}^k \text{Poisson}(\theta_i).$$

O resultado seguinte enuncia uma propriedade reprodutiva da distribuição multinomial.

Resultado 2.3 - Se $X_1, \dots, X_r \mid \theta \sim \prod_{q=1}^r M_k(n_q; \theta)$ então

$$\sum_{q=1}^r X_q \mid \theta \sim M_k(\sum_{q=1}^r n_q; \theta).$$

Demonstração: Pelo Resultado 2.2, para $q = 1, \dots, r$,

$$X_{q0}, X_{q1}, \dots, X_{qk} | \theta \sim \bar{X}_{q0}, \bar{X}_{q1}, \dots, \bar{X}_{qk} \left| \sum_{i=0}^k X_{qi} = n_q, \theta, \right.$$

$$X_{q0} \amalg \dots \amalg X_{qk} | \theta \text{ e } X_{qi} | \theta_i \sim \text{Poisson}(\theta_i), i = 0, 1, \dots, k.$$

Além disso, por hipótese, $X_1 \amalg \dots \amalg X_r | \theta$. Logo

$$X_{11} \amalg \dots \amalg X_{rk} | \theta.$$

Por outro lado, temos

$$\sum_{q=1}^r (X_{q0}, X_{q1}, \dots, X_{qk}) | \theta \sim \sum_{q=1}^r \bar{X}_{q0}, \sum_{q=1}^r \bar{X}_{q1}, \dots, \sum_{q=1}^r \bar{X}_{qk} | n_0, n_1, \dots, n_k,$$

$$\sum_{q=1}^r \bar{X}_{q0} \amalg \dots \amalg \sum_{q=1}^r \bar{X}_{qk} | \theta,$$

com

$$\sum_{q=1}^r \bar{X}_{qi} | \theta_i \sim \text{Poisson}(r\theta_i), i = 0, 1, \dots, k.$$

Aplicando o Resultado 2.2 novamente, agora com $\lambda_i = r\theta_i$, $i = 0, 1, \dots, k$ e $n = \sum_{i=0}^k n_i$, obtemos o resultado proposto.

O resultado seguinte trata das distribuições marginais de uma distribuição multinomial.

Resultado 2.4 - Se $X | \theta \sim M_k(n; \theta)$ e A_0, A_1, \dots, A_r é uma partição do conjunto de índices $\{0, 1, 2, \dots, k\}$ então

$$\left(\sum_{A_1} X_i, \dots, \sum_{A_r} X_i \right) | \theta \sim M_r(n; \theta'),$$

onde

$$\theta' = \left(\sum_{A_1} \theta_i, \dots, \sum_{A_r} \theta_i \right).$$

Demonstração: Pelo Resultado 2.2 podemos escrever

$$(X_0, X_1, \dots, X_k) | \theta \sim \bar{X}_0, \bar{X}_1, \dots, \bar{X}_k \left| \sum_{i=0}^k \bar{X}_i = n, \theta, \right.$$

com

$$\bar{X}_0 \perp\!\!\!\perp \dots \perp\!\!\!\perp \bar{X}_k | \theta \quad \text{e} \quad \bar{X}_i | \theta_i \sim \text{Poisson}(\theta_i), \quad i = 0, 1, \dots, k.$$

Como consequência,

$$\left(\sum_{A_0} X_i, \dots, \sum_{A_r} X_i \right) | \theta \sim \left(\sum_{A_0} \bar{X}_i, \dots, \sum_{A_r} \bar{X}_i \right) \left| \sum_q \sum_{A_q} \bar{X}_i = n, \theta, \right.$$

com

$$\sum_{A_0} \bar{X}_i \perp\!\!\!\perp \dots \perp\!\!\!\perp \sum_{A_r} \bar{X}_i | \theta \quad \text{e} \quad \sum_{A_q} \bar{X}_i \sim \text{Poisson} \left(\sum_{A_q} \theta_i \right), \quad q = 0, \dots, r.$$

Como $\sum_{q=0}^r \sum_{A_q} \theta_i = 1$, segue-se do Resultado 2.2 novamente, o resultado proposto.

Exemplo 2.3 - Suponha que $X | \theta \sim M_3(10; \theta)$, com $\theta = (.20, .30, .15)$, como no Exemplo 2.1. Tomemos $A_0 = \{0, 1\}$, $A_1 = \{2, 3\}$. Então

$$\sum_{A_1} X_i | \theta = X_2 + X_3 | \theta \sim M_1(10; \theta_2 + \theta_3),$$

ou

$$X_2 + X_3 | \theta \sim M_1(10; .45).$$

Analogamente

$$X_0 + X_1 | \theta \sim M_1(10; .55),$$

$$X_1 + X_3 | \theta \sim M_1(10; .35),$$

$$X_2 | \theta \sim M_1(10; .30).$$

Note que, em geral, se $X | \theta \sim M_k(n; \theta)$ então $X_i | \theta \sim M_1(n; \theta_i)$; $i = 1, \dots, k$.

Resultado 2.5 - Se $X|\theta \sim M_k(n;\theta)$ então

$$(X_1, \dots, X_q) | x_{q+2}, \dots, x_k, x_0, \theta \sim M_q(n - x_{q+2} - \dots - x_k - x_0; \theta'),$$

com

$$\theta'_i = \frac{\theta_i}{\theta_1 + \dots + \theta_{q+1}}, \quad i = 1, \dots, q,$$

$$\theta'_0 = \frac{\theta_{q+1}}{\theta_1 + \dots + \theta_{q+1}}.$$

Demonstração: Do Resultado 2.4, temos

$$(X_{q+2}, \dots, X_k, X_0) | \theta \sim M_{k-q}(n; \theta'),$$

com

$$\theta' = (\theta_{q+2}, \dots, \theta_k, \theta_0),$$

enquanto que $X \sim M_k(n; \theta)$. Dividindo a expressão para $M_k(n; \theta)$ pela expressão de $M_{k-q}(n; \theta')$ obtemos a distribuição procurada.

Exemplo 2.4 - Tabelas de Contingência 3x3.

Suponha que $X|\theta \sim M_8(n; \theta)$, como numa tabela de contingências 3x3:

x_{11}	x_{12}	x_{13}	$x_{1.}$
x_{21}	x_{22}	x_{23}	$x_{2.}$
x_{31}	x_{32}	x_{33}	$x_{3.}$
$x_{.1}$	$x_{.2}$	$x_{.3}$	n

Aplicando o Resultado 2.4 obtemos

$$(X_1, X_2) | \theta \sim M_2(n; \theta'), \quad \theta' = (\theta_1, \theta_2), \quad \theta'_0 = \theta_3.$$

Este mesmo Resultado, diz-nos que

$$(X_{i1}, X_{i2}, X_{i3}) | \theta \sim M_3(n; \theta'_i) ,$$

com

$$\theta'_i = (\theta_{i1}, \theta_{i2}, \theta_{i3}), \quad \theta'_{0i} = 1 - \theta_{i.}, \quad i = 1, 2, 3 .$$

Podemos então aplicar o Resultado 2.5 para obter a distribuição de probabilidade de cada linha da Tabela de Contingências, condicional na respectiva soma, ou na soma das outras duas linhas.

Temos então

$$(X_{i1}, X_{i2}) | x_{i.}, \quad \theta \sim M_2(x_{i.}; \theta'_i) ,$$

com

$$\theta'_i = (\theta_{i1}, \theta_{i2}) / \theta_{i.}, \quad \theta'_{0i} = \theta_{i3} / \theta_{i.} .$$

O seguinte resultado, expressa a distribuição de $X | \theta$ em termos das distribuições condicionais, de cada linha da tabela, na sua respectiva soma, e em termos das distribuições destas somas.

Resultado 2.6 - Se $X | \theta \sim M_{r-1}(n; \theta)$, como numa Tabela de contingências $r \times r$, então $P(X | \theta)$ pode ser escrita como

$$P(X | \theta) = \left[\prod_{i=1}^r P(X_{i1}, \dots, X_{i,r-1} | x_{i.}, \theta) \right] P(X_{1.}, \dots, X_{r-1.} | \theta) .$$

Demonstração: Temos:

$$P(X | \theta) = n! \pi \frac{\theta_i^{x_i}}{x_i!} = n! \frac{\theta_{11}^{x_{11}} \dots \theta_{rr}^{x_{rr}}}{x_{11}! \dots x_{rr}!} =$$

$$= \left[\prod_{i=1}^r \frac{x_{i.}!}{x_{i1}! \dots x_{ir}!} \left(\frac{\theta_{i1}}{\theta_{i.}} \right)^{x_{i1}} \dots \left(\frac{\theta_{ir}}{\theta_{i.}} \right)^{x_{ir}} \right] \times \frac{n!}{x_{1.}! \dots x_{r.}!} \theta_{1.}^{x_{1.}} \dots \theta_{r.}^{x_{r.}}$$

Dos Resultados 2.4 e 2.5, como no exemplo anterior, reconhecemos cada um dos primeiros r fatores acima como as probabilidades de cada linha da Tabela, condicional na sua respectiva soma, e reconhecemos o último fator como sendo a distribuição de probabilidade conjunta para as somas destas r linhas.

Corolário 2.1 - Se $X|\theta \sim M_{r^2-1}(n;\theta)$, como no Resultado 2.6, então

$$P(X|x_{1.}, \dots, x_{r-1.}, \theta) = \prod_{i=1}^r P(X_{i1}, \dots, X_{i,r-1} | x_{i.}, \theta)$$

e portanto, conhecidos $\theta, x_{1.}, \dots, x_{r-1.}$,

$$(X_{11}, \dots, X_{1,r-1}) \perp\!\!\!\perp \dots \perp\!\!\!\perp (X_{r1}, \dots, X_{r,r-1}).$$

Demonstração: Como

$$P(X|\theta) = P(X|x_{1.}, \dots, x_{r-1.}, \theta) P(X_{1.}, X_{2.}, \dots, X_{r-1.} | \theta),$$

do Resultado 2.6 obtemos a igualdade proposta.

O seguinte resultado será utilizado mais a frente e confere ao Resultado 2.6 uma forma de representação canônica para $P(X|\theta)$.

Resultado 2.7 - Se $X|\theta \sim M_{r^2-1}(n;\theta)$, como no Resultado 2.6, então a transformação

$$T: (\theta_{11}, \dots, \theta_{1r}, \dots, \theta_{r1}, \dots, \theta_{r,r-1}) \rightarrow \\ \rightarrow (\lambda_{11}, \dots, \lambda_{1,r-1}, \dots, \lambda_{r1}, \dots, \lambda_{r,r-1}, \eta_1, \dots, \eta_{r-1})$$

dada por

$$\lambda_{11} = \frac{\theta_{11}}{\theta_{1.}}, \dots, \lambda_{1,r-1} = \frac{\theta_{1,r-1}}{\theta_{1.}}$$

⋮

$$\lambda_{r1} = \frac{\theta_{r1}}{\theta_{r.}}, \dots, \lambda_{r,r-1} = \frac{\theta_{r,r-1}}{\theta_{r.}}$$

$$\eta_1 = \theta_{1.}, \eta_2 = \theta_{2.}, \dots, \eta_{r-1} = \theta_{r-1.}$$

é uma transformação biunívoca definida definida em $\{0 < \theta_{11} + \dots + \theta_{r,r-1} < 1; 0 < \theta_{ij} < 1\}$ sobre o cubo unitário com dimensão $r^2 - 1$. Além disso, o Jacobiano da transformação T é

$$J = \eta_1^{r-1} \eta_2^{r-1} \dots \eta_{r-1}^{r-1} (1 - \eta_1 - \dots - \eta_{r-1})^{r-1}.$$

A demonstração é recomendada ao leitor, como exercício.

Exemplo 2.5 - Vejamos o caso de uma Tabela de Contingências 2x2:

x_{11}	x_{12}
x_{21}	x_{22}

n

θ_{11}	θ_{12}
θ_{21}	θ_{22}

1

Para obtermos a representação canônica de $P(X|\theta)$ usamos a transformação T no caso $r = 2$:

$$\lambda_{11} = \theta_{11} / (\theta_{11} + \theta_{12}),$$

$$\lambda_{21} = \theta_{21} / (\theta_{21} + \theta_{22}),$$

$$\eta_1 = \theta_{11} + \theta_{12},$$

de modo que

$$P(X|\theta) = \binom{x_1}{x_{11}} \lambda_{11}^{x_{11}} (1-\lambda_{11})^{x_1-x_{11}} \binom{x_2}{x_{21}} \lambda_{21}^{x_{21}} (1-\lambda_{21})^{x_2-x_{21}} \binom{n}{x_1} \eta_1^{x_1} (1-\eta_1)^{n-x_1},$$

$$0 < \lambda_{11} < 1, \quad 0 < \lambda_{21} < 1, \quad 0 < \eta_1 < 1.$$

2.4 - DISTRIBUIÇÃO HIPERGEOMÉTRICA MULTIVARIADA

Na seção 2.2, mostramos como a distribuição hipergeométrica pode ser gerada de um processo permutável de variáveis de Bernoulli. A generalização natural desse fato aparece ao considerarmos um processo permutável onde seus elementos são vetores aleatórios que tomam valores no conjunto

$$\{e_0, e_1, \dots, e_k\}$$

onde

$$e_0 = (0, 0, \dots, 0), \quad e_1 = (1, 0, \dots, 0), \quad e_2 = (0, 1, \dots, 0), \dots$$

e

$$e_k = (0, 0, \dots, 1)$$

são vetores de \mathbb{R}^k .

Contudo, para uma melhor ilustração, apresentamos essa distribuição em um contexto mais intuitivo.

Uma amostra de tamanho n é tomada de uma população finita de tamanho N ($>n$), a qual é particionada em $k+1$ ($\leq N$) categorias. As frequências populacionais (número de elementos em cada categoria) são representadas por $\psi_0, \psi_1, \dots, \psi_k$ e além disso $N = \psi_0 + \psi_1 + \dots + \psi_k$. Baseado na amostra, deseja-se inferir sobre $(\psi_0, \psi_1, \dots, \psi_k)$. Correspondendo a cada ψ_i ($i=0, 1, \dots, k$), x_i é a frequência amostral de i -ésima categoria onde $n = x_0 + x_1 + \dots + x_k$.

Uma outra forma de descrever esse problema é considerar uma caixa com N bolas de $k+1$ diferentes cores que são denominadas c_0, c_1, \dots, c_k . O número de bolas com a i -ésima cor ($i = 0, 1, \dots, k$) é denotado por ψ_i onde $N = \psi_0 + \dots + \psi_k$. Suponha que as N bolas são separadas em duas outras caixas de tal forma que uma caixa (caixa número 1) contém n bolas e a outra (caixa número 2) $N - n$ bolas. O estatístico pode observar a composição da caixa 1 representada pelo vetor (x_1, \dots, x_k) de frequências amostrais. Agora, a quantidade desconhecida para o estatístico é o vetor $(\psi_1 - x_1, \dots, \psi_k - x_k)$ que representa a composição da caixa 2. É importante lembrar que $\psi_0 = N - (\psi_1 + \dots + \psi_k)$ e $x_0 = n - (x_1 + \dots + x_k)$.

Vamos supor que, a cada (bola) unidade j ($j = 1, \dots, N$) da população $P = \{1, 2, \dots, N\}$ (de bolas), associamos um vetor de incidência U_j o qual é igual a e_i se a unidade j tem cor c_i . O parâmetro de interesse é o vetor

$$\psi = (\psi_1, \psi_2, \dots, \psi_k) = U_1 + U_2 + \dots + U_N.$$

Dizer que a seleção é feita ao acaso, corresponde a dizer que a seleção é feita no conjunto P de rótulos que são não informativos sobre os valores de U_1, U_2, \dots, U_N .

Assim, seleção ao acaso corresponde a considerar-se que (U_1, U_2, \dots, U_N) é uma secção finita de um processo permutável e dessa forma qualquer vetor de n componentes extraídas de $\{U_1, U_2, \dots, U_N\}$ tem a mesma distribuição de (U_1, U_2, \dots, U_n) .

