ESTATÍSTICA

Michelle Hanne Soares de Andrade

michellehanne.andrade@gmail.com

Métodos de Estimação Pontual

Métodos de Estimação Pontual

Os estimadores considerados até agora foram sugeridos por intuição. Seria desejável, entretanto, um método ou princípio que levasse sempre a bons estimadores. Infelizmente não existe um método geral aplicável a todas as situações.

Métodos de Estimação Pontual

- **As definições de não-enviesamento e outras propriedades são muito úteis para avaliar a qualidade de um estimador, contudo as mesmas não oferecem nenhum dispositivo para a construção de tais estimadores.**
- Nem sempre as quantidades desconhecidas de um problema estão associadas a parâmetros de fácil interpretação como média e variância.

Métodos de Estimação Pontual

- Nestes casos, a obtenção de estimadores adequados é não-trivial.
- Duas abordagens são bastante comuns na estatística clássica frequentista: **método dos momentos, método da máxima verossimilhança e dos mínimos quadrados**.

Métodos de Estimação Pontual

mais fácil de calcular

1. Método do Momentos

2. Método da Máxima Verossimilhança

mais eficiente

Método dos Momentos

- É o método mais antigo, determinado por K. Pearson em 1894. Se existem “K” parâmetros para serem estimados, o método consiste em expressar os primeiros “k” momentos da população em termos dos “K” parâmetros através de “K” equações.
- A solução do sistema formado pro estas equações leva a determinação dos estimadores.

Método dos Momentos

As estimativas obtidas desta forma são claramente funções dos momentos da amostra. Uma vez que os momentos da amostra são estimativas consistentes dos momentos da população, os parâmetros estimados serão geralmente coerentes.

Método dos Momentos

- Definição. Seja $X_1, X_2, \ldots, X_n$ uma amostra aleatória de uma variável X com função densidade (ou função de probabilidade) $f(x)$. O k-ésimo momento populacional do modelo é $E(X^k)$, $k \in \mathbb{N}$. O k-ésimo momento amostral é $\frac{1}{n} \sum_{i=1}^{N} X_i^k$, $k \in \mathbb{N}$.
- Definição. Sejam $X_1, X_2, \ldots, X_n$ e X como na definição anterior e suponha que $f(x)$ possua m parâmetros $\theta_1, \theta_2, \ldots, \theta_m$ desconhecidos. Estimadores $\hat{\Theta}_1, \hat{\Theta}_2, \ldots, \hat{\Theta}_m$ podem ser obtidos igualando-se os m primeiros momentos populacionais aos m primeiros momentos amostrais.

Método dos Momentos

População com distribuição de probabilidade *f(x)*

Amostragem aleatória: $X_1, X_2, \ldots, X_n$

densidade de probabilidade contínua ou função de massa de probabilidade discreta.

k-ésimo momento da população: $E(X^k) \quad k = 1,2,\ldots$

k-ésimo momento amostral: $$M_k = \frac{1}{n}\sum_{i=1}^{n} X_i^k$$

Método dos Momentos

Idéia geral por trás do método de momentos:

- *igualar os momentos da população aos momentos amostrais.*

- *os momentos da população são funções de parâmetros desconhecidos.*

- *estas equações são resolvidas para os estimadores dos parâmetros desconhecidos.*

Método dos Momentos

Primeiro momento da população: $E(X) = \mu$

Primeiro momento amostral: $M_1 = \overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$

Igualando os momentos da amostra e da população temos

$$\hat{\mu} = \overline{X}$$

Portanto, a média amostral é o estimador de momento da média da população

Em geral, momentos da população são funções de parâmetros desconhecidos da população:

$$\theta_1, \theta_2, \ldots, \theta_m$$

Método dos Momentos

Amostragem aleatória: $X_1, X_2, \ldots, X_n$

Parâmetros desconhecidos da população: $\theta_1, \theta_2, \ldots, \theta_m$

Os estimadores de momento $\Theta_1, \Theta_2, \ldots, \Theta_m$

são encontrados igualando os m primeiros momentos da população aos primeiros m momentos amostrais e resolvendo as equações resultantes para os parâmetros desconhecidos.

