Inferência Comparada e Planejamento Adaptativo Aplicado em Ensaios Clínicos

Caroline de Freitas Sakamoto

Monografia - PIBIC/CNPq

Departamento de Estatística, IMECC - UNICAMP.

Área de Concentração: **Estatística**

Orientadora: **Profª. Dra. Laura Leticia Ramos Rifo**

- Campinas, Junho de 2009 -

Sumário

1 Prefácio **5**

2 Inferência Bayesiana **7**

2.0.1 Interpretação da Priori 8

2.0.2 Função de Verossimilhança 9

2.0.3 Distribuição Posterior 14

3 Teoria de Decisão **33**

3.0.4 Fator de Bayes 35

3.0.5 Teste da Razão de Verossimilhança 37

4 Ensaios Clínicos **39**

4.0.6 Aplicação 41

4.0.7 Outras Regras de Parada 47

Capítulo 1

Prefácio

Um dos problemas mais recorrentes encontrados na prática estatística é a determinação do tamanho da amostra a ser analisada.

Em particular, este é um componente rotineiro no planejamento de ensaios clínicos, em que os custos humanos e éticos devem ser levados em conta, exercendo um papel fundamental na avaliação das propostas apresentadas e nas decisões subseqüentes.

Neste trabalho foi apresentado um exemplo de um ensaio clínico, onde comparamos os resultados obtidos pelos métodos Clássico e Bayesiano quanto à determinação do tamanho da amostra.

Qualquer que seja o ponto de vista adotado, a solução deste problema exige um equilíbrio entre o custo do experimento e a qualidade da análise final.

Primeiramente, descreveremos as definições utilizadas, fazendo uma revisão detalhada dos diversos critérios.

Serão vistos na Seção 2, Inferência Bayesiana: Interpretação da Priori, Função de Verossimilhança, Princípio de Verossimilhança, Distribuição Posterior, Estatística Suficiente, Família Exponencial.

Na Seção 3 será apresentado alguns métodos utilizados para os testes estatísticos, como: Teoria de Decisão, Fator de Bayes e Razão de Verossimilhança.

A aplicação apresentada na Seção 4 é um exemplo de ensaio clínico, com o objetivo de testar uma hipótese precisa. Aplicamos os métodos apresentados nas Seções 2 e 3, para a análise dos dados com tamanho de amostra pré-fixado e com planejamento seqüencial.

Capítulo 2

Inferência Bayesiana

Seja $X_1, X_2, \ldots, X_n$ amostra aleatória de um distribuição com função de densidade ou probabilidade $f(x|\theta)$, onde $\theta \in \Omega$. Em muitos problemas, antes das observações serem avaliadas, é possível ter uma informação prévia a respeito de θ. Esta informação pode ser quantificada em termos de uma função de probabilidade para θ. Esta probabilidade é chamada de probabilidade priori do parâmetro. A inferência bayesiana atualiza as probabilidades prioris das hipóteses testadas baseada nas observações obtidas. A fórmula de Bayes fornece uma relação entre a probabilidade priori, a verossimilhança e a probabilidade posterior de uma hipótese. Probabilidades empíricas ou subjetivas podem ser utilizadas como probabilidades prioris e várias hipóteses podem ser avaliadas simultaneamente.

Esse capítulo está baseado no DeGroot[5].

2.0.1 Interpretação da Priori

Considere um problema de inferência estatística com observações com função de dendidade de probabilidade (f.d.p.) ou função de probabilidade (f.p.) $f(x|\theta)$, onde θ é um parâmetro de valor desconhecido, e está contido no espaço paramétrico Ω. Como já citado, em mutios problemas, antes das observações de $f(x|\theta)$ serem avaliadas, o estatístico tem uma informação prévia do valor de θ, obtida empiricamente de estudos anteriores, ou por considerações físicas ou teóricas do problema, por exemplo, ele acredita que θ tem probabilidade maior de estar em certas regiões de Ω do que em outras. Essa distribuição é chamada de distribuição priori de θ, e será denotado por $\xi(\theta)$.

Distribuições Prioris Discretas e Contínuas

O parâmetro θ pode ter apenas um número finito de diferentes valores ou uma sequência infinita de diferentes valores, dessa maneira uma distribuição discreta é a melhor opção. O parâmetro θ também pode ser um valor real ou algum intervalo real, e portanto uma distribuição contínua é atribuída para θ. A f.p. ou f.d.p. $\xi(\theta)$ é chamada distribuição priori para θ.

Exemplo 1 *Moeda balanceada ou duas-caras*
Seja θ a probabilidade de obter uma cara quando uma certa moeda é lançada, e suponha que sabemos que a moeda é balanceada ou que tem cara em um dos lados. Portanto, os valores possíveis para θ são somente $\theta = 1/2$ e $\theta = 1$. Se a probabilidade priori de que a moeda seja justa for p, então $\xi(1/2) = p$ e $\xi(1) = 1-p$.

Exemplo 2 *Proporção de itens defeituosos*

Suponha que a proporção de itens defeituosos fabricados em grande lote é dada por θ desconhecido e que a distribuição priori atribuída para θ é uma distribuição uniforme no intervalo $(0,1)$. Então a f.d.p. priori de θ é

$$\xi(\theta) = \begin{cases} 1 & para \quad 0 < \theta < 1, \\ 0 & caso\ contrário \end{cases} \tag{2.1}$$

Exemplo 3 *Parâmetro com distribuição exponencial*

Suponha que a vida útil das lâmpadas fluorescentes de um certo tipo estão sendo observadas, e a distribuição de uma lâmpada em particular é uma distribuição exponencial com parâmetro β. Suponha que o valor exato de β é desconhecido e que a partir de experiências anteriores a distribuição priori de β é uma distribuição gama com média 0,0002 e desvio padrão 0,0001. Podemos determinar a f.d.p. priori de β.

Suponha que a distribuição priori de β é uma distribuição gama com parâmetros α_0 e β_0. Sabemos que a média da distribuição é α_0/β_0 e a variância é α_0/β_0^2. Então, $\alpha_0/\beta_0 = 0,0002$ e $\alpha_0^{1/2}/\beta_0 = 0,0001$. Assim, podemos encontrar que $\alpha_0 = 4$ e $\beta_0 = 20000$. A f.d.p. priori de β é dada a seguir:

$$\xi(\beta) = \begin{cases} \frac{(20000)^4}{3!}\beta^3 e^{-20000\beta} & para \quad \beta > 0, \\ 0 \quad para \quad \beta \leq 0 \end{cases} \tag{2.2}$$

2.0.2 Função de Verossimilhança

A função de verossimilhança é baseada na f.d.p. de uma dada distribuição, consideremos a f.d.p. genérica

$$f(\mathbf{x}|\theta)$$

Se $X_1, \ldots, X_n$ são independentes, podemos dizer que a função de verossimilhança é dada por

$$f_n(\mathbf{x}|\theta) = \prod_{i=1}^{n} f(x_i|\theta)$$

A função $f_n(\mathbf{x}|\theta)$ de θ, para valores observados de $\mathbf{x}$, é a verossimilhança ou probabilidade de θ.

Essa função de verossimilhança é utilizada para calcular a distribuição posterior, que será vista na seção 2.3.

Definições Básicas

Nesta seção foram apresentados alguns conceitos que utilizaremos nas próximas seções. Estes conceitos estão de acordo com Donald Berry[3].

- Experimento: é considerado um método de investigar as relações casuais entre variáveis. É utilizado em ciências sociais e naturais, e pode ser usado para ajudar a resolver problemas práticos e apoiar ou negar hipóteses teóricas.

- Experimento equivalente:

- Experimentos Independentes: são independentes se os resultados de um experimento são independentes com o resultado dos outros, isso vale para qualquer par de resultados, um de cada experimento.

- Experimentos Permutáveis: A definição de experimentos independentes, dado acima, pode ser muito restritiva. Por exemplo, observações repetidas geradas de uma amostra de uma população não são independentes (condicionando que a primeira observação muda a probabilidade da segunda. Uma

suposição que não é tão restritiva é a permutabilidade. Dois experimentos são permutáveis se obedecer as seguintes regras:

(1) Os possíveis resultados são os mesmos em ambos.

(2) A probabilidade de um em cada resultado é o mesmo que nos outros.

(3) As probabilidades condicionais para o segundo experimento, dados os resultados do primeiro, são as mesmas probabilidades condicionais que para a primeira, dados os resultados do segundo.

Experiementos permutáveis são simétricos. As implicações de observações para dois experimentos permutáveis são os mesmo, independentemente de qual experimento produziu a observação. Observar um 12 na primeira observação e um 27 na segunda é o mesmo que observar 27 na primeira e 12 na segunda.

Princípio de Verossimilhança

Os valores estimados tanto pelo EMV (Estimador de Máxima Verossimilhança) como pelo estimador de Bayes para o parâmetro θ dependem dos valores observados na amostra só através da função de verossimilhança $f_n(\mathbf{x}|\theta)$. Dessa forma, se dois diferentes conjuntos de valores observados determinam a mesma função de verossimilhança, então a mesma estimativa para θ será obtida por ambos osconjuntos de valores observados.

Considere que dois diferentes conjuntos de valores observados $\mathbf{x}$ e $\mathbf{y}$ determinam funções de verossimilhança que são proporcionais uma com a outra, ou seja, elas

diferem somente por um fator que pode depender de $\mathbf{x}$ e $\mathbf{y}$ mas não depende de θ.

$$f_n(\mathbf{x}|\theta) \propto g_n(\mathbf{y}|\theta).$$

Neste caso, pode ser verificado que o EMV de θ será o mesmo, independente do fato de $\mathbf{x}$ ou $\mathbf{y}$ ser considerado. Igualmente, para uma dada f.d.p. priori de θ, o estimador de Bayes de θ será o mesmo. Ambos os estimadores, EMV e de Bayes, são compatíveis com o seguinte princípio de inferência estatística, o qual é conhecido como *Princípio da Verossimilhança*.

