# Introdução aos Processos Estocásticos - Estimadores

Eduardo M. A. M. Mendes

DELT - UFMG
Programa de Pós-Graduação em Engenharia Elétrica
Universidade Federal de Minas Gerais
emmendes@cpdee.ufmg.br

# Introdução

Problema: Amostras retiradas de uma distribuição conhecida onde os parâmetros que caracterizam a distribuição são desconhecidos.

- Em geral, a pdf de uma variável aleatória $X$ é $f(x|\theta)$ onde $\theta$ é o vetor de parâmetros que caracterizam a pdf. O vetor de parâmetros $\theta$ é definido no espaço de parâmetros $\Theta$.
- Para cada valor de $\theta \in \Theta$, existe uma pdf diferente.
- Para obter valores para o vetor de parâmetros, amostras aleatórias são retiradas da população e uma estatística denominada estimador é construída.
- Os valores dos estimadores são chamados estimativas pontuais.

# Métodos baseados na Média Amostral

Sabemos que

$$\bar{X} = \frac{1}{N}\sum_{i=1}^{N} X_i \qquad X \text{ é } IID$$

Podemos calcular o valor esperado

$$\begin{aligned} E(\bar{X}) &= E\left(\frac{1}{N}\sum_{i=1}^{N} X_i\right) \\ &= \frac{1}{N}\sum_{i=1}^{N} E(X_i) \\ &= \frac{1}{N}\sum_{i=1}^{N} E(X) \\ &= E(X) \end{aligned}$$

# Métodos baseados na Média Amostral (cont.)

e a variância

$$\begin{aligned} Var(\bar{X}) &= Var\left(\frac{1}{N}\sum_{i=1}^{N} X_i\right) \\ &= \frac{1}{N^2}\sum_{i=1}^{N} Var(X) \\ &= \frac{Var(X)}{N} \end{aligned}$$

# Propriedades dos Estimadores

Considere o experimento que produz os valores amostrais de uma variável aleatória $X$.

- Um número indefinido de VAs $X_1, X_2, \ldots$, todas com o mesmo modelo probabilístico de $X$.
- Assuma que $\theta$ é parâmetro do modelo probabilístico.
- As observações $X_1, X_2, \ldots$ são usadas para produzir uma sequência de estimadores de $\theta$. As estimativas $\hat{T}_1, \hat{T}_2, \ldots$ são variáveis aleatórias.

$$
\begin{array}{lll}
T_1 & \text{depende de} & X_1 \\
T_2 & \text{depende de} & X_1 \text{ e } X_2 \\
\vdots & & \\
T_N & \text{depende de} & X_1 \quad \cdots \quad X_N
\end{array}
$$

# Estimador Consistente

Estimador Consistente: A sequência de estimativas $\hat{T}_1, \hat{T}_2, \ldots$ do parâmetro $\theta$ é consistente se para qualquer $\epsilon > 0$

$$\lim_{N \to \infty} P\left[|\hat{T}_N - \theta| \geq \epsilon\right] = 0$$

Exemplo: Seja $\{X_1, X_2, \ldots, X_N\}$ uma amostra aleatória de tamanho $N$ de uma distribuição com média $\mu$ e variância $\sigma^2$. Mostre que $\bar{X}_N$ é um estimador consistente de $\mu$.

Solução: Para que $\bar{X}_N$ seja um estimador de $\mu$, precisamos mostrar que

$$\lim_{N \to \infty} P\left[|\bar{X}_N - \mu| \geq \epsilon\right] = 0$$

Usando a inequação de Chebyshev e o fato de que $E(\bar{X}) = \mu$ e $Var(\bar{X}_N) = \frac{\sigma^2}{N}$, temos

# Estimador Consistente (cont.)

$$P\left[|\bar{X}_N - \mu| \geq \epsilon\right] \leq \frac{\sigma^2}{N\epsilon^2}$$

