CE-210
Inferência Estatística II

Ricardo S. Ehlers

Versão Revisada em 19 de agosto de 2003

Sumário

Capítulo 1

Estimação por Intervalos

A principal restrição da estimação pontual é que quando estimamos um parâmetro através de um único valor numérico toda a informação presente nos dados é resumida através deste número. É importante encontrar também um intervalo de valores plausíveis para o parâmetro.

A idéia é construir um intervalo em torno da estimativa pontual de modo que ele tenha uma probabilidade conhecida de conter o verdadeiro valor do parâmetro. Tipicamente as distribuições amostrais de estimadores dos parâmetros desconhecidos serão utilizadas. Antes de descrever o procedimento geral veremos um exemplo simples de construção do intervalo de confiança.

Exemplo 1.1 : Seja $X_1, \dots, X_n$ uma amostra aleatória da distribuição $N(\theta, \sigma^2)$, com σ^2 conhecido. Para fazer inferências sobre θ nos baseamos na média amostral $\overline{\boldsymbol{X}}$ e sabemos que

$$U = \frac{\sqrt{n}(\overline{\boldsymbol{X}} - \theta)}{\sigma} \sim N(0, 1).$$

Note que a estatística U é uma função da amostra e também de θ, o parâmetro de interesse, mas sua distribuição de probabilidades não depende de θ. Usando uma tabela da distribuição normal padronizada podemos obter o valor do percentil $z_{\alpha/2}$ tal que

$$P(-z_{\alpha/2} \leq U \leq z_{\alpha/2}) = 1 - \alpha$$

e assim, após isolar θ, obtemos que

$$P\left(\overline{\boldsymbol{X}} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}} \leq \theta \leq \overline{\boldsymbol{X}} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right) = 1 - \alpha.$$

Esta última igualdade pode dar margem a interpretações errôneas, o que aliás acontece com bastante frequência. O parâmetro θ é desconhecido mas fixo e portanto não é passível de descrição probabilística, ou seja não se trata de um intervalo de probabilidade para θ. Na verdade os limites do intervalo é que são variáveis aleatórias e após a amostra ser observada dizemos que

$$\overline{\boldsymbol{x}} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}} \leq \theta \leq \overline{\boldsymbol{x}} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$$

com confiança $100(1-\alpha)\%$.

Vale notar também que, para um dado valor de $1-\alpha$, é possível construir muitos intervalos de confiança diferentes para θ. Na verdade, quaisquer constantes c_1 e c_2 tais que $P(c_1 \leq U \leq c_2) = 1-\alpha$ podem ser usadas para construir um intervalo com limites

$$\overline{\boldsymbol{x}} - c_2\frac{\sigma}{\sqrt{n}} \quad \text{e} \quad \overline{\boldsymbol{x}} - c_1\frac{\sigma}{\sqrt{n}}.$$

No entanto, pode-se mostrar que dentre todos os intervalos de confiança com esta característica, aquele definido acima que é simétrico em torno do média amostral $\overline{\boldsymbol{x}}$ é o de menor comprimento.

1.1 Procedimento Geral

O procedimento geral para construção de intervalos de confiança consiste nos seguintes passos,

1. Obter uma estatística que depende de θ, $U = G(X,\theta)$, mas cuja distribuição não depende de θ.

2. Usando a distribuição de U, encontrar as constantes a e b tais que $P(a \leq U \leq b) \geq 1-\alpha$.

3. Definir $\{\theta : a \leq G(x,\theta) \leq b\}$ como o intervalo (ou região) de confiança 100(1-α)% para θ.

A exigência de que a probabilidade no item 2 acima possa ser maior do que o nível de confiança é essencialmente técnica pois queremos que o intervalo seja o menor possível, o que em geral implica em usar uma igualdade. A desigualdade será útil principalmente no caso de distribuições discretas onde nem sempre é possível satisfazer a igualdade.

Note que a variável aleatória U, comumente denominada quantidade pivotal ou *pivot*, é fundamental para o funcionamento do método. Idealmente ela deve depender da amostra através de estatísticas suficientes minimais e ter distribuição conhecida.

É importante notar também que este intervalo não pode ser interpretado como um intervalo de probabilidade para θ já que a aleatoriedade presente é devida a amostra X. Ou seja, o procedimento leva a construção de um intervalo probabilístico para U e não para θ.

Tecnicamente, dizemos que $100(1-\alpha)\%$ de todos os intervalos de confiança que construirmos conterão o verdadeiro valor do parâmetro (dado que todas as suposições envolvidas estejam corretas). Por exemplo se $1-\alpha = 0,95$ então, em média, somente 5 a cada 100 intervalos não conterão θ. A probabilidade $1-\alpha$ é denominada nível de confiança e sua escolha depende da precisão com que queremos estimar o parâmetro, sendo que 0,90, 0,95 e 0,99 são os valores mais comuns na prática.

1.2 Estimação no Modelo Normal

Nesta seção serão discutidos os casos em que os dados provém de uma distribuição normal. Inicialmente veremos o caso em que temos uma única amostra de uma distribuição normal e queremos estimar sua média e sua variância. Na Seção 1.2.2 estudaremos o caso de duas amostras tomadas de distribuições normais independentes.

1.2.1 O caso de uma amostra

No exemplo 1.1, se σ^2 for desconhecido não podemos usar a mesma quantidade pivotal já que ela depende de σ. Ou seja, precisamos obter uma outra quantidade pivotal que depende apenas de $\boldsymbol{X}$ e de θ e com uma distribuição que seja conhecida e não dependa de nenhum parâmetro desconhecido. No modelo normal isto será possível usando os resultados a seguir.

Teorema 1.1 *Seja $X_1, \dots, X_n$ uma amostra aleatória da distribuição $N(\theta, \sigma^2)$ e sejam $\overline{X}$ e S^2 a média e a variância amostrais. Então, condicionado em θ e σ^2, $\overline{X}$ e S^2 são independentes com distribuições amostrais*

$$\frac{\sqrt{n}(\overline{\boldsymbol{X}}-\theta)}{\sigma} \sim N(0,1) \quad e \quad \frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{n-1}.$$

Lema 1.1 *Se $U \sim N(0,1)$ e $W \sim \chi^2_\nu$ e se U e W são independentes então*

$$\frac{U}{\sqrt{\frac{W}{\nu}}} \sim t_\nu(0,1).$$

Prova. A prova é deixada como exercício.

Corolário 1.1 *Seja $X_1, \dots, X_n$ uma amostra aleatória da distribuição $N(\theta, \sigma^2)$ e sejam $\overline{X}$ e S^2 a média e a variância amostrais. Então, condicionado em θ e σ^2, $\overline{X}$ tem distribuição amostral*

$$\frac{\sqrt{n}(\overline{X} - \theta)}{S} \sim t_{n-1}(0, 1)$$

Prova. Aplicação direta do lema acima com $U = \sqrt{n}(\overline{X} - \theta)/\sigma$, $W = (n-1)S^2/\sigma^2$ e $\nu = n - 1$.

Estes resultados nos permitem definir quantidades pivotais para construção de intervalos de confiança para θ e σ^2. No caso da média θ, o valor desconhecido de σ é substituido pelo seu estimador S levando a uma quantidade pivotal com distribuição t com $n - 1$ graus de liberdade. Assim, podemos obter o percentil $t_{\alpha/2,n-1}$ tal que

$$P\left(-t_{\alpha/2,n-1} \leq \frac{\sqrt{n}(\overline{X} - \theta)}{S} \leq t_{\alpha/2,n-1}\right) = 1 - \alpha$$

e, após isolar θ, obtemos que

$$P\left(\overline{\boldsymbol{X}} - t_{\alpha/2,n-1}\frac{S}{\sqrt{n}} \leq \theta \leq \overline{\boldsymbol{X}} + t_{\alpha/2,n-1}\frac{S}{\sqrt{n}}\right) = 1 - \alpha.$$

Note que, mesmo se S pudesse estimar σ sem erro, esta substituição implica em um aumento da amplitude do intervalo de confiança pois $t_{\alpha,n} > z_\alpha$ para n pequeno.

Finalmente, após observar a amostra substituimos as estimativas e dizemos que

$$\overline{\boldsymbol{x}} - t_{\alpha/2,n-1}\frac{s}{\sqrt{n}} \leq \theta \leq \overline{\boldsymbol{x}} + t_{\alpha/2,n-1}\frac{s}{\sqrt{n}}$$

com confiança $100(1 - \alpha)\%$.

Para obter estimativas da variância populacional σ^2 usamos uma quantidade pivotal com distribuição qui-quadrado com $n - 1$ graus de liberdade. Devemos então obter os percentis $\underline{\chi}^2_{\alpha/2,n-1}$ e $\overline{\chi}^2_{\alpha/2,n-1}$ desta distribuição tais que

$$P\left(\underline{\chi}^2_{\alpha/2,n-1} \leq \frac{(n-1)S^2}{\sigma^2} \leq \overline{\chi}^2_{\alpha/2,n-1}\right) = 1 - \alpha$$

e após observar a amostra o intervalo de confiança de $100(1-\alpha)\%$ para σ^2 é dado por

$$\left(\frac{(n-1)s^2}{\overline{\chi}^2_{\alpha/2,n-1}}, \frac{(n-1)s^2}{\underline{\chi}^2_{\alpha/2,n-1}}\right).$$

1.2.2 O caso de duas amostras

Nesta seção vamos assumir que $X_{11}, \ldots, X_{1n_1}$ e $X_{21}, \ldots, X_{2n_2}$ são amostras aleatórias das distribuições $N(\theta_1, \sigma_1^2)$ e $N(\theta_2, \sigma_2^2)$ respectivamente e que as amostras são independentes.

Podemos comparar as médias populacionais estimando a diferença $\beta = \theta_1 - \theta_2$. A estimação é baseada na diferença entre médias amostrais, i.e. $\overline{X}_1 - \overline{X}_2$ que é o estimador de máxima verossimilhança de β. Se as variâncias populacionais forem conhecidas então a distribuição amostral é dada por

$$\overline{X}_1 - \overline{X}_2 \sim N(\overline{x}_1 - \overline{x}_2, \sigma^2)$$

onde

$$\sigma^2 = \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}.$$

Assim, o intervalo de confiança de $100(1-\alpha)\%$ para a diferença entre médias é dado por

$$\left(\overline{x}_1 - \overline{x}_2 - z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}\ ;\ \overline{x}_1 - \overline{x}_2 + z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} \right).$$

No caso de variâncias populacionais desconhecidas porém iguais, i.e. $\sigma_1^2 = \sigma_2^2 = \sigma^2$ podemos combinar os duas variâncias amostrais para formar uma estimativa combinada da variância. Atribuímos mais peso às amostras maiores e esta variância combinada é dada por

$$S_p^2 = \frac{(n_1-1)S_1^2 + (n_2-1)S_2^2}{n_1+n_2-2},$$

isto é, a média ponderada das variâncias amostrais com pesos dados por $n_1 - 1$ e $n_2 - 1$. Agora podemos calcular o erro padrão das diferenças nas médias como

$$S_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}.$$

Do exposto acima, um intervalo de confiança para a diferença entre médias $\theta_1 - \theta_2$ assumindo desvios padrão iguais pode ser construído usando-se a quantidade pivotal

$$\frac{\hat{\beta} - \beta}{S_p\sqrt{\dfrac{1}{n_1} + \dfrac{1}{n_2}}} \sim t_\nu(0,1)$$

onde $\nu = n_1 + n_2 - 2$ e $\hat{\beta} = \overline{X}_1 - \overline{X}_2$. Assim, o intervalo de confiança de $100(1-\alpha)\%$ para a diferença fica,

$$\left(\overline{x}_1 - \overline{x}_2 - t_{\alpha/2,\nu} s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} \; ; \; \overline{x}_1 - \overline{x}_2 + t_{\alpha/2,\nu} s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} \right)$$

Analogamente ao caso de uma amostra, o intervalo de confiança para σ^2 é construído usando-se a quantidade pivotal

$$\frac{\nu S_p^2}{\sigma^2} \sim \chi^2_\nu.$$

Então devemos obter os quantis $\alpha/2$ inferior e superior desta distribuição qui-quadrado e o intervalo de confiança de $100(1-\alpha)\%$ para a variância populacional fica

$$\left(\frac{\nu s_p^2}{\overline{\chi}^2_{\alpha/2,\nu}} \; ; \; \frac{\nu s_p^2}{\underline{\chi}^2_{\alpha/2,\nu}} \right)$$

1.2.3 Variâncias desiguais

Até agora assumimos que as variâncias populacionais desconhecidas eram iguais (ou pelo menos aproximadamente iguais). A violação desta suposição leva a problemas teóricos e práticos uma vez que não é trivial encontrar uma quantidade pivotal para β com distribuição conhecida. Na verdade, se existem grandes diferenças de variabilidade entre as duas populações pode ser mais apropriado analisar conjuntamente as consequências das diferenças entre as médias e as variâncias. Assim, caso o pesquisador tenha interesse no parâmetro β deve levar em conta os problemas de ordem teóricas introduzidos por uma diferença substancial entre σ_1^2 e σ_2^2.

