Estimação Pontual

tradução do Jay Davore

índice

Primeiras ideias

Inferência Estatística tem por objetivo fazer generalizações sobre uma população com base nos valores amostrais.

Existem dois problemas básicos:

1. Estimação de parâmetros
2. Teste de hipóteses sobre parâmetros

Primeiras ideias

Inferência Estatística tem por objetivo fazer generalizações sobre uma população com base nos valores amostrais.

Existem dois problemas básicos:

1. Estimação de parâmetros
2. Teste de hipóteses sobre parâmetros

exemplo

Uma amostra de $n = 500$ pessoas de uma cidade é escolhida e a cada um é perguntado se concorda ou não com uma solução a um problema do estado. A resposta pode ser SIM (concorda) ou NÃO (discorda). Deseja-se estimar a proporção de pessoas na população favoráveis à solução proposta.

Se 300 pessoas responderam SIM à questão, uma estimativa natural para a proporção populacional seria $300/500$ ou 60%. (Supondo que a amostra é representativa).

exemplo

Uma amostra de $n = 500$ pessoas de uma cidade é escolhida e a cada um é perguntado se concorda ou não com uma solução a um problema do estado. A resposta pode ser SIM (concorda) ou NÃO (discorda).
Deseja-se estimar a proporção de pessoas na população favoráveis à solução proposta.
Se 300 pessoas responderam SIM à questão, uma estimativa natural para a proporção populacional seria $300/500$ ou 60%. (Supondo que a amostra é representativa).

exemplo

Sejam $X_1, ..., X_n$ tais que

$$X_i = \begin{cases} 1, & \text{Se a i-ésima pessoa da amostra responde SIM} \\ 0, & \text{Se a i-ésima pessoa da amostra responde NÃO} \end{cases}$$

Seja $p = P(sucesso)$ (resposta SIM), podemos definir $Y = \sum_{i=1}^{n} X_i$, sabemos que $Y \sim bin(n, p)$ queremos portanto, estimar p.

exemplo

Sejam $X_1, ..., X_n$ tais que

$$X_i = \begin{cases} 1, & \text{Se a i-ésima pessoa da amostra responde SIM} \\ 0, & \text{Se a i-ésima pessoa da amostra responde NÃO} \end{cases}$$

Seja $p = P(sucesso)$ (resposta SIM), podemos definir $Y = \sum_{i=1}^{n} X_i$, sabemos que $Y \sim bin(n, p)$ queremos portanto, estimar p.

Um possível estimador de p é

$$\hat{p} = \frac{Y_n}{n} = \frac{\sum_{i=1}^{n} X_i}{n}$$

exemplo

Sejam $X_1, ..., X_n$ tais que

$$X_i = \begin{cases} 1, & \text{Se a i-ésima pessoa da amostra responde SIM} \\ 0, & \text{Se a i-ésima pessoa da amostra responde NÃO} \end{cases}$$

Seja $p = P(sucesso)$ (resposta SIM), podemos definir $Y = \sum_{i=1}^{n} X_i$, sabemos que $Y \sim bin(n, p)$ queremos portanto, estimar p.
Um possível estimador de p é

$$\hat{p} = \frac{Y_n}{n} = \frac{\sum_{i=1}^{n} X_i}{n}$$

Se $Y_n = k$, obteremos: $\hat{p} = \frac{k}{n}$ como uma estimativa de p.

exemplo

Sejam $X_1, ..., X_n$ tais que

$$X_i = \begin{cases} 1, & \text{Se a i-ésima pessoa da amostra responde SIM} \\ 0, & \text{Se a i-ésima pessoa da amostra responde NÃO} \end{cases}$$

Seja $p = P(sucesso)$ (resposta SIM), podemos definir $Y = \sum_{i=1}^{n} X_i$, sabemos que $Y \sim bin(n, p)$ queremos portanto, estimar p.
Um possível estimador de p é

$$\hat{p} = \frac{Y_n}{n} = \frac{\sum_{i=1}^{n} X_i}{n}$$

Se $Y_n = k$, obteremos: $\hat{p} = \frac{k}{n}$ como uma estimativa de p.

