Inferência Estatística II

Prof. Fernando de Souza Bastos
fernando.bastos@ufv.br

Departamento de Estatística
Programa de Pós-Graduação em Estatística Aplicada e Biometria
Universidade Federal de Viçosa
Campus UFV - Viçosa

Sumário

Definição Uma variável aleatória $Q(X_1, \ldots, X_n; \theta) = Q(\boldsymbol{X}; \theta)$ é dita ser uma quantidade Pivotal para o parâmetro θ se sua distribuição for independente de θ.

Exemplo 1

Considere $\boldsymbol{X} = (X_1, \ldots, X_n)$ amostra aleatória de $X \sim N(\theta, 1)$, $\theta \in \mathbb{R}$. Segue que, são quantidades pivotais:

$$Q_1(\boldsymbol{X}, \theta) = X_1 - \theta \sim N(0, 1)$$

Exemplo 1

Considere $\boldsymbol{X} = (X_1, \ldots, X_n)$ amostra aleatória de $X \sim N(\theta, 1)$, $\theta \in \mathbb{R}$. Segue que, são quantidades pivotais:

$$Q_1(\boldsymbol{X}, \theta) = X_1 - \theta \sim N(0,1)$$

$$Q_2(\boldsymbol{X}, \theta) = \frac{(\bar{X} - \theta)}{\sqrt{\frac{1}{n}}} = \sqrt{n}(\bar{X} - \theta) \sim N(0,1)$$

Exemplo 1

Considere $\boldsymbol{X} = (X_1, \ldots, X_n)$ amostra aleatória de $X \sim N(\theta, 1)$, $\theta \in \mathbb{R}$. Segue que, são quantidades pivotais:

$$Q_1(\boldsymbol{X}, \theta) = X_1 - \theta \sim N(0, 1)$$

$$Q_2(\boldsymbol{X}, \theta) = \frac{(\bar{X} - \theta)}{\sqrt{\frac{1}{n}}} = \sqrt{n}(\bar{X} - \theta) \sim N(0, 1)$$

$$Q_3(\boldsymbol{X}, \theta) = \sqrt{n}\frac{(\bar{X} - \theta)}{S^2_{n-1}} \sim t_{n-1}$$

Exemplo 2

Sejam $X_1, \ldots, X_n$ uma amostra aleatória da distribuição da variável aleatória X, com densidade

$$f(x|\theta) = \theta e^{-\theta x}, \quad \theta > 0, \ x > 0. \tag{1}$$

Como vimos anteriormente, a estatística $T = \sum_{i=1}^{n} X_i$ é suficiente para θ. Mas, como a distribuição de T é Gama(n, θ), temos que T não é uma quantidade pivotal para θ.

Por outro lado, a densidade de $Q(X;\theta) = 2\theta\sum_{i=1}^{n} X_i$ é dada por

$$f_Q(y) = \frac{y^{n-1}e^{-y/2}}{2^n\Gamma[n]}, \quad y > 0, \tag{2}$$

que corresponde à densidade de uma distribuição qui-quadrado com $2n$ graus de liberdade, denotada por χ^2_{2n}. Portanto, $Q(X;\theta)$ pode ser considerada como uma quantidade pivotal, pois sua distribuição é independente de θ.

Exemplo 3

$\boldsymbol{X} = (X_1, \ldots, X_n)$ amostra aleatória de $X \sim N(\mu, \sigma^2)$ em que $\theta = (\mu, \sigma^2)$.

$$Q(\mu, \boldsymbol{X}) = \sqrt{n}\frac{\bar{X} - \mu}{\sqrt{S^2_{n-1}}} \sim t_{(n-1)} \text{ é uma quantidade pivotal para } \mu$$

$$Q(\sigma^2, \boldsymbol{X}) = (n-1)\frac{S^2_{n-1}}{\sigma^2} \sim \chi^2_{(n-1)} \text{ é uma quantidade pivotal para } \sigma^2$$