Este fato implica que $X = (X_1, \dots, X_k) = U_1 + U_2 + \dots + U_n$ tem a mesma distribuição do vetor de frequências de uma amostra casual de tamanho n .

Como no caso univariado, o nosso objetivo é encontrar a distribuição de $(X_1, \dots, X_k) | (\psi_1, \dots, \psi_k)$. Isto é, a distribuição

de $(U_1 + \dots + U_n) | (U_1 + \dots + U_N)$. Novamente, usando-se os resultados de De Finetti, existe um vetor $(\theta_0, \theta_1, \dots, \theta_k)$, onde $\theta_i \geq 0$ ($i = 0, 1, \dots, k$) e $\theta_0 + \theta_1 + \dots + \theta_k = 1$, tal que $\prod_{j=0}^N U_j | (\theta_0, \dots, \theta_k)$ e $P(U_j = e_i) = \theta_i$, $\forall i = 0, 1, \dots, k$ e $j = 1, 2, \dots, N$. Usando a notação $\theta = (\theta_1, \dots, \theta_k)$, esses resultados produzem os seguintes fatos:

- (i) $\psi | \theta \sim M_k(N; \theta)$
- (ii) $X | \theta \sim M_k(n; \theta)$
- (iii) $(\psi - X) | \theta \sim M_k(N - n; \theta)$
- (iv) $(\psi - X) \perp\!\!\!\perp X | \theta$

Usando-se os resultados da seção anterior e seguindo as mesmas etapas do caso univariado, obtemos a seguinte expressão para a função de probabilidade condicional de (X_1, \dots, X_k) dado (ψ_1, \dots, ψ_k) :

$$p(x_1, \dots, x_k | \psi) = \frac{\binom{\psi_0}{x_0} \binom{\psi_1}{x_1} \dots \binom{\psi_k}{x_k}}{\binom{N}{n}},$$

onde $\psi_0 + \psi_1 + \dots + \psi_k = N$, $x_0 + x_1 + \dots + x_k = n$ e $\max\{0; n - (N - \psi_i)\} \leq x_i \leq \{\min n; \psi_i\}$. Essa distribuição é denominada Hipergeométrica multivariada e a seguinte notação será usada na sequência:

$$X | \psi \sim H_k(N, n, \psi).$$

Sugerimos que o leitor encontre $E\{X_i | \psi_i\}$, $\text{Var}\{X_i | \psi_i\}$ e $\text{Cov}\{X_i, X_j | (\psi_i, \psi_j)\}$. Um bom treino seria encontrar a distribuição de $a_0 X_0 + \dots + a_k X_k | (\psi_1, \dots, \psi_k)$ onde (a_0, a_1, \dots, a_k) é um vetor fixado de zeros e uns.

2.5 A DISTRIBUIÇÃO DIRICHLET

Na Seção 2 deste Capítulo apresentamos a distribuição Multinomial $M_k(n; \theta)$. Nesta Seção apresentaremos a classe de distribuições Dirichlet definidas nos pontos θ , tais como os parâmetros

do modelo Multinomial.

Definição 2.2 - A função

$$g(\theta_1, \dots, \theta_k) = \Gamma\left(\sum_{i=0}^k \alpha_i\right) \prod_{i=0}^k \frac{\theta_i^{\alpha_i - 1}}{\Gamma(\alpha_i)},$$

definida para os pontos $\theta_0, \theta_1, \dots, \theta_k$, tais que $\sum_{i=0}^k \theta_i = 1$, $0 < \theta_i < 1$, $i = 0, 1, \dots, k$, onde $\alpha_i > 0$, $i = 0, 1, \dots, k$ são parâmetros, chama-se função de densidade Dirichlet com parâmetros $\alpha_0, \alpha_1, \dots, \alpha_k$.

Notação: O símbolo $\theta \sim D_k(\alpha; \alpha_0)$ será lido como "θ tem distribuição Dirichlet, com dimensão k, parâmetros α e parâmetro complementar α_0 " e representa a distribuição de probabilidade conjunta de $\theta = (\theta_1, \dots, \theta_k)$ definida acima, cujos parâmetros são $\alpha = (\alpha_1, \dots, \alpha_k)$ e α_0 . Salientamos α_0 como "complementar" apenas no sentido de que $\alpha_0 - 1$ é o expoente de $\theta_0 = 1 - \sum_{i=1}^k \theta_i$.

Resultado 2.8 - A função dada pela Definição 2.2 é uma função de densidade.

Antes de demonstrarmos este resultado, vejamos como a distribuição Dirichlet está naturalmente "conjugada" com a distribuição Multinomial. Para isso, lembremos que a notação $X|\theta \sim M_k(n; \theta)$ indica que a distribuição conjunta de $X = (X_1, \dots, X_k)$ condicional em $\theta = (\theta_1, \dots, \theta_k)$ é multinomial (dimensão k) com parâmetros n e probabilidades de ocorrência

$$\theta_1, \dots, \theta_k, \theta_0 = 1 - \sum_{i=1}^k \theta_i.$$

A partir da distribuição condicional de X e da distribuição marginal de θ podemos calcular a distribuição conjunta de X e θ , bem como a distribuição condicional de θ dado X. Temos o seguinte resultado:

Resultado 2.9 - Se $\theta \sim D_k(a; a_0)$ e $X|\theta \sim M_k(n; \theta)$ então

$$\theta|x \sim D_k(a+x; a_0+x_0).$$

Demonstração: Pela fórmula de Bayes, temos:

$$f(\theta|x) = P(x|\theta)g(\theta) / \int P(x|\theta)g(\theta)d\theta.$$

Como $P(x|\theta)g(\theta) = x! \prod_0^k \frac{\theta_i^{x_i}}{x_i!} \Gamma(\sum_0^k a_i) \prod_0^k \frac{\theta_i^{a_i-1}}{\Gamma(a_i)}$, segue-se que o denominador de $f(\theta|x)$ é dado por

$$\frac{x! \Gamma(\sum_0^k a_i)}{\prod_0^k x_i! \Gamma(a_i)} \int_0^k \prod_0^k \theta_i^{a_i+x_i-1} d\theta.$$

Note que o fator multiplicando o resultado da integral acima aparece em ambos os lados do quociente $f(\theta|x)$, simplificando o resultado para

$$f(\theta|x) = \prod_0^k \theta_i^{a_i+x_i-1} / \int_0^k \prod_0^k \theta_i^{a_i+x_i-1} d\theta.$$

Pelo Resultado 2.6, a distribuição condicional de X dado N é Dirichlet (dimensão k) com parâmetros a_1+x_1, \dots, a_k+x_k e parâmetro complementar a_0+x_0 ; ou seja:

$$X|x \sim D_k(a+x; a_0+x_0).$$

Exemplo 2.6 - Suponha $k=3$,

$$X|\theta \sim M_3(20; \theta),$$

com

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

$$\theta_0 = 1 - .6 = .4.$$

Além disso

$$\theta \sim D_3(a; a_0)$$

com

$$a = (2, 2, 1) ,$$

$$a_0 = 2 .$$

Então, pelo resultado anterior, a distribuição a posteriori de θ dado x , quando

$$x = (1, 5, 2)$$

$$x_0 = 20 - 8 = 12$$

é ainda um membro da família de distribuições Dirichlet com parâmetros

$$a' = a + x = (2, 2, 1) + (1, 5, 2) = (3, 7, 3)$$

e

$$a'_0 = a_0 + x_0 = 2 + 12 = 14 .$$

Vejamos em seguida a demonstração do Resultado 2.6, do qual devemos mostrar que

$$\int_S \prod_{i=0}^k \theta_i^{a_i-1} d\theta = \frac{\prod_{i=0}^k \Gamma(a_i)}{\Gamma(\sum_{i=0}^k a_i)} ,$$

para todo $a_i > 0$, $i = 0, 1, \dots, k$, $k > 1$ inteiro, e $S = \{(\theta_1, \dots, \theta_k); 0 < \sum_{i=1}^k \theta_i < 1, 0 < \theta_i < 1, i = 1, \dots, k\}$. Tal resultado depende apenas da seguinte transformação de variáveis:

$$\theta_1 = v_1 ,$$

$$\theta_2 = v_2 (1 - v_1) ,$$

$$\theta_3 = v_3 (1 - v_1) (1 - v_2) ,$$

$$\vdots$$

$$\theta_k = v_k (1 - v_1) (1 - v_2) \dots (1 - v_{k-1}) ,$$

que relaciona univocamente o conjunto S e o cubo unitário

$(0,1) \times \dots \times (0,1)$ em \mathbb{R}^k . A matriz $d\theta/dv$ é triangular, de modo que o seu determinante J é o produto dos elementos da diagonal.

Ou seja

$$J = 1 \times (1-v_1)(1-v_1)(1-v_2) \dots (1-v_1) \dots (1-v_{k-1}) \\ (1-v_1)^{k-1} (1-v_2)^{k-2} \dots (1-v_{k-1}) > 0,$$

já que

$$(v_1, \dots, v_k) \in (0,1) \times \dots \times (0,1)$$

Note ainda que

$$\begin{aligned} \theta_0 &= 1 - \theta_1 - \dots - \theta_k \\ &= 1 - v_1 - v_2(1-v_1) - v_3(1-v_1)(1-v_2) - \dots - v_k(1-v_1)(1-v_2) \dots (1-v_{k-1}) \\ &= (1-v_1)(1-v_2)(1-v_3) \dots (1-v_k). \end{aligned}$$

Substituindo os valores de θ_i por seus transformados e multiplicando pelo Jacobiano da transformação, obtemos

$$\begin{aligned} & \int_S \prod_{i=1}^k \theta_i^{a_i-1} d\theta = \\ & \int_0^1 \dots \int_0^1 v_1^{a_1-1} (v_2(1-v_1))^{a_2-1} \dots (v_k(1-v_{k-1}) \dots (1-v_1))^{a_k-1} \times \\ & \quad ((1-v_1)(1-v_2) \dots (1-v_k))^{a_0-1} \times \\ & \quad (1-v_1)^{k-1} (1-v_2)^{k-2} \dots (1-v_{k-1}) dv \\ &= \int_0^1 \dots \int_0^1 v_1^{a_1-1} (1-v_1)^{a_0+a_2+\dots+a_k-1} v_2^{a_2-1} (1-v_2)^{a_0+a_3+\dots+a_k-1} \dots v_k^{a_k-1} (1-v_k)^{a_0-1} dv \\ &= \frac{\Gamma(a_1)\Gamma(a_0+a_2+\dots+a_k)}{\Gamma(\sum_1^k a_i)} \times \frac{\Gamma(a_2)\Gamma(a_0+a_3+\dots+a_k)}{\Gamma(a_0+a_2+\dots+a_k)} \times \dots \times \frac{\Gamma(a_k)\Gamma(a_0)}{\Gamma(a_k+a_0)} = \frac{\prod_0^k \Gamma(a_i)}{\Gamma(\sum_0^k a_i)}, \end{aligned}$$

concluindo a demonstração.

Exemplo 2.7 - Suponha $k=1$, de modo que a distribuição de $\theta \sim D_1(a; a_0)$ é dada por

$$g(\theta) = \frac{\Gamma(a_1 + a_0)}{\Gamma(a_1)\Gamma(a_0)} \theta^{a_1-1} (1-\theta)^{a_0-1}, \quad 0 < \theta < 1,$$

para $a > 0, a_0 > 0$. Neste caso dizemos que X tem uma distribuição Beta com parâmetros a_1 e a_0 . Usaremos a notação $X \sim B(a_1; a_0)$ para indicar a distribuição Beta. Em particular, note que a distribuição Beta (1,1) coincide com a distribuição uniforme em (0,1).

Exemplo 2.8 - Suponha ainda $k=1$. Portanto, como vimos no Resultado 2.7, se $X|\theta \sim M_1(n; \theta_1)$ e $\theta \sim D_1(a_1; a_0)$ então $\theta|x \sim D_1(x_1 + a_1; x_0 + a_0)$. Isto é, conjugando-se a distribuição Binomial $M_1(n; \theta_1)$ com a distribuição Beta (a_1, a_0) obtemos a posteriori uma distribuição Beta com novos parâmetros $a'_1 = a_1 + x_1$, $a'_0 = a_0 + x_0$.

Resultado 2.10 - Momentos de θ .

Se $\theta \sim D_k(a; a_0)$ então

$$E(\theta_1^{r_1} \dots \theta_k^{r_k}) = \frac{\Gamma(a_0)}{\Gamma(a_0 + \sum_{i=1}^k a_i + \sum_{i=1}^k r_i)} \prod_{j=1}^k \frac{\Gamma(a_j + r_j)}{\Gamma(a_j)},$$

para todo conjunto de inteiros não negativos r_1, \dots, r_k .

Demonstração: O resultado acima segue diretamente da Definição 2.2.

Corolário 2.2 - Se $\theta \sim D_k(a; a_0)$ então

$$a) \quad E(\theta_j) = \frac{a_j}{\sum_{i=1}^k a_i}, \quad j = 1, \dots, k,$$

$$b) \text{Var}(\theta_j) = \frac{a_j \sum_0^k a_i - a_j^2}{\left(\sum_0^k a_i\right)^2 (1 + \sum_0^k a_i)}, \quad j = 1, \dots, k,$$

$$c) \text{Cov}(\theta_i, \theta_j) = \frac{-a_i a_j}{\left(\sum_0^k a_i\right) (1 + \sum_0^k a_i)}, \quad i \neq j.$$

Demonstração: Os resultados acima seguem diretamente do Resultado 2.8, escolhendo-se r_1, \dots, r_k apropriadamente.

Corolário 2.3 - Se $\theta \sim D_k(a; a_0)$ então a distribuição conjunta de $\theta_1, \dots, \theta_k$ é idêntica à distribuição conjunta de v_1, \dots, v_k onde

$$v_i \sim D_1(a_i; a_0 + \sum_{j>i} a_j)$$

e v_1, \dots, v_k são independentes.

Demonstração: Basta definir v_1, \dots, v_k a partir de $\theta_1, \dots, \theta_k$ pela transformação vista, na prova do Resultado 2.6.

O Corolário 2.1 mostra-nos, em particular, que a distribuição marginal de qualquer θ_i é Dirichlet com dimensão unitária (ou seja, uma distribuição Beta). Mais precisamente, se $\theta \sim D_k(a; a_0)$ então

$$\theta_i \sim D_1(a_i; a_0 + \sum_{j \neq i} a_j),$$

pois $v_1 \equiv \theta_1$ na transformação construída na prova do Resultado 2.6 e tal resultado vale simetricamente para qualquer outro índice escolhido.

Em seguida, vejamos como a distribuição Dirichlet pode ser caracterizada a partir da distribuição Gama. Este resultado será utilizado logo em seguida, quando estudaremos a distribuição conjunta para subgrupos de X_1, \dots, X_k , ou para certas transformações destas variáveis. Para isso, lembremos que a distribuição

Gama com parâmetros a e c é dada por

$$g(v) = \frac{c}{\Gamma(a)} (cv)^{a-1} e^{-cv},$$

para $v \geq 0$, $c > 0$ e $a > 0$.

O leitor pode verificar, por exemplo, que

$$E(v) = \frac{a}{c}, \quad \text{Var}(v) = \frac{a}{c^2}.$$

Resultado 2.11 - Suponha que $v_i \sim \text{Gama}(a_i, c)$, $i = 0, 1, \dots, k$, independentes, e defina

$$\theta_j = \frac{v_j}{v_0 + \sum_{i=1}^k v_i}, \quad j = 1, \dots, k.$$

Então $(\theta_1, \dots, \theta_k) \sim D_k(a; a_0)$, onde $a = (a_1, \dots, a_k)$. Reciprocamente, se $\theta \sim D_k(a; a_0)$ então para quaisquer distribuições Gama, $v_i \sim \text{Gama}(a_i; c)$, $i = 0, 1, \dots, k$ independentes, temos que a distribuição conjunta de $v_j / (v_0 + \sum_{i=1}^k v_i)$, $j = 1, \dots, k$ é $D_k(a; a_0)$.