observe que $\hat{p}$ não é uma v.a., já k/n é um número ou uma estimativa.

exemplo

Sejam $X_1, ..., X_n$ tais que

$$X_i = \begin{cases} 1, & \text{Se a i-ésima pessoa da amostra responde SIM} \\ 0, & \text{Se a i-ésima pessoa da amostra responde NÃO} \end{cases}$$

Seja $p = P(sucesso)$ (resposta SIM), podemos definir $Y = \sum_{i=1}^{n} X_i$, sabemos que $Y \sim bin(n, p)$ queremos portanto, estimar p.
Um possível estimador de p é

$$\hat{p} = \frac{Y_n}{n} = \frac{\sum_{i=1}^{n} X_i}{n}$$

Se $Y_n = k$, obteremos: $\hat{p} = \frac{k}{n}$ como uma estimativa de p.

observe que $\hat{p}$ não é uma v.a., já k/n é um número ou uma estimativa.

Definição de estimador

Seja uma amostra $(X_1, ..., X_n)$ de uma v.a. que descreve uma característica de interesse da população. Seja θ o parâmetro que desejamos estimar (exemplo a média μ ou a variância σ^2)

Um estimador T de θ é qualquer função das observações amostrais, i.e, $T = g(X_1, ..., X_n)$.

Uma estimativa é o valor assumido pelo estimador em uma particular amostra.

Definição de estimador

Seja uma amostra $(X_1, ..., X_n)$ de uma v.a. que descreve uma característica de interesse da população. Seja θ o parâmetro que desejamos estimar (exemplo a média μ ou a variância σ^2)

Um estimador T de θ é qualquer função das observações amostrais, i.e, $T = g(X_1, ..., X_n)$.

Uma estimativa é o valor assumido pelo estimador em uma particular amostra.

Estimador pontual

- Um estimador pontual de um parâmetro θ, é um número que pode ser considerado como um valor "sensato"para θ.
- Um estimador pontual pode ser obtido selecionando uma "estatística conveniente" e calculando o seu valor a partir dos dados amostrais

estimador não viesado

- um estimador pontual $\hat{\theta}$ é dito ser um estimador não viesado de θ se $E(\hat{\theta}) = \theta$, para cada possível valor de θ.
- se $\hat{\theta}$ for viesado, a diferença $E(\hat{\theta} - \theta)$ é chamado de Viês de θ

estimador pontual

As f.d.p.'s de um estimador viesado $\hat{\theta}_1$ e um estimador não viesado $\hat{\theta}_2$ para um parâmetro θ:

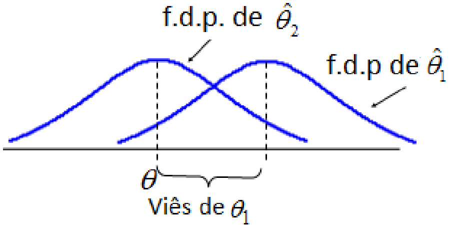


Figura: f.d.p's de estimadores

estimador pontual

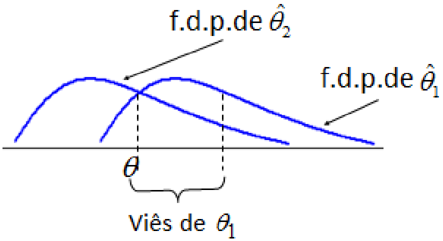


Figura: f.d.pťs de estimadores

estimador não viesado de uma binomial

quando X é uma v.a. binomial com parâmetros n e p, a proporção amostral $\hat{p} = \frac{X}{n}$ é um estimador não viesado de p.

Foi visto na aula anterior que:

$$\hat{p} \sim N(p, \frac{p(1-p)}{n})$$

estimador não viesado de uma binomial

quando X é uma v.a. binomial com parâmetros n e p, a proporção amostral $\hat{p} = \frac{X}{n}$ é um estimador não viesado de p.

Foi visto na aula anterior que:

$$\hat{p} \sim N(p, \frac{p(1-p)}{n})$$

isto ajuda a verificar que $\hat{p}$ "em média acerta" p. Portanto dizemos que $\hat{p}$ é um estimador não viesado para p.

estimador não viesado de uma binomial

quando X é uma v.a. binomial com parâmetros n e p, a proporção amostral $\hat{p} = \frac{X}{n}$ é um estimador não viesado de p.

