Variância e Valor Esperado

$$Var(X) = E[(X - EX)^2] = E(X^2) + E(X)^2$$

$$Var(aX + b) = a^2 Var(X)$$

$$E[aX + b] = aE[X] + b$$

$$E[x] = \int_{\infty}^{\infty} xf(x)dx$$

Convergência

Definição: Dizemos que uma sequência de variáveis aleatórias Z_n converge para b, se para todo $\epsilon > 0$, temos que:

$$\lim \ Pr(|Z_n - b| < \epsilon) = 1$$

Teorema de Bayes e Probabilidade Condicional

Probabilidade Condicional de B dado A:

$$Pr(B|A) = Pr(A \cap B)Pr(B)$$

$$Pr(B|A) = \frac{Pr(A|B)Pr(B)}{Pr(A|B)Pr(B) + Pr(A|\bar{B})Pr(\bar{B})}$$

Teorema Central do Limite (TCL)

Para qualquer distribuição de X, com média μ e variância σ^2, então para $n \to \infty$:

$$(\bar{X_n} - \mu)/\sigma_{\bar{X}} \to N(0, 1)$$

$N(0, 1)$: Normal padronizada ou Distribuição Normal Padrão.

$$\bar{X_n} \tilde{} N(\mu, \sigma^2/n)$$

Ps:. X pode ser qualquer variável aleatória, seja ela discreta ou contínua.

Lei dos Grandes Números (LGN)

Sejam $(X_1, X_2, \ldots, X_n)$ variáveis aleatórias independentes e identicamente distribuidas com média μ, então:

$$\bar{X_n} = \frac{1}{n} \sum x_i \to \mu$$

$$Pr(\lim_{n \to \infty} \bar{X_n} = \mu) = 1$$

Inferência Frequentista

Em uma abordagem frequencista à inferência, parâmetros desconhecidos são frequentemente, mas nem sempre tratados como se tivessem valores fixos, mas desconhecidos que não podem ser tratados como variados aleatórios em qualquer sentido e assim não há como associar probabilidades a eles.

Inferência Bayesiana

Trata tudo como variável aleatória:

$$\xi(\theta|\bar{x}) = \frac{\xi(\theta) f_n(\bar{x}|\theta)}{f(\bar{x})} = \frac{\xi(\theta) \prod_{i=1}^{n} f(x_i|\theta)}{\int \xi(\theta) \prod_{i=1}^{n} f(x_i|\theta) d\theta},\ \bar{x}(x_1, x_2, \ldots, x_n)$$

Com variáveis discretas:

$$\xi(\theta|\bar{x}) = \frac{\xi(\theta) f_n(\bar{x}|\theta)}{\sum_\theta \xi(\theta) \prod_i f_n(\bar{x}|\theta)}$$

Priori e Posteriori: $\xi(\theta)$ & $\xi(\theta|x)$

$$\xi(\theta|x) = \frac{\xi(\theta) \prod_{i=1}^{n} f_n(x_i|\theta)}{g_n(x)}$$

Função de Verossimilhança: $f(x_1, x_2, \ldots, x_n|\theta)$

$$\xi(\theta|x) \propto f(x|\theta)\xi(\theta)$$

Distribuição Uniforme

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

$$E[x] = \frac{a+b}{2}$$

$$Var[x] = \frac{(b-a)^2}{12}$$

$$f_n(x|\theta_1, \theta_2) = \frac{1}{(\theta_2 - \theta_1)^n}$$

Distribuição Binomial

k = número de sucessos

$$f(x|\theta, k) = \binom{n}{k} \theta^k (1-\theta)^{n-k}$$

$$E[x] = p \leftrightarrow \theta$$

$$E[x^2] = n^2p^2 + np(1-p) \leftrightarrow n^2\theta^2 + n\theta(1-\theta)$$

$$Var[x] = np(1-p) \leftrightarrow n\theta(1-\theta)$$

$$f_n(x|\theta) = \prod_{i=1}^{n} \binom{n}{x_i} \theta^y (1-\theta)^{n-y},\ y = \sum_{1}^{n} x_i$$