Como $N \rightarrow \infty$ e $\sigma^2$ é finito, logo

$$\lim_{N\rightarrow\infty} P\left[|\bar{X}_N - \mu| \geq \epsilon\right] = 0$$

# Estimador Não-Polarizado

Estimador Não-Polarizado: Uma estimativa, $\hat{T}$, do parâmetro $\theta$ é não polarizada se

$$E(\hat{T}) = \theta$$

Caso contrário é polarizada. Podemos, então, definir polarização (bias) como

$$bias = E(\hat{T}) - \theta$$

Exemplo: Suponha $X \sim Pois(\lambda)$ $(E(X) = Var(X) = \lambda)$ onde $\lambda$ é desconhecido. Mostrar que

(a) $\bar{X}$ é um estimador não-polarizado de $\lambda$. Para mostrar isso, considere

$$E(\bar{X}) = E\left(\sum_{i=1}^{N} \frac{X_i}{N}\right) = \sum_{i=1}^{N} \frac{E(X_i)}{N} = \frac{N\lambda}{N} = \lambda$$

# Estimador Não-Polarizado (cont.)

(b) $2\bar{X}$ é um estimador não-polarizado de $2\lambda$.

$$E(2\bar{X}) = 2E(\bar{X}) = 2\lambda$$

(c) $\bar{X}^2$ é um estimador polarizado de $\lambda^2$.

$$E(\bar{X}^2) = Var(\bar{X}) + \mu_{\bar{X}}^2 = \frac{\lambda}{N} + \lambda^2 \neq \lambda^2$$

# Estimador Não-Polarizado assintoticamente

Estimador Não-Polarizado assintoticamente: A sequência de estimadores $\hat{T}_N$ do parâmetro $\theta$ é assintoticamente não-polarizada se

$$\lim_{N\to\infty} E(\hat{T}_N) = \theta$$

Exemplo: No exemplo anterior temos

$$E(\bar{X}^2) = Var(\bar{X}) + \mu_{\bar{X}}^2 = \frac{\lambda}{N} + \lambda^2 \underset{N\to\infty}{\longrightarrow} \lambda^2$$

# Erro Quadrático Médio (MSE)

Erro Quadrático Médio (MSE): O erro quadrático médio do estimador $\hat{T}$ é

$$e = E\left((\hat{T} - \theta)^2\right)$$

- O MSE consiste de dois componentes não-negativos: a variância do estimador $\hat{T}$ e a polarização (bias $= E(\hat{T}) - \theta$) ao quadrado do mesmo estimador:

$$\begin{aligned}
MSE(\hat{T}) &= E\left((\hat{T} - E(\hat{T}) + E(\hat{T}) - \theta)^2\right) \\
&= E\left((\hat{T} - E(\hat{T}))^2\right) + E\left((E(\hat{T}) - \theta)^2\right) + 2E\left((\hat{T} - E(\hat{T}))(E(\hat{T}) - \theta)\right) \\
&= Var(\hat{T}) + (E(\hat{T}) - \theta)^2 + 2\left(E(\hat{T}) - E(\hat{T})\right)\left(E(\hat{T}) - \theta\right) \\
&= Var(\hat{T}) + (E(\hat{T}) - \theta)^2 \\
&= Var(\hat{T}) + (\text{bias}(\hat{T}))^2
\end{aligned}$$

# Erro Quadrático Médio (MSE) (cont.)

- Quando $\hat{T}$ é uma estimativa não polarizada de $\theta$ e $E(\hat{T}) = \theta$, MSE é simplesmente a variância de $\hat{T}$.
- Para uma sequência de estimadores não polarizados, é suficiente mostrar que o MSE vai para zero para provar que o estimador é consistente.

Teorema: Se a sequência de estimativas não polarizadas $\hat{T}_1, \hat{T}_2, \ldots$ do parâmetro $\theta$ tem MSE com

$$e_N = Var(\hat{T}_N)$$

satisfazendo $\lim_{N\to\infty} e_N = 0$, então a sequência é consistente.