Axioma 1 (Princípio de Verossimilhança) *Suponha que dois diferentes conjuntos de valores observados* $\mathbf{x}$ *e* $\mathbf{y}$ *são obtidos a partir do mesmo experimento ou de dois experimentos diferentes, e suponha que* $\mathbf{x}$ *e* $\mathbf{y}$ *determinam a mesma função de verossimilhança para um certo paramêtro* θ *ou determinam funções de verossimilhança proporcionais uma com a outra. Então* $\mathbf{x}$ *e* $\mathbf{y}$ *fornecem a mesma informação sobre o valor desconhecido de* θ*, e devem determinar a mesma estimativa de* θ*.*

Analisando o axioma acima, constatamos que duas funções de verossimilhança para θ, provenientes do mesmo experimento ou de experimentos diferentes, contêm a mesma informação a respeito de θ se elas são proporcionais entre si.

De acordo com Lurdes Inoue[7], o desenvolvimento do princípio da Verossimilhança está ligado ao desenvolvimento de uma metodologia de inferência estatística, baseada nesse princípio, que pretendia substituir as metodologias clássica

e bayesiana. Entretanto, observamos que este princípio é compatível à metodologia bayesiana.

A utilização deste princípio é um grande ponto de divergência entre bayesianos e frequentistas. A sua violação tem levado diversos frequentistas a abandonar a sua visão em favor do ponto de vista bayesiano.

Exemplo 4 *Suponha que devemos estimar a proporção θ desconhecida de itens defeituosos de uma remessa em uma grande fábrica. Consideremos que as observações são Bernoulli com parâmetro θ, permutabilidade independentes e identicamente distribuídas.*

Em uma amostra de 10 itens, 2 têm defeitos e 8 não, com vetor de observações é $\mathbf{x} = (0, 0, 0, 1, 0, 0, 0, 0, 0, 1)$.

Suponha as seguintes regras de parada que poderiam definir o experimento:

E_1 : *O pesquisador havia decidido tomar exatamente 10 observações,*

E_2 : *O pesquisador havia decidido prosseguir a amostragem até obter 2 itens defeituosos.*

As funções de verossimilhança para θ geradas em E_1 e E_2 são, respectivamente

$$\textit{Distribuição Binomial:} \quad f_n^1(\mathbf{x}|\theta) = C_2^{10}\theta^2(1-\theta)^8,$$

$$\textit{Distribuição Binomial Negativa:} \quad f_n^2(\mathbf{x}|\theta) = C_1^{9}\theta^2(1-\theta)^8.$$

Podemos observar que as funções de verossimilhança são proporcionais e a informação sobre θ é a mesma tanto em E_1 quanto em E_2.

Os estimadores não viciados para θ, respectivamente em E_1 e E_2 são

$$\hat{\theta}^{(1)} = \bar{X}_n = \sum_{i=1}^{n} X_i/n$$

$$\hat{\theta}^{(2)} = (k-1)/(N-1),$$

onde k é o número de itens defeituosos, e N o número de observações necessárias até obter k itens defeituosos.

Obtemos as estimativas, $\hat{\theta}^{(1)} = 2/10 = 0,2$ e $\hat{\theta}^{(2)} = 1/9 = 0,111$, isto é, estimativas diferentes de θ . Elas evidenciam a dependência sobre a regra de parada do experimento e portanto a violação do Princípio de Verossimilhança.

A violação do Princípio de Verossimilhança fica evidente neste exemplo de estimação não viciada, um mesmo conjunto de dados leva a diferentes estimativas dependendo da regra de parada.

Portanto, esta metodologia de análise não é consistente com o Princípio de Verossimilhança, pois ao considerar as regras de parada após a observação dos dados podemos chegar a conclusões distintas.

Vale lembrar que a análise bayesiana, que interpreta uma função de probabilidade como uma opinião particular, jamais viola o Princípio de Verossimilhança, pois a priori é uma opinião e independe de aspectos do planejamento do experimento.

2.0.3 Distribuição Posterior

Considere que temos n variáveis aleatórias $X_1, \ldots, X_n$ com f.d.p. ou f.p. $f(x|\theta)$, que o valor do parâmetro θ é desconhecido e que a f.d.p. ou f.p. priori é $\xi(\theta)$. Para simplicidade, vamos assumir que $f(x|\theta)$ é uma f.d.p., em vez de f.p. (mas a conclusão pode ser facilmente adaptada para um problema com $\xi(\theta)$ ou $f(x|\theta)$ de

um f.p.).

As variáveis aleatórias $X_1, \ldots, X_n$ com f.d.p. ou f.p. $f(x|\theta)$ tem a f.d.p. conjunta $f_n(x_1, \ldots, x_n|\theta)$ que é dada pela seguinte equação:

$$f_n(x_1, \ldots, x_n|\theta) = f(x_1|\theta) \cdots f(x_n|\theta).$$

Se usarmos o vetor $\mathbf{x} = (x_1, \ldots, x_n)$ então a f.d.p. conjunta acima pode ser escrita como $f_n(\mathbf{x}|\theta)$.

A f.d.p. conjunta $f_n(\mathbf{x}|\theta)$ pode ser considerada com uma probabilidade condicional conjunta de $X_1, \ldots, X_n$ dado o valor de θ. Para obtermmos a f.d.p. marginal de $X_1, \ldots, X_n$ integramos $f_n(\mathbf{x}|\theta)\xi(\theta)$ em todos os valores de θ. Assim, a f.d.p. marginal n-dimensional de $X_1, \ldots, X_n$ pode ser escrita da seguinte forma:

$$g_n(\mathbf{x}) = \int_\Omega f_n(\mathbf{x}|\theta)\xi(\theta)d\theta. \tag{2.3}$$

Além disso, a f.d.p. condicional de θ dado que $X_1 = x_1, \ldots, X_n = x_n$, $\xi(\theta|x)$, deve ser igual à f.d.p. conjunta de $X_1, \ldots, X_n$ e θ dividido pela f.d.p. marginal de $X_1, \ldots, X_n$. Obtemos então,

$$\xi(\theta|\mathbf{x}) = \frac{f_n(\mathbf{x}|\theta)\xi(\theta)}{g_n(\mathbf{x})} \quad \text{para} \quad \theta \in \Omega. \tag{2.4}$$

Essa distribuição de probabilidade dada na Equação(2.4) é chamada de *distribuição posterior* de θ depois dos valores de $X_1, \ldots, X_n$ forem observados. Podemos reescrever esta equação,

$$\xi(\theta|\mathbf{x}) = f_n(\mathbf{x}|\theta)\xi(\theta). \tag{2.5}$$

Exemplo 5 *Proporção de itens defeituosos.*

Suponha agora, como no Exemplo 2, que a proporção θ de itens defeituosos em

um lote de uma grande fábrica é desconhecido e que a distribução priori de θ é uma Uniforme$(0,1)$. Suponha que uma amostra aleatória de tamanho n foi obtida do lote e para $i = 1, \ldots, n, X_i = 1$ se o i-ésimo item é defeituoso ou $X_i = 0$ caso contrário. Então $X_1, \ldots, X_n$ formam n ensaios Bernoulli(θ). Devemos determinar a f.d.p. posterior de θ.

A f.p. de cada observação X_i é

$$f(x|\theta) = \begin{cases} \theta^x(1-\theta)^{1-x} & \textit{para} \quad x = 0, 1, \\ 0 & \textit{caso contrário} \end{cases}$$

Se $y = \sum_{i=1}^{n} x_i$, então a f.p. conjunta de $X_1, \ldots, X_n$ pode ser escrita como

$$f_n(\mathbf{x}|\theta) = \theta^y(1-\theta)^{n-y} \quad 0 \leq y \leq n$$

Com a f.d.p. priori dada na Equação(2.1), temos

$$f_n(\mathbf{x}|\theta)\xi(\theta) = \theta^y(1-\theta)^{n-y} \quad \textit{para} \quad 0 < \theta < 1. \tag{2.6}$$

Observemos que, exceto pelo fator constante, a Equação(2.6) é a densidade de uma distribuição Beta com parâmetros $\alpha = y + 1$ e $\beta = n - y + 1$. Como a f.d.p. posterior $\xi(\theta|x)$ é proporcional à Equação(2.6),

$$\xi(\theta|\mathbf{x}) = \frac{\Gamma(n+2)}{\Gamma(y+1)\Gamma(n-y+1)}\theta^y(1-\theta)^{n-y} \quad \textit{para} \quad 0 < \theta < 1.$$

Exemplo 6 *Parâmetro com distribuição exponencial*

Suponha agora, como no Exemplo 3, que a distribuição da vida útil de uma lâmpada fluorescente de uma amostra aleatória de tamanho n foi observada uma distribuição exponencial com parâmetro β, e que a distribuição priori de β é uma Gamma com f.d.p. $\xi(\beta)$ dada pela Equação(2.2). Suponha também que a vida útil $X_1, \ldots, X_n$

da amostra aleatória de tamanho n de um certo tipo de lâmpadas foi observada. Vamos determinar a f.d.p. posterior de β dado que $X_1 = x_1, \ldots, X_n = x_n$.