Distribuição Geométrica

k = fracasso

$$f(x|\theta) = (1-\theta)^k\theta$$

$$E[x] = \frac{1-p}{p} \leftrightarrow \frac{1-\theta}{\theta}$$

$$Var[x] = \frac{1-p}{p^2} \leftrightarrow \frac{1-\theta}{\theta^2}$$

Distribuição de Bernoulli

$$f(x|\theta) = \theta^x(1-\theta)^{1-x},\ x \in \{0,1\}$$

$$E[x] = p \leftrightarrow \theta$$

$$Var[x] = p(1-p) \leftrightarrow \theta(1-\theta)$$

$$f_n(x|\theta) = \theta^y(1-\theta)^{n-y},\ y = \sum_1^n x_i$$

Distribuição Exponencial

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

$$E[x] = \frac{1}{\lambda} \leftrightarrow \frac{1}{\theta}$$

$$Var[x] = \frac{1}{p^2} \leftrightarrow \frac{1}{\theta^2}$$

$$f_n(x|\theta) = \theta^n e^{-\theta y},\ y = \sum_1^n x_i$$

Distribuição de Poisson

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

$$E[x] = \lambda \leftrightarrow \theta$$

$$Var[x] = \lambda \leftrightarrow \theta$$

$$f_n(x|\theta) = \frac{e^{-n\theta}\theta^y}{\prod x_i!},\ y = \sum_1^n x_i$$

Distribuição Normal(μ, σ^2)

$$f(x|\mu,\sigma^2) = \frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

$$E[x] = \mu$$

$$Var(x) = \sigma^2$$

$$Z = \frac{X - \mu}{\sigma}$$

$Z \sim (0, 1)$

$$Pr(Z \leq c) = \Phi(c)$$

Seja $X_1, X_2, \ldots X_i$ i.i.d onde $X_i \sim (\mu, \sigma^2)$
Então $\bar{X_n} = \frac{1}{n} \sum_{i=1}^{n} X_i$ tem distribuição:

$$E\left(\frac{1}{n} \sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} \frac{1}{n} E(X_i) = n \frac{1}{n} \mu = \mu$$

$$Var\left(\sum_{i=1}^{n} \frac{1}{n} X_i\right) = \sum_{i=1}^{n} \left(\frac{1}{n}\right)^2 Var(X_i) = \sum_{i=1}^{n} \left(\frac{1}{n}\right)^2 \sigma^2 = n\left(\frac{1}{n}\right)^2 \sigma^2 = \frac{\sigma^2}{n}$$

$$\bar{X_n} \sim \left(\mu, \frac{\sigma^2}{n}\right)$$

Distribuição Gamma(α,β)

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

$$\xi(\theta) \propto \theta^{\alpha-1} e^{-\beta\theta}$$

$$E[x] = média = \frac{\alpha}{\beta}$$

Distribuição Gamma-Inversa(α, β)

$$f(x|\alpha, \beta) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{-\alpha-1} e^{-\frac{\beta}{x}}$$

Distribuição Beta(α,β)

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

$$\xi(\theta) \propto \theta^{\alpha-1} (1-\theta)^{\beta-1}$$

$$E[x] = média = \frac{\alpha}{\alpha+\beta}$$

Funções de Perda: $L(\theta, \delta)$

Perda Quadrática:

$$L(\theta, \delta) = (\delta - \theta)^2$$

Perda Absoluta:

$$L(\theta, \delta) = |\delta - \theta|$$

Estimador de Bayes: $\theta_{bayes} = E[\theta|\bar{x}]$

Definição: Estimador que minimiza a função de perca a posteriori.

$$E[L(\theta, \delta(\bar{x}))|\bar{x}] = \int_{\Omega} L(\theta, \delta)\xi(\theta, x)$$

- Na perda quadrátca: θ_{bayes} = média da posteriori
- Na perda absoluta: θ_{bayes} = mediana da posteriori

Exemplos:

$Gamma(\alpha, \beta)$:

$$\delta^* = \frac{\alpha + n\bar{x}}{\beta + n} = \frac{\alpha + y}{\beta + n},\ y = \sum_{1}^{n} x_i$$

$Beta(\alpha, \beta)$:

$$\delta^* = \frac{\alpha}{\alpha + \beta}$$

Se $x_1, \ldots, x_n$ é uma distribuição de poisson $x_i|\theta \sim P(\theta)$ e a priori for uma distribuição Gamma $\theta \sim G(\alpha, \beta)$ então a posteriori também é uma distribuição Gamma.
Se $x_1, \ldots, x_n$ é uma distribuição uniforme $x_i|\theta \sim U(0, \theta)$ e a priori for uma distribuição de Pareto $\theta \sim Pa(\theta_0, a)$ então a posteriori também é uma distribuição de Pareto.