A literatura estatística apresenta vários métodos para resolver este problema mas nenhum deles é completamente satisfatório. Um procedimento possível consiste em utilizar a estatística

$$t = \frac{\hat{\beta} - \beta}{\sqrt{\dfrac{S_1^2}{n_1} + \dfrac{S_2^2}{n_2}}}.$$

No entanto, a distribuição exata de t depende da razão σ_1^2/σ_2^2, que é desconhecida. Se n_1 e n_2 forem grandes t tem distribuição aproximadamente normal padrão, mas quando eles são ambos pequenos uma solução simples é utilizar uma distribuição t de Student com $k-1$ graus de liberdade onde $k = \min(n_1, n_2)$. Outra solução aproximada (método aproximado de Aspin-Welch) consiste em utilizar a estatística acima com distribuição t de Student e número de graus de liberdade dado por

$$\nu = \frac{(w_1 + w_2)^2}{\dfrac{w_1^2}{n_1 - 1} + \dfrac{w_2^2}{n_2 - 1}}$$

onde

$$w_1 = \frac{s_1^2}{n_1} \quad \text{e} \quad w_2 = \frac{s_2^2}{n_2}.$$

1.2.4 Comparação de variâncias

Outra situação de interesse é a comparação das duas variâncias populacionais. Neste caso, faz mais sentido utilizar a razão de variâncias ao invés da diferença já que elas medem a escala de uma distribuição e são sempre positivas. Para obter a distribuição amostral apropriada usaremos o teorema a seguir.

Teorema 1.2 *Sejam as variáveis aleatórias U e W independentes com distribuições qui-quadrado com ν_1 e ν_2 graus de liberdade respectivamente. Então a variável aleatória dada por*

$$X = \frac{U/\nu_1}{W/\nu_2}$$

tem distribuição F com ν_1 e ν_2 graus de liberdade.

Usaremos a notação $X \sim F(\nu_1, \nu_2)$ e dos teoremas 1.1 e 1.2 não é difícil mostrar que

$$\frac{S_1^2}{S_2^2}\frac{\sigma_2^2}{\sigma_1^2} \sim F(n_1 - 1, n_2 - 1).$$

Embora sua função de distribuição não possa ser obtida analiticamente os valores estão tabelados em muitos livros de estatística e também podem ser obtidos na maioria dos pacotes computacionais. Os percentis podem então ser utilizados na construção de intervalos de confiança para a razão de variâncias.

Uma propriedade bastante útil para calcular probabilidade com a distribuição F vem do fato de que se $X \sim F(\nu_2, \nu_1)$ então $X^{-1} \sim F(\nu_1, \nu_2)$ por simples inversão na razão de distribuições qui-quadrado independentes. Assim, denotando os quantis α e $1 - \alpha$ da distribuição $F(\nu_1, \nu_2)$ por $\underline{F}_\alpha(\nu_1, \nu_2)$ e $\overline{F}_\alpha(\nu_1, \nu_2)$ respectivamente segue que

$$\underline{F}_\alpha(\nu_1, \nu_2) = \frac{1}{\overline{F}_\alpha(\nu_2, \nu_1)}.$$

Note que é usual que os livros forneçam tabelas com os percentis superiores da distribuição F para várias combinações de valores de ν_1 e ν_2 devido à propriedade acima. Por exemplo, se temos os valores tabelados dos quantis 0,95 podemos obter também um quantil 0,05. Basta procurar o quantil 0,95 inverterndo os graus de liberdade.

1.3 Intervalos de confiança para uma proporção

Seja $X_1, \dots, X_n$ uma amostra aleatória de uma distribuição de Bernoulli com parâmetro θ. Assim,

$$Y = \frac{1}{n}\sum_{i=1}^{n} X_i$$

é a proporção amostral de sucessos e será o nosso estimador pontual da verdadeira probabilidade de sucesso θ. Vamos considerar agora a construção de um intervalo de confiança para θ.

Pelo teorema central do limite, para n grande e θ não muito próximo de 0 ou 1, a distribuição de Y será aproximadamente normal com média θ e um desvio padrão dado por

$$\sqrt{\frac{\theta(1-\theta)}{n}}.$$

já que $E(X_i) = \theta$ e $V(X_i) = \theta(1-\theta)$. Ou seja, a quantidade pivotal será dada por

$$\frac{Y-\theta}{\sqrt{\dfrac{\theta(1-\theta)}{n}}} \sim N(0,1).$$

Assim, após observar a amostra o intervalo de confiança de 100(1-α)% para θ fica

$$\left(y - z_{\alpha/2}\sqrt{\frac{\theta(1-\theta)}{n}}, y + z_{\alpha/2}\sqrt{\frac{\theta(1-\theta)}{n}}\right).$$

Note que os limites do intervalo dependem do valor desconhecido de θ e aqui duas abordagens são possíveis. Podemos usar o fato de que o valor máximo de $\theta(1-\theta)$ é atingido para $\theta = 1/2$, logo $\theta(1-\theta) \leq 1/4$, ou equivalentemente $\sqrt{\theta(1-\theta)/n} \leq 1/\sqrt{4n}$. Neste caso, um intervalo de confiança *conservativo* é dado por

$$\left(y - z_{\alpha/2}\sqrt{\frac{1}{4n}}, y + z_{\alpha/2}\sqrt{\frac{1}{4n}}\right).$$

No entanto, se o verdadeiro valor de θ estiver afastado do seu valor máximo e estiver próximo de 0 ou de 1 então este intervalo tem amplitude desnecessariamente grande porque substituimos $\theta(1-\theta)$ pelo seu valor máximo. Um enfoque mais otimista consiste em substituir θ pela sua estimativa de máxima verossimilhança, i.e. a proporção amostral de sucessos y e utilizar o intervalo

$$\left(y - z_{\alpha/2}\sqrt{\frac{y(1-y}{n}}, y + z_{\alpha/2}\sqrt{\frac{y(1-y)}{n}}\right).$$

Note que, para n e $1-\alpha$ fixos a amplitude do intervalo conservativo será a mesma para todas as possíveis amostras de tamanho n. Por outro lado, usando-se esta última expressão o intervalo terá amplitude $2z_{\alpha/2}\sqrt{y(1-y)/n}$ que varia de amostra para amostra.

1.4 Intervalos de Confiança Assintóticos

Utilizando os conceitos do método da quantidade pivotal e a propriedade de normalidade assintótica dos estimadores de máxima verossimilhança podemos construir intervalos de confiança para θ. Antes porém precisamos da definição da medida de informação de Fisher.

Definição 1.1 *Considere uma única observação X com função de (densidade) de probabilidade $p(x|\theta)$. A medida de informação esperada de Fisher de θ através de X é definida como*

$$I(\theta) = E\left[-\frac{\partial^2 \log p(x|\theta)}{\partial\theta^2}\right].$$

No caso de um vetor paramétrico $\boldsymbol{\theta} = (\theta_1, \dots, \theta_k)$ define-se a matriz de informação esperada de Fisher de $\boldsymbol{\theta}$ através de X como

$$\boldsymbol{I}(\boldsymbol{\theta}) = E\left[-\frac{\partial^2 \log p(x|\boldsymbol{\theta})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}\right].$$

Note que o conceito de informação aqui está sendo associado a uma espécie de curvatura média da função de verossimilhança no sentido de que quanto maior a curvatura mais precisa é a informação contida na verossimilhança, ou equivalentemente maior o valor de $I(\theta)$. Em geral espera-se que a curvatura seja negativa e por isso seu valor é tomado com sinal trocado. Note também que a esperança matemática é tomada em relação à distribuição amostral $p(x|\theta)$.

Podemos considerar então $I(\theta)$ uma medida de informação global enquanto que uma medida de informação local é obtida quando não se toma o valor esperado na definição acima. A medida de informação observada de Fisher $J(\theta)$ fica então definida como

$$J(\theta) = -\frac{\partial^2 \log p(x|\theta)}{\partial\theta^2}.$$

Lema 1.2 *Seja $\boldsymbol{X} = (X_1, \dots, X_n)$ uma coleção de variáveis aleatórias independentes com distribuições $p_i(x|\theta)$, $i = 1, \dots, n$ e sejam $I(\theta)$ e $I_i(\theta)$ as medidas de informação de θ obtidas através de $\boldsymbol{X}$ e de X_i, respectivamente. Então,*

$$I(\theta) = \sum_{i=1}^{n} I_i(\theta).$$

Prova. A prova é simples e será deixada como exercício.

O lema nos diz então que a informação total contida em observações independentes é igual a soma das informações individuais. Um caso particular importante é quando as observações são também identicamente distribuidas já que neste caso $I_i(\theta)$ é constante e assim a informação total é simplesmente $nI(\theta)$.

Outra estatística importante no estudo da função de verossimilhança e que será útil na construção de intervalos de confiança assintóticos é a função escore.

Definição 1.2 *A função escore de X denotada por $U(X;\theta)$ é dada por*

$$U(X;\theta) = \frac{\partial \log p(X|\theta)}{\partial \theta}.$$

No caso de um vetor paramétrico $\boldsymbol{\theta} = (\theta_1, \dots, \theta_k)$ a função escore será um vetor $\boldsymbol{U}(X;\boldsymbol{\theta})$ com componentes $U_i(X;\boldsymbol{\theta}) = \partial \log p(X|\boldsymbol{\theta})/\partial \boldsymbol{\theta}_i$.

Além disso, pode-se mostrar que o valor esperado da função escore é zero e sua variância é dada por $I(\theta)$.

Vimos em estimação pontual que, para grandes amostras, o estimador de máxima verossimilhança $\hat{\theta}_n$ para um parâmetro θ tem distribuição aproximadamente normal com média θ sob condições de regularidade gerais. Assim, mesmo que $\hat{\theta}_n$ seja viesado para n fixo ele será assintoticamente não viesado. A variância assintótica é dada por $1/nI(\theta)$. Ou seja, para n grande $\hat{\theta}_n$ tem distribuição aproximadamente $N(\theta, (nI(\theta))^{-1})$ e podemos construir intervalos de confiança aproximados para θ. Neste caso, $(\hat{\theta}_n - \theta)\sqrt{nI(\theta)}$ pode ser tratado como uma quantidade pivotal aproximada e se for possível isolar θ na desigualdade

$$-z_{\alpha/2} < (\hat{\theta}_n - \theta)\sqrt{nI(\theta)} < z_{\alpha/2}$$

teremos um intervalo de confiança com coeficiente de confiança aproximado igual a $1 - \alpha$.

Exemplo 1.2 : Seja $X_1, \dots, X_n$ uma amostra aleatória da distribuição exponencial com parâmetro θ. A função de densidade conjunta é dada por

$$p(\boldsymbol{x}|\theta) = \theta^n e^{-\theta t}, \quad \theta > 0, \quad \text{onde } t = \sum_{i=1}^{n} x_i.$$

Tomando-se o logaritmo obtém-se

$$\log p(\boldsymbol{x}|\theta) = n \log(\theta) - \theta t$$

de modo que as derivadas de primeira e segunda ordem são

$$\frac{\partial \log p(|\theta)}{\partial \theta} = \frac{n}{\theta} - t \quad \text{e} \quad \frac{\partial^2 \log p(|\theta)}{\partial \theta^2} = -\frac{n}{\theta^2}$$

e a informação esperada de Fisher baseada na amostra é $nI(\theta) = n/\theta^2$. Sabemos também que o estimador de máxima verossimilhança de θ é $1/\overline{\boldsymbol{X}}$ e portanto, para n grande, $1/\overline{\boldsymbol{X}}$ tem distribuição aproximadamente normal com média θ e variância θ^2/n. Assim, o intervalo de confiança aproximado é obtido fazendo-se

$$P\left(-z_{\alpha/2} < \frac{1/\overline{\boldsymbol{X}} - \theta}{\sqrt{\theta^2/n}} < z_{\alpha/2}\right) \approx 1 - \alpha.$$

Isolando θ obtemos que

$$P\left(\frac{1/\overline{\boldsymbol{X}}}{1 + z_{\alpha/2}} < \theta < \frac{1/\overline{\boldsymbol{X}}}{1 - z_{\alpha/2}}\right) \approx 1 - \alpha.$$

Um fato importante é que, em geral, na distribuição assintótica normal do estimador de máxima verossimilhança a sua variância $(nI(\theta))^{-1}$ pode ser substituida pelo seu estimador $(nI(\hat{\theta}))^{-1}$ sem afetar muito a acurácia da aproximação. Este fato, que não será provado aqui, simplifica bastante a conversão das desigualdades para obtenção de intervalos de confiança aproximados. Assim,

$$P\left(-z_{\alpha/2} < (\hat{\theta} - \theta)\sqrt{nI(\hat{\theta})} < z_{\alpha/2}\right) \approx 1 - \alpha$$

é facilmente convertido para

$$P\left(\hat{\theta} - z_{\alpha/2}\sqrt{nI(\hat{\theta})} < \theta < \hat{\theta} + z_{\alpha/2}\sqrt{nI(\hat{\theta})}\right) \approx 1 - \alpha.$$

Note que este resultado foi utilizado na Seção 1.3 para construir um intervalo de confiança aproximado para uma proporção. Naquele caso, $\theta(1-\theta)/n$ era a variância de $\overline{\boldsymbol{X}}$ que foi substituida pelo seu estimador de máxima verossimilhança.