Prova: Como $E\left(\hat{T}_N\right) = \theta$, a desigualdade de Chebyshev pode ser aplicada a $\hat{T}_N$. Para a constante $\epsilon > 0$

# Erro Quadrático Médio (MSE) (cont.)

$$P\left(|\hat{T}_N - \theta| > \epsilon\right) \leq \frac{var(\hat{T}_N)}{\epsilon^2}$$

No limite

$$\lim_{N\to\infty} P\left(|\hat{T}_N - \theta| > \epsilon\right) \leq \lim_{N\to\infty} \frac{var(\hat{T}_N)}{\epsilon^2} = 0$$

Exemplo: Em um intervalo de $k$ segundos, o número $N_k$ de pacotes passando por um roteador é uma VA de Poisson com valor esperado $E[N_k] = k\theta$ pacotes. Seja $\hat{T}_k = \frac{N_k}{k}$ uma estimativa de $\theta$. Cada estimativa $\hat{T}_k$ é uma estimativa não polarizada de $\theta$? Qual é o MSE?

# Erro Quadrático Médio (MSE) (cont.)

Solução:

$$E\left(\hat{T}_k\right) = E\left(\frac{N_k}{k}\right) = \frac{E(N_k)}{k} = \theta$$

Como $N_k$ é Poisson, $Var(N_k) = k\theta$:

$$Var\left(\hat{T}_k\right) = Var\left(\frac{N_k}{k}\right) = \frac{Var(N_k)}{k^2} = \frac{\theta}{k}$$

Como $\hat{T}_k$ é não polarizada

$$\text{MSE} = \frac{\theta}{k}$$

Repare que

$$\lim_{k\to\infty} Var\left(\hat{T}_k\right) = 0, \quad \text{então}$$

a sequência de estimadores $\hat{T}_k$ é consistente.

Considerando que $\theta = E(X)$, o estimador usado é

$$\hat{T}_N = \bar{X} \quad \text{(Média Amostral)}$$

- Como $E(\bar{X}) = E(X)$, a média amostral é não-polarizada.

$$e_N = E\left((\bar{X} - E(X))^2\right) = Var(\bar{X}) = \frac{Var(X)}{N}$$

Teorema: Se $X$ tem variância finita, então a média amostral $\bar{X}$ é uma sequência de estimadores consistentes de $E(X)$.

Prova

$$\lim_{N\to\infty} Var(\bar{X}) = \lim_{N\to\infty} \frac{Var(X)}{N} = 0$$

então $\bar{X}$ é uma sequência consistente.

# Estimativas da Variância

Considerando

$$\theta = Var(X) = E\left((X - \mu_X)^2\right)$$

temos que analisar dois casos: quando $E(X)$ é conhecido e quando $E(X)$ não é conhecido.

- Suponha que $X$ tenha média zero, logo $Var(X) = E(X^2)$.

$$\begin{array}{ccl} Y = X^2 & \rightarrow & E(Y) \\ & & \downarrow \\ & & \bar{Y} = \dfrac{1}{N}\left(X_1^2 + X_2^2 + \ldots + X_N^2\right) \end{array}$$

Se $Var(Y)$ existe, pela Lei dos Grandes Números, implica que $\bar{Y}$ é um estimador consistente de $E(X^2) = Var(X)$.

# Estimativas da Variância (cont.)

- Quando $E(X)$ é conhecido, ou seja, $E(X) = \mu_X$, implica que $Var(X) = E\left((X - \mu_X)^2\right)$. Faz-se

$$W = (X - \mu_X)^2$$

e

$$\bar{W} = \frac{1}{N}\sum_{i=1}^{N}(X_i - \mu_X)^2$$

se $Var(W)$ existe, $\bar{W}$ é um estimador não polarizado e consistente de $Var(X)$.

- Quando o valor esperado de $X$ não é conhecido, a situação é mais complicada pois a variância de $X$ depende de $\mu_X$.

# Estimativas da Variância (cont.)

- A variância amostral de um conjunto de $N$ observações independentes da variável $X$ é

$$V_N(X) = \frac{1}{N}\sum_{i=1}^{N}\left(X_i - \bar{X}\right)^2$$

então $V_N(X)$ é um estimador polarizado de $Var(X)$.