Estimador Consistente

Definição: Um estimador δ é dito consistente quando: $\delta_n \to \theta$, para $n \to \infty$

EMV ou MLE: θ_{EMV}

Definição: O estimador de máxima verossimilhança (EMV) de uma amostra é aquele que maximiza a função de verossimilhança $L(\theta) = f_n(\bar{x}|\theta)$.

$$\hat{\theta} = argmax\ L_x(\theta)$$

Ps:. Maximizar $L(\theta)$ é o mesmo que maximizar $logL(\theta)$: "deriva e iguala a 0".

$$(log\ L(\theta))' = 0$$

Ps:. Maximizar $L(\theta)$ também é o mesmo que minimizar $-L(\theta)$.

Exemplos:

Uniforme:

$$\hat{\theta_2} = max\{x_1, x_2, \ldots, x_n\}$$

$$\hat{\theta_1} = min\{x_1, x_2, \ldots, x_n\}$$

Quando $\theta_1 = 0$: intervalo $[0, \theta]$

$$\hat{\theta} = max(X_1, X_2, \ldots, X_n)$$

Prova:

$$L(\theta) = f_n(\bar{x}|\theta_1, \theta_2) = \frac{1}{(\theta_2 - \theta_1)^n}$$

$L(\theta)$ é maximizado quando $(\theta_2 - \theta_1)$ é mínimo.
$\hat{\theta_1} = max\{x_1, x_2, \ldots, x_n\}$
$\hat{\theta_2} = min\{x_1, x_2, \ldots, x_n\}$

Quando $\theta_1 = 0$:

$$L(\theta) = f_n(\bar{x}|\theta) = \frac{1}{(\theta)^n \ \mathbb{I}(\{x_1, x_2, \ldots, x_n\} \leq \theta)}$$

$L(\theta)$ é maximizado quando θ é máximo.

Binomial:

$$\hat{\theta} = \bar{x_n}$$

Prova:

$$L(\theta) = f_n(\bar{x}|\theta) = \prod_{i=1}^{n} \binom{n}{x_i} \theta^y (1-\theta)^{n-y}, \ y = \sum_{1}^{n} x_i$$

$$log \ L(\theta) = \sum_{i=1}^{n} \log\left(\binom{n}{x_i}\right) + y \log(\theta) + (n - \theta)\log(1-\theta)$$

$$(log \ L(\theta))' = 0 \leftrightarrow= \frac{y}{\theta} - (n-y)\frac{1}{1-\theta} = 0$$

$$y(1-\theta) = \theta(n-y) \leftrightarrow y - \cancel{\theta y} = \theta n - \cancel{\theta y}$$

$$y = \theta n \leftrightarrow \theta = \frac{y}{n} = \frac{\sum_1^n x_i}{n} = \bar{x_n}$$

Exponencial:

$$\hat{\theta} = \frac{1}{\bar{x_n}}$$

Prova:

$$L(\theta) = f_n(\bar{x}|\theta) = \theta^n e^{-\theta y}, \ y = \sum_{1}^{n} x_i$$

$$log \ L(\theta) = n \ log(\theta) - \theta y = n \ log(\theta) - \theta \sum_{1}^{n} x_i$$

$$(log \ L(\theta))' = 0 \leftrightarrow \frac{n}{\theta} - \sum_{1}^{n} x_i = 0$$

$$\frac{n}{\theta} = \sum_{1}^{n} x_i \leftrightarrow \theta = \frac{n}{\sum_1^n x_i} = \frac{1}{\bar{x_n}}$$