Em algumas situações não se tem uma forma explícita para o estimador de máxima verossimilhança e neste caso a função escore será particularmente útil. Lembrando que a função escore de X tem média zero e variância igual a $I(\theta)$ então temos pelo teorema central do limite que $\sum_{i=1}^n U(X_i;\theta)$ converge em distribuição para uma $N(0, nI(\theta))$. Podemos usar este resultado para fazer inferência aproximada sobre θ e assim o intervalo de confiança aproximado de $100(1-\alpha)\%$ é obtido fazendo-se

$$P\left(\left|\frac{1}{\sqrt{nI(\theta)}}\sum_{i=1}^n U(X_i;\theta)\right| < z_{\alpha/2}\right) \approx 1 - \alpha.$$

Finalmente, vale ressaltar que todos os resultados desta seção podem ser extendidos para o caso de um vetor paramétrico $\boldsymbol{\theta} = (\theta_1, \dots, \theta_k)$. Neste caso, a distribuição assintótica do estimador de máxima verossimilhança será normal multivariada com vetor de médias $\boldsymbol{\theta}$ e matriz de variância-covariância igual a $\boldsymbol{I}^{-1}(\boldsymbol{\theta})$ onde $\boldsymbol{I}(\boldsymbol{\theta})$ é a matriz de informação de Fisher.

1.5 Intervalos Baseados na Função Deviance

1.6 Problemas

1. Os pulsos em repouso de 920 pessoas sadias foram tomados, e uma média de 72,9 batidas por minuto (bpm) e um desvio padrão de 11,0 bpm foram obtidos. Construa um intervalo de confiança de 95% para a pulsação média em repouso de pessoas sadias com base nesses dados.

2. Tendo sido medido o eixo maior de 9 grãos de quartzo de um corpo arenoso em uma lâmina de arenito, obteve-se um comprimento amostral médio de 1,5mm e um desvio padrão de 0,3mm. Deseja-se construir um intervalo de confiança para o comprimento médio dos grãos de quartzo do corpo arenoso.

3. O tempo médio, por operário, para executar uma tarefa, tem sido de 100 minutos com desvio padrão de 15 minutos. Foi introduzida uma modificação para reduzir este tempo e após alguns meses foi selecionada uma amostra de 16 operários medindo-se o tempo de execução de cada um. Obteve-se um tempo médio amostral de 90 minutos e um desvio padrão de 16 minutos.

   (a) Estime o novo tempo médio de execução por um intervalo com 95% de confiança.

   (b) Inteprete o I.C. obtido no item anterior. Você diria que a modificação surtiu efeito? (Justifique).

   (c) Estime a nova variância populacional por um intervalo com 98% de confiança.

4. Os QIs de 181 meninos com idades entre 6-7 anos de Curitiba foram medidos. O QI médio foi 108,08, e o desvio padrão foi 14,38.

   (a) Calcule um intervalo de confiança de 95% para o QI médio populacional dos meninos entre 6-7 anos de idade em Curitiba usando estes dados.

   (b) Interprete o intervalo de confiança com palavras.

   (c) Foi necessário assumir que os QIs têm distribuição normal neste caso? Por quê?

5. Para decidir se uma moeda é balanceada (honesta) ela é lançada 40 vezes e cara aparece 13 vezes. Construa um intervalo de 95% de confiança para a verdadeira proporção de caras p. O que você conclui?

6. Numa pesquisa eleitoral, 57 dentre 150 entrevistados afirmaram que votariam no candidato X. Com uma confiança de 90%, o que você pode dizer acerca da proporção real de votos aquele candidato terá?

7. Dentre 100 peixes capturados num certo lago, 18 não estavam apropriados para consumo devido aos níveis de poluição do ambiente. Construa um intervalo de confiança de 99% para a verdadeira proporção de peixes contaminados.

8. Uma indústria compra componentes eletrônicos dos fornecedores A e B, mas o fornecedor A garante que o tempo médio de vida (em horas) do seu produto supera o da marca B em 300 horas. Para testar esta afirmação foram selecionadas duas amostras de 5 e 4 componentes, das marcas A e B respectivamente. As médias amostrais foram 1492 e 1182 e as variâncias amostrais foram 770 e 3892.

   (a) Compare as variâncias dos tempos de vida através de um intervalo de confiança de 98%.

   (b) Construa um intervalo de confiança de 95% para a diferença entre os tempos médios de vida.

   (c) Este intervalo dá alguma indicação sobre a afirmação do fornecedor A? Explique.

9. Seja $X_1, \ldots, X_n$ uma amostra aleatória da distribuição normal com média μ desconhecida e variância σ^2 conhecida. Qual deve ser o tamanho da amostra tal que exista um intervalo de confiança para μ com coeficiente de confiança 0,95 e comprimento menor do que $0{,}01\sigma$?

10. Seja $X_1, \ldots, X_n$ uma amostra aleatória da distribuição exponencial com média θ desconhecida. Descreva um método para construir um intervalo de confiança de $100(1-\alpha)\%$ para θ. (Sugestão: Determine as constantes c_1 e c_2 tais que $P(c_1 < (1/\theta)\sum_{i=1}^{n} X_i < c_2) = 1 - \alpha$).

11. Seja $X_1, \ldots, X_n$ uma amostra aleatória da distribuição $Beta(\theta, 1)$. Obtenha o intervalo de confiança aproximado de $100(1-\alpha)\%$ baseando-se na distribuição assintótica da função escore.

12. Se uma variável aleatória X tem distribuição de Poisson com média θ obtenha a informação esperada de Fisher $I(\theta)$ através de X.

13. Suponha que uma variável aleatória X tem distribuição normal com média zero e desvio-padrão desconhecido σ. Obtenha a informação esperada de Fisher $I(\sigma)$ através de X. Suponha agora que a variância seja o parâmetro de interesse e obtenha a informação de Fisher de σ^2 através de X.

14. Seja $X_1, \ldots, X_n$ uma amostra aleatória da distribuição $N(0, \sigma^2)$. Construa um intervalo de confiança aproximado para o desvio-padrão σ baseado no seu estimador de máxima verossimilhança.

Capítulo 2

Testes de Hipóteses

2.1 Introdução e notação

Em geral, intervalos de confiança são a forma mais informativa de apresentar os achados principais de um estudo. Contudo, algumas vezes existe um particular interesse em verificar determinadas afirmações ou conjecturas. Por exemplo, podemos estar interessados em determinar se uma moeda é honesta, se certas quantidades são independentes, ou se populações distintas são similares do ponto de vista probabilístico. Cada uma destas afirmações constitui uma hipótese que pode ser associada a um modelo, i.e. pode ser parametrizada.

O material deste capítulo é fortemente baseado em DeGroot (1989), Gamerman e Migon (1993) e Migon e Gamerman (1999). A teoria clássica de testes de hipóteses é apresentada a um nível mais formal em Lehman (1986).

Chamamos de *hipótese estatística* qualquer afirmação que se faça sobre um parâmetro populacional desconhecido. A idéia básica é que a partir de uma amostra da população iremos estabelecer uma *regra de decisão* segundo a qual rejeitaremos ou aceitaremos a hipótese proposta. Esta regra de decisão é chamada de *teste*. Normalmente existe uma hipótese que é mais importante para o pesquisador que será denotada por H_0 e chamada *hipótese nula*. Qualquer outra hipótese diferente de H_0 será chamada de *hipótese alternativa* e denotada por H_1. Veremos mais adiante que intervalos de confiança e testes de hipóteses estão intimamente relacionados.

Exemplo 2.1 : Um professor aplica um teste do tipo certo-errado com 10 questões. Queremos testar a hipótese de que o aluno está advinhando.

Denotando por p a probabilidade do aluno acertar cada questão a hipótese estatística de interesse pode ser formulada como $H_0 : p = 1/2$. Neste caso, a hipótese alternativa mais adequada é $H_1 : p > 1/2$ indicando que o aluno tem algum conhecimento sobre o assunto. Temos então 10 repetições do experimento com p constante, portanto X="número de acertos" tem distribuição binomial com parâmetros $n = 10$ e p desconhecido. Suponha que adotamos a seguinte regra de decisão: o aluno não está advinhando se acertar 8 ou mais questões. Isto equivale a rejeitar H_0 se $X \geq 8$ (*região de rejeição* ou *região crítica*) e aceitar H_0 se $X < 8$ (*região de aceitação*).

No entanto, é possível que um aluno acerte 8 ou mais questões e esteja advinhando, isto é podemos rejeitar H_0 quando ela é verdadeira. A probabilidade de que isto ocorra é

$$P(X \geq 8 \mid p = 1/2) = \binom{10}{8} 0,5^{10} + \binom{10}{9} 0,5^{10} + \binom{10}{10} 0,5^{10} = \frac{7}{128} \approx 0,054.$$

Esta probabilidade é chamada *nível de significância* e será denotada por α.

Exemplo 2.2 : Um fornecedor garante que 90% de sua produção não apresenta defeito. Para testar esta afirmação selecionamos ao acaso 10 itens de um lote e contamos o número de defeituosos. Decidimos não comprar o lote se o número observado de não defeituosos for muito pequeno (mas quão pequeno?).

Definindo X="número de não defeituosos na amostra de 10 itens" temos então uma distribuição binomial com parâmetros $n = 10$ e p desconhecido, e queremos testar $H_0 : p = 0,90$. Aqui p é a proporção de itens não defeituosos no lote e portanto a hipótese alternativa deve ser $H_1 : p < 0,90$. Suponha que decidimos manter $\alpha < 0,025$ e a partir deste valor vamos estabelecer a nossa regra de decisão. Para isto vamos calcular α para diferentes regiões críticas, assim

$$\begin{aligned} P(X \leq 5 \mid p = 0,90) &= 0,001 \\ P(X \leq 6 \mid p = 0,90) &= 0,012 \\ P(X \leq 7 \mid p = 0,90) &= 0,069. \end{aligned}$$

Portanto, para que o nível de significância máximo seja 0,025 devemos usar a região crítica $X \leq 6$. Isto é, vamos rejeitar o lote se o número de itens defeituosos na amostra for maior do que 6.

Nestes dois exemplos os testes são chamados de *unilaterais* porque somente valores de um lado do espaço amostral foram utilizados para construir a região crítica. Podemos ter também testes *bilaterais* aonde os dois extremos do espaço amostral são usados como região crítica.

No caso geral então temos uma amostra aleatória $\boldsymbol{X} = (X_1, \ldots, X_n)$ tomada de uma distribuição que envolve um parâmetro θ desconhecido, definido em um espaço paramétrico Θ. Assim, as hipóteses podem ser definidas como $H_0 : \theta \in \Theta_0$ e $H_1 : \theta \in \Theta_1$ onde Θ_0 e Θ_1 são subconjuntos disjuntos de Θ. Um teste é especificado particiondo-se o espaço amostral em dois subconjuntos. Um sobconjunto contem os valores de $\boldsymbol{X}$ para os quais H_0 será rejeitada e é chamado região crítica do teste, e o outro contem os valores de $\boldsymbol{X}$ para os quais H_0 será aceita e é chamado região de aceitação do teste. Em resumo, um teste fica determinado quando especificamos sua região crítica.

Além disso, uma hipótese pode ser classificada da seguinte maneira. Se o subconjunto Θ_i, $i = 0$ ou $i = 1$ contém um único valor então H_i é uma hipótese simples. Caso contrário, se Θ_i contém mais de um valor então H_i é uma hipótese composta. Nos exemplos 2.1 e 2.2 H_0 é uma hipótese simples enquanto H_1 é composta.

2.1.1 Tipos de Decisão

Ao tomar uma decisão a favor ou contra uma hipótese existem dois tipos de erros que podemos cometer. Podemos rejeitar a hipótese nula quando de fato ela é verdadeira (erro tipo I) ou podemos falhar em rejeitar H_0 quando de fato ela é falsa (erro tipo II). Frequentemente denotamos as probabilidades destes dois tipos de erro como α e β respectivamente.

Existe um balanço entre esses dois tipos de erros, no sentido de que ao tentar-se minimizar α, aumenta-se β. Isto é, não é possível minimizar estas duas probabilidades simultaneamente e na prática é costume fixar um valor (pequeno) para α. Na Tabela 2.1 estão descritos as decisões que podemos tomar e os tipos de erro associados.

Tabela 2.1: Tipos de decisão e tipos de erro associados a testes de hipóteses.

	Decisão	
Verdade	Aceitar H_0	Rejeitar H_0
H_0 verdadeira	Decisão correta (probabilidade $1-\alpha$)	Erro Tipo I (probabilidade α)
H_0 falsa	Erro Tipo II (probabilidade β)	Decisão correta (probabilidade $1-\beta$)

2.1.2 A Função Poder

As características probabilísticas de um teste podem ser descritas através de uma função que associa a cada valor de θ a probabilidade $\pi(\theta)$ de rejeitar H_0. A função $\pi(\theta)$ é chamada função de poder (ou potência) do teste. Assim, denotando por C a região crítica a função de poder é definida como

$$\pi(\theta) = P(\boldsymbol{X} \in C \mid \theta), \quad \forall\, \theta \in \Theta.$$

A função de poder é a ferramenta utilizada para verificar a adequação de um teste ou para comparar dois ou mais testes. É claro que uma função de poder ideal seria tal que $\pi(\theta) = 0$ para θ satisfazendo H_0 e $\pi(\theta) = 1$ para θ satisfazendo H_1. Em um problema prático no entanto raramente existirá um teste com estas características. Na Figura 2.1 abaixo está representada a função poder para o exemplo 2.2, i.e. $P(X \leq 6 \mid p)$, para $0 < p < 1$ onde $X \sim binomial(10, p)$. Note que neste exemplo se p for maior do que digamos 0,8 então o teste quase certamente aceitará H_0, indicando que o teste é adequado. Por outro lado, para valores de p entre 0,7 e 0,8 o teste ainda rejeita H_0 com probabilidade baixa.