Prova: Substituindo $\bar{X} = \frac{1}{N}\sum_{i=1}^{N} X_i$ em

$$V_N(X) = \frac{1}{N}\sum_{i=1}^{N}\left(X_i - \bar{X}\right)^2$$

resulta

$$V_N = \frac{1}{N}\sum_{i=1}^{N} X_i^2 - \frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N} X_i X_j$$

como $X_i$ é IID, então

$$\begin{aligned} E(X_i^2) &= E(X^2), \quad \forall i \quad \text{e} \\ E(X_i)E(X_j) &= \mu_X^2 \end{aligned}$$

Sabe-se que

$$\begin{aligned} E(X_i X_j) &= cov(X_i, X_j) + E(X_i)E(X_j) \\ &= cov(X_i, X_j) + \mu_X^2 \end{aligned}$$

Logo

$$\begin{aligned} E(V_N) &= E(X^2) - \frac{1}{N^2} \sum_{i=1}^{N} \sum_{j=1}^{N} \left( cov(X_i, X_j) + \mu_X^2 \right) \\ &= Var(X) - \frac{1}{N^2} \sum_{i=1}^{N} \sum_{j=1}^{N} cov(X_i, X_j) \end{aligned}$$

# Estimativas da Variância (cont.)

Como os dois únicos termos diferentes de zero no somatório são $cov(X_i, X_i) = Var(X_i)$

$$
\begin{aligned}
E(V_N) &= Var(X) - \frac{1}{N^2}\left(NVar(X)\right) \\
&= \frac{N-1}{N} Var(X)
\end{aligned}
$$

Embora $V_n(X)$ seja um estimador polarizado, ele é assintoticamente não polarizado pois

$$\lim_{N\to\infty} E\left(V_n(X)\right) = \lim_{N\to\infty} \frac{N-1}{N} Var(X) = Var(X)$$

# Estimativas da Variância (cont.)

- Teorema: A estimativa

$$S^2 = V'_N(X) = \frac{1}{N-1}\sum_{i=1}^{N}(X_i - \bar{X})^2$$

é um estimador não polorizado.

Prova:

$$V'_N(X) = \frac{N}{N-1} V_N(X)$$

logo

$$E\left(V'_N\right) = \frac{N}{N-1} E\left(V_N(X)\right) = Var(X)$$

# Estimativas da Variância (cont.)

Repare que se $N = 1$, temos

$$\bar{X} = X_1 \qquad \text{e} \qquad \begin{aligned} V_1(X) &= 0 \\ V_1'(X) &= \text{indefinido} \end{aligned}$$

# Eficiência

- Além de ser não-polarizado, uma outra propriedade importante de um bom estimador é possuir uma variância pequena, ou seja, estimadores que apresentem *MSE* reduzido.
- Um meio de comparar os *MSE*s dos estimadores é calcular a eficiência relativa. Dados dois estimadores, $\hat{T}_1$ e $\hat{T}_2$, a eficiência de $\hat{T}_1$ relativo a $\hat{T}_2$ é

$$eff(\hat{T}_1, \hat{T}_2) = \frac{MSE(\hat{T}_2)}{MSE(\hat{T}_1)}$$

- Caso o estimador seja não-polarizado, a eficiência de $\hat{T}_1$ relativo a $\hat{T}_2$ pode ser escrito como

$$eff(\hat{T}_1, \hat{T}_2) = \frac{Var(\hat{T}_2)}{Var(\hat{T}_1)}$$

# Eficiência (cont.)

- O estimador $\hat{T}_1$ é mais eficiente do que $\hat{T}_2$, se $MSE(\hat{T}_1) < MSE(\hat{T}_2)$ ou

$$eff(\hat{T}_1, \hat{T}_2) > 1$$

- Se $T = \hat{\theta}$ é um estimador não-polarizado de $\theta$ e as amostras de tamanho $N$, $X_1, X_2, \ldots, X_N$, têm PDF $f(x|\theta)$, então a variância do estimador não-polarizado $\hat{\theta}$ deve satisfazer a seguinte desigualdade

$$Var(\hat{\theta}) \geq \frac{1}{N \times E\left(\left(\frac{\partial ln(f(X|\theta))}{\partial \theta}\right)^2\right)}$$

onde $f(X|\theta)$ é a PDF de interesse tomada em $X$. A desigualdade acima é denominada Desigualdade de Cramér-Rao.