Método dos Momentos – Exemplo 1

- Suponha que $X \sim exp(\lambda)$. Dada uma amostra $X_1, X_2, ..., X_n$, coletada deste modelo populacional, proponha um estimador para λ a partir do método dos momentos.
- Como temos um único parâmetro, basta igualar o primeiro momento populacional $E(X) = \dfrac{1}{\lambda}$ ao primeiro momento amostral $\dfrac{1}{n}\sum_{i=1}^{n} X_i = \overline{X}$.
- Neste caso,

$$E(X) = \frac{1}{\lambda} = \overline{X} \Longrightarrow \hat{\lambda} = \frac{1}{\overline{X}}.$$

Método dos Momentos – Exemplo 2

- Seja $X \sim N(\mu, \sigma^2)$. Dada uma amostra aleatória $X_1, X_n, \ldots, X_n$ de X, obtenha estimadores para μ e σ^2 a partiri do método dos momentos.
- Como o modelo tem dois parâmetros desconhecidos, devemos igualar os dois primeiros momentos populacionais aos dois primeiros momentos amostrais.
- Ou seja,

$$E(X) = \mu = \overline{X} \Longrightarrow \hat{\mu} = \overline{X}$$

e

$$E(X^2) = \mu^2 + \sigma^2 = \frac{1}{n}\sum_{i=1}^{n} X_i^2$$

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

Método dos Momentos – Exemplo 3

Problema específico: O tempo até a falha de um módulo eletrônico utilizado em um controlador de motor automotivo é testado a uma temperatura alta, de modo acelerar o mecanismo de falha. Suponhamos que o tempo até a falha é exponencialmente distribuído.

8 unidades são selecionadas aleatóriamente e testedas, resultando nos seguintes tempos de falha:

x	11.96	5.03	67.40	16.07	31.50	7.73	11.10	22.38

Como $\bar{x} = 21.65$, a estimação do momento de λ é

$$\hat{\lambda} = \frac{1}{\bar{x}} = \frac{1}{21.65} = 0.0462$$

Método dos Momentos

- Os estimadores de momento $\hat{\theta}_1, \ldots, \hat{\theta}_m$ são obtidos igualando-se os primeiros m momentos amostrais aos m momentos populacionais correspondentes e resolvendo para $\theta_1, \ldots, \theta_m$.

DISTRIBUIÇÃO	E[X]	VAR[X]	m
Binomial [$X \sim Bin(n,p)$]	$n.p$	$n.p.(1-p)$	1
Poisson [$X \sim Poi(\lambda)$]	λ	λ	1
Normal [$X \sim N(\mu, \sigma^2)$]	μ	σ^2	2
Uniforme [$X \sim Uni(a, b)$]	$(a+b)/2$	$(b-a)^2/12$	2
Exponencial [$X \sim Exp(\lambda)$]	$1/\lambda$	$1/\lambda^2$	1
Gamma [$X \sim Gamma(a, b)$]	$a.b$	$a.b^2$	2

Obs: n é conhecido

Método da Máxima Verossimilhança

Lei da Verossimilhança

Considere uma variável aleatória X, cujo comportamento pode ser explicado por duas hipóteses (hipóteses A e B) que se deseja comparar. Foi realizado um estudo e se obteve uma observação de X, cujo valor foi x.

O que as hipótese dizem a respeito dessa observação?

- A hipótese A implica que $X = x$ seria observado com probabilidade $p_A(x)$, enquanto

- A hipótese B implica que $X = x$ seria observado com probabilidade $p_B(x)$.

Lei da Verossimilhança

No processo de investigação científica, no entanto, o que interessa é a pergunta: "O que a observação de $X = x$ diz a respeito das hipóteses A e B?" A **Lei da Verossimilhança** afirma que a observação $X = x$ é uma evidência que favorece a hipótese A sobre a hipótese B se e somente se

$$p_A(x) > p_B(x).$$

Mais ainda, a Lei da Verossimilhança implica que a **Razão de Verossimilhança**

$$\frac{p_A(x)}{p_B(x)}$$

mede a ***força de evidência*** em favor da hipótese A sobre a hipótese B.