Demonstração: Para obtermos a distribuição conjunta de $\theta_1, \dots, \theta_k$ começamos com a transformação

$$v_1 = \theta_1 (v_0 + \sum_{i=1}^k v_i)$$

$$\vdots$$

$$v_k = \theta_k (v_0 + \sum_{i=1}^k v_i)$$

$$v_0 = w$$

Como

$$\sum_{i=1}^k v_i = \sum_{i=1}^k \theta_i (v_0 + \sum_{i=1}^k v_i)$$

segue-se que

$$\sum_{i=1}^k v_i = v_0 \frac{\sum_{i=1}^k \theta_i}{1 - \sum_{i=1}^k \theta_i}$$

Portanto

$$v_j = \bar{\theta}_j \left(v_0 + v_0 \frac{\sum_{i=1}^k \theta_i}{1 - \sum_{i=1}^k \theta_i} \right)$$

$$= v_0 \theta_j / (1 - \sum_{i=1}^k \theta_i), \quad j = 1, \dots, k,$$

de modo que a transformação acima pode ser reescrita na forma

$$v_1 = v_0 \theta_1 / (1 - \sum_{i=1}^k \theta_i)$$

$$v_2 = v_0 \theta_2 / (1 - \sum_{i=1}^k \theta_i)$$

$$\vdots$$

$$v_k = v_0 \theta_k / (1 - \sum_{i=1}^k \theta_i)$$

$$v_0 = \omega$$

Calculando $\partial(\theta_1, \dots, \theta_k, \theta_0) / \partial(\theta_1, \dots, \theta_k, \omega)$ obtemos

$$v_0 \begin{bmatrix} \frac{1 - \sum \theta_i + \theta_1}{(1 - \sum \theta_i)^2} & \frac{\theta_1}{(1 - \sum \theta_i)^2} & \dots & \frac{\theta_1}{(1 - \sum \theta_i)^2} & 0 \\ \frac{\theta_2}{(1 - \sum \theta_i)^2} & \frac{1 - \sum \theta_i + \theta_2}{(1 - \sum \theta_i)^2} & \dots & \frac{\theta_2}{(1 - \sum \theta_i)^2} & 0 \\ \vdots & \vdots & & \vdots & \\ \frac{\theta_k}{(1 - \sum \theta_i)^2} & \frac{\theta_k}{(1 - \sum \theta_i)^2} & \dots & \frac{1 - \sum \theta_i + \theta_k}{(1 - \sum \theta_i)^2} & 0 \\ 0 & 0 & & 0 & \frac{1}{v_0} \end{bmatrix}$$

Chamando $\theta_0 = 1 - \sum_{i=1}^k \theta_i$, podemos calcular o determinante J da matriz acima;

$$J = \det \frac{v_0}{\theta_0^2} \begin{bmatrix} \theta_0 + \theta_1 & \theta_1 & & \theta_1 & 0 \\ \theta_2 & \theta_0 + \theta_2 & \dots & \theta_2 & 0 \\ \vdots & \vdots & & \vdots & \vdots \\ \theta_k & \theta_k & \dots & \theta_0 + \theta_k & 0 \\ 0 & 0 & \dots & 0 & \theta_0^2/v_0 \end{bmatrix}$$

$$= \left(\frac{v_0}{\theta_0^2}\right)^{k+1} \frac{\theta_0^2}{v_0} \det \begin{bmatrix} \theta_0 + \theta_1 & \theta_0 & \dots & \theta_0 \\ \theta_2 & -\theta_0 & \dots & 0 \\ \vdots & \vdots & & \vdots \\ \theta_k & 0 & \dots & -\theta_0 \end{bmatrix}$$

$$= \left(\frac{v_0}{\theta_0^2}\right)^{k+1} \frac{\theta_0^2}{v_0} \det \begin{bmatrix} \theta_0 + \sum_{i=1}^k \theta_i & 0 & \dots & 0 \\ \theta_2 & -\theta_0 & \dots & 0 \\ \vdots & \vdots & & \vdots \\ \theta_k & 0 & & -\theta_0 \end{bmatrix}$$

$$= \left(\frac{v_0}{\theta_0^2}\right)^{k+1} \frac{\theta_0^2}{v_0} \theta_0^{k-1} (-1)^{k-1}$$

Portanto, o Jacobiano $|J|$ da transformação é

$$|J| = \frac{v_0^k}{\theta_0^{k+1}} = \frac{\omega^k}{\theta_0^{k+1}} .$$

Por outro lado, a distribuição conjunta de v_0, v_1, \dots, v_k é

$$f(v_0, v_1, \dots, v_k) = \prod_0^k c a_i^{-c} v_i^{a_i-1} \Gamma(a_i), \quad a_i > 0, \quad c > 0,$$

e, como consequência, a distribuição conjunta de $\theta_1, \dots, \theta_k, \omega$ é

$$\begin{aligned} g(\theta_1, \dots, \theta_k, \omega) &= f(v_0, v_1, \dots, v_k) |J| \\ &= c^{\sum_0^k a_i} e^{-c(\omega + \frac{1}{\theta_0} \sum_1^k \theta_i)} \frac{\omega^{a_0-1}}{\Gamma(a_0)} \left[\prod_1^k \left(\frac{\theta_i}{\theta_0} \right)^{a_i-1} \frac{1}{\Gamma(a_i)} \right] \frac{\omega^k}{\theta_0^{k+1}} d\theta_1 \dots d\theta_k d\omega \\ &= c^{\sum_0^k a_i} e^{-\frac{c\omega}{\theta_0} \left(\frac{v_0}{\theta_0} \right)} \prod_1^k a_i^{-k} \left[\prod_1^k \frac{\theta_i^{a_i-1}}{\Gamma(a_i)} \right] \frac{\omega^{a_0-1}}{\Gamma(a_0)} \frac{\omega^k}{\theta_0^{k+1}} \\ &= \left[\prod_1^k \frac{\theta_i^{a_i-1}}{\Gamma(a_i)} \right] c^{\sum_0^k a_i} e^{-c \frac{\omega}{\theta_0} \left(\frac{\omega}{\theta_0} \right)} \prod_0^k a_i^{-1} \frac{\omega^{a_0-1}}{\Gamma(a_0)} \theta_0^{-1}. \end{aligned}$$

Logo, a distribuição marginal de $\theta_1, \dots, \theta_k$ é

$$\prod_0^k \frac{\theta_i^{a_i-1}}{\Gamma(a_i)} \int c^{\sum_0^k a_i} e^{-c \frac{\omega}{\theta_0} \left(\frac{\omega}{\theta_0} \right)} \prod_0^k a_i^{-1} \theta_0^{-1} d\omega.$$

Entretanto, a integral acima, vista com argumento $v = \omega/\theta_0$ e comparada com a integral de uma distribuição Gama $\left(\sum_0^k a_i, c \right)$, tem como valor $\Gamma\left(\sum_0^k a_i\right)$. Neste caso, obtemos o resultado desejado, ou seja

$$\theta_1, \dots, \theta_k \sim \Gamma\left(\sum_0^k a_i\right) \prod_0^k \frac{\theta_i^{a_i-1}}{\Gamma(a_i)};$$

$$\theta \sim D_k(a; a_0).$$

Reciprocamente, suponha que $\theta \sim D_k(a; a_0)$. Então, fixado $c > 0$ qualquer e tomando $v_i \sim \text{Gama}(a_i; c)$, $i = 0, 1, \dots, k$ independentes, temos

$$\frac{v_1}{v_0 + \sum_1^k v_i}, \dots, \frac{v_k}{v_0 + \sum_1^k v_i} \sim D_k(a; a_0).$$

Vejamos em seguida uma aplicação imediata desta caracterização:

Resultado 2.12 - Se $\theta \sim D_k(a; a_0)$ e I_1, \dots, I_r é uma partição dos índices $\{1, \dots, k\}$ então

$$\left(\sum_{j \in I_1} \theta_j, \dots, \sum_{j \in I_r} \theta_j \right) \sim D_r \left(\sum_{j \in I_1} a_j, \dots, \sum_{j \in I_r} a_j; a_0 \right).$$

Demonstração: Do Resultado 2.9 sabemos que

$$(\theta_1, \dots, \theta_k) \sim \left(\frac{v_1}{\sum_0^k v_i}, \dots, \frac{v_k}{\sum_0^k v_i} \right), \text{ com}$$

$v_i \sim \text{Gama}(a_i; c)$, independentes, $i = 0, \dots, k$.

Logo

$$\left(\sum_{j \in I_1} \theta_j, \dots, \sum_{j \in I_r} \theta_j \right) \sim \left(\frac{\sum_{j \in I_1} v_j}{\sum_0^k v_i}, \dots, \frac{\sum_{j \in I_r} v_j}{\sum_0^k v_i} \right), \text{ com}$$

$\sum_{j \in I_t} v_j \sim \text{Gama} \left(\sum_{j \in I_t} a_j; c \right)$, independentes, $t = 1, \dots, r$.

e independentes de v_0 , já que I_1, \dots, I_r é uma partição dos índices $\{1, \dots, k\}$ e v_0, v_1, \dots, v_k são independentes.

Aplicando novamente o Resultado 2.9 para a nova coleção de $r+1$ distribuições Gama, independentes, notando ainda que

$$\sum_0^k v_i = v_0 + \sum_{t=1}^r \sum_{j \in I_t} v_j,$$

obtemos o resultado proposto.

Corolário 2.4 - Se $\theta \sim D_k(a; a_0)$ e $I \subset \{1, \dots, k\}$ então

$$\sum_{i \in I} \theta_i \sim D_1\left(\sum_{i \in I} a_i; a_0 + \sum_{j \in I^c} a_j\right).$$

Com efeito, aplicando o Resultado 2.10 à partição (I, I^c) obtemos

$$\left(\sum_I \theta_i, \sum_{I^c} \theta_i\right) \sim D_2\left(\sum_I a_i, \sum_{I^c} a_i; a_0\right).$$

Do Corolário 2.2 segue-se que

$$\sum_{i \in I} \theta_i \sim D_1\left(\sum_{i \in I} a_i; a_0 + \sum_{i \in I^c} a_i\right).$$

Resultado 2.13 - Se $\theta \sim D_k(a; a_0)$ e I_1, \dots, I_r é uma partição de $\{0, 1, \dots, k\}$ com, digamos,

$\{1\} \subset I_1, \{2\} \subset I_2, \dots, \{r\} \subset I_r$, então

$$\frac{\theta_j}{\sum_{i \in I_j} \theta_i} \sim D_1\left(a_j; \sum_{\substack{i \in I_j \\ i \neq j}} a_i\right), \quad j = 1, \dots, r, \text{ independentes.}$$

Antes de demonstrarmos o resultado acima, vejamos o significado da notação no enunciado proposto:

Exemplo 2.9 - Suponha $k = 8$ e

$$I_1 = \{1, 2, 3\},$$

$$I_2 = \{4, 5, 6\},$$

$$I_3 = \{7, 8, 0\}.$$

Então, digamos,

$$\frac{\theta_1}{\theta_1 + \theta_2 + \theta_3} \sim D_1(a_1; a_2 + a_3) ,$$

$$\frac{\theta_4}{\theta_4 + \theta_5 + \theta_6} \sim D_1(a_4; a_5 + a_6) ,$$

$$\frac{\theta_7}{\theta_7 + \theta_8 + \theta_0} \sim D_1(a_7; a_8 + a_0) ,$$

a as três novas distribuições são independentes.

Vejamos a prova do resultado.

Demonstração: Pelo Resultado 2.9 sabemos que

$$(\theta_1, \dots, \theta_k) \sim \left(\frac{v_1}{\sum_0^k v_i}, \dots, \frac{v_k}{\sum_0^k v_i} \right) ,$$

onde $v_j \sim \text{Gama}(a_j, c)$, $j = 0, 1, \dots, k$, independentes.

Portanto,

$$\left(\frac{\theta_1}{\sum_{I_1} \theta_i}, \dots, \frac{\theta_r}{\sum_{I_r} \theta_i} \right) \sim \left(\frac{v_1}{\sum_{I_1} v_i}, \dots, \frac{v_r}{\sum_{I_r} v_i} \right) .$$

Como existe pelo menos um outro índice j^* em I_j além do índice j , aplicando o Resultado 2.9 para as distribuições Gama em I_j podemos concluir que a distribuição conjunta de

$$\frac{v_i}{v_{j^*} + \sum_{\substack{\ell \in I_j \\ \ell = j^*}} v_\ell} , \quad i \in I_j , \quad i = j^*$$

é

$$D_{\#I_j - 1}(a; a_{j^*}) .$$

onde $\# I_j$ é o número de elementos em I_j , e α o vetor dos respectivos parâmetros $\alpha_i, i \in I_j, i \neq j^*$.

Do Corolário 2.3, obtemos a distribuição marginal de $v_j / \sum_{i \in I_j} v_i$;

$$\frac{v_i}{\sum_{i \in I_j} v_i} \sim D_1(\alpha_j; \sum_{\substack{i \in I_j \\ i \neq j}} \alpha_i),$$

que é independente de qualquer outro quociente

$$\frac{v_t}{\sum_{i \in I_t} v_i}$$

já que v_0, v_1, \dots, v_k são independentes e I_1, \dots, I_r é uma partição de $\{0, 1, \dots, k\}$, demonstrando o resultado proposto. Note, em particular, que neste caso

$$\left(\frac{x_1}{\sum_{I_1} x_i}, \dots, \frac{x_r}{\sum_{I_r} x_i} \right) \sim \prod_{j=1}^r D_1(\alpha_j; \sum_{\substack{i \in I_j \\ i \neq j}} \alpha_i).$$

O resultado seguinte está relacionado com a representação canônica para a distribuição multinomial, como no Resultado 2.6, e será aplicado durante o próximo capítulo no estudo de algumas Tabelas de Contingências.

Resultado 2.14 - Se $\theta \sim D_{r-1}(\alpha; \alpha_{rr})$, como numa Tabela de Contingências $r \times r$, e T é a transformação definida no Resultado 2.7, então

$$(\lambda_1, \dots, \lambda_r, \eta_1, \dots, \eta_{r-1}) \sim \left[\prod_{i=1}^r D_{r-1}(\alpha_i; \alpha_{ir}) \right] D_{r-1}(\alpha_1, \dots, \alpha_{r-1}; \alpha_r.)$$

onde

$$\lambda_i = (\lambda_{i1}, \dots, \lambda_{i,r-1}), \quad \alpha_i = (\alpha_{i1}, \dots, \alpha_{i,r-1}), \quad i = 1, \dots, r.$$

Demonstração: Com efeito,

$$(\theta_{11}, \dots, \theta_{r,r-1}) \sim \Gamma(a_{11} + \dots + a_{rr}) \times$$

$$\prod_{i=1}^r \frac{\theta_{i1}^{a_{i1}-1}}{\Gamma(a_{i1})} \dots \frac{\theta_{i,r-1}^{a_{i,r-1}-1}}{\Gamma(a_{i,r-1})} \frac{\theta_{ir}^{a_{ir}-1}}{\Gamma(a_{ir})},$$

e portanto, a distribuição das novas variáveis é dada por

$$\begin{aligned} & (\lambda_1, \dots, \lambda_r, \eta_1, \dots, \eta_{r-1}) \sim \\ & \Gamma(a_{i.}) \frac{\lambda_{i1}^{a_{i1}-1}}{\Gamma(a_{i1})} \dots \frac{\lambda_{i,r-1}^{a_{i,r-1}-1}}{\Gamma(a_{i,r-1})} \frac{(1-\lambda_{i1} - \dots - \lambda_{i,r-1})^{a_{ir}-1}}{\Gamma(a_{ir})} \eta_1^{a_{i1} + \dots + a_{ir} - r} \times \\ & \times J \times \frac{\Gamma(a_{1.} + \dots + a_{r.})}{\Gamma(a_{1.}) \dots \Gamma(a_{r.})}, \end{aligned}$$

onde

$$J = \prod_{i=1}^r \eta_i^{r-1}, \quad \eta_r = 1 - \eta_1 - \dots - \eta_{r-1}$$

é o Jacobiano da transformação T. Agrupando os termos em $\lambda_1, \dots, \lambda_r, \eta_1, \dots, \eta_{r-1}$ segue-se o resultado proposto.