A f.d.p. de cada observação X_i é

$$f(x|\beta) = \begin{cases} \beta e^{-\beta x} & para \quad x > 0, \\ 0 & caso\ contrário. \end{cases}$$

Se $y = \sum_{i=1}^{n} x_i$, então a f.p. conjunta de $X_1, \ldots, X_n$ pode ser escrita como

$$f_n(\mathbf{x}|\beta) = \beta^n e^{-\beta y} \quad para \quad x_i > 0 (i = 1, \ldots, n).$$

Com a f.d.p. priori dada na Equação(2.2), temos

$$f_n(\mathbf{x}|\beta)\xi(\beta) \propto \beta^{n+3} e^{-(y+20000)\beta} \quad para \quad \beta > 0 \tag{2.7}$$

A f.d.p. posterior $\xi(\beta|\mathbf{x})$ é proporcional à Equação(2.7), dessa forma temos que $\xi(\beta|\mathbf{x})$ tem uma distribuição Gamma com parâmetros $n+4$ e $y+20000$. Portanto,

$$\xi(\beta|x) = \frac{(y+20000)^{n+4}}{(n+3!)} \beta^{n+3} e^{-(y+20000)\beta} \quad para \quad \beta > 0.$$

Observações Sequenciais

Em alguns experimentos, as observações $X_1, \ldots, X_n$ que constituem a amostra aleatória podem ser obtidas sequencialmente, isto é, uma por vez. Nesses experimentos o valor de X_1 é o primeiro a ser observado, X_2 é o segundo, e assim por diante. Considere que a f.d.p. priori do parâmetro θ é $\xi(\theta)$. Depois do valor $X_1 = x_1$ ser observado, a f.d.p. posterior $\xi(\theta|x)$ pode ser calculado de maneira usual pela relação

$$\xi(\theta|x_1) \propto f(x_1|\theta)\xi(\theta). \tag{2.8}$$

Essa f.d.p. serve como a f.d.p. priori de θ para o valor X_2 a ser observado. Depois de ser observado $X_2 = x_2$, a f.d.p. posterior $\xi(\theta|x_1, x_2)$ pode ser calculado pela relação

$$\xi(\theta|x_1, x_2) \propto f(x_2|\theta, x_1)\xi(\theta|x_1),$$

podemos assumir que dado θ, x_1 é independente de x_2, então a fórmula anterior pode ser reescrita como

$$\xi(\theta|x_1, x_2) \propto f(x_2|\theta)\xi(\theta|x_1), \tag{2.9}$$

E assim por diante, calculando a f.d.p. posterior de θ e depois usando-a como a f.d.p. priori da próxima observação. A f.d.p. posterior $\xi(\theta|x_1, \ldots, x_{n-1})$ depois que os valores $x_1, \ldots, x_{n-1}$ terem sido observados acabará por ser a f.d.p. priori de θ para o valor final observado X_n. A f.d.p. posterior depois de todos os n valores de $x_1, \ldots, x_n$ forem observados deverá ser especificado pela relação

$$\xi(\theta|\mathbf{x}) \propto f(x_n|\theta)\xi(\theta|x_1, \ldots, x_{n-1}). \tag{2.10}$$

Alternativamente, depois de todos os n valores $x_1, \ldots, x_n$ observados, podemos calcular a f.d.p. posterior $\xi(\theta|\mathbf{x})$, combinando a f.d.p conjunta $f_n(\mathbf{x}|\theta)$ com a f.d.p. priori original $\xi(\theta)$, como indicado na Equação(2.4).

Observação: Ter isso permite uma amostragem sem tamanho da amostra fixo.

Estatística Suficiente

De acordo com DeGroot[5], temos a seguinte definição.

Definição 1 *Uma estatística é uma função real $T = r(X_1, \ldots, X_n)$ das observações na amostra aleatória, $(X_1, \ldots, X_n)$.*

Considere um problema de estimação específico em que dois observadores A e B devem estimar o valor de um parâmetro θ, supondo que A tem acesso à amostra completa $\mathbf{x} = (x_1, \ldots, x_n)$ e o observador B somente conhece $t = r(x_1, \ldots, x_n)$.
Em geral, o observador A poderá definir melhores estimadores para θ do que o B, já que A pode utilizar toda a amostra $\mathbf{x}$.
No entanto, em certos casos, para certas estatísticas $t = r(\mathbf{x})$, o observador B é capaz de construir estimadores tão bons quanto os de A. Neste caso, dizemos que $T = r(\mathbf{X})$ é uma estatística suficiente.

Definição 2 (Estatística Suficiente) *Seja $X_1, \ldots, X_n$ amostra aleatória de $f(\mathbf{x}|\theta), \theta \in \Omega$ desconhecido. Seja $T = r(X_1, \ldots, X_n)$ estatística. Dizemos que T é uma estatística suficiente para θ se $f(x_1, \ldots, x_n|T = t)$ não depender de θ, qualquer que seja o valor de t.*

Exemplo 7 *Distribuição Bernoulli.*
Seja X_1, X_2, X_3 uma amostra de tamanho $n = 3$ de uma distribuição de Bernoulli(θ), isto é, dado $\theta \in (0, 1)$,

$$f(x_1, x_2, x_3|\theta) = \theta^{x_1+x_2+x_3}(1-\theta)^{3-x_1-x_2-x_3} \prod_{i=1}^{3} I_{0,1}(x_i).$$

Considere as estatísticas

$$T = X_1 + X_2 + X_3,$$

$$S = X_1X_2 + X_3.$$

As densidades condicionais de $(X_1, X_2, \ldots, X_n)$ dado $T = t$ e dado $S = s$ são como na seguinte tabela.

Por exemplo,

(X_1, X_2, X_3)	T	S	$f(x_1, x_2, x_3 \mid T = t)$	$f(x_1, x_2, x_3 \mid S = s)$
(0,0,0)	0	0	1	$\frac{(1-\theta)}{(1+\theta)}$
(0,0,1)	1	1	1/3	$\frac{(1-\theta)}{(1+2\theta)}$
(0,1,0)	1	0	1/3	$\frac{\theta}{(1+\theta)}$
(1,0,0)	1	0	1/3	$\frac{\theta}{(1+\theta)}$
(0,1,1)	2	1	1/3	$\frac{\theta}{(1+2\theta)}$
(1,0,1)	2	1	1/3	$\frac{\theta}{(1+2\theta)}$
(1,1,0)	2	1	1/3	$\frac{\theta}{(1+2\theta)}$
(1,1,1)	3	2	1	1

Tabela 2.1:

$$f(X_1 = 0, X_2 = 1, X_3 = 0 | T = 1) = \frac{P(X_1 = 0, X_2 = 1, X_3 = 0, T = 1)}{P(T = 1)}$$

$$= \frac{P(X_1 = 0, X_2 = 1, X_3 = 0)}{P(T = 1)} = \frac{(1-\theta)\theta(1-\theta)}{3\theta(1-\theta)^2} = \frac{1}{3}$$

e

$$f(X_1 = 1, X_2 = 0, X_3 = 1 | S = 1) = \frac{P(X_1 = 0, X_2 = 1, X_3 = 0, S = 1)}{P(S = 1)}$$

$$= \frac{P(X_1 = 0, X_2 = 1, X_3 = 0)}{P(S = 1)} = \frac{\theta^2(1-\theta)}{\theta(1-\theta)^2 + 3\theta^2(1-\theta)} = \frac{\theta}{1-\theta+3\theta} = \frac{\theta}{1+2\theta}.$$

Pois,

$$\begin{aligned} P(S = 1) &= P(X_1 = 0, X_2 = 0, X_3 = 1) + P(X_1 = 0, X_2 = 1, X_3 = 1) \\ &\quad + P(X_1 = 1, X_2 = 1, X_3 = 0) + P(X_1 = 1, X_2 = 0, X_3 = 1) \\ &= \theta(1-\theta)^2 + \theta^2(1-\theta) + \theta^2(1-\theta) + \theta^2(1-\theta) \\ &= \theta(1-\theta)^2 + 3\theta^2(1-\theta). \end{aligned}$$

A distribuição condicional de (X_1, X_2, X_3) *dado* $T = t$ *é independente de* θ *para todo* t*, desta forma,* $T = \sum_{i=1}^{3} X_i$ *é uma estatística suficiente para* θ*.*
A distribuição condicional de (X_1, X_2, X_3) *dado* $S = s$ *depende de* θ *para* $s = 0$ *ou* $s = 1$*, desta forma,* $S = X_1X_2 + X_3$ *não é uma estatística suficiente.*

Uma estatística suficiente T é vista como um resumo de todas as informações relevantes sobre θ contidas na amostra $X_1, \ldots, X_n$.
Seja $X_1, \ldots, X_n$ uma amostra aleatória com f.d.p. ou f.p. $f(\mathbf{x}|\theta)$, com parâmetro $\theta \in \Omega$. Suponhamos que θ seja um vetor de valores reais. Neste caso, precisamos encontrar duas ou mais estatísticas $T_1, \ldots, T_k$, que serão estatísticas conjuntamente suficientes para θ, conforme Mood[8].

Definição 3 (Estatística Conjuntamente Suficiente) *Seja* $X_1, \ldots, X_n$ *uma amostra aleatória de tamanho* n *com densidade conjunta*

$$f(\mathbf{x}|\theta_1, \ldots, \theta_k),$$

para $\theta = (\theta_1, \ldots, \theta_k) \in \Omega$*. As estatísticas* $T_1, \ldots, T_k$ *são ditas conjuntamente suficientes se, e somente se, a distribuição condicional de* $X_1, \ldots, X_n$ *dado* $T_1 = t_1, \ldots, T_k = t_k$ *não depender de* θ*.*

Exemplo 8 *Distribuição Uniforme.*
Seja $X_1, \ldots, X_n$ *uma amostra aleatória de tamanho* n *com distribuição Uniforme*(θ_1, θ_2)*, com* θ_1 *e* θ_2 *desconhecidos. Para encontrar estatísticas*

conjuntamente suficientes para θ_1 e θ_2., consideremos

$$f(x|\theta) = \frac{1}{\theta_2 - \theta_1} I_{(\theta_1,\theta_2)}(x).$$

A f.d.p. conjunta é dada por

$$f(\mathbf{x}|\theta_1, \theta_2) = \prod_{i=1}^{n} \frac{1}{\theta_2 - \theta_1} I_{(\theta_1,\theta_2)}(x).$$

Definindo as estatísticas de ordem como $X_{(1)}, \ldots, X_{(n)}$, temos que $X_{(1)}$ é o mínimo valor de X, e $X_{(n)}$ é o valor máximo de X. Neste caso, como as estatísticas de ordem também estão entre θ_1 e θ_2, podemos dizer que $x_{(n)} < \theta_2 < \infty$ e $-\infty < \theta_1 < x_{(1)}$. Dessa forma, para $\mathbf{x} = (x_1, \ldots, x_n)$

$$f(\mathbf{x}|\theta_1, \theta_2) = (\theta_2 - \theta_1)^{-n} I_{(x_{(n)},\infty)}(\theta_2) I_{(-\infty,x_{(1)})}(\theta_1).$$

Podemos notar que a f.d.p. conjunta depende de $\mathbf{x}$ através dos valores $t_1 = x_{(1)}$ e $t_2 = x_{(n)}$. Assim, as estatísticas $(T_1, T_2) = (x_{(1)}, x_{(n)})$ são candidatas à estatísticas conjuntamente suficientes para (θ_1, θ_2).