Lei da Verossimilhança - Exemplo

Deseja-se saber o número médio de plântulas numa parcela de regeneração em floresta nativa. Para isso se utilizou uma parcela circular de 3 m de raio (28.3 m^2). Há duas hipóteses competindo:

- Hipótese A: o número médio de plântulas na parcela é 16 (5700 ind/ha);
- Hipótese B: o número médio de plântulas na parcela é 35 (12500 ind/ha);

Lei da Verossimilhança - Exemplo

Deseja-se saber o número médio de plântulas numa parcela de regeneração em floresta nativa. Para isso se utilizou uma parcela circular de 3 m de raio (28.3 m^2). Há duas hipóteses competindo:

- Hipótese A: o número médio de plântulas na parcela é 16 (5700 ind/ha);
- Hipótese B: o número médio de plântulas na parcela é 35 (12500 ind/ha);

Lei da Verossimilhança - Exemplo

Foi medida uma parcela de regeneração e observou-se 24 plântulas (8470 ind/ha).

Neste exemplo, a variável aleatória X é o número de plântulas por parcela e a observação tomada foi de $X = 24$. Faz-se necessário um modelo (distribuição estatística) para se calcular as probabilidades de se observar $X = 24$ sob as duas hipóteses. Como se trata de dados de contagem, a distribuição *Poisson* é uma candidata "natural". Se a variável aleatória X tem distribuição Poisson, sua função de densidade é

$$P(X = x) = \frac{e^{-\mu}\mu^{x}}{x!}$$

Lei da Verossimilhança - Exemplo

onde μ (o parâmetro) é o número médio de plântulas.

Portanto as probabilidades de acordo com as hipóteses são:

- Hipótese A: $\mu = 16$

$$p_A(24) = \frac{e^{-16} 16^{24}}{24!} = 0.01437018$$

- Hipótese B: $\mu = 35$

$$p_B(24) = \frac{e^{-35} 35^{24}}{24!} = 0.01160434$$

Lei da Verossimilhança - Exemplo

Logo, a observação $X = 24$ favorece a hipótese A ($\mu = 16$) sobre a hipótese B ($\mu = 35$). A *força de evidência* em favor de A sobre B é:

$$\frac{p_A(24)}{p_B(24)} = \frac{0.01437018}{0.01160434} = 1.238345.$$

Ou seja, pode se dizer que a observação $X = 24$ é evidência que a hipótese A é aproximadamente 1.3 vêzes mais *verossímil* que a hipótese B.

Lei da Verossimilhança - Exemplo

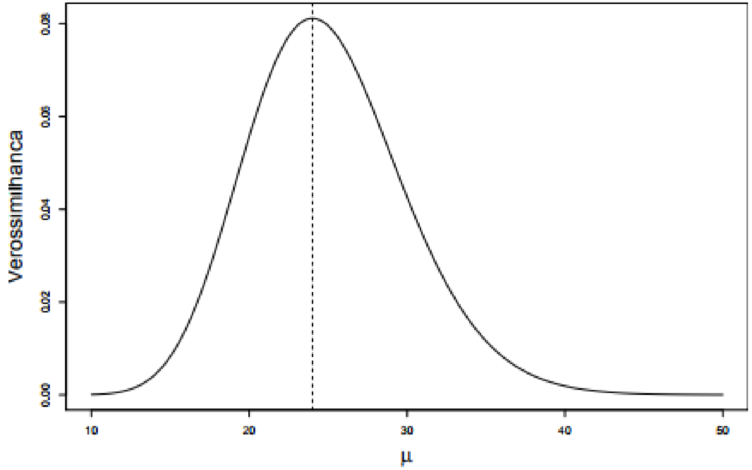


Figura 1: Função de verossimilhança da distribuição Poisson, quando se obtem uma observação $X = 24$.