Corolário 2.5 - $\lambda_1 \amalg \dots \amalg \lambda_r \amalg (\eta_1, \dots, \eta_{r-1})$.

Demonstração: Basta notar que

$$\lambda_1 \sim D_{r-1}(a_1; a_{1r}), \dots, \lambda_r \sim D_{r-1}(a_r; a_{rr}),$$

$$(\eta_1, \dots, \eta_{r-1}) \sim D_{r-1}(a_1, \dots, a_{r-1}; a_r).$$

Exemplo 2.10 - Representação canônica para Tabelas de Contingências 2×2 .

θ_{11}	θ_{12}
θ_{21}	θ_{22}

1

$$\lambda_{11} = \theta_{11} / (\theta_{11} + \theta_{12}) ,$$

$$\lambda_{21} = \theta_{21} / (\theta_{21} + \theta_{22}) ,$$

$$\eta_1 = \theta_{11} + \theta_{12} .$$

Se $\theta \sim D_3(a; a_{22})$, então

$$(\lambda_{11}, \lambda_{21}, \eta_1) \sim D_1(a_{11}; a_{12}) D_1(a_{21}; a_{22}) D_1(a_{11} + a_{12}; a_{21} + a_{22}) ,$$

$$\lambda_{11} \sim D_1(a_{11}; a_{12}) ,$$

$$\lambda_{21} \sim D_1(a_{21}; a_{22}) ,$$

$$\eta_1 \sim D_1(a_{11} + a_{12}; a_{21} + a_{22}) ,$$

e

$$\lambda_{11} \perp\!\!\!\perp \lambda_{21} \perp\!\!\!\perp \eta_1 .$$

Exemplo 2.11 - Representação canônica para Tabelas de Contingências 3×3 .

θ_{11}	θ_{12}	θ_{13}
θ_{21}	θ_{22}	θ_{23}
θ_{31}	θ_{32}	θ_{33}

1

$$\lambda_{11} = \frac{\theta_{11}}{\theta_{1.}} , \quad \lambda_{12} = \frac{\theta_{12}}{\theta_{1.}} ,$$

$$\lambda_{21} = \frac{\theta_{21}}{\theta_{2.}} , \quad \lambda_{22} = \frac{\theta_{22}}{\theta_{2.}} ,$$

$$\lambda_{31} = \frac{\theta_{31}}{\theta_{3.}} , \quad \lambda_{32} = \frac{\theta_{32}}{\theta_{3.}} .$$

$$\eta_1 = \theta_{1.} , \quad \eta_2 = \theta_{2.} .$$

$$0 < \lambda_{i1} + \lambda_{i2} < 1 , \quad 0 < \eta_1 + \eta_2 < 1 , \quad i = 1, 2, 3 .$$

Se $\theta \sim D_8(a; a_{33})$ então.

$$(\lambda, \eta) \sim \left[\prod_{i=1}^3 D_2(a_{i1}, a_{i2}; a_{i3}) \right] D_2(a_1, a_2; a_3),$$

onde

$$\lambda = (\lambda_{11}, \lambda_{12}, \lambda_{21}, \lambda_{22}, \lambda_{31}, \lambda_{32}), \quad \eta = (\eta_1, \eta_2).$$

Além disso

$$(\lambda_{i1}, \lambda_{i2}) \sim D_2(a_{i1}, a_{i2}; a_{i3}), \quad i = 1, 2, 3,$$

$$(\eta_1, \eta_2) \sim D_2(a_1, a_2; a_3),$$

e

$$(\lambda_{11}, \lambda_{12}) \perp\!\!\!\perp (\lambda_{21}, \lambda_{22}) \perp\!\!\!\perp (\lambda_{31}, \lambda_{32}) \perp\!\!\!\perp (\eta_1, \eta_2).$$

O resultado seguinte caracteriza a independência das distribuições marginais de $\theta \sim D_{r^2-1}(a; a_0)$ como numa Tabela de Contingências em termos da reparametrização canônica.

Resultado 2.15 - Sob a representação canônica,

$$\theta_{ij} = \theta_i \cdot \theta_j, \quad i, j = 1, \dots, r \iff \lambda_{1\ell} = \lambda_{2\ell} = \dots = \lambda_{r\ell}, \quad \ell = 1, \dots, r-1.$$

Demonstração: A tabela dos θ_{ij} , segundo a transformação canônica T , é representada pelas r linhas.

$$\lambda_{i1} \eta_i \quad \lambda_{i2} \eta_i \quad \dots \quad (1 - \lambda_{i1} - \lambda_{i2} - \dots - \lambda_{i,r-1}) \eta_i, \quad i = 1, \dots, r$$

onde

$$\eta_r = 1 - \eta_1 - \dots - \eta_{r-1}$$

Chamando a soma dos componentes da coluna ℓ por g_ℓ , obtemos para $i = 1, \dots, r$; $j = 1, \dots, r-1$,

$$\begin{aligned} \theta_{ij} = \theta_{i.} \theta_{.j} &\longleftrightarrow \lambda_{1\ell} \eta_1 = g_\ell \eta_1, \quad \eta_1 \neq 0 \\ &\lambda_{2\ell} \eta_2 = g_\ell \eta_2, \quad \eta_2 \neq 0 \\ &\vdots \\ &\lambda_{r\ell} (1 - \eta_1 - \dots - \eta_{r-1}) = g_\ell (1 - \eta_1 - \dots - \eta_{r-1}) \\ &\longleftrightarrow \lambda_{1\ell} = \lambda_{2\ell} = \dots = \lambda_{r\ell}, \quad \ell = 1, \dots, r-1. \end{aligned}$$

Para $j = r$, note que a igualdade dos λ 's implica em

$$\theta_{i.} \theta_{.r} = \eta_i (1 - \lambda_{i1} - \dots - \lambda_{i,r-1}) = \theta_{ir}, \quad i = 1, \dots, r-1,$$

e

$$\theta_{r.} \theta_{.r} = (1 - \eta_1 - \dots - \eta_{r-1}) (1 - \lambda_{r1} - \dots - \lambda_{r,r-1}) = \theta_{rr},$$

concluindo a demonstração.

2.6 DISTRIBUIÇÃO DIRICHLET-MULTINOMIAL

Quando os dados seguem um modelo multinomial, normalmente a distribuição de Dirichlet é usada como distribuição a priori. Dessa forma, o estudo da mistura da distribuição multinomial por uma de Dirichlet torna-se relevante pois, a distribuição marginal dos dados é nesse caso uma mistura desse tipo. Especificamente, seja (X_0, X_1, \dots, X_k) um vetor de inteiros não negativos onde $X_0 + X_1 + \dots + X_k = n$ é fixado e seja $(\theta_0, \theta_1, \dots, \theta_k)$ um vetor de reais não negativos tais que $\theta_0 + \theta_1 + \dots + \theta_k = 1$. Com a notação anterior, seja $\theta = (\theta_1, \dots, \theta_k)$, $X = (X_1, \dots, X_k)$ e $a = (a_1, \dots, a_k)$.

Definição 2.3 - Se $\theta \sim D_k(a, a_0)$ e $X|\theta \sim M_k(n; \theta)$, então a distribuição (marginal) de X é chamada de Dirichlet-Multinomial (DM) com parâmetro $(n; a_1, a_2, \dots, a_k, a_0)$ escreve-se $X \sim DM_k(n; a, a_0)$ ou $X \sim DM_k(n; a_1, \dots, a_k, a_0)$ e quando $k=1$, escrevemos $X \sim Bb(n; a_1, a_0)$ no lugar de $X \sim DM_1(n; a_1, a_0)$, para indicar que X tem distribuição Beta-Binomial.

Não é difícil mostrar que a função de probabilidade associada com essa distribuição tem a seguinte expressão:

$$p(x_1, \dots, x_k) = \frac{n! \Gamma(A)}{\Gamma(A+n)} \prod_{i=0}^k \frac{\Gamma(a_i + x_i)}{x_i! \Gamma(a_i)}$$

onde $A = a_0 + a_1 + \dots + a_k$.

Apresentamos a seguir algumas propriedades importantes da distribuição DM. Seja $X = (X_1, \dots, X_k) \sim DM_k(n; a, a_0)$.

Resultado 2.16 - Se (i_0, i_1, \dots, i_k) é uma permutação de $(0, 1, \dots, k)$, então

$$(X_{i_1}, \dots, X_{i_k}) \sim DM(n; a_{i_1}, \dots, a_{i_k}, a_{i_0}).$$

Resultado 2.17 - Se $m \in \{0, 1, \dots, k\}$ é fixado então, para $b_m = a_1 + a_2 + \dots + a_m$,

$$(X_1, \dots, X_m) \sim DM(n; a_1, a_2, \dots, a_m, A - b_m) \text{ e}$$

$$X_1 + \dots + X_m \sim Bb(n; b_m, A - b_m).$$

Esses dois resultados são consequências imediatas de propriedades análogas das distribuições Multinomial e Dirichlet e as demonstrações são deixadas a cargo do leitor.

Resultado 2.18 - Acrescentando $n_m = X_0 + X_1 + \dots + X_m$ a notação acima, temos que,

$$(X_1, \dots, X_m) | n_m \sim DM(n_m; a_1, \dots, a_m, a_0)$$

Demonstração: Basta notar que a função de probabilidade de $(X_1, \dots, X_m) | n_m$ é obtida dividindo-se a função de probabilidade de (X_1, \dots, X_m) pela função de probabilidade de n_m .

O próximo resultado é uma caracterização importante da distribuição DM.

Seja (X_0, X_1, \dots, X_k) um vetor aleatório de inteiros não negativos com $n = X_0 + \dots + X_k$ fixado. Escolha um inteiro $m \in \{2, \dots, k-1\}$ e, como antes, denote $n_m = X_0 + \dots + X_m$. Considere o seguinte conjunto de condições:

- (i) $(X_0, X_1, \dots, X_m) \perp\!\!\!\perp (X_{m+1}, \dots, X_k) | n_m$,
- (ii) $(X_1, \dots, X_m) | n_m \sim DM(n_1; a_1, \dots, a_m, a_0)$
- (ii)' $(X_{m+1}, X_{m+2}, \dots, X_{k-1}) | n_m \sim DM(n - n_m; a_{m+1}, \dots, a_{k-1}, a_k)$ e
- (iii) $n_m \sim Bb(n; b_m, A - b_m)$.

Resultado 2.19 - O conjunto de condições acima é necessário e suficiente para

- (iv) $(X_1, \dots, X_k) \sim DM(n; a_1, \dots, a_k, a_0)$.

Demonstração: Pelos resultados 2.16, 2.17 e 2.18, (iv) implica (ii), (ii)' e (iii). Para provar as implicações remanescentes necessita-se apenas notar que a função de probabilidade DM pode ser fatorada como:

$$p(x_1, \dots, x_k) = \left\{ \frac{n! \Gamma(A)}{(A+n) n_m! (n-n_m)! \Gamma(b_m) \Gamma(A-b_m)} \frac{\Gamma(b_m + n_m) (A - b_m + n - n_m)}{\Gamma(b_m) \Gamma(A - b_m)} \right\} \times$$

$$\times \left\{ \frac{n_m! \Gamma(b_m)}{\Gamma(b_m + n_m)} \prod_{L=0}^m \frac{\Gamma(a_i + x_i)}{x_i! \Gamma(a_i)} \right\} \times \left\{ \frac{(n-n_m)! \Gamma(A-b_m)}{\Gamma(A-b_m + n - n_m)} \prod_{L=m+1}^k \frac{\Gamma(a_i + x_i)}{x_i! \Gamma(a_i)} \right\}$$

onde, como antes, $A = a_0 + \dots + a_k$ e $b_m = a_0 + \dots + a_m$.

O primeiro fator é a função de probabilidade de uma $Bb(n; b_m, A-b_m)$, o segundo é a função de probabilidade de uma $DM(n_m; a_1, \dots, a_k, a_0)$ e o terceiro é a função de probabilidade de uma $DM(n-n_m; a_{m+1}, \dots, a_k)$.

Uma propriedade interessante da DM é apresentada abaixo onde considera-se uma sequência (n_0, n_1, \dots, n_k) com $n_m = X_0 + \dots + X_m$, como anteriormente. É evidente que, $n_0 = X_0$ e $n_k = n$.

Corolário 2.6 - Se $(X_1, \dots, X_k) \sim DM(n; a_1, \dots, a_k, a_0)$, então (n_0, n_1, \dots, n_k) forma uma cadeia de Markov.

A demonstração é deixada a cargo do leitor.

Para apresentar o vetor de médias e a matriz de covariância da distribuição DM, introduzimos o vetor

$a = (a_0, a_1, \dots, a_k)$, e a matriz

$$A = \begin{bmatrix} a_0 & 0 & \dots & 0 \\ 0 & a_1 & \dots & 0 \\ \vdots & & & \\ 0 & 0 & \dots & a_k \end{bmatrix}$$

Com os resultados dessa seção, temos que $X_i \sim Bb(n; a_i, A-a_i)$ e $X_i + X_j \sim Bb(n; a_i + a_j, A-a_i - a_j)$ $\forall i, j = 0, 1, \dots, k$ e $i \neq j$. Com simples cálculos e usando as propriedades de esperança condicional temos que $E\{X_i\} = n \frac{a_i}{A}$, $\text{Var}\{X_i\} = \left[a_i - \frac{a_i^2}{A} \right] \frac{A+n}{A(A+1)} n$, e

$$\begin{aligned} \text{Var}\{X_i + X_j\} &= \left[a_i + a_j - \frac{(a_i + a_j)^2}{A} \right] \frac{A+n}{A(A+1)} n = \\ &= \text{Var}\{X_i\} + \text{Var}\{X_j\} + 2\text{Cov}\{X_i, X_j\} . \end{aligned}$$

Esta última igualdade mostra que

$$\text{Cov}\{X_i, X_j\} = - \frac{a_i a_j}{A} \frac{A+n}{A(A+1)} n.$$

Finalmente, o vetor médio e a matriz de covariância são dados por:

$$E\{X\} = \frac{n}{A} a \quad e$$

$$\text{Var}\{X\} = \left[A - \frac{1}{A} a' a \right] \frac{A+n}{A(A+1)} n$$

onde a' é a transposta de a .

Para finalizar este capítulo, recomendamos que o leitor verifique todos os resultados aqui discutidos inclusive, verificando que a estatística de Bose-Einstein (veja Feller, 1968) da Mecânica Estatística é um caso particular da distribuição DM.

CAPÍTULO 3

INFERÊNCIAS PARA MODELOS DISCRETOS

3.1 DISTRIBUIÇÕES A PRIORI, A POSTERIORI

Lembremos, do capítulo anterior, que a distribuição a posteriori de θ dado $X = x$, quando a priori $\theta \sim D_k(a; a_0)$ e $X|\theta \sim M_k(n; \theta)$ é

$$\theta | X = x \sim D_k(a + x, a_0 + x_0).$$

Nesta seção, inicialmente estudaremos com mais detalhe a distribuição a priori $D_k(a; a_0)$, na qual $a = (1, \dots, 1)$ com k componentes e $a_0 = 1$. Escreveremos $\theta \sim D_k(1; 1)$, embora, de fato, $D(1; 1)$ seja a distribuição uniforme em $0 < \sum_{i=1}^k \theta_i < 1$, $0 < \theta_i < 1$.