O seguinte resultado nos permite identificar estatísticas suficientes a partir da densidade conjunta.

Teorema 4 *Seja $f(x|\theta_1, \ldots, \theta_m), a < x < b$, uma densidade, onde a e b não envolvem o vetor $\theta = (\theta_1, \ldots, \theta_m)$. Se a densidade conjunta de uma amostra aleatória dessa distribuição puder ser fatorada como*

$$f(x_1, \ldots, x_n|\theta_1, \ldots, \theta_m) = g(t_1 = r_1(x_1, \ldots, x_n), \ldots, t_k = r_k(x_1, \ldots, x_n), \theta) h(x_1, \ldots, x_n)$$

onde $h(x_1, \ldots, x_n)$ não depende de θ, então $T_1, \ldots, T_k$ são um conjunto de k estatísticas suficientes.

Exemplo 9 *Estatísticas conjuntamente suficientes para os parâmetros de uma distribuição Normal.*

Suponha que $X_1, \ldots, X_n$ formam uma amostra aleatória de uma distribuição normal com média μ e variância σ^2 desconhecidas. A f.d.p. conjunta de $X_1, \ldots, X_n$ é dada por

$$\begin{aligned} f_n(\mathbf{x}|\mu, \sigma^2) &= \prod_{i=1}^{n} \frac{1}{(2\pi)^{1/2}\sigma} \exp\left(-\frac{(x_i-\mu)^2}{2\sigma^2}\right) \\ &= \frac{1}{(2\pi)^{n/2}\sigma^n} \exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^{n} x_i^2\right) \exp\left(\frac{\mu}{\sigma^2}\sum_{i=1}^{n} x_i - \frac{n\mu^2}{2\sigma^2}\right). \end{aligned}$$

Podemos notar que esta f.d.p. conjunta depende de $\mathbf{x}$ através do valores $t_1 = \sum_{i=1}^{n} x_i$ e $t_2 = \sum_{i=1}^{n} x_i^2$,

$$f_n(\mathbf{x}|\mu, \sigma^2) = \frac{1}{(2\pi)^{n/2}\sigma^n} \exp\left(-\frac{1}{2\sigma^2} t_2(\mathbf{x})\right) \exp\left(\frac{\mu}{\sigma^2} t_1(\mathbf{x}) - \frac{n\mu^2}{2\sigma^2}\right).$$

Portanto, as estatística $(T_1, T_2) = (\sum_{i=1}^{n} x_i, \sum_{i=1}^{n} x_i^2)$ são estatísticas conjuntamente suficientes para (μ, σ^2).

Para uma densidade qualquer, podem existir vários conjuntos de estatísticas suficientes. A amostra inteira, por exemplo, é uma estatística conjuntamente suficiente. Mas estamos interessados no menor conjunto, no sentido que este conjunto é uma função de todos os outros conjuntos de estatísticas suficientes. Este menor conjunto é chamado de *estatística suficiente minimal*.

Exemplo 10 *Dsitribuição Bernoulli.*

Seja $X_1, \ldots, X_n$ uma amostra de tamanho n de uma distribuição de Bernoulli(θ), dado $\theta \in (0, 1)$,

$$f_n(x_1, \ldots, x_n|\theta) = \theta^{\sum_{i=1}^{n} x_i}(1-\theta)^{n-\sum_{i=1}^{n} x_i}$$

$$= \left(\frac{\theta}{1-\theta}\right)^{\sum_{i=1}^{n} x_i} (1-\theta)^n.$$

Pelo Teorema 4, temos que $T(\mathbf{X}) = \sum_{i=1}^{n} X_i$ *é uma estatística suficiente para* θ*.*

Fatorando $f_n(x_1, \ldots, x_n|\theta)$ *novamente, podemos obter outras estatísticas suficientes*

$$f_n(x_1, \ldots, x_n|\theta) = \left(\frac{\theta}{1-\theta}\right)^{x_1} \left(\frac{\theta}{1-\theta}\right)^{\sum_{i=2}^{n} x_i} (1-\theta)^n.$$

Desta forma, obtemos um conjunto $S(\mathbf{X}) = (X_1, \sum_{i=2}^{n} X_i)$ *de estatísticas suficientes para* θ*.*

Observamos que a estatística suficiente $T(\mathbf{X})$ *é função do conjunto* $S(\mathbf{X})$*. Portanto, podemos dizer que* $T(\mathbf{X})$ *é estatística suficiente minimal.*

Suficiência e Modelo Bayesiano

Em Bickel[4], temos a seguinte definição de estatística suficiente no contexto bayesiano..

Definição 5 *Dizemos que uma estatística* $T(\mathbf{X}) = r(\mathbf{X})$ *é uma estatística suficiente bayesiana para a distribuição priori* π*, se a distribuição posterior de* θ *dado* $\mathbf{X} = \mathbf{x}$ *é a mesma que a distribuição posterior (condicional) de* θ *dado* $T(\mathbf{X}) = r(\mathbf{x})$ *para todo* x*:*

$$\xi(\theta|\mathbf{x}) = \xi(\theta|T = r(\mathbf{x})).$$

Equivalentemente, $T(\mathbf{X})$ é uma estatística suficiente para θ se θ e $\mathbf{X}$ forem condicionalmente independentes dado $T(\mathbf{X})$.

De fato, observamos que se θ e $\mathbf{X}$ forem condicionalmente independentes dado $T(\mathbf{X})$ então

$$\xi(\theta, \mathbf{X}|T(\mathbf{X})) = \xi(\theta|T(\mathbf{X}))\xi(\mathbf{X}|T(\mathbf{X}))$$

onde $\xi(\theta|T(\mathbf{X}))$ é função de θ dado $T(\mathbf{X})$ e $\xi(\mathbf{X}|T(\mathbf{X}))$ é função apenas de $\mathbf{X}$. Desta maneira, pelo Teorema 4 podemos dizer que $T(\mathbf{X})$ é estatística suficiente para θ.

Estimadores de Máxima Verossimilhança e Estimadores de Bayes como Estatísticas Suficientes

Suponha agora que $X_1, \ldots, X_n$ formam uma amostra aleatória com f.d.p. ou f.p. $f(x|\theta)$, onde o valor do parâmetro θ é desconhecido, e que $T = r(X_1, \ldots, X_n)$ é uma estatística suficiente para θ.

Proposição 6 *O Estimador de Máxima Verossimilhança (EMV), $\hat{\theta}$, depende das observações $X_1, \ldots, X_n$ somente através da estatística T.*

Prova. Pelo Teorema 4, fatorando a função de verossimilhança obtemos,

$$f_n(\mathbf{x}|\theta) = g(T = r(\mathbf{x}), \theta)h(\mathbf{x}).$$

O EMV , $\hat{\theta}$, é o valor de θ onde $f_n(\mathbf{x}|\theta)$ é máximo. Pela fatoração acima, podemos dizer que $\hat{\theta}$ será o valor de θ onde $g(T = r(\mathbf{x}), \theta)$ é máximo. Assim, $g(T = r(\mathbf{x}), \theta)$ depende do vetor observado $\mathbf{x}$ somente através da função $r(\mathbf{x})$, e implica que $\hat{\theta}$ dependerá de $\mathbf{x}$ somente através da função $r(\mathbf{x})$. Desta forma, o estimador $\hat{\theta}$ é uma função de $T = r(X_1, \ldots, X_n)$. $\bullet$

Como o estimador $\hat{\theta}$ é uma função das observações $X_1, \ldots, X_n$ e não é uma função do parâmetro θ, o estimador $\hat{\theta}$ é uma estatística. Em muitos problemas, $\hat{\theta}$ é

uma estatística suficiente. Como O estimador $\hat{\theta}$ sempre será função de alguma estatística suficiente, temos o seguinte resultado.

Teorema 7 *Se o EMV $\hat{\theta}$ for uma estatística suficiente, então $\hat{\theta}$ é uma estatística suficiente minimal.*

Esta propriedade pode ser estendida para um vetor de parâmetros θ. Em muitos problemas, os EMV $\hat{\theta}_1, \ldots, \hat{\theta}_k$ formam um conjunto de estatísticas conjuntamente suficientes, e neste caso, eles serão estatísticas conjuntamente suficientes minimais. Esta discussão também é válida para o estimador de Bayes. Considere que um parâmetro θ deve ser estimado e que uma f.d.p. priori $\xi(\theta)$ é atribuída para θ. Se uma estatística $T = r(X_1, \ldots, X_n)$ for uma estatística suficiente, então pela Equação (2.5) e pela fatoração da função de verossimilhança, a f.d.p. posterior $\xi(\theta|\mathbf{x})$ satisfaz a seguinte relação,

$$\xi(\theta|\mathbf{x}) \propto g(T = r(\mathbf{x}), \theta)\xi(\theta).$$

Observamos nesta relação que a f.d.p. posterior dependerá do vetor de observações $\mathbf{x}$ apenas através do valor $t = r(\mathbf{x})$. Como o estimador de Bayes de θ com respeito a alguma função perda especificada é calculado pela f.d.p. posterior, a estimativa dependerá do vetor de observações $\mathbf{x}$ somente através do valor de $t = r(\mathbf{x})$. Em outras palavras, o estimador de Bayes é uma função de $T = r(\mathbf{X})$.
Como o estimador de Bayes é uma estatística e ele é uma função de alguma estatística suficiente T, temos o seguinte resultado.