# Eficiência (cont.)

- Se $\hat{\theta}$ é um estimador não-polarizado e

$$Var(\hat{\theta}) = \underbrace{\frac{1}{N \times E\left(\left(\frac{\partial ln(f(X|\theta))}{\partial \theta}\right)^2\right)}}_{\text{Cramér-Rao Bound (CRLB)}}$$

então $\hat{\theta}$ é um estimador de variância mínima não-polarizado de $\theta$.

Exemplo: Mostre que $\bar{X}$ é o estimador não-polarizado de mínima variância da média $\lambda$ de população de Poisson.

Solução: Se $X \sim Pois(\lambda)$ com $E(X) = Var(X) = \lambda$ e PDF

$$p(X = x|\lambda) = \frac{\lambda^x e^{-\lambda}}{x!}$$

Como

$$E(\bar{X}) = \sum_{i=1}^{N} \frac{X_i}{N} = \frac{N\lambda}{N} = \lambda$$

$\bar{X}$ é um estimador não-polarizado de $\lambda$ com variância $\frac{\lambda}{N}$, pois

$$Var(\bar{X}) = Var\left(\sum_{i=1}^{N} \frac{X_i}{N}\right) = \frac{1}{N^2}\sum_{i=1}^{N} Var(X_i) = \frac{N\lambda}{N^2} = \frac{\lambda}{N}$$

consequentemente, se o CRLB é $\frac{\lambda}{N}$, $\bar{X}$ é o estimador não-polarizado de variância mínima de $\lambda$. Para mostrar isso, considere

$$ln(p(x|\lambda)) = xln(\lambda) - \lambda - ln(x!)$$

Calculando a derivada da função acima em relação a $\lambda$ resulta em

$$\frac{\partial p(x|\lambda)}{\partial \lambda} = \frac{x}{\lambda} - 1 = \frac{x-\lambda}{\lambda}$$

# Eficiência (cont.)

Logo

$$\begin{aligned} E\left(\left(\frac{\partial p(X|\lambda)}{\partial \lambda}\right)^2\right) &= E\left(\left(\frac{X-\lambda}{\lambda}\right)^2\right) \\ &= \frac{E\left((X-\lambda)^2\right)}{\lambda^2} \\ &= \frac{Var(X)}{\lambda^2} \\ &= \frac{\lambda}{\lambda^2} = \frac{1}{\lambda} \end{aligned}$$

Finalmente o CRLB é

$$\frac{1}{N \times E\left(\left(\frac{\partial ln(f(X|\theta))}{\partial \theta}\right)^2\right)} = \frac{\lambda}{N}$$

# Eficiência (cont.)

Consequentemente, como $\bar{X}$ é não-polarizado e $Var(\bar{X}) = \frac{\lambda}{N}$, o estimador é de mínima variância.

# Estimadores Robustos

- A essência de um estimador robusto é um estimador cuja distribuição amostral não é seriamente afetada pelas violações das suposições feitas ao solucionar o problema de estimação.
- O conceito de robustez também é usado para se referir a habilidade de um estimador em particular de proporcionar estimativas razoáveis quando observações atípicas são encontradas na amostra.

Exemplo: A mediana é um estimador robusto do centro de uma distribuição assimétrica.

Vamos considerar dois métodos clássicos:

1) Método dos Momentos
2) Método da Verossimilhança

Lembrando a notação usada

- A informação nas variáveis aleatórias $X_1, X_2, X_3, \ldots, X_N$ é usada para fazer inferências sobre o parâmetro não-conhecido $\theta$.
- Os valores observados da variável aleatória são $x_1, x_2, x_3, \ldots, x_N$.
- a PDF conjunta é dada por

$$\begin{aligned} f(\mathbf{x}|\theta) &= f(x_1, x_2, x_3, \ldots, x_n|\theta) \\ &= f(x_1|\theta) \times f(x_2|\theta) \times \cdots \times f(x_N|\theta) \\ &= \prod_{i=1}^{N} f(x_i|\theta) \end{aligned}$$