O *tamanho* ou nível de significância α de um teste é definido como

$$\alpha \geq \sup_{\theta \in \Theta_0} \pi(\theta).$$

Assim como no caso de níveis de confiança na Seção 1.1, a desigualdade acima é essencialmente técnica já que estaremos interessados em valores de α tão pequenos quanto possível. Na prática isto implicará em usar uma igualdade e o tamanho do teste então será a probabilidade máxima, para $\theta \in \Theta_0$, de tomar uma decisão errada. A desigualdade será útil principalmente no caso de espaços amostrais discretos.

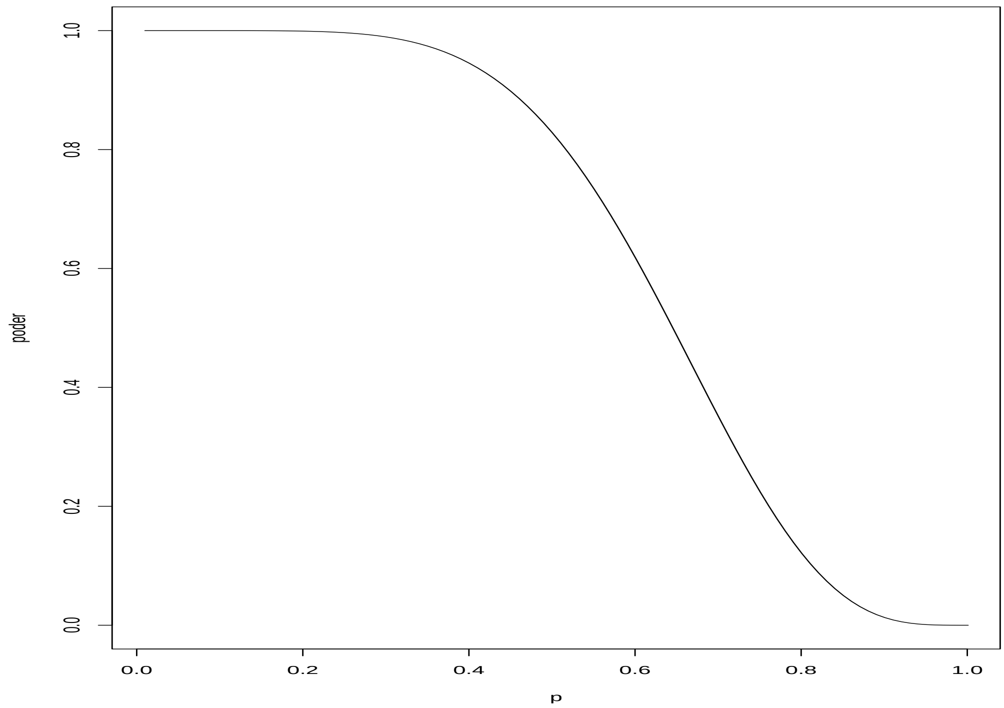
Exemplo 2.3 : Seja $X_1, \dots, X_n$ uma amostra aleatória da distribuição $N(\theta, \sigma^2)$ com $\sigma^2 = 25$ e suponha que queremos testar $H_0 : \theta \leq 17$. Suponha que a regra de decisão consiste em rejeitar H_0 se somente se $\overline{X} > 17 + \sigma/\sqrt{n}$. Neste caso a função poder é dada por

$$\pi(\theta) = P(\text{rejeitar } H_0) = P(\overline{X} > 17 + \sigma/\sqrt{n}) = P\left(Z > \frac{17 + \sigma/\sqrt{n} - \theta}{\sigma/\sqrt{n}}\right)$$

onde $Z \sim N(0, 1)$. Para $n = 25$, calculando esta probabilidade para vários valores de θ podemos construir o gráfico da Figura 2.2 para a função poder do teste. Note que o valor máximo da função quando H_0 é verdadeira ($\theta \leq 17$) é obtido para $\theta = 17$ e portanto o tamanho do teste é dado por

$$\sup_{\theta \leq 17}\left[P\left(Z > \frac{17 + \sigma/\sqrt{n} - \theta}{\sigma/\sqrt{n}}\right)\right] = P(Z > 1) \approx 0,159$$

Figura 2.1: Gráfico da função de poder para o exemplo 2.2.



Exercícios

1. Suponha que $X_1, \ldots, X_n$ é uma amostra aleatória da distribuição $U(0, \theta)$ e queremos testar as hipóteses $H_0 : \theta \geq 2 \times H_1 : \theta < 2$. Seja $Y_n = \max(X_1, \ldots, X_n)$ e um teste que rejeita H_0 se $Y_n \leq 5$.

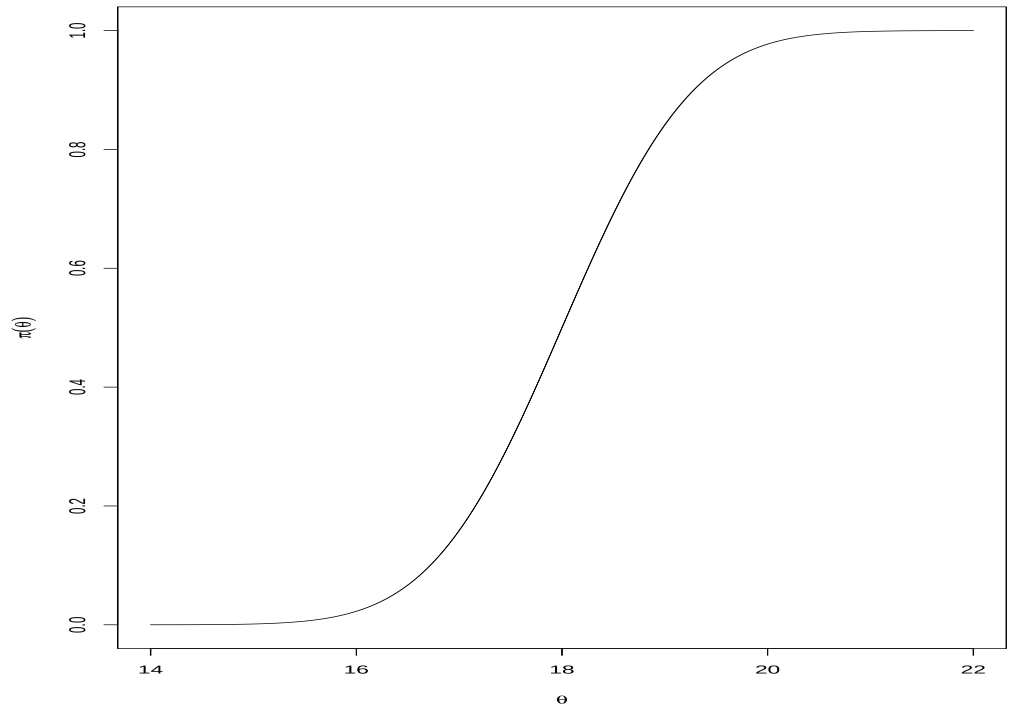
   (a) Determine a função poder do teste.

   (b) Determine o tamanho do teste.

2. Suponha que a proporção p de itens defeituosos em uma população de itens é desconhecida e queremos testar as hipóteses $H_0 : p = 0,2 \times H_1 : p \neq 0,2$. Uma amostra aleatória de 20 itens é tomada desta população e a regra de decisão consiste em rejeitar H_0 se o número amostral de defeituosos for menor ou igual a 1 ou maior ou igual a 7.

   (a) Faça um esboço do gráfico da função poder para $p = 0; 0,1; 0,2, \ldots, 1$

Figura 2.2: Gráfico da função de poder para o exemplo 2.3.



(b) Determine o tamanho do teste.

2.2 Testando Hipóteses Simples

É mais útil começar o estuda da teoria de testes de hipóteses considerando apenas hipóteses simples. Isto equivale a dizer que uma amostra aleatória $X_1, \dots, X_n$ foi tomada de um dentre duas possíveis distribuições e queremos decidir de qual delas vem a amostra. Neste caso o espaço paramétrico Θ contém apenas dois pontos, digamos θ_0 e θ_1 e queremos testar

$$H_0 : \theta = \theta_0 \quad \times \quad H_1 : \theta = \theta_1.$$

As probabilidades dos dois tipo de erro são dadas por

$$\alpha = P(\text{rejeitar } H_0 \mid \theta = \theta_0)$$

$$\beta = P(\text{aceitar } H_0 \mid \theta = \theta_1)$$

e gostariamos de poder construir um teste para o qual estas probabilidades fossem as menores possíveis. Na prática é impossível encontrar um teste que minimize α e β simultaneamente mas pode-se construir testes que minimizam combinações lineares destas probabilidades. Assim, para constantes positivas a e b queremos encontrar um teste δ para o qual $a\alpha(\delta) + b\beta(\delta)$ seja mínima.

Teorema 2.1 (Teste Ótimo) *Seja $X_1, \dots, X_n$ uma amostra aleatória de uma distribuição com função de (densidade) de probabilidade $p(x|\theta)$ e defina $p_i = p(x|\theta_i)$. Se um teste δ^* rejeita H_0 quando $p_0/p_1 < k$, aceita H_0 quando $p_0/p_1 > k$ e nada decide se $p_0/p_1 = k$, então qualquer outro teste δ é tal que*

$$a\alpha(\delta^*) + b\beta(\delta^*) \le a\alpha(\delta) + b\beta(\delta)$$

A razão p_0/p_1 é chamada *razão de verossimilhanças* (RV). O teorema estabelece então que um teste ótimo, no sentido de minimizar $a\alpha(\delta) + b\beta(\delta)$, rejeita H_0 quando a razão de verossimilhanças é pequena e aceita H_0 quando esta razão é grande.

Outro resultado vem do fato de que a hipótese H_0 e o erro tipo I são em geral privilegiados em problemas práticos. Assim, é usual considerar testes tais que $\alpha(\delta)$ não seja maior do que um nível especificado, digamos α_0, e tentar minimizar $\beta(\alpha)$.

Lema 2.1 (Neyman-Pearson) *Seja $X_1, \dots, X_n$ uma amostra aleatória de uma distribuição com função de (densidade) de probabilidade $p(x|\theta)$ e defina $p_i = p(x|\theta_i)$. Se um teste δ^* rejeita H_0 quando $p_0/p_1 < k$, aceita H_0 quando $p_0/p_1 > k$ e nada decide se $p_0/p_1 = k$, então para qualquer outro teste δ tal que $\alpha(\delta) \le \alpha(\delta^*)$, $\beta(\delta) \ge \beta(\delta^*)$. E também, $\alpha(\delta) < \alpha(\delta^*)$ implica em $\beta(\delta) > \beta(\delta^*)$.*

Exemplo 2.4 : Seja $X_1, \dots, X_n$ uma amostra aleatória da distribuição $N(\theta, 1)$ e queremos testar $H_0 : \theta = 0 \times H_1 : \theta = 1$. Neste caso a razão de verossimilhanças é dada por

$$\begin{aligned}
\frac{p_0}{p_1} &= \frac{(2\pi)^{-n/2} \exp(-(1/2)\sum_{i=1}^n x_i^2)}{(2\pi)^{-n/2} \exp(-(1/2)\sum_{i=1}^n (x_i - 1)^2)} \\
&= \exp\left\{-\frac{1}{2}\left[\sum_{i=1}^n x_i^2 - \sum_{i=1}^n (x_i - 1)^2\right]\right\} \\
&= \exp\left[-n\left(\overline{x} - \frac{1}{2}\right)\right].
\end{aligned}$$

Portanto rejeitar H_0 quando $p_0/p_1 < k$ é equivalente a rejeitar H_0 quando $\overline{x} > (1/2) - (1/n)\log k = c$. Não é difícil obter o valor da constante c tal que

$$P(\overline{X} > c \mid \theta = 0) = P(Z > c\sqrt{n}) = \alpha \quad \text{onde} \quad Z \sim N(0,1)$$

Por exemplo para $\alpha = 0,05$ obtemos da tabela da normal padronizada que $c\sqrt{n} = 1,645$ e o teste ótimo (que minimiza β) consiste em rejeitar H_0 se $\overline{X} > 1,645/\sqrt{n}$.