Bernoulli:

$$\hat{\theta} = \bar{x_n}$$

Prova:

$$L(\theta) = f_n(\bar{x}|\theta) = \theta^y(1-\theta)^{n-y},\ y = \sum_1^n x_i$$

$$log\ L(\theta) = y\ log(\theta) + (n-y)(log(1-\theta))$$

$$(log\ L(\theta))' = 0 \leftrightarrow \frac{y}{\theta} - \frac{n-y}{(1-\theta)} = 0$$

$$\frac{y}{\theta} = \frac{n-y}{1-\theta} \leftrightarrow y(1-\theta) = \theta(n-y)$$

$$y - \cancel{\theta y} = n\theta\ -\ \cancel{\theta y}$$

$$\theta = \frac{y}{n} = \frac{\sum_1^n x_i}{n} = \bar{x_n}$$

Poisson:

$$\hat{\theta} = \bar{x_n}$$

Prova:

$$L(\theta) = f_n(\bar{x}|\theta) = \frac{e^{-n\theta}\theta^y}{\prod x_i!},\ y = \sum_1^n x_i$$

$$log\ (L(\theta)) = y\ log(\theta) - n\theta - \log(\prod x_i!)$$

$$(log\ L(\theta))' = 0 \leftrightarrow \frac{y}{\theta} - n = 0$$

$$\frac{y}{\theta} = n \leftrightarrow \frac{\sum_1^n x_i}{n} = n \leftrightarrow \theta = \bar{x_n}$$

Normal:

- $\hat{\theta} = \{\hat{\mu}, \hat{\sigma^2}\}$:

$$\hat{\mu} = \bar{X_n}$$

$$\hat{\sigma^2} = \frac{1}{n}\sum_{i=1}^n (X_i - \mu)^2$$

Risco e Viés

$$R(\theta, \delta) = E_\theta[(\delta(\bar{X}) - \theta)^2]$$

Um estimador bom, terá um risco baixo.

$$R(\theta, \delta) = var_\theta(\delta(\bar{X})) + viés_\theta(\bar{\delta})$$

$$viés(\hat{\theta}) = E(\hat{\theta}) - g(\theta)$$

Estimador θ não envieasado:
$viés(\hat{\theta}) = 0 \rightarrow E(\hat{\theta}) = \theta$
$viés(\hat{\theta}) \rightarrow R(\theta, \delta) = var_{\theta}(\delta(\bar{X}))$

- Lembrete: Nem sempre um estimador não-viesado existe.
- Lembrete 2: Nem sempre um estimador não-viesado é um bom estimador.

Método dos Momentos (MM)

Definição: Estimador que usa os momentos, dependendo da distribuição são usados 1 ou 2 momentos. Caso haja mais parâmetros, os próximos momentos serão informados.

Primeiro Momento:

- $\mu_1(\theta) = E[x] = \mu = média$
- $m_1 = \frac{1}{n}\sum_{i=1}^{n} x_i = \bar{x_n}$

Segundo Momento:

- $\mu_2(\theta) = E[x^2]$
- $m_2 = \frac{1}{n}\sum_{i=1}^{n} x_i^2$

Outros momentos para n parâmetros:

- $\mu_n(\theta) = E[x^n]$
- $m_n = \frac{1}{n}\sum_{i=1}^{n} x_i^n$

Resolver o sistema de n equações:

- $\mu_1(\theta) = m_1$
- $\mu_2(\theta) = m_2$
  .
  .
  .
- $\mu_n(\theta) = m_n$

Suficiência

Definição: Dizemos que uma estatística $T(\theta)$ é *suficiente* para calcular a verossimilhança.
Então também é suficiente para calcular qualquer inferência que dependa apenas dos dados da função de verossimilhança, como EMV e qualquer coisa baseada em distribuição a posteriori.

Em outras palavras, a distribuição condicional da amostra dado o valor da estatística não depende de θ.

$$f(X_1, X_2, \dots, X_n | T = t, \theta) = f(X_1, X_2, \dots, X_n |\ T = t,\ \theta'),\ \forall\theta,\ \theta' \in \Omega$$

Para cada t, a distribuição condicional de $X_1, X_2, \dots, X_n$ dado T = t e θ é a mesma para todos os θ.
Então T é uma estatística suficiente para θ.