Exemplo de Quantidade Pivotal Assintótica

Considere $\boldsymbol{X} = (X_1, \cdots, X_n)$ amostra aleatória de $X \sim \text{Bernoulli}(\theta)$, $\theta \in (0,1)$. Notem que $Q(\boldsymbol{X}, \theta) = \sqrt{n}\dfrac{(\bar{X} - \theta)}{\sqrt{\bar{X}(1 - \bar{X})}}$ é uma quantidade pivotal assintótica, pois, pelo TLC temos que,

$$\sqrt{n}(\bar{X} - \theta) \xrightarrow{D} N(0, \theta(1-\theta)) \Rightarrow \frac{\sqrt{n}(\bar{X} - \theta)}{\sqrt{\theta(1-\theta)}} \xrightarrow{D} N(0,1)$$

Lembrem-se que $\forall\theta \in (0,1)$ $\bar{X} \xrightarrow{P} \theta \Rightarrow \bar{X}(1-\bar{X}) \xrightarrow{P} \theta(1-\theta)$

Lembrem-se que $\forall\theta \in (0,1)$ $\bar{X} \xrightarrow{P} \theta \Rightarrow \bar{X}(1-\bar{X}) \xrightarrow{P} \theta(1-\theta)$

$$\Rightarrow \ \forall\theta \in (0,1) \ \frac{\bar{X}(1-\bar{X})}{\theta(1-\theta)} \xrightarrow{P} 1 \ \Rightarrow \sqrt{\frac{\bar{X}(1-\bar{X})}{\theta(1-\theta)}} \xrightarrow{P} 1$$

Lembrem-se que $\forall \theta \in (0,1)$ $\bar{X} \xrightarrow{P} \theta \Rightarrow \bar{X}(1-\bar{X}) \xrightarrow{P} \theta(1-\theta)$

$$\Rightarrow \ \forall \theta \in (0,1) \ \frac{\bar{X}(1-\bar{X})}{\theta(1-\theta)} \xrightarrow{P} 1 \ \Rightarrow \sqrt{\frac{\bar{X}(1-\bar{X})}{\theta(1-\theta)}} \xrightarrow{P} 1$$

Pelo teorema de Slutsky,

$$\sqrt{n}\frac{(\bar{X}-\theta)}{\sqrt{\cancel{\theta(1-\theta)}}}\frac{1}{\sqrt{\frac{\bar{X}(1-\bar{X})}{\cancel{\theta(1-\theta)}}}} \xrightarrow{D} N(0,1)$$

Intervalos de Confiança com o Uso de Quantidades Pivotais

Notemos que uma quantidade Pivotal não é uma estatística, pois ela depende de um parâmetro θ desconhecido. Podemos, então, para cada $\gamma = 1 - \alpha$ fixado, encontrar λ_1 e λ_2 na distribuição de $Q(\boldsymbol{X};\theta)$ de modo que

$$P[\lambda_1 \leq Q(\boldsymbol{X};\theta) \leq \lambda_2] = \gamma. \quad (3)$$

Sendo a distribuição de $Q(\boldsymbol{X};\theta)$ independente de θ, λ_1 e λ_2 também não dependem de θ. Além disso, se para cada $\boldsymbol{X}$ existirem estatísticas $t_1(\boldsymbol{X})$ e $t_2(\boldsymbol{X})$ tais que $\lambda_1 \leq Q(\boldsymbol{X};\theta) \leq \lambda_2$ se e somente se $t_1(\boldsymbol{X}) \leq \theta \leq t_2(\boldsymbol{X})$, então,

$$P[t_1(\boldsymbol{X}) \leq \theta \leq t_2(\boldsymbol{X})] = \gamma \tag{4}$$

de modo que $[t_1(\boldsymbol{X}); t_2(\boldsymbol{X})]$ é um intervalo (aleatório) que contém θ com probabilidade (coeficiente de confiança) $\gamma = 1 - \alpha$.