Do capítulo anterior, obtemos

$$\theta_i \sim D_1(1; k)$$

$$E(\theta_i) = 1/(k+1),$$

$$\text{Var}(\theta_i) = \frac{k}{(k+1)^2(k+2)}, \quad i = 0, 1, \dots, k.$$

Note ainda que a função de densidade de θ_i é

$$g(\theta_i) = k(1 - \theta_i)^{k-1},$$

de modo que

$$P(\theta_i \leq c) = 1 - (1 - \theta_i)^k, \quad 0 < c < 1.$$

Por exemplo,

$$\theta_i \in (1 - \sqrt[k]{.95}, 1 - \sqrt[k]{.05})$$

com probabilidade 90%.

A função de densidade conjunta de $\theta_i, \theta_j, i \neq j$, como $(\theta_i, \theta_j) \sim D_2(1; k-1)$, é dada por

$$g(\theta_i, \theta_j) = k(k-1)(1 - \theta_i - \theta_j)^{k-2}, \quad 0 < \theta_i + \theta_j < 1, \quad k > 1.$$

Assim,

$$P(\theta_i \leq \theta_j) = 1/2.$$

Outros aspectos desta distribuição a priori para θ podem ser verificados a partir dos resultados do capítulo anterior, e permitem que você decida quanto a coerência entre tais resultados numéricos e a sua crença inicial sobre θ . No decorrer deste capítulo teremos a oportunidade de utilizar outras distribuições a priori, além da distribuição uniforme $D_k(1;1)$.

Vejam os alguns aspectos da distribuição a posteriori para θ , quando utilizamos a distribuição a priori $D_k(1;1)$.

Temos

$$\theta | x \sim D_k(1+x; 1+x_0),$$

e, portanto

$$\theta_i | x \sim D_1(1+x_i; k + \sum_{j \neq i} x_j),$$

ou, se

$$n = \sum_{j=0}^k x_j,$$

$$\theta_i | x \sim D_1(1+x_i; k+n-x_i),$$

$$E(\theta_i | x) = \frac{1+x_i}{n+k+1},$$

$$\text{Var}(\theta_i | x) = \frac{(k+n-x_i)(1+x_i)}{(k+1+n)^2(k+2+n)}.$$

A função de densidade de $\theta_i | x$ é, portanto,

$$g(\theta_i | x) = \theta_i^{x_i} (1 - \theta_i)^{k+n-x_i} / B(1+x_i, k+n-x_i), \quad i=0,1,\dots,k$$

onde, como sabemos, $B(a,b) = \Gamma(a)\Gamma(b)/\Gamma(a+b)$.

Exemplo 3.1

Se $\theta \sim D_1(1,1)$ e observamos $x = 8$ sucessos em $n = 10$ eventos de Bernoulli, então $X|\theta \sim M_1(10;\theta)$ e $\theta|X = 8 \sim D_1(9;3)$. Além disso, podemos calcular

$$E(\theta|X = 8) = \frac{9}{12} = 0.75,$$

$$\text{Var}^{1/2}(\theta|X = 8) = \left(\frac{3 \times 9}{12^2 \times 13}\right)^{1/2} = (0.014)^{1/2} = 0.12.$$

Na Figura 3.1 podemos ver os gráficos para $D_1(1,1)$ e $D_1(9;3)$, para alguns valores de θ .

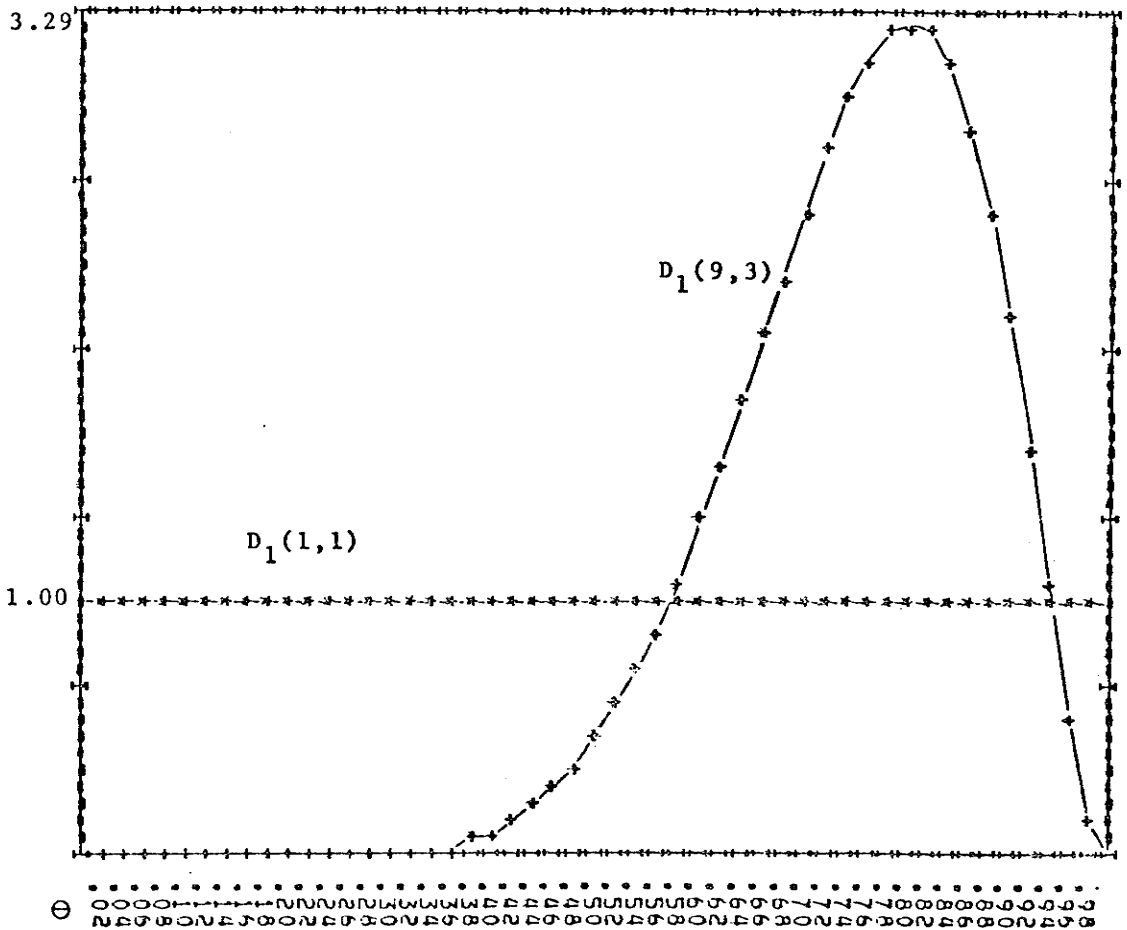


Figura 3.1

Exemplo 3.2

Numa sequência de 92 domicílios de um bairro do Rio de Janeiro, Sant'Anna (1982) observou a seguinte distribuição conjunta para os eventos $M \equiv$ mofo no domicílio, $F \equiv$ tabagismo domiciliar, $R \equiv$ problemas respiratórios nos infantes menores de 7 anos, (1 = presente, 0 = ausente).

M	F	R	Probabilidade	X	Freqüência
1	1	1	θ_1	X_1	09
1	1	0	θ_2	X_2	05
1	0	1	θ_3	X_3	01
1	0	0	θ_4	X_4	04
0	1	1	θ_5	X_5	21
0	1	0	θ_6	X_6	34
0	0	1	θ_7	X_7	04
0	0	0	θ_0	X_0	14

Temos, neste caso, $k=7$, e tomemos $\theta \sim D_7(1;1)$ como distribuição a priori. Isto significa que concordo, em particular, com as seguintes afirmações:

$$\theta_i \sim D_1(1;7),$$

$$E(\theta_i) = 1/8,$$

$$\text{Var}(\theta_i) = 0.012,$$

$$\theta_i \in (0.0073, 0.34) \text{ com probabilidade } 90\%.$$

Por outro lado, temos $X|\theta \sim M_7(92;\theta)$ e portanto $\theta|x \sim D_7(1+x;1+x_0)$. Vejamos alguns casos em particular:

1) π_1 : problemas respiratórios presentes:

$$\pi_1 = P(R=1) = \theta_1 + \theta_3 + \theta_5 + \theta_7$$

$$\pi_1|x \sim D_1(4+x_1+x_3+x_5+x_7; 1+x_0+3+x_2+x_4+x_6),$$

ou seja

$$\pi_1 | x \sim D_1(39; 61).$$

- 2) π_2 : Problemas respiratórios presentes, dado que mofo e tabagismo estão presentes.

$$\pi_2 = P(R = 1 | M = 1, F = 1) = \frac{\theta_1}{\theta_1 + \theta_2}$$

$$\pi_2 | x \sim D_1(1 + x_1; 1 + x_2),$$

ou seja

$$\pi_2 | x \sim D_1(10; 06).$$

- 3) π_3 : Problemas respiratórios presentes, dado que mofo está ausente e tabagismo presente.

$$\pi_3 = P(R = 1 | M = 0, F = 1) = \frac{\theta_5}{\theta_5 + \theta_6}$$

$$\pi_3 | x \sim D_1(1 + x_5; 1 + x_6),$$

ou seja

$$\pi_3 | x \sim D_1(22; 35).$$

- 4) π_4 : Problemas respiratórios presentes, dado que mofo está presente e tabagismo ausente.

$$\pi_4 = P(R = 1 | M = 1, F = 0) = \frac{\theta_3}{\theta_3 + \theta_4}$$

$$\pi_4 | x \sim D_1(1 + x_3; 1 + x_4),$$

ou seja

$$\pi_4 | x \sim D_1(2; 5).$$

- 5) π_5 : Problemas respiratórios presentes, dado que mofo e tabagismo estão ausentes.

$$\pi_5 = P(R = 1 | M = 0, F = 0) = \frac{\theta_7}{\theta_7 + \theta_0}$$

$$\pi_5 | x \sim D_1(1+x_7; 1+x_0), \text{ ou seja,}$$

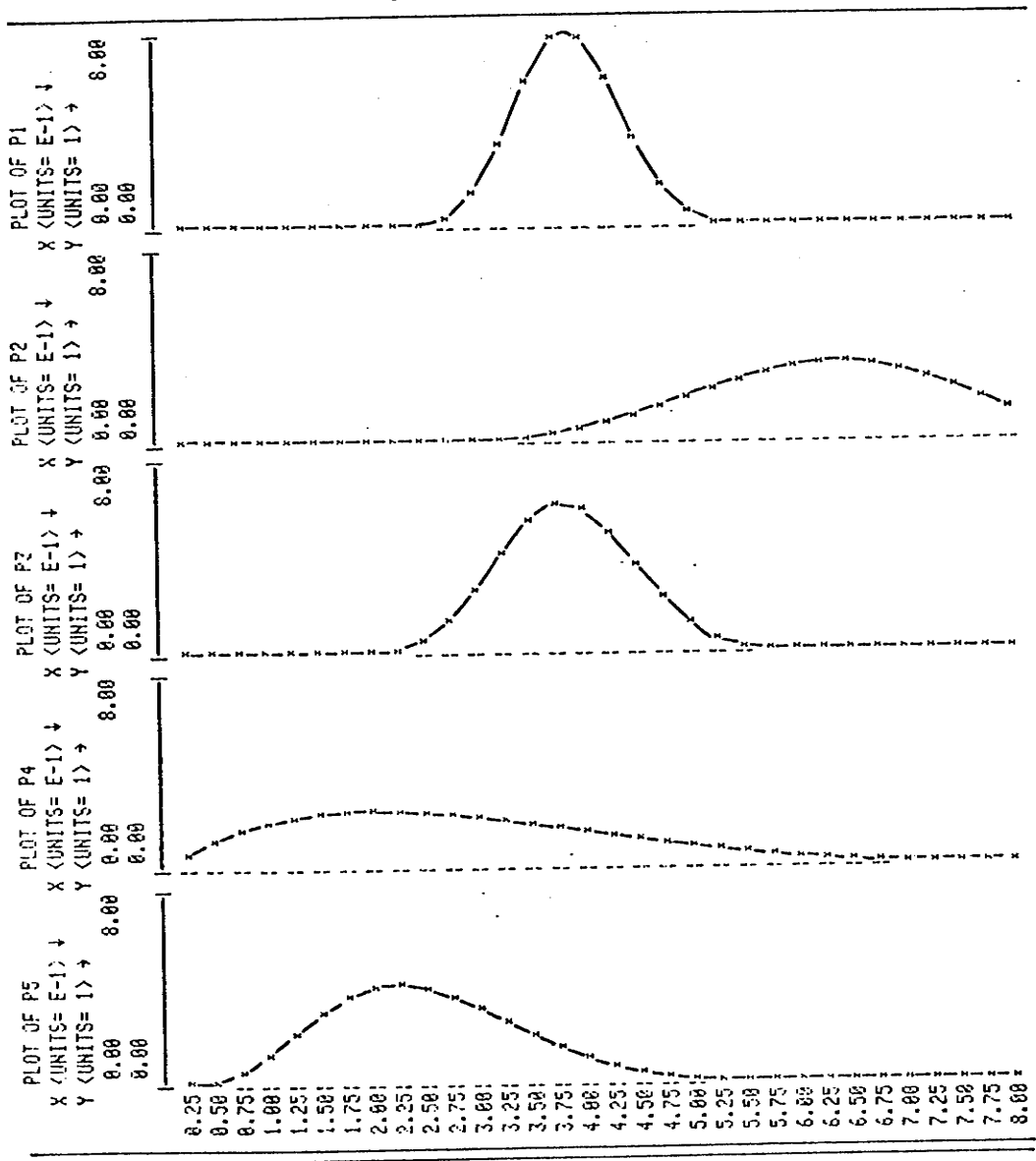
$$\pi_5 | x \sim D_1(5; 15).$$


Figura 3.1

Funções de densidade a posteriori para os eventos do Exemplo 3.1, para alguns valores de π .

3.2 REGIÕES DE CREDIBILIDADE

Nesta seção construiremos regiões R para uma variável aleatória $\theta \sim D_k(a; a_0)$ satisfazendo $P(\theta \in R) = r$, onde r é um valor especificado, tal como 80%, digamos. Em particular, calcularemos tais regiões para $\theta | X = x$ de modo que poderemos fazer afirmações probabilísticas sobre o estado da Natureza após termos observado o valor x de X .

Definição 3.1 - Dizemos que R é uma região de credibilidade r para θ se $P(\theta \in R) = r$, $0 < r < 1$. Quando além disso, existe um valor c tal que $\sup_{\theta \notin R^*} f(\theta) \leq c \leq \inf_{\theta \in R^*} f(\theta)$, onde $f(\theta)$ é a função de densidade de θ , dizemos que R^* é uma região de credibilidade r com densidade máxima.

Temos o seguinte resultado:

Resultado 3.1 - Se R^* é uma região de credibilidade com densidade máxima, então R^* tem volume mínimo dentre todas as regiões R de igual credibilidade.

Demonstração: Indiquemos $P(\theta \in R)$ por $P(R)$ e o volume de R por $V(R)$. Suponha que $V(R) < V(R^*)$. Como R^* é de densidade máxima, temos

$$f(\theta') \leq c \leq f(\theta), \quad \forall \theta \in R^*, \quad \forall \theta' \notin R^*.$$

Inicialmente, suponha que $R \cap R^* = \emptyset$. Então

$$\begin{aligned} P(R) &\leq V(R) \sup_{\theta \in R} f(\theta) \\ &\leq V(R)c, \text{ pois } R \cap R^* = \emptyset \\ &< V(R^*)c, \text{ por hipótese,} \\ &\leq V(R^*)c + \int_{R^*} (f(\theta) - c) d\theta = P(R^*). \end{aligned}$$

Caso $R \cap R^* \neq \emptyset$, note que, indicando por \bar{R} o conjunto complementar de R em Θ ,

$$P(R) = P(R \cap \bar{R}^*) + P(R \cap R^*),$$

$$P(R^*) = P(R^* \cap \bar{R}) + P(R \cap R^*).$$

Além disso, como $R \cap \bar{R}^*$ e $R^* \cap \bar{R} \subset R^*$ são disjuntos e $V(R \cap \bar{R}^*) < V(R^* \cap \bar{R})$, segue-se, como no caso anterior, que $P(R \cap \bar{R}^*) < P(R^* \cap \bar{R})$. Logo, $P(R) < P(R^*)$. Vimos, portanto, que

$$V(R) < V(R^*) \rightarrow P(R) < P(R^*),$$

e portanto,

$$P(R) = P(R^*) \rightarrow V(R^*) \leq V(R),$$

como queríamos demonstrar.