Teorema 8 *Se o estimador de Bayes for uma estatística suficiente, então ele é uma estatística suficiente minimal.*

Famílias Exponenciais

Estudaremos nesta seção, baseada em Bickel[4], distribuições pertencentes à família exponencial, cujas propriedades permitem determinar estatísticas suficientes minimais rapidamente.

Modelos de probabilidade que fazem parte desta família incluem normal, binomial, Poisson, gama, beta, e modelos de regressão múltipla usados para descrever a variável resposta Y de acordo com um conjunto de variáveis preditoras. Em geral, estas famílias formam a base de uma importante classe de modelos, chamada de modelos lineares generalizados.

Caso Uniparamétrico

Considere uma família de distribuições de um modelo $P_\theta, \theta \in \Omega$, com densidade $f(x|\theta)$. Esta família é dita pertencer à família exponencial uniparamétrica, se existirem funções de valores reais, com $T(x)$ e $h(x) \in R^n$ e $\eta(\theta), B(\theta) \in \Omega$, tais que a função de densidade $f(x|\theta)$ pode ser escrita como

$$f(x|\theta) = h(x)\exp(\eta(\theta)T(x) - B(\theta)) \tag{2.11}$$

Na família exponencial uniparamétrica a variável aleatória $T(X)$ é uma estatística suficiente para θ.

Exemplo 11 *Distribuição de Poisson.*

Seja $X_1, \ldots, X_n$ uma amostra aleatória de tamanho n de distribuição de Poisson com média θ desconhecida. Então, para $x \in$ 0, 1, 2,. . .,

$$f(x, \theta) = \frac{\theta^x e^{-\theta}}{x!} =, \quad \theta > 0.$$

A f.d.p. conjunta é dada por

$$f_n(\mathbf{x}|\theta) = \prod_{i=1}^{n} \frac{\theta_i^x e^{-\theta}}{x_i!} = \frac{\theta^{\sum_{i=1}^n x_i} e^{-n\theta}}{\prod_{i=1}^n x_i!} = \frac{1}{\prod_{i=1}^n x_i!} \exp(\sum_{i=1}^n x_i \log\theta - n\theta).$$

Daqui, podemos identificar

$$\eta(\theta) = \log\theta, B(\theta) = n\theta, T(x) = \sum_{i=1}^n x_i, h(x) = \frac{1}{\prod_{i=1}^n x_i!}.$$

Desta forma, $T(X) = \sum_{i=1}^n X_i$ é estatística suficiente para θ.

Exemplo 12 *Família Binomial.*

Suponha que $X_1, \ldots, X_n$ é uma amostra aleatória de tamanho n com distribuição $Bin(m, \theta), \quad 0 < \theta < 1$. Então, para $x \in 0, 1, \ldots, m$,

$$f(x, \theta) = C_x^m \theta^x (1-\theta)^{m-x}.$$

A f.d.p. conjunta é

$$f(\mathbf{x}, \theta) = \prod_{i=1}^n C_{x_i}^m \theta^{x_i} (1-\theta)^{m-x_i} = \prod_{i=1}^n (C_{x_i}^m) \theta^{\sum_{i=1}^n x_i} (1-\theta)^{nm - \sum_{i=1}^n x_i}$$

$$= \prod_{i=1}^n (C_{x_i}^m) \exp\left[\sum_{i=1}^n x_i \log\left(\frac{\theta}{1-\theta}\right) + nm\log(1-\theta)\right].$$

Esta família pertence à família exponencial uniparamétrica com

$$\eta(\theta) = \log\left(\frac{\theta}{1-\theta}\right), B(\theta) = nm\log(1-\theta), T(x) = \sum_{i=1}^n x_i, h(x) = \prod_{i=1}^n (C_{x_i}^m).$$

Uma estatística suficiente é dada por $T(X) = \sum_{i=1}^n X_i$.

Caso Multiparamétrico

Considere uma família de distribuições de um modelo $P_\theta, \theta \in \Omega$, $\Omega \subset R^k$, com

densidade $f(x|\theta)$. Esta família é dita pertencer à família exponencial com k parâmetros, se existirem funções de valores reais, com $T_1(x), \ldots, T_k(x)$ e $h(x) \in R^n$ e $\eta_1(\theta), \ldots, \eta_k(\theta), B(\theta) \in \Omega$, tais que a função de densidade $f(x|\theta)$ pode ser escrita como

$$f(x|\theta) = h(x) \exp\left(\sum_{j=1}^{k} \eta_j(\theta) T_j(x) - B(\theta)\right). \tag{2.12}$$

O vetor $\mathbf{T}(X) = (T_1(X), \ldots, T_k(X))$ são estatísticas suficientes para $\theta = (\theta_1, \ldots, \theta_k)$, baseado no Teorema 4.

Exemplo 13 *Distribuição Normal.*

Seja $X_1, \ldots, X_n$ uma amostra aleatória com distribuição Normal(μ, σ^2), com $\theta = (\mu, \sigma^2)$. A f.d.p. conjunta pode ser escrita como

$$f_n(\mathbf{x}|\theta) = \prod_{i=1}^{n} \frac{1}{(2\pi)^{1/2}\sigma} \exp\left(-\frac{(x_i - \mu)^2}{2\sigma^2}\right)$$

$$= \exp\left[\frac{\mu}{\sigma^2}\sum_{i=1}^{n} x_i - \frac{\sum_{i=1}^{n} x_i^2}{2\sigma^2} - \frac{1}{2}\left(\frac{n\mu^2}{\sigma^2} + \log((2\pi)^{n/2}\sigma^n)\right)\right].$$

Desta forma temos,

$$\eta_1(\theta) = \frac{\mu}{\sigma^2}, T_1(\mathbf{x}) = \sum_{i=1}^{n} x_i, \eta_2(\theta) = -\frac{1}{2\sigma^2}, T_2(\mathbf{x}) = \sum_{i=1}^{n} x_i^2,$$

$$B(\theta) = \frac{1}{2}\left(\frac{n\mu^2}{\sigma^2} + \log((2\pi)^{n/2}\sigma^n)\right), h(\mathbf{x}) = 1.$$

Portanto, as estatísticas $\mathbf{T}(X) = (T_1(\mathbf{x}) = \sum_{i=1}^{n} x_i, T_2(\mathbf{x}) = \sum_{i=1}^{n} x_i^2)$ são conjuntamente suficientes para $\theta = (\theta_1 = \mu, \theta_2 = \sigma^2)$.

Distribuições Prioris Conjugadas

Certas distribuições prioris são particularmente conveniente para usar com amostras de algumas outras distribuições. Por exemplo, suponha que uma

amostra aleatória tenha distribuição Bernoulli com parâmetro θ desconhecido. Se a distribuição priori de θ é Beta, então, após os dados serem observados, a distribuição posterior também é Beta. De fato, temos o seguinte resultado

Teorema 9 *Suponha que $X_1, \ldots, X_n$ formam uma amostra aleatória com distribuição Bernoulli com parâmetro θ desconhecido $(0 < \theta < 1)$. Suponha que a distribuição priori de θ é Beta com parâmetros α e β $(\alpha > 0, \beta > 0)$. Então a distribuição posterior de θ dado $X = x$ é uma distribuição Beta com parâmetros $\alpha + \sum_{i=1}^{n} x_i$,$\beta + n - \sum_{i=1}^{n} x_i$.*

Prova. Seja $y = \sum_{i=1}^{n} x_i$ e $\mathbf{x} = (x_1, \ldots, x_n)$. Então $f_n(\mathbf{x}|\theta)$ de $X_1, X_2, \ldots, X_n$ é dada pela equação

$$f_n(\mathbf{x}|\theta) = \begin{cases} \theta^y (1-\theta)^{n-y} & \text{para} \quad \theta = (0,1), \\ 0 & \text{caso contrário} \end{cases}$$

A f.d.p. priori $\xi(\theta)$ satisfaz a seguinte relação

$$\xi(\theta) \propto \theta^{\alpha-1}(1-\theta)^{\beta-1} \quad \text{para} \quad 0 < \theta < 1.$$

A f.d.p. posterior $\xi(\theta|\mathbf{x})$ é proporcional ao produto $f_n(x|\theta)\xi(\theta)$, dessa forma:

$$\xi(\theta|\mathbf{x}) \propto \theta^{\alpha+y-1}(1-\theta)^{\beta+n-y-1} \quad \text{para} \quad 0 < \theta < 1.$$

Observamos que o lado direito da relação é igual a f.d.p. da Beta com parâmetros $\alpha + y$ e $\beta + n - y$, exceto pelo fator constante. Assim, a distribuição posterior de θ é dado pelo teorema acima. •

Exemplo 14 *A variância e a Distribuição Beta.*

Suponha que a proporção θ de itens defeituosos de uma grande remessa é desconhecido, e que a distribuição priori de θ é uma distribuição uniforme no intervalo $(0,1)$. Os itens foram selecionados de uma amostra dessa remessa e foi inspecionado a variância da distribuição posterior de θ, essa variância deve ser igual ou menor que 0,01. Vamos determinar o número total de itens defeitosos e não defeitosos que pode ser obtido antes do processo amostral ser parado. A distribuição uniforme no intervalo $(0,1)$ é uma distribuição Beta com parâmetros $\alpha = 1$ e $\beta = 1$. Portanto, depois que foi obtido y itens defeituosos e z itens não defeituosos, a disstribuição posterior de θ será uma distribução Beta com $\alpha = y+1$ e $\beta = z+1$. A variância da distribuição Beta com parâmetros α e β é $\frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$. Assim a variância Var da distribuição posterior de θ será:

$$Var = \frac{(y+1)(z+1)}{(y+z+2)^2(y+z+3)}$$

A amostragem pára quando o número de defeitos y e o de não defeitos z que foram obtidos são tais que $Var \leq 0,01$. Se considerarmos $y + z = 22$, obtemos que $Var = 0,01$, portanto não serão necessários selecionar mais que 22 itens.