Exemplo 2.5 : Seja $X_1, \dots, X_n$ uma amostra aleatória da distribuição exponencial com parâmetro θ e queremos testar $H_0 : \theta = \theta_0 \ \times \ H_1 : \theta = \theta_1$. A razão de verossimilhanças é dada por

$$\frac{p_0}{p_1} = \left(\frac{\theta_0}{\theta_1}\right)^n \exp\left[-(\theta_0 - \theta_1)\sum_{i=1}^n x_i\right]$$

então, pelo lema de Neyman-Pearson, o teste mais poderoso (teste ótimo) rejeita H_0 se $p_0/p_1 < k$ ou equivalentemente se

$$\sum_{i=1}^n x_i > \frac{1}{\theta_0 - \theta_1}\log\left[k\left(\frac{\theta_1}{\theta_0}\right)^n\right] = c$$

A constante c é obtida fixando-se o valor de α, i.e. $\alpha = P(\sum_{i=1}^n X_i > c \mid \theta = \theta_0)$ e onde $\sum_{i=1}^n X_i \sim Gama(n, \theta_0)$.

2.2.1 Probabilidade de significância (P-valor)

Vimos que a escolha do nível de significância do teste é completamente arbitrária. Além disso, quando a distribuição da estatística de teste é discreta, como no exemplo da binomial, o nível escolhido pode nem mesmo ser atingido. Por outro lado, a decisão de aceitar ou rejeitar H_0 claramente depende desta escolha. Na maioria das aplicações práticas o valor escolhido é 0,05 ou 0,01 mas não há nada que justifique formalmente o uso destes valores em particular.

Um enfoque alternativo consiste em calcular o menor nível de significância para o qual H_0 é rejeitada, para o valor observado da estatística de teste. Esta quantidade é chamada *nível critico*, *probabilidade de significância* ou *p-valor*. A idéia é que, após calcular o p-valor o pesquisador pode escolher o seu próprio nível de significância como sendo a probabilidade máxima tolerável para um erro tipo I. Em geral, se T é uma estatística de teste e H_0 é rejeitada por exemplo para $T > c$ então o p-valor é a probabilidade $P(T > t \mid H_0)$ onde t é o valor observado de T.

Exemplo 2.6 : No exemplo 2.1 suponha que o número observado de questões certas foi $X = 9$. Então o p-valor será

$$P(X \geq 9 \mid p = 1/2) = \binom{10}{9} 0,5^{10} + \binom{10}{10} 0,5^{10} = 0,0107$$

e rejeitaremos H_0 para todo nível de significância maior do que este valor. Por exemplo, rejeitaremos H_0 para $\alpha = 0,025$ ou $\alpha = 0,05$.

Exemplo 2.7 : No exemplo 2.2 suponha que o número observado de não defeituosos foi $X = 4$. Neste caso o p-valor é dado por

$$P(X \leq 4 \mid p = 0,90) = 0,000146$$

ou seja, rejeitaremos H_0 para praticamente todos os níveis de significância usuais.

Portanto, o p-valor é a probabilidade de observar resultados tão extremos quanto os obtidos se a hipótese nula for verdadeira. A idéia é que se o p-valor for grande ele fornece evidência de que H_0 é verdadeira, enquanto que um p-valor pequeno indica que existe evidência nos dados contra H_0. As seguintes interpretações de p-valores (P) podem ser úteis,

$P \geq 0,10$	Não existe evidência contra H_0
$P < 0,10$	Fraca evidência contra H_0
$P < 0,05$	Evidência significativa ...
$P < 0,01$	Evidência altamente significativa ...
$P < 0,001$	Evidência extremamente significativa ...

2.3 Testes Uniformemente mais Poderosos

Na Seção 2.2 foram definidos testes ótimos para testar hipóteses simples. Nesta seção os resultados serão generalizados para hipóteses compostas. Considere então um teste em que H_0 pode ser uma hipótese simples ou composta e H_1 é sempre uma hipótese composta.

Definição 2.1 *Um teste δ de $H_0 : \theta \in \Theta_0 \times H_1 : \theta \in \Theta_1$ é dito ser uniformemente mais poderoso (UMP) de tamanho α se e somente se*

$$\sup_{\theta \in \Theta_0} \pi(\theta) = \alpha$$

e para qualquer outro teste δ^* que satisfaça esta igualdade

$$\pi(\theta|\delta) \geq \pi(\theta|\delta^*), \quad \forall\ \theta \in \Theta_1.$$

Assim, de acordo com esta definição, precisamos especificar um teste cuja probabilidade máxima de rejeitar H_0 quando ela é verdadeira seja α e que ao mesmo tempo maximize a probabilidade de rejeitar H_0 quando ela é falsa. Veremos a seguir que os testes UMP só existem em situações especiais, por exemplo quando a distribuição pertence à família exponencial.

A família exponencial inclui muitas das distribuições de probabilidade mais comumente utilizadas em Estatística, tanto contínuas quanto discretas. Uma característica essencial desta família é que existe uma estatística suficiente com dimensão fixa.

Definição 2.2 *A família de distribuições com função de (densidade) de probabilidade $p(x|\theta)$ pertence à família exponencial a um parâmetro se podemos escrever*

$$p(\boldsymbol{x}|\theta) = a(\boldsymbol{x}) \exp\{T(\boldsymbol{x})\phi(\theta) + b(\theta)\}.$$

Note que pelo critério de fatoração de Neyman $T(\boldsymbol{x})$ é uma estatística suficiente para θ.

Teorema 2.2 *Se $X_1, \dots, X_n$ é uma amostra aleatória de um membro da família exponencial e ϕ for estritamente crescente em θ então o teste UMP de nível α para testar $H_0 : \theta \leq \theta_0 \times H_1 : \theta > \theta_0$ rejeita H_0 se $T(\boldsymbol{x}) > c$. Se as hipóteses forem invertidas ou ϕ for estritamente decrescente em θ então o teste UMP rejeita H_0 se $T(\boldsymbol{x}) < c$. Se ambas as condições ocorrerem o teste fica inalterado.*

Um fato importante é que, em qualquer condição estes testes têm função poder crescente em θ. Assim a constante c acima é obtida de modo que $P(\text{rejeitar } H_0 \mid \theta = \theta_0) \leq \alpha$, com igualdade no caso contínuo.

Exemplo 2.8 : Seja $X_1, \dots, X_n$ uma amostra aleatória da distribuição de Bernoulli com parâmetro θ e queremos testar $H_0 : \theta \leq \theta_0 \times H_1 : \theta > \theta_0$. Então, definindo $t(\boldsymbol{x}) = \sum_{i=1}^n x_i$

$$\begin{aligned} p(\boldsymbol{x}|\theta) &= \theta^{t(\boldsymbol{x})}(1-\theta)^{n-t(\boldsymbol{x})} = \exp[t(\boldsymbol{x}) \log \theta + (n - t(\boldsymbol{x})) \log(1-\theta)] \\ &= \exp\left\{ t(\boldsymbol{x}) \log\left(\frac{\theta}{1-\theta}\right) + n \log(1-\theta) \right\}. \end{aligned}$$

Logo, a distribuição pertence à família exponencial e $\phi(\theta) = \log(\theta/(1-\theta))$ é uma função estritamente crescente de θ. Assim, um teste UMP deve rejeitar H_0 se $\sum_{i=1}^n X_i > c$ onde c é tal que $P(\sum_{i=1}^n X_i > c \mid \theta = \theta_0) \leq \alpha$.

Exemplo 2.9 : Seja $X_1, \dots, X_n$ uma amostra aleatória da distribuição exponencial com parâmetro θ e queremos testar $H_0 : \theta \leq \theta_0 \times H_1 : \theta > \theta_0$. Definindo $t(\boldsymbol{x}) = \sum_{i=1}^n x_i$ a densidade conjunta é

$$p(\boldsymbol{x}|\theta) = \theta^n e^{-\theta\, t(\boldsymbol{x})} = \exp(n \log \theta - \theta\, t(\boldsymbol{x})).$$

Portanto a distribuição pertence à família exponencial e $\phi(\theta) = -\theta$ é uma função estritamente decrescente de θ. Então pelo teorema 2.2 o teste UMP deve rejeitar H_0 se $\sum_{i=1}^n X_i < c$. Fixando o valor de α a constante c é a solução da equação $P(\sum_{i=1}^n X_i < c \mid \theta = \theta_0) = \alpha$ onde $\sum_{i=1}^n X_i \sim Gama(n, \theta_0)$.

A propriedade que garante a existência de testes UMP na família exponencial pode ser extendida a famílias de distribuições com razão de verossimilhança monótona.

Definição 2.3 *A família de distribuições com função de (densidade) de probabilidade $p(\boldsymbol{x}|\theta)$ é dita ter razão de verossimilhança monótona se existe uma estatística $T(\boldsymbol{X})$ tal que $\forall\, \theta_1, \theta_2 \in \Theta$, com $\theta_1 < \theta_2$, a razão $p(\boldsymbol{x}|\theta_2)/p(\boldsymbol{x}|\theta_1)$ é uma função monótona em $t(\boldsymbol{x})$.*

Intuitivamente, quanto maior for a razão de verossimilhança mais plausível é o valor θ_2 em relação a θ_1. Assim, se queremos testar $H_0 : \theta \leq \theta_0 \times H_1 : \theta > \theta_0$ e se a RV for uma função crescente de $T(\boldsymbol{X})$ então é razoável rejeitar H_0 para valores grandes de $T(\boldsymbol{X})$. Pode-se mostrar que neste caso o teste UMP rejeita H_0 se $T(\boldsymbol{X}) > c$. Analogamente, se as hipóteses forem invertidas ou se a RV for uma função decrescente de $T(\boldsymbol{X})$ então o teste UMP rejeita H_0 se $T(\boldsymbol{X}) < c$. Se ambas as condições ocorrerem o teste fica inalterado.

Em qualquer destas condições o fato importante é que a função poder é sempre crescente em θ. Portanto, a constante c acima é obtida de modo que $P(\text{rejeitar } H_0 \mid \theta = \theta_0) \leq \alpha$, com igualdade no caso contínuo.

Exemplo 2.10 : Seja $X_1, \dots, X_n$ uma amostra aleatória da distribuição de Bernoulli com parâmetro θ e queremos testar $H_0 : \theta \leq \theta_0 \times H_1 : \theta > \theta_0$. Então, definindo $t(\boldsymbol{x}) = \sum_{i=1}^n x_i$ temos que

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

e para $\theta_1 < \theta_2$ a razão de verossimilhança fica

$$\frac{\theta_2^{t(\boldsymbol{x})}(1-\theta_2)^{n-t(\boldsymbol{x})}}{\theta_1^{t(\boldsymbol{x})}(1-\theta_1)^{n-t(\boldsymbol{x})}} = \left[\frac{\theta_2(1-\theta_1)}{\theta_1(1-\theta_2)}\right]^t \left(\frac{1-\theta_2}{1-\theta_1}\right)^n = \alpha^t \beta^n.$$

Como $\theta_2 > \theta_1$ e $1-\theta_1 > 1-\theta_2$ então $\alpha > 1$ e a RV é uma função crescente em t. Portanto, o teste UMP rejeita H_0 se $\sum_{i=1}^n X_i > c$ confirmando assim o resultado no exemplo 2.8.

Exercícios

1. Para cada uma das distribuições abaixo considere uma amostra aleatória $X_1, \ldots, X_n$ e obtenha o teste UMP para testar as hipóteses $H_0 : \theta \leq \theta_0 \times H_0 : \theta > \theta_0$.
   (a) Poisson com parâmetro θ.
   (b) Normal com média conhecida e variância desconhecida.
   (c) Gama com parâmetro α desconhecido e β conhecido.
   (d) Gama com parâmetro α conhecido e β desconhecido.
2. Seja $X_1, \ldots, X_n$ uma amostra aleatória da distribuição $N(0, \sigma^2)$ com σ^2 desconhecido. Obtenha o teste UMP para testar as hipóteses $H_0 : \sigma^2 \leq 2 \ \times \ H_0 : \sigma^2 > 2$ com $n = 10$ e $\alpha = 0,05$.
3. Suponha que $X_1, \ldots, X_n$ seja uma amostra aleatória da distribuição exponencial com parâmetro θ e queremos testar $H_0 : \theta \geq 1/2 \ \times \ H_0 : \theta < 1/2$. Obtenha o teste UMP para estas hipóteses com $n = 10$ e $\alpha = 0,05$.
4. Suponha que $X_1, \ldots, X_n$ seja uma amostra aleatória da distribuição de Poisson com parâmetro θ e queremos testar $H_0 : \theta \leq 1 \times H_0 : \theta > 1$. Obtenha o teste UMP para estas hipóteses com $n = 10$ e $\alpha = 0,05$.
5. Seja $X_1, \ldots, X_n$ é uma amostra aleatória da distribuição com função de densidade $p(x|\theta) = \theta x^{\theta-1}$, para $x \in (0,1)$ e $\theta > 0$ desconhecido. Encontre o teste UMP para as hipóteses $H_0 : \theta \leq 1 \times H_1 : \theta > 1$ com nível de significância $\alpha = 0,05$.
6. A proporção p de itens defeituosos em um grande lote de manufaturados é desconhecida. Uma amostra aleatória de 20 itens foi selecionada e inspecionada, e queremos testar as hipóteses $H_0 : p \leq 0,1 \times H_1 : p > 0,1$. Obtenha o teste UMP.
7. Suponha que $X_1, \ldots, X_n$ seja uma amostra aleatória da distribuição de Poisson com média λ desconhecida e queremos testar $H_0 : \lambda \geq 1 \times H_0 : \lambda < 1$. Para $n = 10$, verifique para quais níveis de significância no intervalo $0 < \alpha < 0,03$ existem testes UMP.