Nesta seção construiremos algumas regiões de credibilidade, sem no entanto procurar a região de volume mínimo. De início, consideremos o caso no qual $\theta \sim D_1(a_1; a_0)$. Queremos determinar um intervalo $(b, c) \subset (0, 1)$ tal que $\theta \in (b, c)$ com probabilidade r especificada. Para isso, avaliaremos a integral

$$P(\theta \leq c) = \int_0^c \frac{\Gamma(a_1 + a_0)}{\Gamma(a_1)\Gamma(a_0)} \theta^{a_1-1} (1-\theta)^{a_0-1} d\theta$$

com auxílio das tabelas para a distribuição F .

Fazendo a transformação

$$\omega = \frac{a_0}{a_1} \frac{\theta}{1-\theta}, \quad 0 < \theta < 1,$$

obtemos

$$\theta^{a_1-1} (1-\theta)^{a_0-1} d\theta = \frac{\omega^{(2a_1-2)/2} d\omega}{(1 + \frac{2a_1}{2a_0} \omega)^{(2a_1+2a_0)/2}},$$

ou seja,

$$W = F(2a_1, 2a_0).$$

Portanto

$$P(F_{\alpha}(2a_1, 2a_0) \leq W \leq F_{1-\alpha}(2a_1, 2a_0)) = 1 - 2\alpha,$$

ou

$$P\left(\frac{1}{F_{1-\alpha}(2a_0, 2a_1)} \leq W \leq F_{1-\alpha}(2a_1, 2a_0)\right) = 1 - 2\alpha,$$

para facilitar o acesso às tabelas usuais da distribuição F. Re-escrevendo os limites do intervalo para θ , obtemos

$$\frac{b}{1-b} = \frac{a_1}{a_0} \frac{1}{F_{1-\alpha}(2a_0, 2a_1)},$$

$$\frac{c}{1-c} = \frac{a_1}{a_0} F_{1-\alpha}(2a_1, 2a_0).$$

Em resumo, podemos enunciar o seguinte resultado:

Resultado 3.2 - Se $\theta \sim D_1(a_1; a_0)$ então:

$$P\left(\frac{a_1}{a_1 + a_0 F} < \theta < \frac{a_1}{a_1 + a_0 / F'}\right) = r,$$

onde

$$F = F_{\frac{1+r}{2}}(2a_0, 2a_1); \quad F' = F_{\frac{1+r}{2}}(2a_1, 2a_0).$$

Note que o intervalo acima é tal que

$$P(\theta < b) = P(\theta > c) = (1 - r)/2$$

Exemplo 3.3 - Suponha que $\pi \sim D_1(9; 3)$, como no Exemplo 3.1 deste capítulo. Para construirmos um intervalo de credibilidade 80% para π consultamos uma tabela da distribuição F, obtendo:

$$F = F_{90}(2a_0, 2a_1) = F_{90}(6; 18) = 2.13,$$

$$F' = F_{90}(2a_1, 2a_0) = F_{90}(18; 6) = 2.85.$$

Aplicando o Resultado 3.2 concluímos que

$$P(0.585 < \theta < 0.895) = 80\%.$$

O leitor pode verificar a posição deste intervalo na Figura 3.1 para $D_1(9;3)$. Note que a função de densidade assume valores diferentes nos extremos do intervalo, não sendo este portanto, de densidade máxima. Na Tabela 3.1 exibimos alguns valores da função de densidade e da distribuição $D_1(9;3)$ e verifica-se que o intervalo $(0.62, 0.92)$ é de (aproximadamente) mesma credibilidade, 80%, tendo entretanto, densidade máxima. Seu comprimento é apenas ligeiramente menor que o comprimento do intervalo $(0.58, 0.89)$.

Exemplo 3.4 - Retomemos o Exemplo 3.1. Para calcular intervalos de credibilidade 90% para as distribuições a posteriori $\pi_i | x$, $i = 1, \dots, 5$, aplicamos o Resultado 3.2 em cada caso, e como resultado obtemos:

<u>Distribuição</u>	<u>Intervalo de Credibilidade 90%</u>
$\pi_1 x \sim D_1(39, 61)$	(0.31, 0.49)
$\pi_2 x \sim D_1(10, 06)$	(0.39, 0.79)
$\pi_3 x \sim D_1(22, 35)$	(0.28, 0.51)
$\pi_4 x \sim D_1(2, 5)$	(0.11, 0.70)
$\pi_5 x \sim D_1(5, 15)$	(0.14, 0.43)

Em seguida, construiremos regiões bivariadas de credibilidade para $\theta \sim D_2(a_1; a_0)$. Para isso, aplicaremos o Corolário 5.1 do Capítulo 2, de modo que a transformação

$$T: S \rightarrow Q.$$

com

$$S = \{(\theta_1, \theta_2); 0 < \theta_1 + \theta_2 < 1, \theta_1 > 0, \theta_2 > 0\},$$

$$Q = \{(\eta_1, \eta_2); 0 < \eta_1 < 1, 0 < \eta_2 < 1\},$$

θ	Função de Densidade	Distribuição	θ	Função de Densidade	Distribuição
.41	.1376E+00	.7194E+02	.71	.2688E+01	.3390E+00
.42	.1612E+00	.8686E+02	.72	.2803E+01	.3665E+00
.43	.1880E+00	.1044E+03	.73	.2910E+01	.3951E+00
.44	.2181E+00	.1246E+03	.74	.3009E+01	.4247E+00
.45	.2518E+00	.1480E+03	.75	.3097E+01	.4552E+00
.46	.2894E+00	.1751E+03	.76	.3173E+01	.4866E+00
.47	.3311E+00	.2066E+03	.77	.3236E+01	.5186E+00
.48	.3772E+00	.2414E+03	.78	.3283E+01	.5512E+00
.49	.4279E+00	.2816E+03	.79	.3312E+01	.5842E+00
.50	.4834E+00	.3271E+03	.80	.3322E+01	.6174E+00
.51	.5439E+00	.3783E+03	.81	.3311E+01	.6506E+00
.52	.6097E+00	.4361E+03	.82	.3278E+01	.6836E+00
.53	.6808E+00	.5006E+03	.83	.3222E+01	.7161E+00
.54	.7573E+00	.5722E+03	.84	.3141E+01	.7479E+00
.55	.8393E+00	.6537E+03	.85	.3035E+01	.7788E+00
.56	.9269E+00	.7440E+03	.86	.2909E+01	.8095E+00
.57	.1020E+01	.8337E+03	.87	.2746E+01	.8398E+00
.58	.1118E+01	.9244E+03	.88	.2563E+01	.8634E+00
.59	.1222E+01	.1018E+04	.89	.2358E+01	.8880E+00
.60	.1330E+01	.1118E+04	.90	.2131E+01	.9104E+00
.61	.1443E+01	.1223E+04	.91	.1885E+01	.9305E+00
.62	.1561E+01	.1329E+04	.92	.1626E+01	.9481E+00
.63	.1682E+01	.1440E+04	.93	.1357E+01	.9630E+00
.64	.1806E+01	.1514E+04	.94	.1085E+01	.9755E+00
.65	.1932E+01	.1600E+04	.95	.8210E+00	.9843E+00
.66	.2060E+01	.2201E+04	.96	.5713E+00	.9917E+00
.67	.2189E+01	.2413E+04	.97	.3492E+00	.9963E+00
.68	.2317E+01	.2639E+04	.98	.1685E+00	.9983E+00
.69	.2444E+01	.2877E+04	.99	.4568E+01	.9998E+00
.70	.2568E+01	.3127E+04			.9999E+00

Tabela 3.1

Alguns valores da função de densidade e da distribuição $D_1(9;3)$

é tal que

$$\eta_1 \sim D_1(a_1; a_0 + a_2),$$

$$\eta_2 \sim D_1(a_2; a_0),$$

e

$$\eta_1 \perp\!\!\!\perp \eta_2.$$

A transformação T produz como imagem inversa da região $Q^* \subset Q$ a região $S^* \subset S$ como ilustrado pela Figura 3.2.

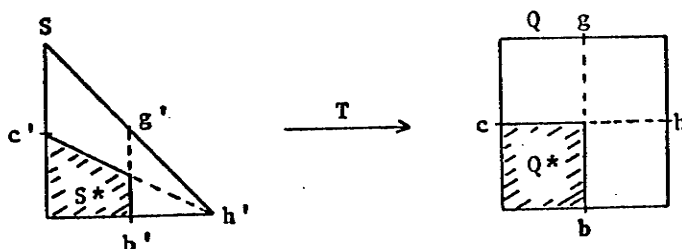


Figura 3.1

Regiões associadas pela transformação $T(\theta_1, \theta_2) = (\theta_1, \theta_2 / (1 - \theta_1))$.

Como consequência,

$$P(\theta \in S^*) = P(\eta \in Q^*) =$$

$$P(\eta_1 \leq b)P(\eta_2 \leq c),$$

sendo que podemos determinar b e c com auxílio das tabelas para a distribuição F , como fizemos no início desta seção. Para obtermos uma região de credibilidade 90% para θ podemos também escolher b_1, c_1 e b_2, c_2 tais que

$$P(b_1 \leq \eta_1 \leq c_1) = P(b_2 \leq \eta_2 \leq c_2) = (0.90)^{1/2}.$$

Em resumo, podemos enunciar o seguinte resultado:

Resultado 3.3 - Se $\theta \sim D_2(a_1; a_0)$ então $\theta \in S^*$ com probabilidade r , onde S^* é a região cuja imagem pela transformação

$$\eta_1 = \theta_1, \quad \eta_2 = \theta_2 / (1 - \theta_1)$$

é a região

$$Q^* = (b_1, c_1) \times (b_2, c_2), \text{ e}$$

$$P(b_1 \leq \eta_1 \leq c_1)P(b_2 \leq \eta_2 \leq c_2) = r,$$

com

$$\eta_1 \sim D_1(a_1; a_0 + a_2), \quad \eta_2 \sim D_1(a_2; a_0).$$

Exemplo 3.5 - Retomemos os dados do Exemplo 3.1, e suponha que se deseje construir uma região de credibilidade 90% para

$$\pi_6 = P(F = 1, R = 1) = \theta_1 + \theta_5$$

e

$$\pi_7 = P(F = 1, R = 0) = \theta_2 + \theta_6.$$

Primeiro, note que

$$(\pi_6, \pi_7) | x \sim D_2(2 + x_1 + x_5, 2 + x_2 + x_6; 1 + x_0 + 3 + x_3 + x_4 + x_7),$$

ou seja, após substituirmos os valores de x ,

$$(\pi_6, \pi_7) | x \sim D_2(a; a_0), \text{ com}$$

$$a = (a_1, a_2), \quad a_1 = 32, \quad a_2 = 41, \quad a_0 = 27.$$

Como consequência, aplicando o Resultado 3.3,

$$\eta_1 \sim D_1(32; 68),$$

$$\eta_2 \sim D_1(41; 27).$$

Aplicando o Resultado 3.2 com $r = (0.9)^{1/2} = 0.95$ obtemos, para η_1

$$F_{0.975}(136;64) \approx 1.58,$$

$$F_{0.975}(64;136) \approx 1.53,$$

e como consequência

$$P(0.22 < \eta_1 < 0.40) = 0.95.$$

Para η_2 ,

$$F_{0.975}(54;82) \approx 1.6,$$

$$F_{0.975}(82;54) \approx 1.7,$$

de modo que

$$P(0.50 < \eta_2 < 0.73) = 0.95.$$

Para obtermos a região no espaço de (π_6, π_7) construímos o quadrado $Q^* = (0.22, 0.40) \times (0.50, 0.73)$ no espaço (η_1, η_2) e o transformamos de volta ao espaço (π_6, π_7) , como indica a Figura 3.3. S^* é a região procurada.

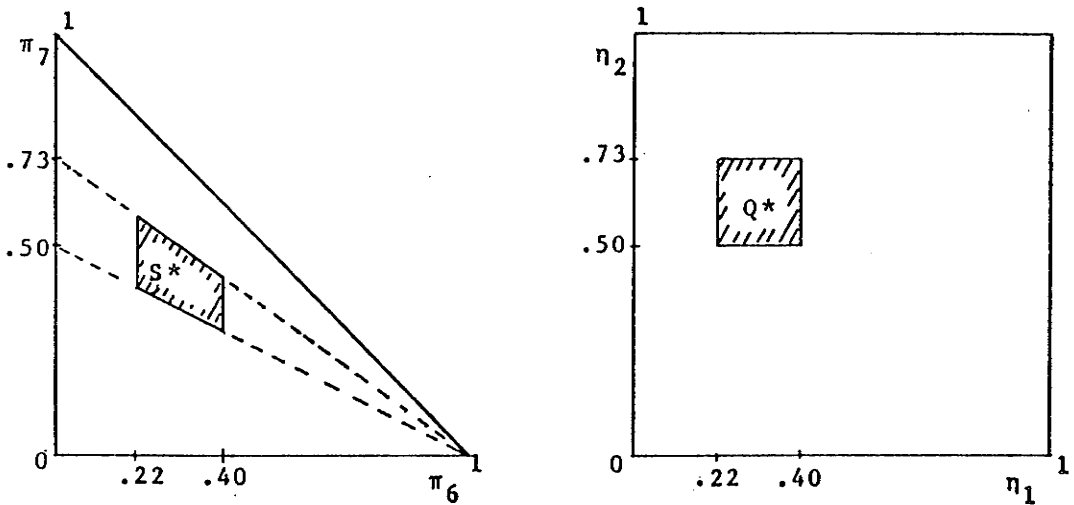


Figura 3.2

Região de credibilidade 90% para $(\pi_6, \pi_7) | x$.

3.3 ESTIMADORES DE BAYES, FATOR DE BAYES

Na seção anterior vimos como construir intervalos de credibilidade para $\theta|x$. Em alguns casos interessa-nos também obter uma estimativa pontual para $\theta|x$. Estimativas para θ obtidas a partir da distribuição a posteriori são geralmente referidas na literatura como estimativas de Bayes. A média, a moda, e a mediana da distribuição de $\theta|x$ são alguns exemplos, cada um gozando de certas propriedades ótimas num contexto de Teoria da Decisão, estritamente falando. No nosso estudo, usaremos a média de $\theta|x$ como estimativa de Bayes para θ . Do Corolário 5.3, Capítulo 1, obtemos imediatamente o seguinte resultado.