Nos casos em que a distribuição da variável aleatória X é discreta, em geral, não se consegue determinar λ_1 e λ_2 de tal forma que (3) esteja satisfeita exatamente. Em tais casos, podemos escolher λ_1 e λ_2 tal que (3) esteja satisfeita para um coeficiente de confiança maior ou igual a γ (o mais próximo possível).

Quando n é razoavelmente grande, uma alternativa seria considerar os intervalos de confiança baseados na distribuição do estimador de máxima verossimilhança. Outro ponto a salientar é que, na maioria dos casos, existem muitos pares (λ_1, λ_2) satisfazendo (3). Sempre que possível, devemos escolher (λ_1, λ_2) que produz o intervalo de menor comprimento. Tal procedimento é facilitado em situações em que a distribuição de $Q(\boldsymbol{X}; \theta)$ é simétrica, como no caso da distribuição normal.

Exemplo 1 - Continuação do Exemplo 2 da seção anterior

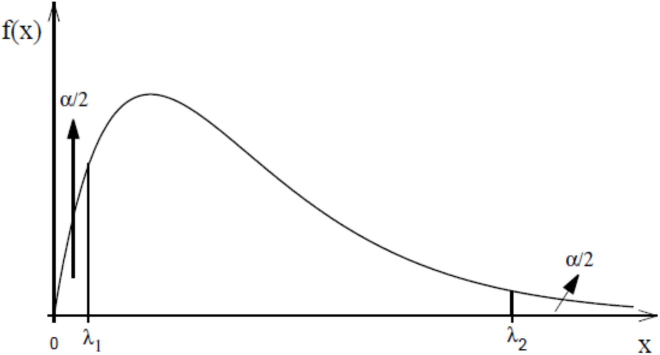
No Exemplo 2 da página 7 dado o coeficiente de confiança $\gamma = 1 - \alpha$, obtemos λ_1 e λ_2 na tabela da distribuição χ^2_{2n}, de modo que

$$P\left(\lambda_1 \leq 2\theta \sum_{i=1}^{n} X_i \leq \lambda_2\right) = \gamma, \tag{5}$$

Logo, um intervalo de confiança para θ com coeficiente de confiança γ é dado por

$$\left[\frac{\lambda_1}{2\sum_{i=1}^{n} X_i}; \frac{\lambda_2}{2\sum_{i=1}^{n} X_i}\right]. \tag{6}$$

existem infinitos pares (λ_1, λ_2) para os quais (5) está verificada. Sempre que possível, (λ_1, λ_2) devem ser escolhidos de modo que o intervalo (6) seja de comprimento mínimo. Tal intervalo existe, mas (λ_1, λ_2) deve ser obtido por métodos computacionais. Uma alternativa é considerarmos intervalos simétricos em que (λ_1, λ_2) são obtidos a partir da distribuição χ^2_{2n}, de modo que a área à esquerda de λ_1 seja igual à área à direita de λ_2 e igual a $\alpha/2$.



Exemplo 2

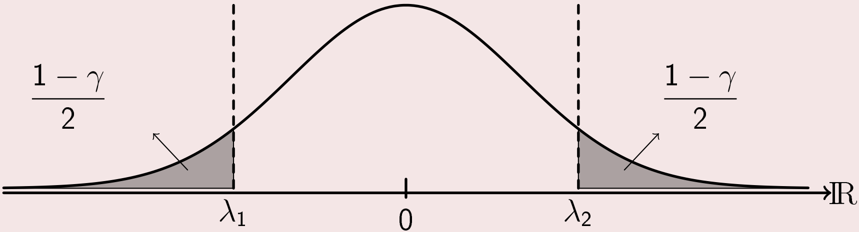
Seja $\boldsymbol{X} = X_1, \ldots, X_n$ amostra aleatória de $X \sim N(\theta, 1), \theta \in \mathbb{R}$.