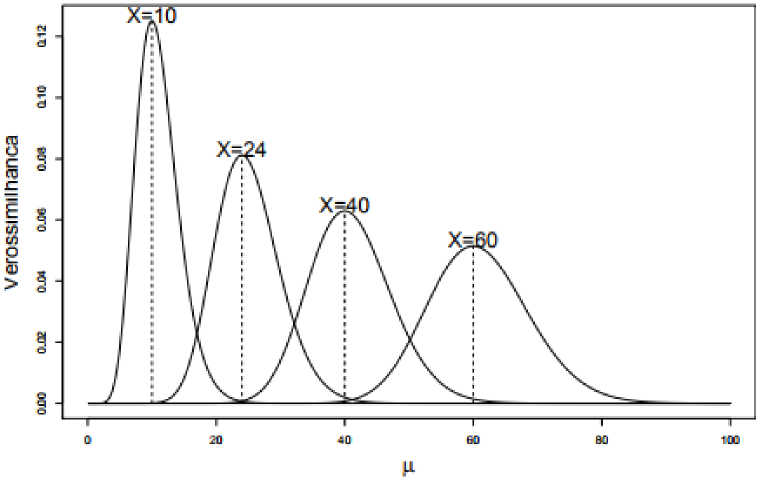


Figura 2: Funções de verossimilhança da distribuição Poisson para diferentes valores observados de X.

Log Verossimilhança Negativa

Para tornar mais fácil a manipulação matemática da verossimilhança se utiliza a função de *log-verossimilhança negativa*, que consistem aplicar a função logaritmo, geralmente logaritmo natural ou neperiano, e transformar o sinal:

$$\mathbf{L}\{\mu|X\} = -\log\left[\,\mathcal{L}\{\mu|X\}\,\right].$$

- Como o valor numérico da verossimilhança é geralmente menor que um, o logaritmo desse valor é negativo.

Log Verossimilhança Negativa

- Assim, se o valor da verossimilhança de uma amostra com muitas observações é um número positivo muito pequeno, o valor da log-verossimilhança negativa será um número positivo numa escala mais fácil de trabalhar.
- Por outro lado, o fato da transformação incluir uma mudança de sinal implica que **o comportamento da função de log-verossimilhança negativa é oposto ao comportamento da função de verossimilhança**. **Isso significa que a hipótese com maior verossimilhança terá menor log-verossimilhança negativa.**