Resultado 3.4 - Se $X|\theta \sim M_k(n;\theta)$ e $\theta \sim D_k(a;a_0)$ então estimamos θ a partir de $\theta|x$ por

$$\hat{\theta} = E(\theta|x) = \frac{a+x}{n + \sum_{i=0}^k a_i}.$$

Exemplo 3.6 - Suponha, ainda no Exemplo 3.1, que $\theta \sim D_7(1;1)$ a priori. Estimemos $\theta_0, \theta_1, \dots, \theta_7$ a partir de $\theta|x$:

$$\hat{\theta}_i = E(\theta_i|x) = \frac{1+x_i}{n+k+1}.$$

Como $n=92$, obtemos imediatamente as seguintes estimativas:

$$\begin{aligned} \hat{\theta}_1 &= 0.10, & \hat{\theta}_5 &= 0.22, \\ \hat{\theta}_2 &= 0.06, & \hat{\theta}_6 &= 0.35, \\ \hat{\theta}_3 &= 0.02, & \hat{\theta}_7 &= 0.05, \\ \hat{\theta}_4 &= 0.05, & \hat{\theta}_0 &= 0.15. \end{aligned}$$

Em outros casos, entretanto, deseja-se uma expressão numérica para a crença a posteriori associada a uma hipótese do tipo

$$H_0: \theta \in \Omega_0 \subset \Theta.$$

Para isso, podemos calcular a partir da distribuição a posteriori para θ , a probabilidade de $\theta \in \Omega_0$ dado o resultado experimental $X = x$:

$$P(H_0|x) = \int_{\Omega_0} f(\theta|x) d\theta,$$

ou seja, a probabilidade a posteriori para a hipótese H_0 .

Em algumas situações, é conveniente expressarmos a probabilidade acima pela Razão de Chances ou Fator de Bayes (K) a favor da hipótese H_0 , dado o resultado experimental $X = x$:

$$K = \frac{P(H_0|x)}{P(H_1|x)},$$

onde

$$P(H_1|x) = 1 - P(H_0|x) = \int_{\Theta - \Omega_0} f(\theta|x) d\theta.$$

Note que $P(H_0|x) = K/(1+K)$, como sempre.

Ao decidirmos a favor de H_0 quando $P(H_0|x) = p$ então, pelo menos, devemos adotar a mesma atitude sempre que $P(H_0|x) \geq p$. Neste caso,

$$X_0 = \{x \in X; P(H_0|x) \geq p\}$$

é a região de aceitação da hipótese H_0 .

Analogamente, se $P(H_0|x) = q$ é suficiente para rejeitarmos H_0 então

$$X_1 = \{x \in X; P(H_0|x) \leq q\}$$

é a região de aceitação de H_1 , a hipótese alternativa.

Se decidirmos a favor da hipótese H_0 estamos sujeitos ao erro, usualmente chamado de Erro 2, ocorrido caso $\theta \in \Omega_1 = \Theta - \Omega_0$. A probabilidade deste evento é:

$$\begin{aligned}
P(\text{Erro 2}) &= P(x \in X_0 | \theta \in \Omega_1) \\
&= \sum_{X_0} P(x | \theta \in \Omega_1) \\
&= \sum_{X_0} P(\theta \in \Omega_1 | x) P(x) / P(\Omega_1) \\
&= \frac{1}{P(\Omega_1)} \sum_{X_0} P(x) \int_{\Omega_1} p(\theta | x) d\theta,
\end{aligned}$$

ou

$$P(\text{Erro 2}) = \frac{1}{P(\Omega_1)} \sum_{X_0} \int_{\Omega_1} f(x|\theta) p(\theta) d\theta.$$

Analogamente, se decidimos a favor da hipótese H_1 , estamos sujeitos ao Erro 1, ocorrido caso $\theta \in \Omega_0$. Neste caso,

$$P(\text{Erro 1}) = \frac{1}{P(\Omega_0)} \sum_{X_1} \int_{\Omega_0} f(x|\theta) p(\theta) d\theta.$$

Será conveniente recordarmos o resultado seguinte, de bastante aplicação em cálculos como o que exemplificaremos em seguida:

Resultado 3.5 - Se $\theta \sim D_1(a; a_0)$ então, para a e a_0 inteiros,

$$P(\theta \leq c) = \sum_{t=a}^{a+a_0-1} \binom{a+a_0-1}{t} c^t (1-c)^{a+a_0-t-1}, \quad 0 < c < 1.$$

Demonstração: Indicando $P(\theta \leq c)$ por $I_c(a, a_0)$, basta notar que

$$\begin{aligned}
\frac{d}{dv} \binom{n}{t} v^t (1-v)^{n-t} &= [t(1-v) - v(n-t)] \binom{n}{t} v^{t-1} (1-v)^{n-t-1}, \quad n > t \\
&= t \binom{n}{t} v^{t-1}, \quad n = t,
\end{aligned}$$

e portanto

$$\begin{aligned}
\binom{n}{t} c^t (1-c)^{n-t} &= I_c(t, n-t+1) - I_c(t+1, n-t), \quad n > t \\
&= I_c(t, 1), \quad n = t.
\end{aligned}$$

As parcelas de

$$\sum_{t=a}^n \binom{n}{t} c^t (1-c)^{n-t}$$

cancelam-se todas exceto a primeira, resultando a igualdade proposta, quando substituímos $n-a+1$ por a_0 .

Exemplo 3.7 - Suponha que $\theta \sim D_1(1;1)$ a priori, $X|\theta \sim M_1(10;\theta)$, observando-se $X=6$. Vejamos qual a probabilidade a posteriori para

$$H_0: \theta \in \Omega_0 = (0, 1/2).$$

Temos $\theta|X=6 \sim D_1(7;5)$, e portanto

$$\begin{aligned} P(H_0|x) &= P(\theta \leq 0.5|x) = \int_0^{0.5} \frac{\theta^6 (1-\theta)^4}{B(7,5)} d\theta = \\ &= \sum_{t=7}^{11} \binom{11}{t} (0.5)^t (0.5)^{11-t} \\ &= \sum_{t=7}^{11} \binom{11}{t} (0.5)^{11} = 0.27. \end{aligned}$$

Portanto, $P(H_0|x) = 27\%$, ou $K = 0.37:1$.

Se decidimos a favor da hipótese H_1 , então devemos fazê-lo também para todo $x \in X_1 = \{6, 7, 8, 9, 10\}$. Assim, estamos sujeitos ao Erro 1, cuja probabilidade é

$$\begin{aligned} P(\text{Erro 1}) &= \frac{1}{P(\Omega_0)} \sum_{x=6}^{10} \int_0^{0.5} \binom{10}{x} \theta^x (1-\theta)^{10-x} d\theta \\ &= \frac{1}{P(\Omega_0)} \sum_{x=6}^{10} \binom{10}{x} B(x+1, 11-x) \int_0^{0.5} \frac{\theta^x (1-\theta)^{10-x}}{B(x+1, 11-x)} d\theta \\ &= \frac{1}{P(\Omega_0)} \sum_{x=6}^{10} \binom{10}{x} B(x+1, 11-x) \sum_{t=x+1}^{11} \binom{11}{t} (0.5)^{11} \\ &= 0.077. \end{aligned}$$

3.4 PROPORÇÕES, TABELAS DE CONTINGÊNCIAS E APLICAÇÕES

Nesta seção exemplificaremos o procedimento Bayesiano na análise de algumas situações experimentais frequentemente propostas ao Estatístico. Tais experimentos são caracterizados pela observação de uma seqüência de variáveis aleatórias

$$Y_\alpha = (e_{0\alpha}, e_{1\alpha}, \dots, e_{t\alpha}), \quad \alpha = 1, 2, \dots$$

cujos possíveis resultados são os vetores Y_α com exatamente uma componente igual a 1 e as demais componentes iguais a zero, como na Seção 2 do Capítulo II.

Exemplo 3.8 - Inferências para uma proporção.

Retornemos os dados do Exemplo 3.2 deste Capítulo, e voltamos nosso interesse para o parâmetro π_1 , por exemplo:

$$\hat{\pi}_1 = P(R=1) = \theta_1 + \theta_3 + \theta_5 + \theta_7$$

Vimos que $\pi_1 | x \sim D_1(39; 61)$. A partir do Resultado 3.4, estimamos π_1 por

$$\hat{\pi}_1 = E(\pi_1 | x) = \frac{39}{39+61} = 39\%.$$

Um intervalo de credibilidade 90% para π_1 pode ser obtido a partir do Resultado 3.2, e obtemos, neste caso

$$P(0.31 \leq \pi_1 \leq 0.49 | x) = 90\%.$$

Note ainda que $P(\pi_1 \leq 0.5 | x) \approx 95\%$, de modo que o fator de Bayes K para $H_0: \pi_1 \leq 0.5$ é aproximadamente igual a 19:1.

O resultado seguinte será aplicado no próximo exemplo, para compararmos duas proporções.

Resultado 3.6 - Se $\alpha \sim D_1(a; a_0)$, $\beta \sim D_1(b; b_0)$ e $\alpha \perp\!\!\!\perp \beta$, então podemos escrever $P(\alpha \leq \beta)$ na forma

$$P(\alpha \leq \beta) = \frac{\Gamma(b+b_0)\Gamma(a+a_0)}{\Gamma(b)\Gamma(b_0)\Gamma(a+a_0+b+b_0-1)} \sum_{t=a}^{a+a_0-1} \frac{\Gamma(a+a_0+b_0-t-1)\Gamma(t+b)}{\Gamma(a+a_0-t)\Gamma(t+1)},$$

com a e a_0 inteiros positivos.

Demonstração: Basta aplicarmos o Resultado 3.6 na expressão de $P(\alpha \leq b)$ e integrarmos tal expressão com relação a $f(b)db$, onde $f(\cdot)$ é a função de densidade para a variável aleatória β .

Exemplo 3.9 - Desigualdade entre duas proporções.

Suponha, no Exemplo 3.2, que se deseje comparar as duas proporções

$$\pi_3 = P(R = 1 | M = 0, F = 1) = \theta_5 / (\theta_5 + \theta_6),$$

$$\pi_5 = P(R = 1 | M = 0, F = 0) = \theta_7 / (\theta_7 + \theta_0).$$

Vimos que

$$\pi_3 | x \sim D_1(22; 35), \quad \pi_5 | x \sim D_1(5; 15),$$

e além disso,

$$\pi_3 \perp\!\!\!\perp \pi_5 | x.$$

Para avaliarmos a probabilidade a posteriori a favor da hipótese

$$H_0: \pi_5 \leq \pi_3$$

podemos calcular o fator de Bayes para H_0 :

$$K = \frac{P(\pi_5 \leq \pi_3 | x)}{P(\pi_5 > \pi_3 | x)},$$

e para isso aplicamos o resultado anterior.

Note, entretanto, que condicional em $X = x$,

$$\begin{aligned} P(\pi_5 \leq \pi_3) &= 1 - P(\pi_3 \leq \pi_5) \\ &= 1 - P(1 - \pi_3 \geq 1 - \pi_5), \end{aligned}$$

$$1 - \pi_3 \sim D_1(35; 22), \quad 1 - \pi_5 \sim D_1(15; 5).$$

Nesta forma, podemos tomar $\alpha \equiv 1 - \pi_5$, $\beta \equiv 1 - \pi_3$,

$$a = 15, \quad a_0 = 5, \quad b = 35, \quad b_0 = 22$$

obtendo um somatório com apenas $a_0 = 5$ parcelas. Temos então, da do $X = x$,

$$P(1 - \pi_5 < 1 - \pi_3) = \frac{\Gamma(57)\Gamma(20)}{(35)(22)(76)} \sum_{t=15}^{19} \frac{\Gamma(41-t)\Gamma(t+35)}{\Gamma(20-t)\Gamma(t+1)},$$

que calculamos com o auxílio de uma tabela de log-fatoriais, obtendo uma probabilidade de 12.2%. Logo, $P(\pi_5 \leq \pi_3) = 87.8\%$ a posteriori, resultando num fator de Bayes

$$K = \frac{87.8}{12.2} = 7.2 : 1$$

a favor da hipótese H_0 .

Em alguns casos, o pesquisador está interessado em avaliar a probabilidade a posteriori associada a uma hipótese da forma

$$H_0: \theta \in \Omega_0,$$

embora o subconjunto Ω_0 não acumule qualquer probabilidade positiva se utilizamos uma função de densidade tal como a uniforme em Θ . Nos parágrafos seguintes, construiremos medidas de probabilidade a priori que concentrem alguma massa de probabilidade positiva nos subconjuntos Ω_0 e Ω_1 especificados pelas hipóteses H_0 e H_1 . Faremos isto para algumas funções de verossimilhança associadas a experimentos multinomiais e, em particular, faremos uso da forma canônica de representação destas funções, como no Resultado 2.7 do Capítulo 2. Para isto, sejam

$$S(k) = \{(\alpha_1, \dots, \alpha_k); 0 < \sum \alpha_i < 1, \alpha_i > 0, i = 1, \dots, k\},$$

V um espaço paramétrico da forma

$$V = S^h(k) \times S^{h'}(k'), \quad h, h' = 0, 1, \dots; \quad k, k' = 1, 2, \dots,$$

v_x a função de verossimilhança

$$v_x; \quad V \longrightarrow \mathbb{R}, \\ (\lambda, \eta)$$

e ϕ uma função contínua

$$\phi: \quad V \quad \rightarrow \quad V \\ (u, v)$$

tal que $\phi(V) = \Omega_0$. Suponhamos ainda que Ω_0 e Ω_1 formem uma partição do espaço V , sendo nulo o volume de Ω_0 em V . Seja $d(V)$ o volume de V . Podemos então definir, para todo $A \subset V$ mensurável com relação a $d\lambda d\eta$, as medidas de probabilidade

$$\mu_1(A) = \int_{\phi^{-1}(A \cap \Omega_0)} dudv/d(v), \quad \mu_2(A) = \int_A d\lambda d\eta/d(v)$$

e a medida de probabilidade $\mu = q\mu_1 + (1-q)\mu_2$, $0 < q < 1$.

Então $\mu(\Omega_0) = q$, $\mu(\Omega_1) = 1 - q$, e

$$\int_A v_x d\mu = q \int_{\phi^{-1}(A \cap \Omega_0)} (v_x \circ \phi)(u, v) dudv/d(V) + (1-q) \int_A v_x(\lambda, \eta) d\lambda d\eta/d(V),$$

resultando na medida de probabilidade a posteriori

$$\mu(A|x) = \int_A v_x d\mu / \int_V v_x d\mu,$$

e no fator da Bayes a favor da hipótese $(\lambda, \eta) \in \Omega_0$

$$K = \frac{q}{1-q} \frac{\int_V (v_x \circ \phi)(\bar{u}, \bar{v}) dudv}{\int_V v_x(\lambda, \eta) d\lambda d\eta}$$

Vejamos a seguir algumas aplicações:

Exemplo 3.10 - Igualdade entre duas proporções.

Consideremos um experimento descrito por uma tabela de contingência 2×2

x_{11}	x_{12}	$x_{1.}$
x_{21}	x_{22}	$x_{2.}$
$x_{.1}$	$x_{.2}$	

θ_{11}	θ_{12}	$\theta_{1.}$
θ_{21}	θ_{22}	$\theta_{2.}$

e a representação canônica para a função de verossimilhança

$$v_x(\lambda, \eta) \propto \lambda_1^{x_{11}} (1-\lambda_1)^{x_{12}} \lambda_2^{x_{21}} (1-\lambda_2)^{x_{22}} \eta^{x_{1.}} (1-\eta)^{x_{2.}},$$

$$\lambda_1 = \theta_{11}/\theta_{1.}, \quad \lambda_2 = \theta_{21}/\theta_{2.}, \quad \eta = \theta_{1.},$$

e seja $H_0: \lambda_1 = \lambda_2$ a hipótese nula.