$$Q(\boldsymbol{X}; \theta) = \sqrt{n}\frac{(\bar{X} - \theta)}{\sqrt{S_{n-1}^2}} \sim t_{(n-1)}$$

Exemplo 2

Seja $\boldsymbol{X} = X_1, \ldots, X_n$ amostra aleatória de $X \sim N(\theta, 1), \theta \in \mathbb{R}$.

$$Q(\boldsymbol{X}; \theta) = \sqrt{n}\frac{(\bar{X} - \theta)}{\sqrt{S_{n-1}^2}} \sim t_{(n-1)}$$



Notem que

$$\lambda_1 \leq \sqrt{n}\frac{(\bar{X} - \theta)}{\sqrt{S^2_{n-1}}} \leq \lambda_2 \iff \bar{X} - \lambda_2\sqrt{\frac{S^2_{n-1}}{n}} \leq \theta \leq \bar{X} + \lambda_1\sqrt{\frac{S^2_{n-1}}{n}}$$

Notem que

$$\lambda_1 \leq \sqrt{n}\frac{(\bar{X} - \theta)}{\sqrt{S_{n-1}^2}} \leq \lambda_2 \iff \bar{X} - \lambda_2\sqrt{\frac{S_{n-1}^2}{n}} \leq \theta \leq \bar{X} + \lambda_1\sqrt{\frac{S_{n-1}^2}{n}}$$

$$\Rightarrow IC(\theta, \gamma) = \left(\bar{X} - \lambda_2\sqrt{\frac{S_{n-1}^2}{n}}, \bar{X} + \lambda_1\sqrt{\frac{S_{n-1}^2}{n}}\right)$$

Notem que

$$\lambda_1 \leq \sqrt{n}\frac{(\bar{X}-\theta)}{\sqrt{S^2_{n-1}}} \leq \lambda_2 \iff \bar{X} - \lambda_2\sqrt{\frac{S^2_{n-1}}{n}} \leq \theta \leq \bar{X} + \lambda_1\sqrt{\frac{S^2_{n-1}}{n}}$$

$$\Rightarrow IC(\theta,\gamma) = \left(\bar{X} - \lambda_2\sqrt{\frac{S^2_{n-1}}{n}}, \bar{X} + \lambda_1\sqrt{\frac{S^2_{n-1}}{n}}\right)$$

Note que este é um intervalo de confiança exato! Basta encontrar λ_1 e λ_2 a partir da distribuição t com $n-1$ graus de liberdade.

Considere $\boldsymbol{X} = (X_1, \cdots, X_n)$ amostra aleatória de $X \sim \text{Bernoulli}(\theta)$, $\theta \in (0,1)$. Mostramos anteriormente que $Q(\boldsymbol{X}, \theta) = \sqrt{n}\dfrac{(\bar{X} - \theta)}{\sqrt{\bar{X}(1-\bar{X})}} \xrightarrow{D} N(0,1), \forall\theta \in (0,1)$, logo

Considere $\boldsymbol{X} = (X_1, \cdots, X_n)$ amostra aleatória de $X \sim \text{Bernoulli}(\theta)$, $\theta \in (0,1)$. Mostramos anteriormente que $Q(\boldsymbol{X}, \theta) = \sqrt{n}\dfrac{(\bar{X} - \theta)}{\sqrt{\bar{X}(1-\bar{X})}} \xrightarrow{D} N(0,1), \forall\theta \in (0,1)$, logo

$$\bar{X} - \lambda_2\sqrt{\frac{\bar{X}(1-\bar{X})}{n}} \leq \theta \leq \bar{X} + \lambda_1\sqrt{\frac{\bar{X}(1-\bar{X})}{n}}$$

$$\Rightarrow IC(\theta, \gamma) = \left(\bar{X} - \lambda_1\sqrt{\frac{\bar{X}(1-\bar{X})}{n}}, \bar{X} + \lambda_1\sqrt{\frac{\bar{X}(1-\bar{X})}{n}}\right) \cap [0,1],$$