Foi visto na aula anterior que:

$$\hat{p} \sim N(p, \frac{p(1-p)}{n})$$

isto ajuda a verificar que $\hat{p}$ "em média acerta" p. Portanto dizemos que $\hat{p}$ é um estimador não viesado para p.

Observe que a diferença entre $\hat{p}$ e p tenderá a ser pequena, pois quando $n \to \infty$, $Var(\hat{p}) \to 0$

estimador não viesado de uma binomial

quando X é uma v.a. binomial com parâmetros n e p, a proporção amostral $\hat{p} = \frac{X}{n}$ é um estimador não viesado de p.

Foi visto na aula anterior que:

$$\hat{p} \sim N(p, \frac{p(1-p)}{n})$$

isto ajuda a verificar que $\hat{p}$ "em média acerta"p. Portanto dizemos que $\hat{p}$ é um estimador não viesado para p.

Observe que a diferença entre $\hat{p}$ e p tenderá a ser pequena, pois quando $n \to \infty$, $Var(\hat{p}) \to 0$

Princípio do estimador não viesado

Na escolha de diferentes estimadores de θ, selecione-se o que é não viesado

estimador não viesado da média

Seja $X_1, X_2, ..., X_n$ uma a.a. de uma distribuição com média μ. Então $\bar{X}$ é um estimador não viesado de μ.

Esta propriedade foi provada na aula anterior $(E(\bar{X}) = \mu)$

estimador não viesado da média

Seja $X_1, X_2, ..., X_n$ uma a.a. de uma distribuição com média μ. Então $\bar{X}$ é um estimador não viesado de μ.

Esta propriedade foi provada na aula anterior $(E(\bar{X}) = \mu)$

Se em adição a distribuição é contínua e simétrica, então $\tilde{X}$ (mediana) e outras médias centradas, são também estimadores não viesados de μ.

estimador não viesado da média

Seja $X_1, X_2, ..., X_n$ uma a.a. de uma distribuição com média μ. Então $\bar{X}$ é um estimador não viesado de μ.

Esta propriedade foi provada na aula anterior $(E(\bar{X}) = \mu)$

Se em adição a distribuição é contínua e simétrica, então $\tilde{X}$ (mediana) e outras médias centradas, são também estimadores não viesados de μ.

estimador não viesado da variância

Seja $X_1, X_2, ..., X_n$ uma a.a. de uma distribuição com média μ e variância σ^2. Então o estimador:

$$\hat{\sigma^2} = S^2 = \frac{\sum(X_i - \bar{X})^2}{n-1}$$

é um estimador não viesado.

estimador não viesado da variância

A variância populacional é definida como

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

Seja

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

um possível estimador para σ^2 extraído de uma a.a. de tamanho n.

estimador não viesado da variância

A variância populacional é definida como

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

Seja

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

um possível estimador para σ^2 extraído de uma a.a. de tamanho n.

provaremos que este estimador é viesado.

estimador não viesado da variância

A variância populacional é definida como

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

Seja

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

um possível estimador para σ^2 extraído de uma a.a. de tamanho n. provaremos que este estimador é viesado.

estimador não viesado da variância

Escrevemos:

$$\begin{aligned}\sum_{i=1}^{n}(X_i-\bar{X})^2 &= \sum_{i=1}^{n}(X_i-\mu+\mu-\bar{X})^2\\ &= \sum_{i=1}^{n}(X_i-\mu)^2-2\sum_{i=1}^{n}(X_i-\mu)(\bar{X}-\mu)+\sum_{i=1}^{n}(\bar{X}-\mu)^2\\ &= \sum_{i=1}^{n}(X_i-\mu)^2-n(\bar{X}-\mu)^2\end{aligned}$$