Capítulo 3

Teoria de Decisão

Consideremos o contexto do problema

$$\begin{cases} H_0 : \theta = \theta_0 \\ H_1 : \theta = \theta_1, \end{cases} \tag{3.1}$$

Denotemos por d_0 a decisão de aceitar H_0, e por d_1 a decisão de aceitar H_1. Denotemos por K_1 a perda sofrida ao tomar a decisão d_1 quando H_0 for verdadeira, e por K_0, a perda sofrida ao tomar a decisão d_0 quando H_1 for verdadeira.

Suponhamos que a probabilidade priori de que H_0 seja verdadeira é π_0 e, portanto, a probabilidade priori de que H_1 seja verdadeira é $\pi_1 = 1 - \pi_0$.

Assim, a perda esperada, $r(\delta)$, para qualquer procedimento de decisão δ é

$$r(\delta) = \pi_0 E(\text{Perda}|\, \theta = \theta_0) + \pi_1 E(\text{Perda}|\, \theta = \theta_1).$$

Em termos de $\alpha(\delta)$ e $\beta(\delta)$, as probabilidades do Erro Tipo I e Erro Tipo II, respectivamente, temos

$$E(\text{Perda}|\, \theta = \theta_0) \;=\; K_1 Pr(\text{escolher } d_1|\, \theta = \theta_0) = K_1 \alpha(\delta)$$

$$E(\text{Perda}|\,\theta = \theta_1) \quad = \quad K_0 Pr(\text{escolher } d_0|\,\theta = \theta_1) = K_0\beta(\delta).$$

Portanto, temos finalmente, que o erro esperado do procedimento de teste é

$$r(\delta) = \pi_0 K_1 \alpha(\delta) + \pi_1 K_0 \beta(\delta)$$

Um procedimento de teste de Bayes é aquele que minimiza $r(\delta)$. Pelo discutido anteriormente, dados K_1, K_0, π_0 e π_1, fixos, e uma amostra observada $d = (x_1, \ldots, x_n)$, o procedimento de teste bayesiano, $\delta_B^*(d)$, é definido por

$$\delta_B^*(d) = \begin{cases} \text{rejeitar } H_0 & \text{se } \pi_0 K_1 f_n(\mathbf{x}|\theta_0) < \pi_1 K_0 f_n(\mathbf{x}|\theta_1), \\ \text{não rejeitar } H_0 & \text{se } \pi_0 K_1 f_n(\mathbf{x}|\theta_0) > \pi_1 K_0 f_n(\mathbf{x}|\theta_1). \end{cases} \tag{3.2}$$

Exemplo 15 *Proporção de itens defeituosos.*

Como no Exemplo 2, suponha que a proporção θ de itens defeituosos em um lote de uma grande fábrica é desconhecido e que a probabilidade priori de que $H_0 : \theta_0$ seja verdadeira é $\pi_0 = 0,8$ e a probabilidade de $H_1 : \theta_1$ seja verdadeira é $\pi_1 = 0,2$. Queremos testar

$$\begin{cases} H_0 : \theta = 1/4 \\ H_1 : \theta = 1/8. \end{cases}$$

A perda sofrida $K_1 = 5$ e a perda $K_0 = 1$, ou seja, como a perda K_1 é maior que K_0 quer dizer que o erro do tipo I é mais grave que o erro tipo II.

A amostra obtida de $n = 100$ foi de $y = \sum i = 1^1 00 x_i = 20$.

Calculando $f_n(\mathbf{x}|\theta_0)$ e $f_n(\mathbf{x}|\theta_1)$:

$$f_n(\mathbf{x}|\theta_0 = 0,25) = \theta_0^y (1-\theta_0)^{n-y} = 0,25^{20} 0,75^{100-20} = 9,19 \times 10^{-23}$$

$$f_n(\mathbf{x}|\theta_1 = 0,125) = 0,125^2 00,875^{100-20} = 1,98 \times 10^{-23}$$

Dessa forma, temos que

$$\pi_0 K_1 f_n(\mathbf{x}|\theta_0) = 0,8 \times 5 \times 9,19 \times 10^{-23} = 3,676 \times 10^{-22} \quad e$$

$$\pi_1 K_0 f_n(\mathbf{x}|\theta_1) = 0,2 \times 1 \times 1,98 \times 10^{-23} = 3,96 \times 10^{-24}.$$

Concluimos que $\pi_0 K_1 f_n(\mathbf{x}|\theta_0) > \pi_1 K_0 f_n(\mathbf{x}|\theta_1)$*, ou seja, não temos evidência de rejeitar que* θ *seja igual a 0,25.*

3.0.4 Fator de Bayes

Para o problema (3.1), denotemos por π_0 e π_1 as probabilidades prioris das hipóteses testadas: $\pi_i = \mathrm{P}(\theta_i)$, i=0,1.

Obtida a amostra, para testarmos H_0 e H_1, calculamos as probabilidades posteriores $\alpha_0 = P(\theta_0|x)$ e $\alpha_1 = P(\theta_1|x)$ das respectivas hipóteses e decidir entre elas de acordo com essas probabilidades.

Definição 10 *A razão* α_0/α_1 *é chamada de razão de chances posterior a favor de* H_0 *em relação a* H_1*, e* π_0/π_1 *é chamada de razão de chances priori. A quantidade*

$$B(x) = \frac{\textit{razão de chances posterior}}{\textit{razão de chances priori}} = \frac{\alpha_0/\alpha_1}{\pi_0/\pi_1} = \frac{\alpha_0\pi_1}{\alpha_1\pi_0}$$

é chamado Fator de Bayes a favor de H_0 *em relação a* H_1*.*

Assumamos que a função perda é a fornecida pela Tabela 1, onde K_1 é uma constante que indica a perda quando rejeitamos H_0 quando essa hipótese é verdadeira,

ou seja, é a perda do Erro Tipo I. Ela é definida pela comparação com a perda do Erro Tipo II, dada por K_0. Por exemplo, quando o Erro Tipo I é mais grave que o Erro Tipo II, então a constante K_1 é maior que K_0.

	H_0 é verdadeiro	H_1 é verdadeiro
Não Rejeitar H_0	0	K_0
Rejeitar H_0	K_1	0

Tabela 3.1: Função perda: não existe perda se tomarmos a decisão correta, mas existe perda K_0 se não rejeitamos H_0 dado que H_1 é verdadeiro, e perda K_1 se rejeitamos H_0 dado que H_0 é verdadeiro.

Uma regra de decisão de Bayes resulta na rejeição da hipótese nula se e somente se: $\alpha_0 K_1 < \alpha_1 K_0$, ou seja, se a perda esperada da decisão d_1 é menor que a perda esperada da decisão d_0.

Em termos do Fator de Bayes, rejeitamos H_0 se e somente se

$$B(x) = \frac{\alpha_0/\alpha_1}{\pi_0/\pi_1} < \frac{K_0/K_1}{\pi_0/\pi_1}$$

Desta forma, definimos a Região Crítica como

$$RC = \left\{ x : B(x) < \frac{K_0/K_1}{\pi_0/\pi_1} \right\}.$$

Exemplo 16 *Proporção de itens defeituosos.*

Continuando com o Exemplo 7, iremos calcular o Fator de Bayes, para isso, primeiramente iremos calcular as probabilidades posteriores $\alpha_0 = P(\theta_0|x)$ *e* $\alpha_1 = P(\theta_1|x)$.

$$\alpha_0 = f_n(\mathbf{x}|\theta_0 = 0,25)\pi_0 = \theta_0^y(1-\theta_0)^{n-y} \times 0,8 = 7,35 \times 10^{-23}$$

$$\alpha_1 = f_n(\mathbf{x}|\theta_1 = 0,125)\pi_1 = 0,125^2 00,875^{100-20} \times 0,82 = 3,96 \times 10^{-24}$$

A razão de chances posterior α_0/α_1 é 18,56, e a razão de chances priori é 4, dessa forma o fator de Bayes é dado por $B(x) = 18,56/4 = 4,64$. Temos que $\frac{K_0/K_1}{\pi_0/\pi_1} = 0,05$. Assim, $B(x) > \frac{K_0/K_1}{\pi_0/\pi_1}$, isto é, não rejeitamos $H_0 : \theta = 0,25$.

3.0.5 Teste da Razão de Verossimilhança

Considerando novamente o problema (3.1), onde $\theta \in \Omega$, considere que $\mathbf{X} = (X_1, \ldots, X_n)$ tem função de densidade $f_n(\mathbf{x}|\theta)$. O teste estatístico que queremos considerar é a *Razão de Verossimilhança*, dada por

$$\lambda(\mathbf{x}) = \frac{\sup_{\theta\in\Omega_0} f_n(\mathbf{x}|\theta)}{\sup_{\theta\in\Omega} f_n(\mathbf{x}|\theta)}.$$

Note que λ é uma função de $x_1, \ldots, x_n$. Quando as observações são trocadas com as respectivas variáveis aleatórias $X_1, \ldots, X_n$, então temos que

$$\Lambda = \lambda(X_1, \ldots, X_n)$$

é uma estatística.

A estatística Λ é usada para formular o Teste da Razão de Verossimilhança,

$$\delta_B^*(d) = \begin{cases} \text{rejeitar } H_0 & \text{se } \Lambda \leq k, \\ \text{não rejeitar } H_0 & \text{se } \Lambda > k, \end{cases} \tag{3.3}$$

onde k é uma constante fixa $0 < k < 1$.