em que $\lambda_1 = -\lambda_2, \ \forall\gamma \in [0,1]$

Considere $\boldsymbol{X} = (X_1, \cdots, X_n)$ amostra aleatória de $X \sim \text{Bernoulli}(\theta)$, $\theta \in (0,1)$. Mostramos anteriormente que $Q(\boldsymbol{X}, \theta) = \sqrt{n}\dfrac{(\bar{X} - \theta)}{\sqrt{\bar{X}(1-\bar{X})}} \xrightarrow{D} N(0,1), \forall \theta \in (0,1)$, logo

$$\bar{X} - \lambda_2\sqrt{\frac{\bar{X}(1-\bar{X})}{n}} \leq \theta \leq \bar{X} + \lambda_1\sqrt{\frac{\bar{X}(1-\bar{X})}{n}}$$

$$\Rightarrow IC(\theta, \gamma) = \left(\bar{X} - \lambda_1\sqrt{\frac{\bar{X}(1-\bar{X})}{n}}, \bar{X} + \lambda_1\sqrt{\frac{\bar{X}(1-\bar{X})}{n}}\right) \cap [0,1],$$

em que $\lambda_1 = -\lambda_2$, $\forall \gamma \in [0,1]$

Nesse caso temos um intervalo de confiança aproximado, pois $\bar{X} \xrightarrow{P} \theta$.

Exemplo 3

Sejam $X_1, \ldots, X_n$ uma amostra aleatória de tamanho n da distribuição $N(\mu, \sigma^2)$. Assumindo σ^2 conhecido, temos que uma quantidade pivotal para μ, baseada na estatística suficiente

$$\frac{1}{n}\sum_{i=1}^{n} X_i = \bar{X},$$

é dada por

$$Q(X; \mu) = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}},$$

que tem distribuição $N(0, 1)$.

Por outro lado, sendo σ^2 desconhecido, temos que

$$Q(X,\mu) = \frac{\bar{X} - \mu}{S/\sqrt{n}} \sim t_{n-1},$$

que, nesse caso, é uma quantidade pivotal. Então, dado γ, existem λ_1 e λ_2 na distribuição t_{n-1} de modo que

$$P\left(\lambda_1 \leq \frac{\bar{X} - \mu}{S/\sqrt{n}} \leq \lambda_2\right) = \gamma.$$

Quanto a σ^2, considerando μ desconhecido, temos que

$$Q(X, \sigma^2) = \frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{n-1}$$

é uma quantidade pivotal para σ^2. Portanto, dado γ, podemos determinar λ_1 e λ_2 de modo que

$$P\left(\lambda_1 \leq \frac{(n-1)S^2}{\sigma^2} \leq \lambda_2\right) = \gamma.$$

Considerando o intervalo simétrico, ou seja, $\lambda_1 = q_1$ e $\lambda_2 = q_2$, em que $P[\chi^2_{n-1} \geq q_2] = P[\chi^2_{n-1} \leq q_1] = \alpha/2$, temos o intervalo:

$$\left(\frac{(n-1)S^2}{q_2}; \frac{(n-1)S^2}{q_1}\right).$$

Intervalos de Confiança Aproximados

Já vimos que, se $\hat{\theta}$ é o EMV de θ, então

$$\frac{\hat{\theta} - \theta}{\sqrt{(nI(\theta))^{-1}}} \xrightarrow{D} N(0, 1)$$

Intervalos de Confiança Aproximados

Já vimos que, se $\hat{\theta}$ é o EMV de θ, então

$$\frac{\hat{\theta} - \theta}{\sqrt{(nI(\theta))^{-1}}} \xrightarrow{D} N(0, 1)$$

Como $I(\theta)$ pode depender de θ, que não é conhecido, substituindo $I(\theta)$ por $I(\hat{\theta})$, temos que

$$Q(X, \theta) = \frac{\hat{\theta} - \theta}{\sqrt{(nI(\hat{\theta}))^{-1}}} \sim N(0, 1),$$

de modo que $Q(X, \theta)$ é uma quantidade pivotal com distribuição aproximadamente igual à distribuição $N(0, 1)$ em grandes amostras.