8. Seja $X_1, \dots, X_n$ uma amostra aleatória da distribuição $N(\mu, 1)$ com μ desconhecido e queremos testar as hipóteses $H_0 : \mu \leq 0 \ \times \ H_0 : \mu > 0$. Sejam δ^* o teste UMP ao nível $\alpha = 0,025$ e $\pi(\mu|\delta^*)$ função poder do teste.
   (a) Determine o menor valor de n para o qual $\pi(\mu|\delta^*) \geq 0,9$ para $\mu \geq 0,5$.
   (b) Determine o menor valor de n para o qual $\pi(\mu|\delta^*) \leq 0,001$ para $\mu \leq -0,1$.

9. Seja $X_1, \dots, X_n$ uma amostra aleatória da distribuição χ^2 com número de graus de liberdade θ desconhecido, $\theta = 1, 2, \dots$. Suponha que queremos testar as hipóteses $H_0 : \theta \leq 8 \times H_1 : \theta \geq 9$ ao nível de significância α. Mostre que existe um teste UMP que rejeita H_0 se $\sum_{i=1}^n \log X_i > k$ para uma constante k.

2.4 Testes Bilaterais

Suponha agora que queremos testar hipóteses do tipo

$$H_0 : \theta = \theta_0 \times H_1 : \theta \neq \theta_0,$$

ou seja H_0 é uma hipótese simples e H_1 é uma alternativa bilateral. Como veremos nas próximas seções este tipo de teste pode ser útil na comparação de tratamentos. O problema é que neste caso não existe um teste UMP para estas hipóteses, i.e. não é possível construir um teste cuja probabilidade de rejeitar H_0 seja maximizada quando ela é falsa.

Alternativamente poderiamos construir testes tais que as chances de rejeitar H_0 sejam maiores quando ela é falsa do que quando ela é verdadeira. Isto nos leva à definição de testes não viesados a seguir.

Definição 2.4 *Um teste δ é dito ser não viesado para as hipóteses $H_0 : \theta \in \Theta_0 \times H_1 : \theta \in \Theta_1$ se $\forall\ \theta \in \Theta_0$ e $\theta' \in \Theta_1$ então $\pi(\theta) \leq \pi(\theta')$. Caso contrário o teste é dito viesado.*

Ou seja, em testes não viesados a probabilidade de rejeitar H_0 quando ela é falsa é no mínimo tão grande quanto para H_0 verdadeira.

Podemos agora tentar construir testes para hipóteses bilaterais que sejam UMP dentro da classe de testes não viesados. Se a distribuição pertence à família exponencial, pode-se mostrar que se $\phi(\theta)$ for uma função estritamente crescente em θ então o teste UMP não viesado de nível α para $H_0 : \theta = \theta_0 \times H_1 : \theta \neq \theta_0$

aceita H_0 quando $c_1 < T(\boldsymbol{X}) < c_2$. As constantes c_1 e c_2 são obtidas de modo que $P(c_1 < T(\boldsymbol{X}) < c_2 \mid \theta = \theta_0) = 1 - \alpha$.

Note que existe uma infinidade de valores de c_1 e c_2 satisfazendo a esta condição. Em muitas situações é conveniente tomar valores tais que

$$P(T(\boldsymbol{X}) < c_1 \mid \theta = \theta_0) = P(T(\boldsymbol{X}) > c_2 \mid \theta = \theta_0) = \alpha/2$$

e se $T(\boldsymbol{X})$ tem uma distribuição simétrica em torno de um ponto isto implica em escolher c_1 e c_2 simetricamente em relação a este ponto. No entanto, nada impede que outros valores possam ser considerados. Por exemplo, o pesquisador pode considerar mais grave aceitar H_0 quando $\theta < \theta_0$ do que quando $\theta > \theta_0$ e neste caso é melhor considerar testes com função de perda assimétrica.

2.4.1 Testes Gerais

Em muitas situações não é possível obter nem mesmo um teste não viesado. Um procedimento geral para testar $H_0 : \theta \in \Theta_0 \times H_1 : \theta \in \Theta_1$ é baseado na estatística da razão de máxima verossimilhança (RMV) dada por

$$\lambda(\boldsymbol{X}) = \frac{\sup_{\theta \in \Theta_0} p(\boldsymbol{X}|\theta)}{\sup_{\theta \in \Theta_1} p(\boldsymbol{X}|\theta)}$$

Deste modo estaremos comparando o valor máximo atingido pela função de verossimilhança quando $\theta \in \Theta_0$ com o valor máximo atingido quando $\theta \in \Theta_1$. Neste caso, é razoável decidir pela rejeição de H_0 se $\lambda(\boldsymbol{X}) < c$ onde a constante c é obtida de modo que

$$\sup_{\theta \in \Theta_0} P(\lambda(\boldsymbol{X}) < c \mid \theta) \leq \alpha.$$

Novamente, a igualdade será usada sempre que possível ficando a desigualdade para o caso de distribuições discretas.

Além do cálculo de valores máximos da função de verossimilhança existe outra dificuldade associada a estes testes que é a determinação da distribuição amostral de $\lambda(\boldsymbol{X})$. Este problema será discutido quando falarmos de testes assintóticos na Seção 2.6.

2.5 Testes de Hipóteses no Modelo Normal

Os resultados desenvolvidos nas seções anteriores serão aplicados ao modelo normal para testes sobre média e variância em problemas de uma ou mais amostras e em modelos de regressão linear. Nesta seção considere uma amostra aleatória $X_1, \cdots, X_n$ tomada da distribuição $N(\theta, \sigma^2)$.

Suponha que queremos testar $H_0 : \theta = \theta_0 \times H_1 : \theta \neq \theta_0$ e inicialmente vamos assumir que σ^2 é conhecida. Neste caso,

$$\begin{aligned}
p(\boldsymbol{x}|\theta) &= (2\pi\sigma^2)^{-n/2} \exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\theta)^2\right) \\
&= (2\pi\sigma^2)^{-n/2} \exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^{n}x_i^2\right) \exp\left(\overline{\boldsymbol{x}}n\theta - \frac{n\theta^2}{2\sigma^2}\right)
\end{aligned}$$

e como $n\theta$ é uma função estritamente crescente de θ segue que o teste UMP não viesado rejeita H_0 se $\overline{\boldsymbol{X}} < c_1$ ou $\overline{\boldsymbol{X}} > c_2$. Ao nível de significância α podemos obter as constantes c_1 e c_2 tais que

$$P(\overline{\boldsymbol{X}} < c_1 \mid \theta = \theta_0) + P(\overline{\boldsymbol{X}} > c_2 \mid \theta = \theta_0) = \alpha.$$

Conforme discutido anteriormente, existe uma infinidade de valores que satisfazem esta condição. Na maioria dos experimentos envolvendo o modelo normal será conveniente tomar c_1 e c_2 simétricos em relação a $E(\overline{\boldsymbol{X}})$. Assim, usando uma tabela da distribuição normal padronizada podemos obter o valor do percentil $z_{\alpha/2}$ tal que

$$P\left(-z_{\alpha/2} \leq \frac{\sqrt{n}(\overline{\boldsymbol{X}}-\theta_0)}{\sigma} \leq z_{\alpha/2}\right) = 1-\alpha$$

e o teste bilateral UMP não viesado rejeita H_0 se $\overline{\boldsymbol{X}} < \theta_0 - z_{\alpha/2}\sigma/\sqrt{n}$ ou $\overline{\boldsymbol{X}} > \theta_0 + z_{\alpha/2}\sigma/\sqrt{n}$.

No caso em que a variância populacional é também desconhecida o espaço dos parâmetro é $\Theta = \{(\theta,\sigma^2) : \theta \in \mathbb{R}, \sigma^2 > 0\}$ e vamos obter o teste da RMV. Note que, como H_0 é uma hipótese simples então $\Theta_0 = \{(\theta_0,\sigma^2) : \sigma^2 > 0\}$ e não é difícil verificar que o valor de σ^2 que maximiza a verossimilhança para θ_0 fixo é $\hat{\sigma}_0^2 = \sum_{i=1}^{n}(x_i-\theta_0)^2/n$. Portanto,

$$\sup_{(\theta,\sigma^2)\in\Theta_0} p(\boldsymbol{X}|\theta,\sigma^2) = p(\boldsymbol{x}|\theta_0,\hat{\sigma}_0^2).$$

Para $\theta \neq \theta_0$ a função de verossimilhança é maximizada em $(\hat{\theta},\hat{\sigma}^2)$ onde $\hat{\theta} = \overline{\boldsymbol{x}}$ e $\hat{\sigma}^2 = \sum_{i=1}^{n}(x_i-\overline{\boldsymbol{x}})^2/n$. Portanto

$$\sup_{(\theta,\sigma^2)\in\Theta} p(\boldsymbol{X}|\theta,\sigma^2) = p(\boldsymbol{x}|\hat{\theta},\hat{\sigma}^2).$$

Assim, a estatística da RMV é dada por

$$\lambda(\boldsymbol{X}) = \frac{(2\pi\hat{\sigma}_0^2)^{-n/2}\exp\{-\sum_{i=1}^{n}(X_i-\theta_0)^2/2\hat{\sigma}_0^2\}}{(2\pi\hat{\sigma}^2)^{-n/2}\exp\{-\sum_{i=1}^{n}(X_i-\overline{\boldsymbol{X}})^2/2\hat{\sigma}^2\}}$$

e substituindo as somas de quadrados obtemos que $\lambda(\boldsymbol{X}) = (\hat{\sigma}_0^2/\hat{\sigma}^2)^{-n/2}$. Mas,

$$\frac{\hat{\sigma}_0^2}{\hat{\sigma}^2} = \frac{\sum_{i=1}^n (X_i - \overline{\boldsymbol{X}})^2 + n(\overline{\boldsymbol{X}} - \theta_0)^2}{\sum_{i=1}^n (X_i - \overline{\boldsymbol{X}})^2} = 1 + \frac{n(\overline{\boldsymbol{X}} - \theta_0)^2}{(n-1)S^2} = 1 + \frac{T^2}{n-1}$$

onde $T = \sqrt{n}(\overline{\boldsymbol{X}} - \theta_0)/S$ e então podemos reescrever a RMV como

$$\lambda(\boldsymbol{X}) = \left(1 + \frac{T^2}{n-1}\right)^{-n/2}.$$

Finalmente, o teste da RMV rejeita H_0 se $\lambda(\boldsymbol{X}) < c^*$ ou equivalentemente se $T^2 > c$ ou $|T| > c$. Como $T \sim t_{n-1}$ a constante c é simplesmente o percentil $t_{\alpha/2,n-1}$ desta distribuição.

O teste desenvolvido acima é conhecido como teste t e talvez um dos mais utilizados em Estatística. Pode-se mostrar que o teste t é não viesado já que o valor mínimo da função poder ocorre em $\theta = \theta_0$. Além disso, as propriedades do teste não são afetadas pelo valor de σ^2 (parâmetro de distúrbio) já que σ^2 foi substituido pelo seu estimador S^2 e T é uma quantidade pivotal. O teste também é invariante a transformações lineares das observações.

Testes bilaterais do tipo $H_0 : \sigma^2 = \sigma_0^2 \times H_1 : \sigma^2 \neq \sigma_0^2$ para a variância podem ser construídos fazendo-se analogia com intervalos de confiança. Vimos na Seção 1.2.1 do Capítulo 1 que o intervalo de confiança de $100(1-\alpha)\%$ para σ^2 é dado por

$$\left(\frac{(n-1)s^2}{q2}, \frac{(n-1)s^2}{q1}\right)$$

onde q_1 e q_2 são os quantis $\alpha/2$ e $1-\alpha/2$ da distribuição χ^2_{n-1}. Assim, o teste deve aceitar H_0 se e somente se σ_0^2 estiver contido neste intervalo. Será deixado como exercício mostrar que este é o teste da razão de máxima verossimilhança para as hipóteses acima.