Idéia: Igualar os momentos com os momentos amostrais correspondentes.
O $r$-ésimo Momento Amostral, $m_r$, é definido como

$$m_r = \frac{1}{N}\sum_{i=1}^{N} X_i^r$$

O momento de uma variável aleatória foi definido como

$$\alpha_r = E(X^r) = \sum_{i=1}^{\infty} x_i^r p(X = x_i) \text{ ou } \int_{-\infty}^{\infty} x^r f(x)dx$$

Dada uma amostra aleatória $X_1, X_2, \ldots, X_N$ de uma população com PDF $f(x|\theta_1, \theta_2, \ldots, \theta_k)$, no métodos dos momentos, os estimadores $\theta_i$ para $i = 1, \ldots, k$ são determinados igualando os primeiros $k$ momentos com os momentos amostrais correspondentes e resolvendo o sistema resultante.

# Métodos dos Momentos (cont.)

$$\begin{cases} \alpha_1(\theta_1, \ldots, \theta_k) & = & m_1 \\ \alpha_2(\theta_1, \ldots, \theta_k) & = & m_2 \\ & \vdots & \\ \alpha_k(\theta_1, \ldots, \theta_k) & = & m_k \end{cases}$$

Considerações sobre o método:

- Fácil utilização e obtenção de estimadores consistentes.
- Sob certas condições, os momentos amostrais convergem para os momentos da população.
- Os momentos amostrais são estimadores não-polarizados.

Exemplo: Dada uma amostra de tamanho $N$ da distribuição $Gamma(\alpha, \lambda)$, encontrar, pelo método dos momentos, estimadores de $\alpha$ e $\lambda$.

Solução: Sabemos que para a distribuição $Gamma(\alpha, \lambda)$

$$
\begin{aligned}
E(X) &= \frac{\alpha}{\lambda} \\
Var(X) &= \frac{\alpha}{\lambda^2}
\end{aligned}
$$

e que o primeiro e segundo momentos amostrais são

$$
\begin{aligned}
m_1 &= \bar{X} \\
m_2 &= \frac{1}{N} \sum_{i=1}^{N} x_i^2
\end{aligned}
$$

# Métodos dos Momentos (cont.)

Calculando o primeiro e segundo momentos da população, temos:

$$\begin{aligned}\alpha_1(\alpha,\lambda) &= E(X^1) = \frac{\alpha}{\lambda}\\ \alpha_2(\alpha,\lambda) &= E(X^2) = \sigma^2 + E(X)^2 = \frac{\alpha}{\lambda^2} + \frac{\alpha^2}{\lambda^2} = \frac{\alpha(1+\alpha)}{\lambda^2}\end{aligned}$$

Montando o sistema de equações, temos

$$\left\{\begin{array}{lcl}\alpha_1(\alpha,\lambda) &=& \frac{\alpha}{\lambda} = \bar{X} = m_1\\ \alpha_2(\alpha,\lambda) &=& \frac{\alpha(1+\alpha)}{\lambda^2} = \frac{1}{N}\sum_{i=1}^{N} x_i^2 = m_2\end{array}\right.$$

Chamando $S^2 = \frac{\sum_{i=1}^{N}(x_i - \bar{X})^2}{N}$, podemos chegar a seguinte solução do sistema acima

# Métodos dos Momentos (cont.)

$$
\begin{aligned}
\hat{\alpha} &= \frac{\bar{X}^2}{S^2} \\
\hat{\lambda} &= \frac{\bar{X}}{S^2}
\end{aligned}
$$

# Estimadores de Máxima Verossimilhança

Idéia: Selecionar o valor de $\theta$ que faça com que o modelo probabilístico, $f(x|\theta)$, seja o mais "provável" gerador dos dados observados.

- $L(\theta|\mathbf{x})$ é a função de verossimilhança de $\theta$ para $X$

$$L(\theta|\mathbf{x}) = f(\mathbf{x}|\theta) = \prod_{i=1}^{N} f(x_i|\theta) = f(x_1|\theta) \times \cdots \times f(x_n|\theta)$$

- O valor de $\theta$ que maximiza $L(\theta|\mathbf{x})$ é chamado estimativa de máxima verossimilhança (MLE).