Com relação a uma função $g(\theta)$, podemos considerar a variável aleatória

$$Q(X, g(\theta)) = \frac{g(\hat{\theta}) - g(\theta)}{\sqrt{\frac{(g'(\hat{\theta}))^2}{nI(\hat{\theta})}}} \sim N(0, 1),$$

que, para amostras grandes, é uma quantidade pivotal.

Exemplo 1

Sejam $X_1, \ldots, X_n$ uma amostra aleatória da variável aleatória $X \sim$ Bernoulli(θ). Como o estimador de máxima verossimilhança de θ é $\hat{\theta} = \bar{X}$ e $I(\theta) = \frac{1}{\theta(1-\theta)}$, temos que uma quantidade pivotal para θ é dada por

$$Q(X, \theta) = \frac{\bar{X} - \theta}{\sqrt{\frac{\bar{X}(1-\bar{X})}{n}}} \sim N(0, 1),$$

de modo que para valores grandes de n, um intervalo de confiança para θ com coeficiente de confiança aproximadamente γ é dado por

$$\left[\bar{X} - z_{\frac{\alpha}{2}}\sqrt{\frac{\bar{X}(1-\bar{X})}{n}}; \bar{X} + z_{\frac{\alpha}{2}}\sqrt{\frac{\bar{X}(1-\bar{X})}{n}}\right].$$

Suponhamos agora que seja de interesse a obtenção de um intervalo de confiança para $g(\theta) = \theta(1-\theta)$. Como $g'(\theta) = 1 - 2\theta$ e $I(\theta) = \frac{1}{\theta(1-\theta)}$, temos que uma quantidade pivotal para $g(\theta)$ é dada por

$$Q(X, \theta) = \frac{\hat{\theta}(1-\hat{\theta}) - \theta(1-\theta)}{\sqrt{\frac{\hat{\theta}(1-\hat{\theta})(1-2\hat{\theta})^2}{n}}} \sim N(0,1),$$

de modo que um intervalo de confiança aproximado para $g(\theta) = \theta(1-\theta)$ é dado por

$$\left[\bar{X}(1-\bar{X}) - z_{\frac{\alpha}{2}}\sqrt{\frac{\bar{X}(1-\bar{X})(1-2\bar{X})^2}{n}}; \bar{X}(1-\bar{X}) + z_{\frac{\alpha}{2}}\sqrt{\frac{\bar{X}(1-\bar{X})(1-2\bar{X})^2}{n}}\right],$$

em que $z_{\frac{\alpha}{2}}$ é obtido na tabela da distribuição $N(0,1)$.

Exemplo 2

Sejam $X_1, \ldots, X_n$ uma amostra aleatória de tamanho n da variável aleatória $X \sim \text{Exp}(\theta)$, com função densidade

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

Como $I^{-1}(\theta) = \theta^2$ e $\hat{\theta} = \dfrac{1}{\bar{X}}$, segue que uma quantidade pivotal para θ é dada por

$$Q(X, \theta) = \frac{\dfrac{1}{\bar{X}} - \theta}{\sqrt{\dfrac{\hat{\theta}^2}{n}}} \sim N(0, 1),$$

de modo que um intervalo de confiança com coeficiente de confiança aproximado $\gamma = 1 - \alpha$ é dado por

$$\left(\frac{1}{\bar{X}} - z_{\frac{\alpha}{2}} \sqrt{\frac{1}{n\bar{X}^2}}; \frac{1}{\bar{X}} + z_{\frac{\alpha}{2}} \sqrt{\frac{1}{n\bar{X}^2}} \right).$$

Referências I

Bolfarine, Heleno e Mônica Carneiro Sandoval (2001). *Introdução à inferência estatística*. Vol. 2. SBM.