2.5.1 Testes para Várias Médias

Para começar vamos assumir que temos duas amostras aleatórias $X_{11}, \dots, X_{1n_1}$ e $X_{21}, \dots, X_{2n_2}$ das distribuições $N(\theta_1, \sigma_1^2)$ e $N(\theta_2, \sigma_2^2)$ respectivamente e que as amostras são independentes. Neste caso o vetor de parâmetros é $(\theta_1, \theta_2, \sigma_1^2, \sigma_2^2)$ e em geral estaremos interessados em testar as hipóteses

$$\begin{aligned} &H_0 : \theta_1 = \theta_2, \sigma_1^2 > 0, \sigma_2^2 > 0 \\ &H_1 : \theta_1 \neq \theta_2, \sigma_1^2 > 0, \sigma_2^2 > 0 \end{aligned} \tag{2.1}$$

Se pudermos assumir que as variâncias populacionais são iguais, i.e. $\sigma_1^2 = \sigma_2^2 = \sigma^2$, o problema de construção do teste se torna relativamente simples usando a estatística da razão de máxima verossimilhança. Neste caso, como as amostras são independentes, podemos escrever a função de verossimilhança como

$$p(\boldsymbol{x}_1, \boldsymbol{x}_2|\theta_1, \theta_2, \sigma^2) = p(\boldsymbol{x}_1|\theta_1, \sigma^2)p(\boldsymbol{x}_2|\theta_2, \sigma^2)$$

e após algum algebrismo segue que a verossimilhança de $(\theta_1, \theta_2, \sigma^2)$ é dada por

$$(2\pi\sigma^2)^{-(n_1+n_2)/2} \exp\left\{-\frac{1}{2\sigma^2}\left[(n_1-1)S_1^2 + n_1(\theta_1 - \overline{x}_1)^2 + (n_2-1)S_2^2 + n_2(\theta_2 - \overline{x}_2)^2\right]\right\}.$$

Quando $\theta_1 \neq \theta_2$ as estimativas de máxima verossimilhança de θ_1, θ_2 e σ^2 são respectivamente $\overline{x}_1$, $\overline{x}_2$ e

$$\hat{\sigma}^2 = \frac{(n_1-1)S_1^2 + (n_2-1)S_2^2}{n_1+n_2-2}$$

onde S_1^2 e S_2^2 são as variâncias amostrais. Quando $\theta_1 = \theta_2 = \theta$ segue que as estimativas de máxima verossimilhança de θ e σ^2 são

$$\hat{\theta} = \frac{n_1\overline{x}_1 + n_2\overline{x}_2}{n_1+n_2} \quad \text{e} \quad \hat{\sigma}_0^2 = \hat{\sigma}^2 + \frac{n_1 n_2}{(n_1+n_2)^2}(\overline{x}_1 - \overline{x}_2)^2.$$

Substituindo estas expressões na razão de verossimilhanças pode-se mostrar que o teste da RMV rejeita H_0 se

$$|T| = \left|\frac{(\overline{x}_1 - \overline{x}_2)}{\hat{\sigma}\sqrt{\dfrac{1}{n_1} + \dfrac{1}{n_2}}}\right| > c.$$

Pode-se mostrar que T tem distribuição t de Student com $\nu = n_1 + n_2 - 2$ graus de liberdade de modo que a constante c é simplesmente o percentil $t_{\alpha/2,\nu}$ desta distribuição. Este teste é conhecido como teste t para duas amostras.

2.5.2 Variâncias Desconhecidas

O procedimento visto na seção anterior para variâncias iguais pode ser extendido facilmente para o caso de variâncias desconhecidas e desiguais, desde que a razão de variâncias σ_1^2/σ_2^2 seja conhecida. Suponha por exemplo que $\sigma_1^2 = k\sigma_2^2$ onde k é uma constante positiva conhecida. Definindo-se

$$\hat{\sigma}^2 = \frac{(n_1-1)S_1^2 + (n_2-1)S_2^2/k}{n_1+n_2-2}$$

então pode-se mostrar que quando $\theta_1 = \theta_2$ a variável aleatória

$$U = \frac{(\overline{X}_1 - \overline{X}_2)}{\hat{\sigma}\sqrt{\dfrac{1}{n_1} + \dfrac{k}{n_2}}}$$

tem distribuição t de Student com $n_1 + n_2 - 2$ graus de liberdade.

Finalmente, se mesmo a razão de variâncias for desconhecida então o problema de testar as hipóteses 2.1 torna-se bastante complexo. Este problema é conhecido na literatura como o *problema de Behrens-Fisher*. Vários procedimentos de teste já foram propostos e a maioria foi objeto de controvérsia em relação a sua utilidade e correção.

2.5.3 Comparação de Variâncias

Em problemas com duas ou mais amostras de distribuições normais é natural que se tenha interesse em comparar as variâncias populacionais. Neste caso, a distribuição F é utilizada para testar as hipóteses associadas. No caso de duas amostras suponha que queremos testar

$$\begin{aligned} H_0 &: \sigma_1^2 \leq \sigma_2^2 \\ H_1 &: \sigma_1^2 > \sigma_2^2 \end{aligned}$$

Pode-se mostrar que não existe teste UMP para estas hipóteses e é prática comum utilizar-se o chamado teste F. Este teste é não viesado e na verdade é UMP dentro da classe de testes não viesados. Usando a estatística da razão de máxima verossimilhança pode-se mostrar que o teste F rejeita H_0 se

$$\frac{\sum_{i=1}^{n_1}(x_{1i} - \overline{x}_1)^2/(n_1 - 1)}{\sum_{i=1}^{n_2}(x_{2i} - \overline{x}_2)^2/(n_2 - 1)} = \frac{s_1^2}{s_2^2} > c.$$

Vimos na Seção 1.2.4 que

$$\frac{S_1^2}{S_2^2}\frac{\sigma_2^2}{\sigma_1^2} \sim F(n_1 - 1, n_2 - 1).$$

e portanto a constante c pode ser obtida tal que

$$P\left(\frac{S_1^2}{S_2^2}\frac{\sigma_2^2}{\sigma_1^2} > c|\sigma_1^2 = \sigma_2^2\right) = \alpha$$

usando os valores tabelados da distribuição F com $n_1 - 1$ e $n_2 - 1$ graus de liberdade.

No caso de testes bilaterais, i.e.

$$H_0 : \sigma_1^2 = \sigma_2^2$$
$$H_1 : \sigma_1^2 \neq \sigma_2^2$$

o teste F rejeita H_0 se $s_1^2/s_2^2 < c_1$ ou $s_1^2/s_2^2 > c_2$ onde as constantes c_1 e c_2 são mais uma vez obtidas como percentis da distribuição F com $n_1 - 1$ e $n_2 - 1$ graus de liberdade. Analogamente ao teste t, é prática comum escolher c_1 e c_2 tal que as probabilidades nas caudas sejam iguais, i.e. $\alpha/2$.

Exercícios

1. Suponha que $X_1, \dots, X_n$ é uma amostra aleatória da distribuição $N(\mu, 1)$ e queremos testar as hipóteses $H_0 : \mu = \mu_0 \times H_1 : \mu \neq \mu_0$. Considere um teste que rejeita H_0 se $\overline{X} \leq c_1$ ou $\overline{X} \geq c_2$.

   (a) Determine os valores de c_1 e c_2 tais que $\pi(\mu_0) = 0,10$ e $\pi(\mu)$ seja simétrica em torno de μ_0.

   (b) Determine os valores de c_1 e c_2 tais que $\pi(\mu_0) = 0,10$ e o teste seja não viesado.

   (c) Suponha que $c_1 = \mu_0 - 1,96/\sqrt{n}$. Determine c_2 tal que $\pi(\mu_0) = 0,10$.

   (d) Determine o menor valor de n para o qual $\pi(\mu_0) = 0,10$ e $\pi(\mu_0 + 1) = \pi(\mu_0 - 1) \geq 0,95$.

2. Suponha que $X_1, \dots, X_n$ é uma amostra aleatória da distribuição $N(\mu, 1)$ e queremos testar as hipóteses

$$H_0 : 0,1 \leq \mu \leq 0,2$$
$$H_1 : \mu < 0,1 \text{ ou } \mu > 0,2.$$

   Considere um teste que rejeita H_0 se $\overline{X} \leq c_1$ ou $\overline{X} \geq c_2$.

   (a) Para $n = 25$ determine c_1 e c_2 tais que tais que $\pi(0,1) = \pi(0,2) = 0,07$.

   (b) Idem para $\pi(0,1) = 0,02$ e $\pi(0,2) = 0,05$.

3. Os comprimentos de fibras metálicas (em milímetros) produzidas por uma máquina têm distribuição normal com média μ e variância σ^2 desconhecidos. Suponha que queremos testar as seguintes hipóteses

$$H_0 : \mu \leq 5,2$$
$$H_1 : \mu > 5,2.$$

Os comprimentos de 15 fibras selecionadas ao acaso foram medidos e obteve-se a média amostral $\overline{x} = 5,4$ e $\sum_{i=1}^{n}(x_i - \overline{x})^2 = 2,5$.

(a) Construa um teste t ao nível de 0,05 baseado nestes resultados.

(b) Repita o item anterior para as hipóteses $H_0 : \mu = 5,2 \times H_1 : \mu \neq 5,2$. Qual a conclusão do exercício?

4. Suponha que foi selecionada uma amostra aleatória de 9 observações da distribuição $N(\mu, \sigma^2)$ com parâmetros desconhecidos. Obteve-se $\overline{X} = 22$ e $\sum_{i=1}^{n}(X_i - \overline{X})^2 = 72$.

   (a) Teste as hipóteses $H_0 : \mu \leq 20 \times H_1 : \mu > 20$ ao nível de significância 0,05.

   (b) Teste as hipóteses $H_0 : \mu = 20 \times H_1 : \mu \neq 20$ ao nível de significância 0,05. Use um teste simétrico com probabilidade 0,025 em cada cauda.

5. O tempo médio, por operário, para executar uma tarefa, tem sido de 100 minutos com desvio padrão de 15 minutos. Foi introduzida uma modificação para reduzir este tempo e após alguns meses foi selecionada uma amostra de 16 operários medindo-se o tempo de execução de cada um. Obteve-se um tempo médio amostral de 90 minutos e um desvio padrão amostral de 16 minutos. Estabeleça claramente as suposições que precisam ser feitas.

   (a) Verifique se existem evidências, ao nível de significância 0,05, de que a modificação surtiu efeito?

   (b) Verifique se há evidências, ao nível de significância 0,05, de que a modificação alterou a variância populacional.

6. Uma indústria compra componentes eletrônicos dos fornecedores A e B, mas o fornecedor A garante que o tempo médio de vida (em horas) do seu produto supera o da marca B em 300 horas. Para testar esta afirmação foram selecionadas duas amostras de componentes, uma de cada fornecedor, e obteve-se os seguintes tempos de vida:

marca A	1500	1450	1480	1520	1510
marca B	1100	1200	1180	1250	

   Após estabelecer claramente as suposições que precisam ser feitas,

   (a) teste a hipótese de igualdade das variâncias dos tempos de vida, ao nível de significância 0,02;

(b) teste a afirmação do fornecedor A, ao nível de significância 0,05.

7. Uma droga A foi administrada em um grupo de 8 pacientes selecionados ao acaso. Após um período fixo de tempo a concentração da droga em certas células de cada paciente foi medida (em unidades apropriadas). O procedimento foi repetido em um outro grupo de 6 pacientes selecionados ao acaso usando uma droga B. As concentrações obtidas foram

droga A	1,23	1,42	1,41	1,62	1,55	1,51	1,60	1,76
droga B	1,76	1,41	1,87	1,49	1,67	1,81		

Após estabelecer claramente as suposições que precisam ser feitas,

(a) teste a hipótese de que a concentração média de droga A entre todos os pacientes é pelo menos tão grande quanto da droga B;

(b) teste a hipótese de que as concentrações médias das duas drogas são iguais.

2.6 Testes Assintóticos

Vimos que a construção de um teste envolve a obtenção de constantes através da distribuição de probabilidades de uma estatística. Em muitas situações, particularmente para a razão de máxima verossimilhança, estas distribuições não podem ser determinadas de forma exata e precisamos recorrer a resultados aproximados. Nesta seção serão desenvolvidos testes baseados em distribuições assintóticas das estatísticas de teste envolvidas. Iremos nos concentrar em testes baseados na distribuição assintótica da razão de máxima verossimilhança, do estimador de máxima verossimilhança e da função escore.

Suponha que uma amostra aleatória $X_1, \dots, X_n$ é tomada de uma distribuição com parâmetro $\theta \in \Theta \subset \mathbb{R}$ desconhecido e queremos testar $H_0 : \theta = \theta_0$. Expandindo em série de Taylor a função $L(\theta_0) = \log p(\boldsymbol{x}|\theta_0)$ em torno do estimador de máxima verossimilhança $\hat{\theta}$ obtemos

$$L(\theta_0) \approx L(\hat{\theta}) + U(\boldsymbol{x}; \hat{\theta})(\theta_0 - \hat{\theta}) - \frac{1}{2} J(\hat{\theta})(\theta_0 - \hat{\theta})^2$$

onde J é a informação observada de Fisher definida na Seção 1.4 e podemos desprezar os termos de ordem mais alta já que, sob H_0, θ_0 e $\hat{\theta}$ estão próximos para n grande.

Mas função escore avaliada em $\hat{\theta}$ é igual a zero por definição. Além disso, a razão de máxima verossimilhança neste caso é

$$\lambda(\boldsymbol{X}) = \frac{p(\boldsymbol{X}|\theta_0)}{p(\boldsymbol{X}|\hat{\theta})}$$

e podemos escrever então que

$$-2\log\lambda(\boldsymbol{X}) = -2\log\left(\frac{p(\boldsymbol{X}|\theta_0)}{p(\boldsymbol{X}|\hat{\theta})}\right) = -2[L(\theta_0) - L(\hat{\theta})] \approx J(\hat{\theta})(\theta_0 - \hat{\theta})^2.$$

Lembrando que $\hat{\theta}$ é assintóticamente normal com média θ e usando o fato de que $J(\hat{\theta})/n$ converge quase certamente para o seu valor esperado $I(\theta_0)/n$ quando H_0 é verdadeira então a distribuição assintótica de $-2\log\lambda(\boldsymbol{X})$ é χ^2_1. Assim, um teste com nível de significância assintótico α rejeita H_0 se $-2\log\lambda(\boldsymbol{X}) > c$ onde c é tal que $P(-2\log\lambda(\boldsymbol{X}) > c|\theta = \theta_0) = \alpha$.