Suficiência Conjunta

Definição: Quando a estatística suficiente não é representada por um único valor, e sim um vetor de valores. $T = t_1, \ldots$ e $T_k = t_k$ não depende de θ.

- Vale o teorema da fatorização.
- Exemplos: Normal e Uniforme.

Suficiência Mínima

Definição: Uma estatística T é dita *mínima suficiente* se T é suficiente **e** é função de qualquer outra estatística suficiente. Analogamente, um vetor $T = \{T_1, T_2, \ldots, T_k\}$ é dito *minimamente suficiente conjunto* se é função de qualquer outro vetor de estatísticas suficientes.

Definição: Uma estatística T é *mínima suficiente* se para cada estatística T' e para todo $x, y \in X$, $T(x) = T(y)$ sempre que $T'(x) = T'(y)$. Em outras palavras, T é uma função de T'.

Teorema (slide 96): Os estimadores de Bayes e de máxima verossimilhança (EMV) são estatísticas minimamente suficientes.

Prova:

- EMV: Se T é uma estatística suficiente para θ e existe uma EMV única de θ, então a EMV deve ser uma função de T.

$$f_x(x|\theta) \propto v[r(x), \theta]$$

- Bayes: Escrever a perda esperada a posteriori explicitamente usando a verossimilhança na forma do TF.

Teorema da Fatorização

Definição: Se T é suficiente para θ podemos escrever a verossimilhança como o produto entre uma função que não depende de θ e uma função que só depende de X através de T.

$$f_n(x|\theta) = u(x)\ .\ v(T(x, \theta))$$

Em outras palavras:

- u depende apenas de x (e não de θ).
- v depende de θ e x, mas depende de x apenas através de T.

Erro Quadrático Médio (EQM) ou (MSE)

$$EQM(\hat{\theta}) = E_\theta[(\hat{\theta} - \theta)^2]$$

$$EQM(\hat{\theta}) = Var_\theta(\hat{\theta}) + viés(\hat{\theta})^2$$

Estimador Condicionado

$$\delta_0(T) = E_\theta[\delta(X)|T]$$

Teorema de Rao-Blackwell

Definição: O teorema Rao-Blackwell diz que todo estimador condicionado em uma estatística suficiente é admissível.
Seja $\delta(X)$ um estimador, T uma *estatística suficiente* para θ e seja $\delta_0(T)$ um *estimador condicionado*, então:

$$R(\theta, \delta_0) \leq R(\theta, \delta)$$

Melhoramento de Blackwell ou Rao-Blackwellização

O teorema de Rao-Blackwell (às vezes chamado de teorema de Rao-Blackwell-Kolmogorov ou Rao-Blackwellization) é uma maneira de melhorar a eficiência dos estimadores iniciais. Estimadores são variáveis *aleatórias observáveis* usadas para estimar quantidades. Por exemplo, a média amostral (observável) é um estimador para a média populacional (desconhecida).
O melhoramento de Blackwell condiciona a estatística suficiente.

Admissibilidade e Dominância

Relação de dois estimadores: Um estimador δ é dito inadmissível se existe outro estimador δ_0 tal que $R(\theta, \delta_0) \leq R(\theta, \delta)$ para todo $\theta \in \Omega$ e existe $\theta' \in \Omega$ tal que $R(\theta', \delta_0) \leq R(\theta', \delta)$.

- Dizemos que δ_0 *domina* δ : $R(\theta, \delta_0) \leq R(\theta, \delta)$
- O estimador δ_0 é admissível se e somente se não há estimador que o domine.

Informação de Fisher : $I(\theta)$

$$l(x|\theta) = \log f(x|\theta)$$

$f(x|\theta)$ é diferenciável em θ:

$$l'(x|\theta) = \frac{\partial l(x|\theta)}{\partial \theta}$$

$$l''(x|\theta) = \frac{\partial^2 l(x|\theta)}{\partial \theta^2}$$

Fórmula da informação de Fisher:

$$I_n(\theta) = E_\theta[l'(x|\theta)^2] = -E_\theta[l''(x|\theta)] = Var_\theta(l'(x|\theta))$$

- Lembrete: A informação de Fisher sempre é positiva.