$(\bar{X}-\mu)$ é constante, e

$$\sum_{i=1}^{n}(X_i-\mu)=n(\bar{X}-\mu)$$

estimador não viesado da variância

$$E(\hat{\sigma}^2) = \frac{1}{n}\left[\sum_{i=1}^{n} E(X_i - \mu)^2 - nE(\bar{X} - \mu)^2\right]$$

$$= \frac{1}{n}\left[\sum_{i=1}^{n} Var(X_i) - n\,Var(\bar{X})\right]$$

$$= \frac{1}{n}\left[n\sigma^2 - n\frac{\sigma^2}{n}\right] = \frac{n-1}{n}\sigma^2$$

vemos que este estimador é viesado e temos que $E\left(\frac{n}{n-1}\hat{\sigma}^2\right) = \sigma^2$
Logo, se definirmos

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$$

é não viesado

Exemplo

Suponha que X, tempo de reação a certo estímulo, possua distribuição uniforme no intervalo $[0, \theta]$ com θ desconhecido. Deseja-se estimar θ com base em uma a.a. $X_1, ..., X_n$ de tempos de reação. Sendo θ o maior tempo possível de reação, podemos considerar $\hat{\theta}_1 = max\{X_1, ..., X_n\}$ como estimador de θ.

Se $n = 5$ e $x_1 = 4,2$, $x_2 = 1,7$, $x_3 = 2,4$, $x_4 = 3,9$, $x_5 = 1,3$, temos que $max(x_1, .., x_5) = 4,2$

demonstra-se que

$$E(\hat{\theta}_1) = \frac{n}{n+1}\theta < \theta$$

Exemplo

Suponha que X, tempo de reação a certo estímulo, possua distribuição uniforme no intervalo $[0, \theta]$ com θ desconhecido. Deseja-se estimar θ com base em uma a.a. $X_1, ..., X_n$ de tempos de reação. Sendo θ o maior tempo possível de reação, podemos considerar $\hat{\theta}_1 = max\{X_1, ..., X_n\}$ como estimador de θ.

Se $n = 5$ e $x_1 = 4,2$, $x_2 = 1,7$, $x_3 = 2,4$, $x_4 = 3,9$, $x_5 = 1,3$, temos que $max(x_1, .., x_5) = 4,2$

demonstra-se que

$$E(\hat{\theta}_1) = \frac{n}{n+1}\theta < \theta$$

portanto,

$$\hat{\theta}_2 = \frac{n+1}{n} max\{X_1, ..., X_n\}$$

será um estimador não viesado de θ

Exemplo

Suponha que X, tempo de reação a certo estímulo, possua distribuição uniforme no intervalo $[0, \theta]$ com θ desconhecido. Deseja-se estimar θ com base em uma a.a. $X_1, ..., X_n$ de tempos de reação. Sendo θ o maior tempo possível de reação, podemos considerar $\hat{\theta}_1 = max\{X_1, ..., X_n\}$ como estimador de θ.

Se $n = 5$ e $x_1 = 4,2$, $x_2 = 1,7$, $x_3 = 2,4$, $x_4 = 3,9$, $x_5 = 1,3$, temos que $max(x_1, .., x_5) = 4,2$

demonstra-se que

$$E(\hat{\theta}_1) = \frac{n}{n+1}\theta < \theta$$

portanto,

$$\hat{\theta}_2 = \frac{n+1}{n} max\{X_1, ..., X_n\}$$

será um estimador não viesado de θ

Estimadores consistentes

Uma sequência $\{T_n\}$ de estimadores de um parâmetro θ é consistente se para todo $\epsilon > 0$:

$$P\left(|T_n - \theta| > \epsilon\right) \to 0, \quad n \to \infty$$

Estimadores consistentes

Uma sequência $\{T_n\}$ de estimadores de um parâmetro θ é consistente se:

$$\lim_{n\to\infty} E(T_n) = \theta$$
$$\lim_{n\to\infty} Var(T_n) = 0$$

da definição é possível ver que $\hat{p}$ e $\bar{X}$ são estimadores consistentes de p e de μ respectivamente

Estimadores consistentes

Uma sequência $\{T_n\}$ de estimadores de um parâmetro θ é consistente se:

$$\lim_{n\to\infty} E(T_n) = \theta$$
$$\lim_{n\to\infty} Var(T_n) = 0$$

da definição é possível ver que $\hat{p}$ e $\bar{X}$ são estimadores consistentes de p e de μ respectivamente