Este resultado pode ser generalizado para o caso de um vetor de parâmetros $\boldsymbol{\theta} = (\theta, \dots, \theta_k)$ de dimensão k. Neste caso, a estatística $-2\log\lambda(\boldsymbol{X})$ tem distribuição assintótica χ^2_k.

2.6.1 Teste Qui-quadrado

Um caso de particular interesse em Estatística é quando os dados são tais que cada observação pode ser classificada de acordo com um número finito de possíveis categorias. Por isso, observações deste tipo são chamadas *dados categóricos* e estaremos interessados em fazer inferência sobre as probabilidades de cada categoria.

Suponha que uma população consiste de itens que podem ser classificados em k diferentes categorias. Seja θ_i a probabilidade de que um item selecionado ao acaso pertença à categoria i, $i = 1, \dots, k$. Assumimos também que $\theta_i \geq 0$, $i = 1, \dots, k$ e $\sum_{i=1}^n \theta_i = 1$. Sejam agora os valores específicos $\theta_1^0, \dots, \theta_k^0$ tais que $\theta_i^0 > 0$, $i = 1, \dots, k$ e $\sum_{i=1}^n \theta_i^0 = 1$ e queremos testar as hipóteses

$$\begin{aligned} &H_0 : \theta_i = \theta_i^0, \quad i = 1, \dots, k \\ &H_0 : \theta_i \neq \theta_i^0, \quad \text{para ao menor um valor de } i. \end{aligned} \tag{2.2}$$

Suponha agora que uma amostra aleatória de tamanho n é tomada desta população e as hipóteses (2.2) serão testadas com base nesta amostra. Para isto vamos denotar N_i o número amostral de observações na categoria i, i.e. $N_1, \dots, N_k$ são inteiros não negativos tais que $\sum_{i=1}^k N_i = n$. Quando H_0 é verdadeira, o número esperado de observações do tipo i é $n\theta_i$ e a diferença entre o número observado e o número esperado tende a ser menor quando H_0 é verdadeira do que quando ela é falsa. Parece razoável então basear o teste nas magnitudes relativas destas diferenças. Neste caso, usando-se a função escore pode-se mostrar que o teste assintótico rejeita H_0 se

$$Q = \sum_{i=1}^{k} \frac{(N_i - n\theta_i^0)^2}{n\theta_i^0} > c$$

onde a estatística Q tem distribuição assintótica χ^2_{k-1}. Estes testes também são conhecidos na literatura como *testes de qualidade de ajuste* ou *testes de aderência* e estão entre os mais utilizados em Estatística.

Uma observação de ordem prática é que as frequências esperadas $n\theta_i^0$ não devem ser muito pequenas para que a distribuição χ^2 seja uma boa aproximação da distribuição de Q. Especificamente, pode-se mostrar que a aproximação será muito boa se $n\theta_i^0 \geq 5$ e apenas razoável $n\theta_i^0 \geq 1,5$.

Várias aplicações para dados categóricos e métodos não paramétricos que utilizam testes qui-quadrado podem ser vistas por exemplo em DeGroot (1989).

Apêndice A

Lista de Distribuições

Neste apêndice são listadas as distribuições de probabilidade utilizadas no texto para facilidade de referência. São apresentadas suas funções de (densidade) de probabilidade além da média e variância. Uma revisão exaustiva de distribuições de probabilidades pode ser encontrada em Johnson *et al.* (1992, 1994, 1995).

A.1 Distribuição Normal

X tem distribuição normal com parâmetros μ e σ^2, denotando-se $X \sim N(\mu, \sigma^2)$, se sua função de densidade é dada por

$$p(x|\mu, \sigma^2) = (2\pi\sigma^2)^{-1/2} \exp[-(x-\mu)^2/2\sigma^2], \quad -\infty < x < \infty,$$

para $-\infty < \mu < \infty$ e $\sigma^2 > 0$. Quando $\mu = 0$ e $\sigma^2 = 1$ a distribuição é chamada normal padrão. A distribuição log-normal é definida como a distribuição de e^X.

No caso vetorial, $\boldsymbol{X} = (X_1, \dots, X_p)$ tem distribuição normal multivariada com vetor de médias $\boldsymbol{\mu}$ e matriz de variância-covariância Σ, denotando-se $\boldsymbol{X} \sim N(\boldsymbol{\mu}, \Sigma)$ se sua função de densidade é dada por

$$p(x|\boldsymbol{\mu}, \Sigma) = (2\pi)^{-p/2} |\Sigma|^{-1/2} \exp[-(\boldsymbol{x}-\boldsymbol{\mu})'\Sigma^{-1}(\boldsymbol{x}-\boldsymbol{\mu})/2]$$

para $\boldsymbol{\mu} \in \mathbb{R}^p$ e Σ positiva-definida.

A.2 Distribuição Gama

X tem distribuição Gama com parâmetros α e β, denotando-se $X \sim Ga(\alpha, \beta)$, se sua função de densidade é dada por

$$p(x|\alpha, \beta) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x}, \quad x > 0,$$

para $\alpha, \beta > 0$.

$$E(X) = \alpha/\beta \quad \text{e} \quad V(X) = \alpha/\beta^2.$$

Casos particulares da distribuição Gama são a distribuição de Erlang, $Ga(\alpha, 1)$, a distribuição exponencial, $Ga(1, \beta)$, e a distribuição qui-quadrado com ν graus de liberdade, $Ga(\nu/2, 1/2)$.

A.3 Distribuição Gama Inversa

X tem distribuição Gama Inversa com parâmetros α e β, denotando-se $X \sim GI(\alpha, \beta)$, se sua função de densidade é dada por

$$p(x|\alpha, \beta) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{-(\alpha+1)} e^{-\beta/x}, \quad x > 0,$$

para $\alpha, \beta > 0$.

$$E(X) = \frac{\beta}{\alpha - 1} \quad \text{e} \quad V(X) = \frac{\beta^2}{(\alpha-1)^2(\alpha-2)}.$$

Não é difícil verificar que esta é a distribuição de $1/X$ quando $X \sim Ga(\alpha, \beta)$.

A.4 Distribuição Beta

X tem distribuição Beta com parâmetros α e β, denotando-se $X \sim Be(\alpha, \beta)$, se sua função de densidade é dada por

$$p(x|\alpha, \beta) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(1-x)^{\beta-1}, \quad 0 < x < 1,$$

para $\alpha, \beta > 0$.

$$E(X) = \frac{\alpha}{\alpha+\beta} \quad \text{e} \quad V(X) = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}.$$

A.5 Distribuição de Dirichlet

O vetor aleatório $\boldsymbol{X} = (X_1, \ldots, X_k)$ tem distribuição de Dirichlet com parâmetros $\alpha_1, \ldots, \alpha_k$, denotada por $D_k(\alpha_1, \ldots, \alpha_k)$ se sua função de densidade conjunta é dada por

$$p(\boldsymbol{x}|\alpha_1, \ldots, \alpha_k) = \frac{\Gamma(\alpha_0)}{\Gamma(\alpha_1), \ldots, \Gamma(\alpha_k)} x_1^{\alpha_1 - 1} \ldots x_k^{\alpha_k - 1}, \quad \sum_{i=1}^{k} x_i = 1,$$

para $\alpha_1, \ldots, \alpha_k > 0$ e $\alpha_0 = \sum_{i=1}^{k} \alpha_i$.

$$E(X_i) = \frac{\alpha_i}{\alpha_0}, \quad V(X_i) = \frac{(\alpha_0 - \alpha_i)\alpha_i}{\alpha_0^2(\alpha_0 + 1)}, \quad \text{e} \quad Cov(X_i, X_j) = -\frac{\alpha_i \alpha_j}{\alpha_0^2(\alpha_0 + 1)}$$

Note que a distribuição Beta é obtida como caso particular para $k = 2$.

A.6 Distribuição t de Student

X tem distribuição t de Student (ou simplesmente t) com média μ, parâmetro de escala σ e ν graus de liberdade, denotando-se $X \sim t_\nu(\mu, \sigma^2)$, se sua função de densidade é dada por

$$p(x|\nu, \mu, \sigma^2) = \frac{\Gamma((\nu + 1)/2)\nu^{\nu/2}}{\Gamma(\nu/2)\sqrt{\pi}\sigma} \left[\nu + \frac{(x - \mu)^2}{\sigma^2}\right]^{-(\nu+1)/2}, \quad x \in \mathbb{R},$$

para $\nu > 0$, $\mu \in \mathbb{R}$ e $\sigma^2 > 0$.

$$E(X) = \mu, \quad \text{para } \nu > 1 \quad \text{e} \quad V(X) = \frac{\nu}{\nu - 2}, \quad \text{para } \nu > 2.$$

Um caso particular da distribuição t é a distribuição de Cauchy, denotada por $C(\mu, \sigma^2)$, que corresponde a $\nu = 1$.

A.7 Distribuição F de Fisher

X tem distribuição F com ν_1 e ν_2 graus de liberdade, denotando-se $X \sim F(\nu_1, \nu_2)$, se sua função de densidade é dada por

$$p(x|\nu_1, \nu_2) = \frac{\Gamma((\nu_1 + \nu_2)/2)}{\Gamma(\nu_1/2)\Gamma(\nu_2/2)} \nu_1^{\nu_1/2} \nu_2^{\nu_2/2} x^{\nu_1/2 - 1} (\nu_2 + \nu_1 x)^{-(\nu_1 + \nu_2)/2}$$

$x > 0$, e para $\nu_1, \nu_2 > 0$.

$$E(X) = \frac{\nu_2}{\nu_2 - 2}, \quad \text{para } \nu_2 > 2 \quad \text{e} \quad V(X) = \frac{2\nu_2^2(\nu_1 + \nu_2 - 2)}{\nu_1(\nu_2 - 4)(\nu_2 - 2)^2}, \quad \text{para } \nu_2 > 4.$$

A.8 Distribuição Binomial

X tem distribuição binomial com parâmetros n e p, denotando-se $X \sim bin(n,p)$, se sua função de probabilidade é dada por

$$p(x|n,p) = \binom{n}{x} p^x (1-p)^{n-x}, \quad x = 0,\dots,n$$

para $n \geq 1$ e $0 < p < 1$.

$$E(X) = np \quad \text{e} \quad V(X) = np(1-p)$$

e um caso particular é a distribuição de Bernoulli com $n = 1$.

A.9 Distribuição Multinomial

O vetor aleatório $\boldsymbol{X} = (X_1,\dots,X_k)$ tem distribuição multinomial com parâmetros n e probabilidades $\theta_1,\dots,\theta_k$, denotada por $M_k(n,\theta_1,\dots,\theta_k)$ se sua função de probabilidade conjunta é dada por

$$p(\boldsymbol{x}|\theta_1,\dots,\theta_k) = \frac{n!}{x_1!,\dots,x_k!}\theta_1^{x_1},\dots,\theta_k^{x_k}, \quad x_i = 0,\dots,n, \quad \sum_{i=1}^{k} x_i = n,$$

para $0 < \theta_i < 1$ e $\sum_{i=1}^{k} \theta_i = 1$. Note que a distribuição binomial é um caso especial da multinomial quando $k = 2$. Além disso, a distribuição marginal de cada X_i é binomial com parâmetros n e θ_i e

$$E(X_i) = n\theta_i, \quad V(X_i) = n\theta_i(1-\theta_), \quad \text{e} \quad Cov(X_i, X_j) = -n\theta_i\theta_j.$$

A.10 Distribuição de Poisson

X tem distribuição de Poisson com parâmetro θ, denotando-se $X \sim Poisson(\theta)$, se sua função de probabilidade é dada por

$$p(x|\theta) = \frac{\theta^x e^{-\theta}}{x!}, \quad x = 0,1,\dots$$

para $\theta > 0$.

$$E(X) = V(X) = \theta.$$

A.11 Distribuição Binomial Negativa

X tem distribuição de binomial negativa com parâmetros r e p, denotando-se $X \sim BN(r, p)$, se sua função de probabilidade é dada por

$$p(x|r, p) = \binom{r + x - 1}{x} p^r (1 - p)^x, \quad x = 0, 1, \ldots$$

para $r \geq 1$ e $0 < p < 1$.

$$E(X) = r(1 - p)/p \quad \text{e} \quad V(X) = r(1 - p)/p^2.$$

Referências

DeGroot, M. H. (1989). *Probability and Statistics* (2nd ed.). Addison Wesley.

Gamerman, D. and H. S. Migon (1993). *Inferência Estatística: Uma Abordagem Integrada*. Textos de Métodos Matemáticos. Instituto de Matemática, UFRJ.

Johnson, N. L., S. Kotz, and N. Balakrishnan (1994). *Continuous Univariate Distributions* (2nd ed.), Volume 1. John Wiley, New York.

Johnson, N. L., S. Kotz, and N. Balakrishnan (1995). *Continuous Univariate Distributions* (2nd ed.), Volume 2. John Wiley, New York.

Johnson, N. L., S. Kotz, and A. W. Kemp (1992). *Univariate Discrete Distributions* (2nd ed.). John Wiley, New York.

Lehman, E. (1986). *Testing Statistical Hypothesis* (2nd ed.). Wiley.

Migon, H. S. and D. Gamerman (1999). *Statistical Inference: An Integrated Approach*. Arnold.