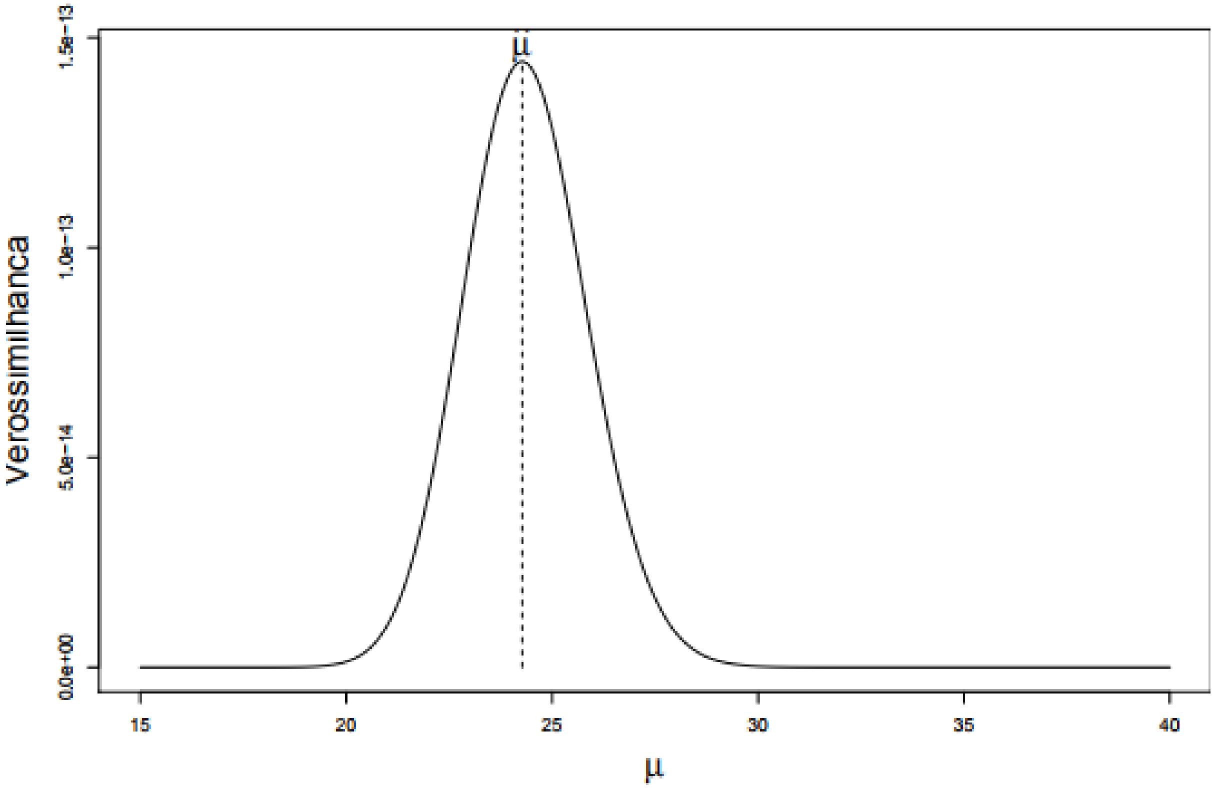
Log Verossimilhança Negativa

Tabela 2: Números de plântulas observados em 10 parcelas de regeneração natural, seguidos dos valores de *log-verossimilhança negativa* calculados segundo a distribuição de Poisson.

PARCELA (i)	NO. DE PLÂNTULAS ($X = x_i$)	LOG-VEROSSIMILHANÇA NEG. ($\mu = 16$)	($\mu = 35$)
1	24	4.2426	4.4564
2	27	5.6976	3.5631
3	23	3.8371	4.8337
4	28	6.2573	3.3400
5	26	5.1744	3.8227
6	24	4.2426	4.4564
7	17	2.3711	8.0642
8	23	3.8371	4.8337
9	24	4.2426	4.4564
10	27	5.6976	3.5631
	$\sum_{i=1}^{10}$	49.8427	49.8459

$$\mathbf{L}\{\mu|X_n\} = n\mu - \log(\mu)\sum_{i=1}^{n} x_i + \sum_{i=1}^{n} \log(x_i!)$$