Procedimento Bayesiano

De acordo com James Berger[**??**], seja π_0' e π_0'' constantes que satisfazem $\pi_0' \leq$

$K_0/(K_0 + K_1) \leq \pi_0''$, onde K_0 e K_1 são dados pela Tabela 3.0.4.

Desta forma, podemos definir a *razão de verossimilhança* para θ_1 e θ_0 como

$$L_n = \frac{\prod_{i=1}^{n} f(x_i|\theta_1)}{\prod_{i=1}^{n} f(x_i|\theta_0)}.$$

Teorema 11 *Seja π_0 a probabiliddae priori de θ_0 e π_1 a probabiliddae priori de θ_1. Denotemos por d_0 a decisão de aceitar H_0, e por d_1 a decisão de aceitar H_1.*

Se $0 < \pi_0 < 1$, então o procedimento de Bayes é da seguinte forma:

se $L_n \leq A$, para a amostra e decide por d_0;

se $L_n \geq B$, para a amostra e decide por d_1;

se $A < L_n < B$, extrair outra observação;

onde $A = \pi_0(1 - \pi_0'')/(\pi_1\pi_0'')$ e $B = \pi_0(1 - \pi_0')/(\pi_1\pi_0')$.

Assumimos que $\pi_0' < \pi_0 < \pi_0''$, de tal forma que estas constantes são fixadas como limites inferior e superior.

Notamos que podemos comparar este procedimento de Bayes utilizando a razão de verossimilhança, com o procedimento clássico, no qual fixamos o nível de significância.

O nível de significância é a probabilidade de rejeição fixada de um teste de hipótese estatística quando a hipótese nula é verdadeira. No procedimento de Bayes apresentado no Teorema 11 podemos reescrever a primeira parte do procedimento, como

$L_n \leq A = \pi_0(1 - \pi_0'')/(\pi_1\pi_0'')$ que equivale dizer $\frac{\pi_1}{\pi_0} L_n \leq \frac{1-\pi_0''}{\pi_0''}$.

Portanto, podemos observar que $\frac{1-\pi_0''}{\pi_0''}$ é um limite fixado para aceitar a decisão d_0.

Capítulo 4

Ensaios Clínicos

Os Ensaios Clínicos apresentam uma fase inicial do estudo. Primeiramente os novos medicamentos são testadas em animais, e se os resultados forem satisfatórios iniciam-se os estudos em seres humanos.

Estes estudos em seres humanos são divididos em 4 fases. A Fase 1 é um estudo piloto que avalia os efeitos do medicamento no organismo e a dose de tolerância. Envolve voluntários "normais"para reduzir o risco de problemas tóxicos e evitar confundimento entre doença e efeito do medicamento. São cerca de 4 a 20 indivíduos que são monitorados intensamente.

Quando os níveis de dose seguros são estabelecidos, então o medicamento começa a ser administrado em pacientes, é a Fase 2: que determina se há evidência suficiente de eficácia preliminar para justificar estudos adicionais.

A Fase 3 é uma extensão da Fase 2 englobando os ensaios clínicos controlados comparativos planejados com o objetivo de se obter informações mais precisas

sobre efeitos adversos e eficácia.

Em uma fase adicional, a Fase 4, consiste em ensaios de pós-aprovação do medicamento por órgão regular competente, os estudos prosseguem com diferentes propósitos: avaliar a incidência de efeitos adversos raros, explorar um específico efeito do medicamento, etc.

Para que um novo medicamento seja utilizado na prática médica é fundamental que sua eficácia seja demonstrada quando comparada com uma terapia usual. Estudos controlados, onde alguns pacientes são tratados como experimentais por utilização do medicamento experimental e outros como controles por omissão do novo medicamento, assim permitem um padrão de comparabilidade. Estes controles podem ser concorrentes, quando o mesmo protocolo é seguido para os grupos controle e experimental, ou históricos.

Focalizando nos ensaios clínicos com controles concorrentes, dois aspectos são merecedores de alguma atenção, a aleatorização e o planejamento.

A aleatorização é um fundamento básico nos ensaios clínicos, tendo como propósito de evitar vícios na atribuição de tratamentos e confundimento de variáveis desconhecidas.

A aleatorização é também crucial para os propósitos de inferências estatísticas frequentistas.

Ainda que os experimentos sejam aleatorizados, se o paciente responde de um jeito diferente por saber que medicamento está utilizando, ou que o médico atribua tratamento diferenciado dependendo do medicamento tomado pelo paciente, pode interferir no estudo. Por este motivo existe o "double-blind"(duplo-cego) onde

nem o paciente nem o médico são informados do tratamento atribuído.
De modo generalizado podemos classificar os planejamentos dos ensaios clínicos em dois grandes grupos:

- Planejamento não sequenciais: inclusão de um número fixo de pacientes para os quais o tratamento é atribuído aleatoriamente.

- Planejamento sequenciais: o monitoramento dos dados pode ser feito continuamente, ou periodicamente. Estes planejamentos permitem que os resultados provisórios afetem a atribuição dos tratamentos para os próximos pacientes, ou que a parada do experimento ocorra mais cedo.

Na próxima Seção será apresentada uma aplicação descrita por Berry[2].

4.0.6 Aplicação

Consideremos o experimento conduzido em vários hospitais das cidades dos Estados Unidos, com início em abril de 1959, com o objetivo de comparar dois medicamentos para o tratamento de leucemia aguda: A: medicamento 6-mercaptopurina e B: placebo (controle). Neste estudo, os pacientes tratados foram agrupados em pares homogêneos. A escolha do grupo foi aleatorizada para os pacientes que compõem o par.
O ensaio modela as observações como sendo de Bernoulli indicando se o medicamento A é mais efetivo no par. Supõe-se que empates na comparação não ocorram.

Logo, podemos representar os resultados deste experimento pelas variáveis:

$$X_i = \begin{cases} 1 & \text{se medicamento A melhor} \\ 0 & \text{se medicamento B (placebo) melhor} \end{cases} \text{, no i-ésimo par, i=1, 2, ..., n}$$

Consideremos θ como a probabilidade de que o medicamento A tenha melhor desempenho do que o medicamento B em um dado par.

Assim, as variáveis X_i são condicionalmente independentes e identicamente distribuídas Bernoulli(θ), dado $\theta \in$ [0,1].

Estamos interessados em testar: $H_0 : \theta = 1/2$ e $H_1 : \theta \neq 1/2$.

Neste estudo foi adotado o planejamento seqüencial, isto é, os dados obtidos foram analisados continuamente até obter evidência suficiente a favor de uma das hipóteses.

O critério de parada do experimento foi definido como a diferença entre o número de sucessos (n_a) e fracassos (n_b) dado pela Figura 1 (Berry (1987-a), p.120).

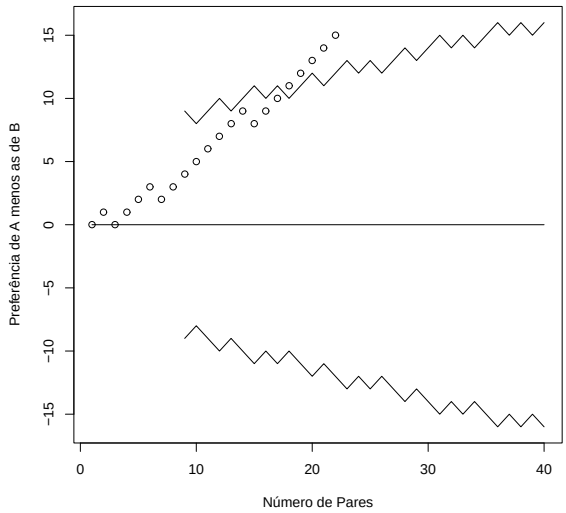

Figura 4.1: Projeto seqüencial de um ensaio clínico comparativo em que a amostragem pára quando a diferença entre n_a e n_b torna-se suficientemente grande (Berry(1987-a), p.120)

A Figura 1 ilustra os limites para decisão (linha contínua). O ensaio pára e H_0 é rejeitada quando o experimento alcança ou o limite inferior ou o limite superior. A característica do plano é que a probabilidade posterior de atingir o limite superior ou o limite inferior quando $\theta = 1/2$ é $0,025$.

Na Tabela 2 comparamos diversas estatísticas para a amostra de n=21 unidades observadas.

Na primeira coluna estão os pares testados. Na segunda está o tratamento preferível em cada par.

Na terceira coluna está a diferença $n_a - n_b$ e na quarta coluna está o nível de significância bilateral igual a 0,05, indicando n=18.

Na 5º coluna calculamos o p-valor bilateral, assumindo a distribuição binomial

com $\theta = 1/2$ e n= 1, 2, ..., 21.

Na 6º coluna determinamos as probabilidades atualizadas de H_0, assumindo como distribuição priori uma Uniforme(0,1) com massa pontual $Pr(\theta) = 1/2$ em $\theta = 1/2$. E finalmente, na última coluna consideramos a probabilidade de B ser melhor, assumindo como distribuição priori uma Uniforme(0,1) e com massa pontual $Pr(\theta) = 0$.