De acordo com a construção anterior, temos

$$S = (0, 1), \quad \lambda = (\lambda_1, \lambda_2), \quad (\lambda, \eta) \in V = S^2 \times S,$$

$$\Omega_0 = \{(\lambda, \eta) \in V; \lambda_1 = \lambda_2\},$$

$$\phi(u_1, u_2, v) = (u_1, u_1, v).$$

Como consequência, o fator de Bayes para H_0 é

$$K = \frac{q}{1-q} \frac{\int_{S^2 \times S} u_1^{x_{1.}} (1-u_1)^{x_{12}} u_2^{x_{21}} (1-u_2)^{x_{22}} \eta^{x_{1.}} (1-\eta)^{x_{2.}} du_1 du_2 d\eta}{\int_{S^2 \times S} \lambda_1^{x_{11}} (1-\lambda_1)^{x_{12}} \lambda_2^{x_{21}} (1-\lambda_2)^{x_{22}} \eta^{x_{1.}} (1-\eta)^{x_{2.}} d\lambda_1 d\lambda_2 d\eta}$$

$$= \frac{q}{1-q} \frac{B(x_{.1}+1, x_{.2}-1)}{B(x_{11}+1, x_{12}+1) B(x_{21}+1, x_{22}+1)},$$

ou seja,

$$K = \frac{q}{1-q} \frac{x_{.1}! x_{.2}! (x_{1.}+1)! (x_{2.}+1)!}{x_{11}! x_{12}! x_{21}! x_{22}!} \frac{1}{(n+1)!}, \quad n = \sum_{i,j} x_{ij}$$

Do Resultado 2.15, Capítulo 2, sabemos que a hipótese H_0 é equivalente à hipótese de independência entre as distribuições marginais de θ ; ou seja, para testarmos a hipótese H_0 de independência entre as distribuições marginais de uma tabela de contingência 2×2 , aplicamos o fator de Bayes obtido acima.

Como ilustração, se observamos

09	21	30
01	04	05
10	25	<u>35</u>

e adotamos $q = 0.5$ então o fator de Bayes a favor de $H_0: \lambda_1 = \lambda_2$ é

$$K = \frac{10!25!31!06!}{09!21!01!04!} \frac{1}{36!} = 2 : 1$$

ou

$$P(H_0 | x) = 66\%.$$

Exemplo 3.11 - Igualdade entre várias proporções.

Consideremos um experimento descrito por uma tabela de contingência $l \times 2$

x_{11}	x_{12}	$x_{1.}$	θ_{11}	θ_{12}	$\theta_{1.}$
\vdots			\vdots		
x_{l1}	x_{l2}	$x_{l.}$	θ_{l1}	θ_{l2}	$\theta_{l.}$
$x_{.1}$	$x_{.2}$	n	$\theta_{.1}$	$\theta_{.2}$	

Neste caso, temos

$$v_x(\lambda, \eta) \propto \prod_{i=1}^{\ell} \lambda_i^{x_{i1}} (1-\lambda_i)^{x_{i2}} \left| \eta_1^{x_1} \eta_2^{x_2} \dots \eta_{\ell-1}^{x_{\ell-1}} (1-\eta_1) \dots \eta_{\ell-1} \right|^{x_{\ell}},$$

$$\lambda = (\lambda_1, \dots, \lambda_{\ell}), \quad \lambda_i = \theta_{i1}/\theta_i,$$

$$\eta = (\eta_1, \dots, \eta_{\ell-1}), \quad \eta_i = \theta_i,$$

e a hipótese H_0 de interesse é

$$H_0: \lambda_1 = \dots = \lambda_{\ell},$$

e de acordo com a notação anterior,

$$(\lambda, \eta) \in V = S(1)^{\ell} \times S(\ell-1),$$

$$\Omega_0 = \{(\lambda, \eta) \in V; \lambda_1 = \dots = \lambda_{\ell}\},$$

e

$$\phi(u_1, u_2, \dots, u_{\ell}, v) = (u_1, u_1, \dots, u_{\ell}, v), \quad v \in S(\ell-1)$$

Como consequência, o fator de Bayes para H_0 é

$$K = \frac{q}{1-q} \frac{\int_V (v_x \circ \phi)(u_1, \dots, u_{\ell}, v) du_1 du_2 \dots du_{\ell} dv}{\int_V v_x(\lambda, \eta) d\lambda d\eta} =$$

$$= \frac{q}{1-q} \frac{B(x_{.1}+1, x_{.2}+1)}{\prod_{i=1}^{\ell} B(x_{i1}+1, x_{i2}+1)},$$

ou seja

$$K = \frac{q}{1-q} \frac{x_{.1}! x_{.2}! \prod_{i=1}^{\ell} (x_{i.}+1)!}{\prod_{i,j} x_{ij}!} \frac{1}{(n+1)!}.$$

Como ilustração, se observamos

02	05	07
01	05	06
02	06	08
05	16	<u>21</u>

e adotamos $q = 0.5$, segue-se que o fator de Bayes a favor de $H_0: \lambda_1 = \lambda_2 = \lambda_3$ é

$$K = \frac{05!16!08!07!09!}{02!05!01!05!02!06!} \frac{1}{22!} = 1.38 : 1 ,$$

ou

$$P(H_0 | x) = 58\% .$$

Para obtermos o fator de Bayes para a hipótese H_0 de independência entre as distribuições marginais numa tabela de contingência $\ell \times \ell$, basta modificarmos ligeiramente a construção acima, partindo de

$$(\lambda, \eta) \in V = S(\ell-1)^\ell \times S(\ell-1)$$

Do Resultado 2.15 sabemos que a hipótese H_0 é definida por

$$H_0: \lambda_1 = \dots = \lambda_\ell , \quad \lambda_i \in S(\ell-1), \quad i = 1, \dots, \ell .$$

Note ainda que

$$\phi(u_1, \dots, u_\ell, v) = (u_1, u_1, \dots, u_1, v), \quad u_i, v \in S(\ell-1), \quad i = 1, \dots, \ell ,$$

e $S(\ell-1)$ tem volume $1/(\ell-1)!$.

Portanto, o fator de Bayes para H_0 é

$$K = \frac{q}{1-q} \frac{B(x_{.1}+1, x_{.2}+1, \dots, x_{.l}+1)}{\prod_{i=1}^l B(x_{i1}+1, x_{i2}+1, \dots, x_{il}+1)} \frac{1}{(\ell-1)!^{\ell-1}},$$

ou seja,

$$K = \frac{q}{1-q} \frac{\prod_{i=1}^l x_{.i} (x_{.i} + \ell - 1)!}{(n + \ell - 1)! \prod_{i,j=1}^l x_{ij}!} \frac{1}{(\ell-1)!^{\ell-1}}.$$

Vejamos duas ilustrações numéricas com $\ell = 3$, e admitindo $q = 0.5$.

05	02	01	08
01	10	02	13
01	03	05	09
07	15	08	<u>30</u>

$$K = 0.038 : 1,$$

ou

$$P(H_0 | x) = 3.6\%.$$

05	02	01	08
10	02	01	13
05	03	01	09
20	07	03	<u>30</u>

$$K = 10.65 : 1,$$

ou

$$P(H_0 | x) = 91\%.$$

Os cálculos acima podem ser rapidamente efetuados a partir de uma tabela de log-fatoriais.

Exemplo 3.12 - Probabilidade dos Erros 1 e 2 no teste da hipótese de independência numa tabela de contingência 2×2 .

Suponhamos a seguinte configuração:

01	02	03
01	00	01
02	02	<u>04</u>

e $q = 0.5$. O fator de Bayes para a hipótese de independência H_0 é

$$K = \frac{q}{1-q} \frac{x_{.1}! x_{.2}! (x_{1.}+1)! (x_{2.}+1)!}{x_{11}! x_{12}! x_{21}! x_{22}!} \frac{1}{(n+1)!}$$

$$= \frac{2! 2! 4! 2!}{1! 2! 1! 0!} \frac{1}{5!} = 0.8 : 1 ,$$

ou

$$P(H_0 | x) = 44\%$$

Se decidimos contra a hipótese de independência estamos sujeitos ao Erro 1, ocorrido caso H_0 seja verdadeira. Neste caso, como vimos anteriormente,

$$P(\text{Erro 1}) = \frac{1}{P(\Omega_0)} \int_{X_1} \int_{\Omega_0} P(x | \lambda_1, \lambda_2, \eta) d\mu ,$$

onde

$$X_1 = \{x; P(H_0 | x) \leq 0.44\} ,$$

pele menos, μ é a medida de probabilidade a priori estudada nos exemplos anteriores e $P(x | \lambda_1, \lambda_2, \eta)$ é a representação canônica para $P(x | \theta)$.

Assim,

$$\int_{\Omega_0} P(x | \lambda_1, \lambda_2, \eta) d\mu = q \int_{S(1)^2 \times S(1)} P(x | u_1, u_1, v) du_1 du_2 dv ,$$

e como

$$P(x|\lambda_1, \lambda_2, \eta) = \\ = \binom{x_{1.}}{x_{11}} \binom{x_{2.}}{x_{21}} \binom{n}{x_{1.}} \lambda_1^{x_{11}} (1-\lambda_1)^{x_{12}} \lambda_2^{x_{21}} (1-\lambda_2)^{x_{22}} \eta^{x_{1.}} (1-\eta)^{x_{2.}}$$

segue-se que

$$P(\text{Erro 1}) = \sum_{X_1} \binom{x_{1.}}{x_{11}} \binom{x_{2.}}{x_{21}} \binom{n}{x_{1.}} B(x_{1.}+1, x_{2.}+1) B(x_{1.}+1, x_{2.}+1),$$

ou, simplificando a expressão acima,

$$P(\text{Erro 1}) = \sum_{X_1} \frac{x_{1.}! x_{2.}! x_{1.}! x_{2.}!}{x_{11}! x_{12}! x_{21}! x_{22}!} \frac{1}{(n+1)!(n+1)}.$$

Para completarmos o exemplo, suponhamos agora a seguinte configuração

02	00	02
02	00	02
04	00	<u>04</u>

mantendo $q = 0.5$. Como resultado,

$$K = 1.80 : 1$$

ou

$$P(H_0|x) = 64\%.$$

Se decidimos a favor da hipótese H_0 devemos tomar

$$X_0 = \{x; P(H_0|x) \geq 0.64\},$$

pelo menos, e calcular

$$P(\text{Erro 2}) = \frac{1}{P(\Omega_1)} \sum_{X_0} \int_{\Omega_1} P(x|\lambda_1, \lambda_2, \eta) d\mu,$$

cujo resultado é

$$P(\text{Erro 2}) = \sum_{X_0} \binom{x_1}{x_{11}} \binom{x_2}{x_{21}} \binom{n}{x_1} B(x_{11}+1, x_{12}+1) B(x_{21}+1, x_{22}+1) B(x_1+1, x_2+1),$$

ou, após simplificarmos a expressão acima

$$P(\text{Erro 2}) = \sum_{X_0} \frac{1}{(n+1)(x_1+1)(x_2+1)}.$$

Na Tabela 3.2 listamos todas as configurações de contingências 2x2 baseadas em $n = 04$ observações, as probabilidades $P(H_0|x)$, $P(x|H_0)$ e $P(x|H_1)$. No primeiro caso mencionado neste exemplo obtivemos $P(H_0|x) = 44\%$, e, portanto, rejeitando H_0 ;

$$P(\text{Erro 1}) = \sum_{X_1} P(x|H_0) = 13.2\% ;$$

tomando $X_1 = \{x; P(H_0|x) \leq 0.44\}$.

No segundo caso calculamos $P(H_0|x) = 64\%$, e como consequência, aceitando H_0 ;

$$P(\text{Erro 2}) = \sum_{X_0} P(x|H_1) = 8\% ,$$

considerando $X_0 = \{x; P(H_0|x) \geq 0.64\}$.

Na Tabela 3.3 exibimos as configurações e respectivas probabilidades para o caso $n = 05$.

3.5 PREDIÇÃO BAYESIANA

Na seção 2.4 descrevemos, através de bolas em urnas, o modelo de populações finitas e a distribuição Hipergeométrica multivariada. As N unidades populacionais representados por bolas eram subdivididos em dois conjuntos (representados pelas duas caixas) um com n (caixa 1) e outro (caixa 2) com $N - n$ unidades. Após observar a composição (X_0, X_1, \dots, X_k) da caixa 1, o que permanece desconhecido é a composição $(\psi_0 - X, \psi_1 - X_1, \dots, \psi_k - X_k)$

da caixa 2. O objetivo desta seção é apresentar o método Bayesiano da predição dessa composição.

A primeira meta de um Bayesiano é encontrar a distribuição a posteriori de $(\psi_1 - X_1, \dots, \psi_k - X_k)$ onde (X_0, X_1, \dots, X_k) são os dados amostrais. Note que antes (a priori) de $X = (X_1, \dots, X_k)$ ser observado, o parâmetro de interesse é $\psi = (\psi_1, \dots, \psi_k)$ pois $\psi_0 = N - \psi_1 - \dots - \psi_k$. Após (a posteriori) X ser observado, o parâmetro de interesse seria $\psi - X$.

A escolha da distribuição a priori para ψ , será restrita a classe das DM. Esta restrição é natural pois, como a distribuição Hipergeométrica é derivada de um processo permutável, temos que $\psi | \theta \sim M_k(n; \theta)$ e ao se considerar $\theta \sim D_k(a; a_0)$, obtêm-se $\psi \sim DM_k(n; a, a_0)$.

O seguinte resultado é o principal desta seção.

Resultado 3.7 - Para a situação da amostragem em populações finitas, descrita acima, se a priori $\psi \sim DM(n; a_1, a_2, \dots, a_k, a_0)$, então,

$$\psi - X | X \sim DM(N - n, a_1 + X_1, a_2 + X_2, \dots, a_k + X_k, a_0 + X_0)$$

a posteriori.

Demonstração: Da análise Bayesiana do caso Multinomial, sabemos que $\theta | X \sim D(a_1 + X_1, \dots, a_k + X_k; a_0 + X_0)$. Por outro lado, a distribuição de $(\psi - X) | X$ pode ser vista como a composição da distribuição de $(\psi - X) | (X, \theta) \sim (\psi - X) | \theta$ pela distribuição de $\theta | X$. Assim, pela definição de DM, concluímos o teorema.

Para finalizar, note que o estimador de Bayes de ψ é dado por

$$\begin{aligned} \hat{\psi} &= E\{\psi | X\} = E\{X | X\} + E\{\psi - X | X\} = X + E\{(\psi - X) | X\} = \\ &= (x_0, \dots, x_k) + \frac{n}{a}(a_0, \dots, a_k). \end{aligned}$$

Os leitores poderiam estar surpresos com o fato de intitularmos esta seção de predição Bayesiana. Contudo é interessante

te lembrarmos que o problema da predição estatística surge quando desejamos adivinhar o resultado de um experimento após observar-se os resultados de uma série de n realizações de experimentos semelhantes. Assim se tomarmos $N = n+1$, fica claro que esse é um caso particular do problema da inferência em populações finitas.

BIBLIOGRAFIA

- Aitchison, J. and I.R. Dunsmore (1975) - Statistical Prediction Analysis - Cambridge University Press, London.
- Altham, P.E. (1969) - Exact Bayesian Analysis of a 2x2 contingency table, and Fisher's "exact" significance test, JRSS B, 31, 261-9.
- Basu, D. (1975) - Statistical Information and Likelihood - Sankhyā, Vol. 37, A, 1-71.
- Basu, D. and C.A.B. Pereira (1982) - On the Bayesian Analysis of Categorical Survey Data: The Problem of Nonresponse. Journal of Statistical Planning and Inference 6, to appear.
- Feller (1968) - An Introduction to Probability Theory and Its Applications, Vol. 1, 3rd ed., Acad. Press, N.Y.
- Good, I.J. (1965) - The Estimation of Probabilities: An Essay on Modern Bayesian Approach - MIT Press, N.Y.
- Heath, D. and W. Suddeth (1976) - De Finetti's Theorem on Exchangeable Variables. The American Statistician, Vol. 30, n^o 3, 188-189.
- Lindley, D.V. (1972) - Bayesian Statistics, A Review, SIAM, Pennsylvania.
- Lindley, D.V. (1978) - The Bayesian Approach. Scandinavian Journal of Statistics, Vol. 5, 1-26
- Lindley D. and L.D. Phillips (1976) - Inference For a Bernoulli Process (a Bayesian View). The American Statistician, Vol.30, n^o 3, 112-119.

Sant'Anna C.C. (1982) - O Tabagismo Passivo e Problemas Respirató
rios em Crianças.