# Estimadores de Máxima Verossimilhança (cont.)

- Devido à dificuldade de manipulação de $L(\theta|\mathbf{x})$, opta-se pela chamada função log de verossimilhança, ou seja, $ln(L(\theta|\mathbf{x}))$. Para achar $\theta$, basta

$$\frac{\partial(ln(L(\theta|\mathbf{x})))}{\partial\theta} = 0$$

Para achar um máximo, seja local ou global, é necessário calcular a derivada segunda

$$\left.\frac{\partial^2(ln(L(\theta|\mathbf{x})))}{\partial\theta^2}\right|_{\theta=\hat{\theta}(x)} < 0$$

Obs.: É necessário também testar as condições de contorno.

Exemplo: Dada uma amostra aleatória de tamanho $N$ de uma distribuição de Bernoulli, $Ber(p)$, determine a estimativa de verossimilhança e a de máxima verossimilhança do parâmetro $p$.

Solução: Sabemos que

$$P(X = x|p) = p^x(1-p)^{1-x}$$

onde $x$ é 1 com probabilidade $p$ e 0 com probabilidade $(1-p)$.

A função de verossimilhança para os $N$ valores observados é

$$L(p|\mathbf{x}) = \prod_{i=1}^{N} p^{x_i}(1-p)^{1-x_i}$$

# Estimadores de Máxima Verossimilhança (cont.)

e

$$\begin{aligned}
ln(L(p|\mathbf{x})) &= ln\left[\prod_{i=1}^{N} p^{x_i}(1-p)^{1-x_i}\right] \\
&= \sum_{i=1}^{N} ln\left[p^{x_i}(1-p)^{1-x_i}\right] \\
&= \sum_{i=1}^{N} [x_i ln(p) + (1-x_i) ln(1-p)]
\end{aligned}$$

Para achar o valor que maximiza a expressão acima, é necessário calcular a derivada parcial de primeira ordem de $ln(L(p|\mathbf{x}))$ e igualar a zero

$$\frac{\partial ln(L(p|\mathbf{x}))}{\partial p} = \frac{\sum_{i=1}^{N} x_i}{p} - \frac{N - \sum_{i=1}^{N} x_i}{1-p} = 0$$

A solução é

$$p = \frac{\sum_{i=1}^{N} x_i}{N} = \bar{x}$$

Para que $p = \bar{x}$ seja máximo, é necessário que a derivada parcial de segunda ordem seja negativa no valor

$$\frac{\partial^2 ln(L(p|\mathbf{x}))}{\partial p^2} = \frac{-\sum_{i=1}^{N} x_i}{p^2} - \left.\frac{N - \sum_{i=1}^{N} x_i}{(1-p)^2}\right|_{p=\bar{x}} = -\frac{N}{\bar{x}} - \frac{N}{1-\bar{x}}$$

que é negativa para $0 \leq \bar{x} \leq 1$ e $N > 0$. Como, para $p = 0$ e $p = 1$, a função de verossimilhança é zero $p = \bar{x}$ é o valor que maximiza tal função.

# Estimadores de Máxima Verossimilhança (cont.)

Exemplo: Suponha $\{X_1, X_2, \ldots, X_N\}$ é uma amostra aleatória de uma distribuição gaussiana com $N(\mu, \sigma)$, onde $\sigma$ é conhecido. Encontre o estimador de verossimilhança de $\mu$.

Soluçao: A PDF de uma variável aleatória $X \sim N(\mu, \sigma)$ é

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

A função de verossimilhança é

$$L(\mu|\mathbf{x}) = \prod_{i=1}^{N} f(x_i) = \prod_{i=1}^{N} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x_i-\mu)^2}{2\sigma^2}}$$

# Estimadores de Máxima Verossimilhança (cont.)

e o log dela

$$ln(L(\mu|\mathbf{x})) = -\frac{N}{2} ln(2\pi) - \frac{N}{2} ln(\sigma^2) - \frac{\sum_{i=1}^{N}(x_i - \mu)^2}{2\sigma^2}$$

Para achar o valor que maximiza a função acima é preciso calcular a primeira derivada parcial e a segunda

$$\begin{aligned} \frac{\partial ln(L(\mu|\mathbf{x}))}{\partial \mu} &= \frac{\sum_{i=1}^{N}(x_i - \mu)}{\sigma^2} = 0 \\ \frac{\partial^2 ln(L(\mu|\mathbf{x}))}{\partial \mu^2} &= -\frac{N}{\sigma^2} < 0 \end{aligned}$$

Como a função de verossimilhança vai para zero quando $x \to \pm\infty$, $\mu = \bar{x}$ é o máximo global.