Seja $X_1, X_2, \ldots, X_n$ uma amostra aleatória e seja $I_n(\theta) = -E_\theta[l''(x|\theta)]$, então:

$$I_n(\theta) = nI(\theta)$$

$I(\theta)$ é a informação de Fisher aplicada em apenas *uma* observação, então usa: $f(x|\theta)$.
$I_n(\theta)$ é a informação de Fisher aplicada para várias observações, então usa: $f_n(x|\theta)$.

Exemplos:

Bernoulli:

$$\frac{1}{p(1-p)} \leftrightarrow \frac{1}{\theta(1-\theta)}$$

Prova:

$$f(x|\theta) = \theta^{x}(1-\theta)^{1-x},\ x = 1\ ou\ x = 0$$

$$l(x|\theta) = \log f(x|\theta) = x\ \log(\theta)\ +\ (1-x)\ \log(1-\theta)$$

$$l'(x|\theta) = \frac{x}{\theta} - \frac{1-x}{1-\theta}$$

$$l''(x|\theta) = -\frac{x}{\theta^2} - \frac{1-x}{(1-\theta)^2}$$

Sabemos que $E[x] = \theta$, então:

$$I(\theta) = -E_\theta[\lambda''(x|\theta)] = \frac{E[x]}{\theta^2} + \frac{1-E[x]}{(1-\theta)^2}$$

$$I(\theta) = \frac{\theta}{\theta^2} + \frac{1-\theta}{(1-\theta)^2} = \frac{1}{\theta} + \frac{1}{1-\theta} = \frac{1-\theta+\theta}{\theta(1-\theta)} = \frac{1}{\theta(1-\theta)}$$

Binomial:

$$\frac{n}{p(1-p)} \leftrightarrow \frac{1}{\theta(1-\theta)}$$

Prova:

$$f_n(x|\theta) = \prod_{i=1}^{n} \binom{n}{x_i} \theta^{y}(1-\theta)^{n-y},\ y = \sum_{1}^{n} x_i$$

$$l(x|\theta) = \log f(x|\theta) = \sum_{i=1}^{n} \log\left(\binom{n}{x_i}\right) + y\log(\theta) + (n-\theta)\log(1-\theta)$$

$$l'(x|\theta) = \frac{y}{\theta} - (n-y)\frac{1}{1-\theta}$$

$$l''(x|\theta) = -\frac{y}{\theta^2} - \frac{n-y}{(1-\theta)^2}$$

$$I_n(\theta) = -E_\theta[\lambda''(x|\theta)] = \frac{E[y]}{\theta^2} + \frac{1}{(1-\theta)^2}E[n-y]$$

$$I_n(\theta) = \frac{n\theta}{\theta^2} + \frac{n-n\theta}{(1-\theta)^2} = \frac{n\theta(1-\theta)^2 + (n-n\theta)\theta^2}{\theta^2(1-\theta)^2}$$

$$I_n(\theta) = \frac{n\theta - 2\theta^2 + \cancel{n\theta^3} + n\theta^2 - \cancel{n\theta^3}}{\theta^2(1-\theta)^2} = \frac{n\cancel{\theta}\ \cancel{(1-\theta)}}{\theta^{\cancel{2}}(1-\theta)^{\cancel{2}}}$$

$$I_n(\theta) = \frac{n}{\theta(1-\theta)} \leftrightarrow I(\theta) = \frac{1}{\theta(1-\theta)}$$

Exponencial:

$$\frac{1}{\lambda^2} \leftrightarrow \frac{1}{\theta^2}$$

Prova:

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

$$l(x|\theta) = \log f(x|\theta) = \log(\theta) - \theta x$$

$$l'(x|\theta) = \frac{1}{\theta} - x$$

$$l''(x|\theta) = -\frac{1}{\theta^2}$$

Sabemos que $E[x] = \frac{1}{\theta}$, mas a função já independe de x então:

$$I(\theta) = -E_\theta[\lambda''(x|\theta)] = -\left(-\frac{1}{\theta^2}\right) = \frac{1}{\theta^2}$$