Log Verossimilhança Negativa



Log Verossimilhança Negativa

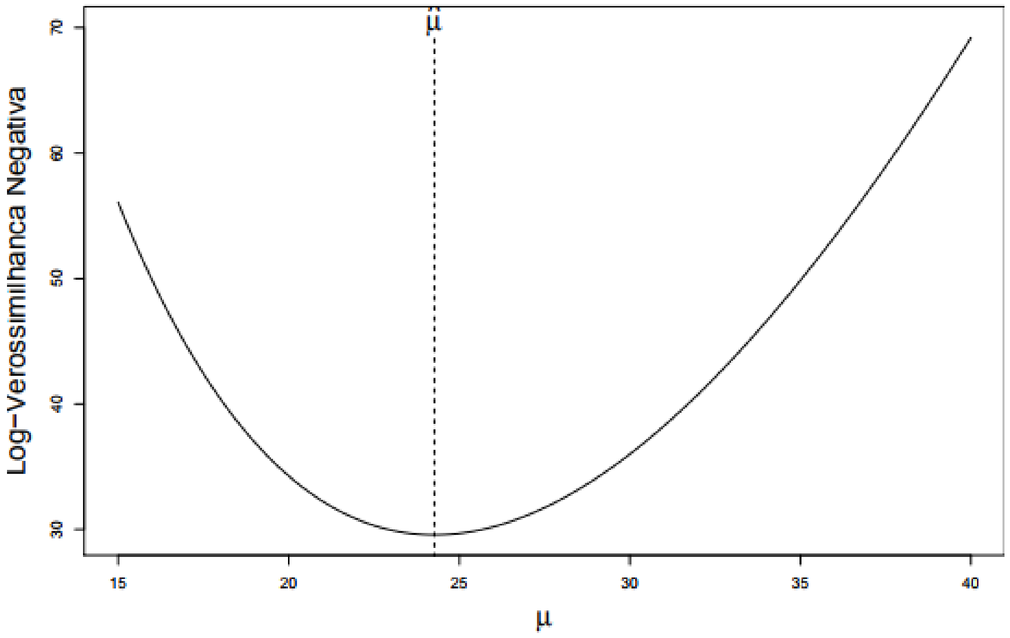


Figura 4: Função de verossimilhança e log-verossimilhança negativa da distribuição Poisson para uma amostra de tamanho 10: $X = \{24, 27, 23, 28, 26, 24, 17, 23, 24, 27\}$. A linha vertical indica a possição da estimativa de máxima verossimilhança.

Método da Máxima Verossimilhança

- **O método da Máxima Verossimilhança** consiste em estimar os parâmetros de um modelo utilizando as estimativas que tornam máximo o valor da função de verossimilhança. Isso é equivalente a encontrar o valor para o parâmetro que torna mínima a função de log-verossimilhança negativa.
- As estimativas de máxima verossimilhança são chamadas em inglês de *maximum likelihood estimates* (MLE).

Método da Máxima Verossimilhança

- Utilizando o cálculo diferencial, podemos encontrar o ponto de mínimo de uma função igualando a zero a primeira derivada da função e solucionando a expressão. **No caso da distribuição Poisson, a função log-verossimilhança negativa é:**

$$\mathbf{L}\{\mu|X_n\} \;=\; n\mu - \log(\mu)\sum_{i=1}^{n} x_i + \sum_{i=1}^{n} \log(x_i!).$$

Encontrando a primeira derivada de $\mathbf{L}\{\mu|X_n\}$ e igualando a zero temos:

$$\frac{d\mathbf{L}\{\mu|X_n\}}{d\mu} \;=\; n - \frac{\sum_{i=1}^{n} x_i}{\mu} = 0 \quad\Rightarrow\quad \widehat{\mu} = \frac{\sum_{i=1}^{n} x_i}{n}$$

Método da Máxima Verossimilhança

Portanto, a estimativa de máxima verossimilhança do parâmetro μ da distribuição de Poisson nada mais é que a média amostral. Assim no exemplo de regeneração de plântulas temos como estimativa de máxima verossimilhança:

$$\widehat{\mu} = \frac{24 + 27 + 23 + 28 + 26 + 24 + 17 + 23 + 24 + 27}{10} = 24.3$$

Método da Máxima Verossimilhança

- O Método da Máxima Verossimilhança é um método para estimação dos parâmetros a partir de uma amostra aleatória proposto por Fisher em 1912.

➢ Def: **Função de Verossimilhança**: seja $X_1,\ldots,X_n$ uma amostra aleatória com f.d.p. conjunta $f(x_1,\ldots,x_n\,;\,\theta_1,\ldots,\theta_m)$ em que $\theta_1,\ldots,\theta_m$ tem valores desconhecidos. Quando $x_1,\ldots,x_n$ são os valores observados e a f.d.p. conjunta é vista como em função de $\theta_1,\ldots,\theta_m$ esta é chamada de função de verossimilhança.

Método da Máxima Verossimilhança

- **Procedimento**: a estimativa de máxima verossimilhança de $\hat{\theta}_1, \ldots, \hat{\theta}_m$ são os valores de $\hat{\theta}_i$ que maximizam a função de verossimilhança.

- Por esse método obtém-se os valores de $\theta_1, \ldots, \theta_m$ que maximizam o valor que torna a amostra observada a "mais provável".