Par	Tratamento com melhor desempenho	$n_a - n_b$	Nível de Significância Bilateral	p-valor Bilateral	Probabilidade Atual de $\theta = 1/2$	Probabilidade Atual de "B"ser melhor
1	A	1	-	1,0000	0,5000	0,2500
2	B	0	-	1,0000	0,6000	0,5000
3	A	1	-	1,0000	0,6000	0,3100
4	A	2	-	0,6300	0,5600	0,1900
5	A	3	-	0,3800	0,4800	0,1100
6	B	2	-	0,6900	0,6200	0,2300
7	A	3	-	0,4500	0,5800	0,1400
8	A	4	-	0,2900	0,5000	0,0900
9	A	5	-	0,1800	0,4100	0,0550
10	A	6	-	0,1100	0,3300	0,0330
11	A	7	-	0,0650	0,2400	0,0190
12	A	8	-	0,0390	0,1700	0,0110
13	A	9	-	0,0220	0,1200	0,0065
14	B	8	-	0,0570	0,2500	0,0180
15	A	9	-	0,0350	0,1800	0,0110
16	A	10	-	0,0210	0,1300	0,0064
17	A	11	-	0,0130	0,0850	0,0038
18	A	12	0,05	0,0075	0,0560	0,0022
19	A	13	-	0,0044	0,0340	0,0013
20	A	14	-	0,0026	0,0220	0,0008
21	A	15	-	0,0015	0,0140	0,0005

Tabela 4.1: Análise de um ensaio seqüencial envolvendo pacientes com leucemia aguda.(página 120 de Berry (1987 - a))

Uma análise bayesiana deste ensaio clínico envolve cálculos de probabilidades posteriores e a parada do processo de amostragem ocorre quando a evidência fornecida pelos dados é conclusiva a favor de uma das hipóteses.

Utilizando como critério de parada probabilidades menores ou iguais a 0,05 para uma das hipóteses, podemos constatar na Tabela 2 que uma decisão é tomada após a 10ª observação a favor do medicamento A, quando a probabilidade de que B seja melhor é 0,033. Decidimos então que a droga A é melhor e atingimos uma decisão antes do procedimento clássico seqüencial que pára somente após a observação do 18º par.

Um dos principais aspectos que devemos considerar neste exemplo é a questão ética envolvida. Um médico que acredita que um tratamento seja mais efetivo que o outro não pode submeter pacientes a tratamentos inferiores. Têm-se a imposição ética de usar os tratamentos superiores tão logo esta superioridade seja revelada.

Dessa forma, é natural que em um ensaio clínico deva-se interromper o experimento assim que alguma diferença significante for detectada.

Na Tabela 3, apresentamos as probabilidades correntes de que o medicamento B seja melhor ($P(\theta < 1/2|\text{dados})$) adotando para θ uma distribuições priori Beta(a,b), para alguns valores de a e b. Estes parâmetros representam opinião favorável ao tratamento A ($a < b$), favorável ao tratamento B ($a > b$), ou indiferente ($a = b$).

Através da Tabela 3, verificamos que o procedimento bayesiano é vantajoso mesmo quando a opinião inicial do médico não é Uniforme(0,1). Em geral, consegue-se

obter resultados conclusivos, com $P(\theta < 1/2|\text{dados}) \leq 0,05$, antes do 14° par e, por isso, encerra-se o ensaio antes do projeto seqüencial clássico, com outras opiniões a priori.

Mesmo com opiniões a priori sendo "desfavoráveis"ao tratamento com o medicamento A, isto é, apresentam valor esperado menor que 1/2, consegue-se concluir o experimento antes do projeto seqüencial clássico, mostrando que a análise bayesiana é realmente mais flexível.

Um ponto relevante deve ser esclarecido sobre análise bayesiana: verificamos que adotando prioris diferentes podemos chegar a resultados conclusivos em instantes diferentes. Como mostra na Tabela 3, se adotamos como priori a densidade da Beta(3,1), obtemos $P(\theta < 1/2|(dados)) \leq 0,05$ no 5º par, e se definimos como priori a densidade Beta(4,2) o resultado conclusivo ocorre no 8º par.

A análise bayesiana não depende de aspectos do planejamento do experimento, mas da opinião a priori do médico a respeito de eficácia dos tratamentos. Os dados atualizam esta informação, confirmando ou não, a opinião inicial do médico.

				Probabilidades	Atualizadas		
Par	Tratamento Preferível	$n_a - n_b$	Beta(3,1)	Beta(4,2)	Beta(2,2)	Beta(1,3)	Beta(2,3)
1	A	1	0,0625	0,1094	0,3125	0,6875	0,5000
2	B	0	0,1875	0,2266	0,5000	0,8125	0,6563
3	A	1	0,1094	0,1445	0,3437	0,6563	0,5000
4	A	2	0,0625	0,0898	0,2266	0,5000	0,3632
5	A	3	0,0352	0,0547	0,1445	0,3632	0,2539
6	B	2	0,0898	0,1132	0,2539	0,5000	0,3769
7	A	3	0,0547	0,0730	0,1718	0,3769	0,2743
8	A	4	**0,0327**	**0,0462**	0,1132	0,2743	0,1921
9	A	5	0,0193	0,0295	0,0730	0,1921	0,1321
10	A	6	0,0113	0,0185	**0,0462**	0,1321	0,0893
11	A	7	0,0065	0,0116	0,0295	0,0893	0,0596
12	A	8	0,0042	0,0072	0,0185	0,0596	**0,0394**
13	A	9	0,0025	0,0044	0,0116	**0,0394**	0,0258
14	B	8	0,0072	0,0109	0,0258	0,0719	0,0492
15	A	9	0,0044	0,0070	0,0168	0,0492	0,0334
16	A	10	0,0027	0,0044	0,0109	0,0034	0,0225
17	A	11	0,0017	0,0028	0,0070	0,0225	0,0151
18	A	12	0,0010	0,0018	0,0044	0,0151	0,0100
19	A	13	0,0006	0,0011	0,0028	0,0100	0,0066
20	A	14	0,0004	0,0007	0,0018	0,0066	0,0043
21	A	15	0,0002	0,0004	0,0011	0,0043	0,0028

Tabela 4.2: Probabilidades atualizadas de que o medicamento B seja melhor para outras prioris

4.0.7 Outras Regras de Parada

Podemos apresentar a aplicação anterior com outras regras de parada para testar a hipótese proposta.

A Tabela 3 fornece a função perda. Para o exemplo do ensaio clínico, faremos

outras duas regras de parada, primeiramente definindo $K_0 = 1$ e $K_1 = 4$, ou seja, a ocorrência de Erro Tipo I é 4 vezes pior que a ocorrência de Erro Tipo II, e depois definindo $K_0 = 1$ e $K_1 = 10$, considerando que ocorrer o Erro Tipo I é 10 vezes pior que ocorrer o Erro Tipo II.

Primeiramente consideremos uma amostra aleatória $X_1, X_2, ..., X_n$ de uma distribuição Bernoulli(θ), e seja $Y = \sum X_i$. Queremos testar $H_0 : \theta = \theta_0$ e $H_1 : \theta = \theta_1$. Para a nossa aplicação, definimos as hipóteses: $H_0 : \theta = 0,5$ e $H_1 : \theta = 0,6$.

Suponhamos que a priori $P(\theta_0) = 1 - P(\theta_1) = \pi$. No exemplo supomos que $\pi = 0,5$, ou seja, não temos evidência maior a favor de $\theta_0 = 0,5$ ou de $\theta_1 = 0,6$. A hipótese nula será rejeitada quando $Y > y_c$, onde y_c é o ponto de corte dado por:

$$y_c = \left(\frac{log(K_1 \frac{\pi}{1-\pi}) - nlog(\frac{1-\theta_1}{1-\theta_0})}{log(\frac{\theta_1}{\theta_0}) - log(\frac{1-\theta_1}{1-\theta_0})} \right)$$

Desta forma, conseguimos fazer limites de decisão para $K_1 = 4$ (linha contínua) e $K_1 = 10$ (linha tracejada), em função do tamanho da amostra n, como mostrado na Figura 2. O valor de Y, em cada novo par, é mostrado pela curva de pontos.

Observamos o processo de amostragem termina antes para $K_1 = 4$, quando o número de pares é 12, já para $K_1 = 10$ o processo de amostragem termina quando o número de pares é 19, tomando como decisão a rejeição da hipótese nula.

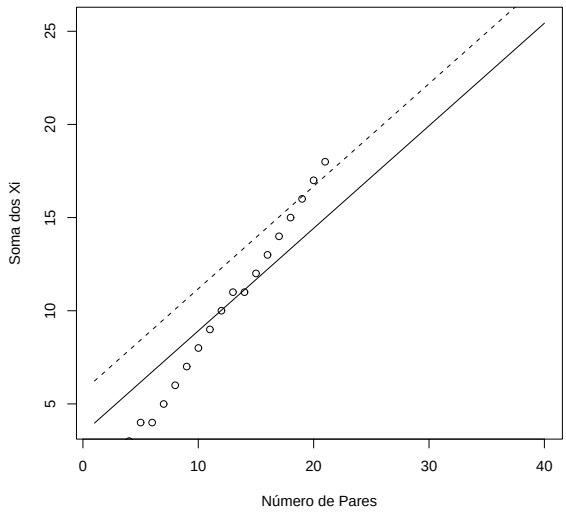

Figura 4.2: Gráfico com outros limites de decisão.

Referências Bibliográficas

[1] Berger, J.O. (1985) Statistical Decision Theory and Bayesian Analysis (2nd ed.).

[2] Berry,Donald A. (1987) Interim Analysis in Clinical Trials: The Role of the Likelihood Principle. The American Statistician, Vol.41, N°2, p.117-122.

[3] Berry,Donald A. (1996) Statistics: A Bayesian Perspective.

[4] Bickel,Peter J. and Doksum,Kjell A. Mathematical Statistics: Basic ideas and selected topics Vol.1 (2nd ed.).

[5] DeGroot, Morris H. (1989)Probability and Statistics Second Edition.

[6] Inoue,Lurdes Y.T. , Berry,Donald A. , and Parmigiani,Giovanni (2005)Relationship Between Bayesian and Frequentist Sample Size Determination. The American Statistician, Vol.59, N°1, p.79-87.

[7] Inoue,Lurdes Y.T. (1995)Desenvolvimento e Implicações do Princípio da Verossimilhança.

[8] Mood,A.M. (1963) Introduction to the Theory of Statistics.

[9] O'Hagan, A. (1994) Kendall's Advanced Theory of Statistics Volume 2B Bayesian Inference.