# Informação de Fisher

- O objetivo é achar alguma medida sobre a variância dos estimadores encontrados.
- A informação de Fisher é uma dessas medidas.
- A informação de Fisher é a quantidade de informação que uma variável aleatória $X$ carrega sobre um parâmetro $\theta$ desconhecido, do qual depende a função de verossimilhança $L(\theta|\mathbf{x})$ de $X$:

$$E\left[\left(\frac{\partial ln(f(\mathbf{X}|\theta))}{\partial\theta}\right)^2\right]$$

repare que a expressão acima é parte do denominador do CRLB.

# Informação de Fisher (cont.)

- O inverso da informação de Fisher nos dá um limite para a variância do melhor estimador não-polarizado. Podemos então determinar duas formas para expressar o valor de informação para uma variável aleatória de tamanho $N$

$$
\begin{aligned}
I_N(\theta) &= E\left[\left(\frac{\partial ln(f(\mathbf{X}|\theta))}{\partial\theta}\right)^2\right] = NI_N(\theta) = NE\left[\left(\frac{\partial ln(f(X|\theta))}{\partial\theta}\right)^2\right] \\
I_N(\theta) &= -E\left[\left(\frac{\partial^2 ln(f(\mathbf{X}|\theta))}{\partial\theta^2}\right)\right] = NI_N(\theta) = -NE\left[\left(\frac{\partial^2 ln(f(X|\theta))}{\partial\theta^2}\right)\right]
\end{aligned}
$$

# Informação de Fisher (cont.)

Exemplo: Dada a PDF de uma distribuição gaussiana com média $\mu$ desconhecida e variância $\sigma^2$ conhecida, encontre a informação de Fisher de $\mu$, dada uma amostra aleatória de tamanho $N$.

Solução: Já sabemos que a função de verossimilhança é

$$L(\mu|\mathbf{x}) = \prod_{i=1}^{N} f(x_i) = \prod_{i=1}^{N} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x_i-\mu)^2}{2\sigma^2}}$$

e o log dela

$$ln(L(\mu|\mathbf{x})) = -\frac{N}{2} ln(2\pi) - \frac{N}{2} ln(\sigma^2) - \frac{\sum_{i=1}^{N}(x_i-\mu)^2}{2\sigma^2}$$

Note que

$$\frac{\partial ln(f(x|\mu))}{\partial \mu} = \frac{(x-\mu)}{\sigma^2}$$

$$\frac{\partial^2 ln(f(x|\mu^2))}{\partial \mu} = -\frac{1}{\sigma^2}$$

Usando as expressões anteriores, podemos escrever

# Informação de Fisher (cont.)

$$
\begin{aligned}
I_N(\mu) &= NE\left[\left(\frac{\partial ln(f(X|\mu))}{\partial\mu}\right)^2\right] \\
&= NE\left[\left(\frac{X-\mu}{\sigma^2}\right)^2\right] \\
&= N\frac{E\left[(X-\mu)^2\right]}{\sigma^4} = \frac{N\sigma^2}{\sigma^4} = \frac{N}{\sigma^2}
\end{aligned}
$$

$$
\begin{aligned}
I_N(\mu) &= -NE\left[\left(\frac{\partial^2 ln(f(X|\mu))}{\partial\mu^2}\right)\right] \\
&= -NE\left[-\frac{1}{\sigma^2}\right] = \frac{N}{\sigma^2}
\end{aligned}
$$

# Informação de Fisher (cont.)

Podemos notar que quanto menor é a variância, mais informação há na variável aleatória de tamanho $N$ sobre o parâmetro $\mu$.