- Definição. Seja X uma v.a. com função densidade (ou função de probabilidade) $f_\theta(x)$, onde θ é um vetor de parâmetros desconhecido. Dada uma amostra observada $x_1, x_2, ..., x_n$ de X, a função de verossimilhança de θ é dada por:

$$L(\theta|x_1, x_2, ..., x_n) = f_\theta(x_1) f_\theta(x_2) \cdots f_\theta(x_n).$$

- Uma vez que os valores $x_1, x_2, ..., x_n$ já foram observados, os mesmos se comportam como constantes. Neste caso $L(\theta|x_1, x_2, ..., x_n)$ é uma função de θ apenas.
- Em outras palavras, o valor $L(\theta_0|x_1, x_2, ..., x_n)$ mede o quaão provável seria observarmos a amostra $x_1, x_2, ..., x_n$ caso o parâmetro real fosse de fato θ_0.
- O método da máxima verossimilhança propõe buscar o valor $\hat{\theta}$ que maximiza tal verossimilhança:

$$\hat{\theta} = \arg\max_\theta L(\theta).$$

Método da Máxima Verossimilhança – Propriedades

Consistência: as MLE são consistentes, i.e., elas *convergem em probabilidade* para o valor do parâmetro. Ou seja, para grandes amostras ($n \to \infty$) as MLE, para efeitos práticos, são não-viciadas (não enviesadas).

Eficiência Assintótica: O Teorema do Limite Inferior de Cramer-Rao afirma que, para um dado parâmetro qualquer, existe um limite inferior para a variância das estimativas não-viciadas. Para grandes amostras, as MLE atingem esse limite e, portanto, têm a menor variância possível dentre as estimativas não-viciadas.

Método da Máxima Verossimilhança – Propriedades

Normalidade Assimptótica: As MLE convergem em distribuição para distribuição Gaussiana. Para grandes amostras, os MLE tem distribuiçõa aproximadamente gaussiana.

Invariância: as MLE são invariantes sob transformações monotônicas. Por exemplo, seja $\widehat{\mu}$ uma MLE que pode ser transformada para:

- $\widehat{\theta}_1 = \log(\widehat{\mu})$,
- $\widehat{\theta}_2 = \sqrt{\widehat{\mu}}$ e
- $\widehat{\theta}_3 = e^{\widehat{\mu}}$,

então as estimativas $\widehat{\theta}_1$, $\widehat{\theta}_2$ e $\widehat{\theta}_3$ também são MLE.

Método da Máxima Verossimilhança – Propriedades

- Uma aspecto muito importante que deve ser frisado é que as três primeiras propriedades são válidas para grandes amostras.

Método da Máxima Verossimilhança – Exemplo

- Objetivo: estimar p de Bernoulli

$$f_{X_i}(x_i) = p^{x_i}(1-p)^{1-x_i}; \qquad X_i = \begin{cases} 1, & P(X_i = 1) = p \\ 0, & P(X_i = 0) = 1 - p \end{cases}$$

$$X_i \sim Bern(p); \qquad i = 1, \ldots, n$$

Método da Máxima Verossimilhança – Exemplo

- Vamos considerar $X_1, ..., X_n$ independentes
- definimos a função de verossimilhança $L(p)$

$$
\begin{aligned}
L(p) &= f_{X_1}(x_1) \times \ldots \times f_{X_n}(x_n) \\
&= p^{\sum_{i=1}^{n} x_i}(1-p)^{n-\sum_{i=1}^{n} x_i}
\end{aligned}
$$

$$
\ln(L(p)) = \left(\sum_{i=1}^{n} x_i\right) \ln(p) + \left(n + \sum_{i=1}^{n} x_i\right) \ln(1-p)
$$

Método da Máxima Verossimilhança – Exemplo

- Propósito: determinar o máximo valor de $L(p)$ ou, equivalentemente, de $\ln(l(p))$

$$\frac{\partial}{\partial p}\ln(L(p)) = \frac{\sum_{i=1}^{n} x_i}{p} - \frac{\left(n - \sum_{i=1}^{n} x_i\right)}{1-p} = 0$$

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

$\Rightarrow \hat{p} = \dfrac{\sum_{i=1}^{n} x_i}{n}$ é o estimador de máxima verossimilhança de p

Método da Máxima Verossimilhança – Exemplo

- Exemplo
  - $n = 4$
  - $(X_1, X_2, X_3, X_4) = (1, 0, 0, 1)$
  - $\hat{p} = \frac{2}{4} = \frac{1}{2} = 0.5$