Poisson

$$\frac{1}{\lambda} \leftrightarrow \frac{1}{\theta}$$

Prova:

$$f_n(\bar{x}|\theta) = \frac{e^{-n\theta}\theta^y}{\prod x_i!},\ y = \sum_1^n x_i,\ ,\ \theta > 0$$

$$l(x|\theta) = \log f_n(x|\theta) = y\ log(\theta) - n\theta - \log(\prod x_i!)$$

$$l'(x|\theta) = -n + \frac{y}{\theta}$$

$$l''(x|\theta) = \frac{-y}{\theta^2}$$

$$I_n(\theta) = -E_\theta[\lambda''(x|\theta)] = \frac{1}{\theta^2} E\left[\sum_1^n x_i\right]$$

Sabemos que $E[x] = \theta$, então:

$$I_n(\theta) = \frac{1}{\theta^2} n\theta = \frac{n}{\theta} \leftrightarrow I(\theta) = \frac{1}{\theta}$$

Normal

μ desconhecido e σ^2 é dado

$$I(\mu) = E[I''(x|\mu)] = \frac{1}{\sigma^2}$$

Prova:

$$l(x|\mu) = log\ f(x|\theta) = -\frac{1}{2} log\ (2\pi\sigma^2)\ - \frac{(x-\mu)^2}{2\sigma^2}$$

$$l'(x|\mu) = \frac{x-\mu}{\sigma^2}$$

$$l''(x|\mu) = -\frac{1}{\sigma^2}$$

Sabemos que $E[x] = \mu$, mas já é idependente de μ então basta trocar o sinal:

$$I(\theta) = -E_\theta[\lambda''(x|\theta)] = \frac{1}{\sigma^2}$$

Cramer-Ráo e Eficiência

O limite de Cramér-Rao afirma que o inverso da informação de Fisher é um limite inferior na variância de qualquer estimador não enviesado de θ. ($\hat{\theta}$)

Information Inequality:

$$Var_\theta(\delta) \geq \frac{[m'(\theta)]^2}{nI(\theta)}$$

$m(\theta) = E_\theta[\hat{\theta}]$

Definição: Um estimador é dito eficiente se ele atinge a cota de Cramer-Ráo.

$$Var_\theta(\delta) = \frac{[m'(\theta)]^2}{nI(\theta)}$$

Exemplo: Seja $X_1, X_2, \ldots, X_n$ uma amostra aleatória de uma distribuição de Poisson, $\bar{x_n}$ é um estimador eficiente de θ.

- Se $\hat{\theta}$ é um estimador não-viesado, então $m(\theta) = \theta \leftrightarrow m'(\theta) = 1$

$$Var_\theta(\delta) \geq \frac{1}{nI(\theta)}$$

Distribuição Assintótica

Propriedades assintóticas são aquelas válidas apenas para grandes amostras, ou para amostras com tamanho $n \to \infty$. A distribuição amostral de um estimador é diferente para tamanhos de amostras diferentes. A distribuição da média amostral para amostras de qualquer população é definida pelo Teorema do Limite Central (TCL).

O Teorema nos garante que quando o tamanho da amostra cresce, a distribuição assintótica será aproximadamente normal. Então, para qualquer que seja a distribuição que tenha gerado os dados amostrais, sabemos que o estimador convergirá para a distribuição normal, sendo esta a distribuição assintótica do estimador.

Definição: $\hat{\theta}$ é um estimador assintoticamente não viesado de θ se $\lim_{n\to\infty} E(\hat{\theta}) = \theta$.

Bônus

Propriedades de Logarítmo

$$\log(mn) = log(m) + log(n)$$

$$\log\left(\frac{m}{n}\right) = \log(m) - \log(n)$$

$$log(m^a) = a\log(m)$$

$$\log e^x = x$$

Propriedades de Derivadas:

$$(x^n)' = n\ x^{n-1}$$

$$(fg)' = f'g + fg'$$

$$(e^x)' = e^x$$

$$(e^{cx})' = ce^{cx}$$

$$(\log x)' = \frac{1}{x}$$

$$(c^x)' = c^x \log c$$