ENVVM

Entre todos os estimadores de θ que são não viesado, escolhe-se aquele que têm variância mínima.
O estimador resultante $\hat{\theta}$ é chamado de estimador não viesado de variância mínima (ENVVM) de θ

Representação de dois estimadores não viesados

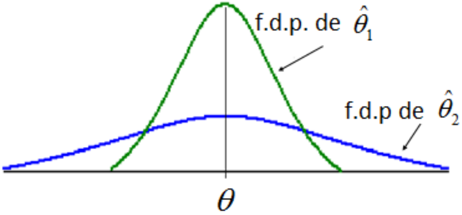


Figura: f.d.p's de estimadores não viesados

ENVVM para uma distribuição Normal

Seja $X_1, X_2, ..., X_n$ uma a.a. de uma distribuição normal com parâmetros μ e σ.
Então o estimador $\hat{\mu} = \bar{X}$ é um ENVVM para μ.

Estimador viesado que é preferível ao ENVVM

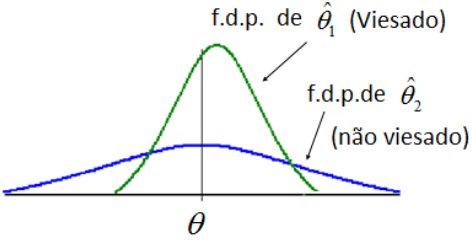


Figura: f.d.p's de um ENVVM e um estimador viesado

estimadores para μ

- Se a a.a. vem de uma distribuição normal, então $\bar{X}$ é o melhor estimador, dado que têm a variância mínima entre todos os estimadores não viesados.
- Se a a.a. vem de uma distribuição de Cauchy, então $\tilde{X}$ é um bom estimador (o ENVVM não é conhecido) e, $\bar{X}$ e $maxX$ não são bons estimadores
- se a a.a. vem de uma uniforme, o melhor estimador é $\bar{X}_e$ (média dos valores extremos), mas ele é influenciavel por valores extremos.

Erro Padrão de um estimador

- Obtida a distribuição amostral de um estimador, podemos calcular a variância,
- se não pudermos obter a distribuição exata, usamos uma aproximação, se estiver disponível. No caso de $\bar{X}$, e a variância do estimador será a variância dessa aproximação.

Por exemplo, para $\bar{X}$, obtida de uma amostra tamanho n, temos que

$$var(\bar{X}) = \frac{\sigma^2}{n}$$

na qual σ^2 é a variância da v.a. X definida sobre a população.
À raiz quadrada dessa variância chamaremos de Erro Padrão de $\bar{X}$ e o denotaremos por

$$EP(\bar{X}) = \frac{\sigma}{\sqrt{n}}$$

Definição de erro padrão

Se $\hat{\theta}$ for o estimador de uma parâmetro θ, chamamos de erro padrão de $\hat{\theta}$ ao seu desvio padrão:

$$EP(\hat{\theta}) = \sigma_{\hat{\theta}} = \sqrt{V(\hat{\theta})}$$

Se o erro padrão envolve parâmetros desconhecidos, cujos valores podem ser estimados, eles serão substituidos em $\sigma_{\hat{\theta}}$ e teremos o erro padrão estimado do estimador, e será denotado por

$$\hat{EP}(\hat{\theta}) = \hat{\sigma}_{\hat{\theta}} \quad \text{ou} \quad s_{\hat{\theta}}$$

no caso do $EP(\bar{X})$ depende de σ, que em geral é desconhecida, obtemos então o erro padrão estimado de $\bar{X}$ por

$$ep(\bar{X}) = \hat{EP}(\bar{X}) = \frac{s}{\sqrt{n}}$$

Exemplo

Foi visto que S^2 é um estimador não viesado para σ^2. É possível demonstrar no caso que $X_1, ..., X_n$ são observações de uma distribuição $N(\mu, \sigma^2)$, que

$$Var(S^2) = \frac{2\sigma^4}{n-1}$$

Como $E(S^2) = \sigma^2$ e $\lim_{n\to\infty} Var(S^2) = 0$, podemos afirmar que S^2 é um estimador consistente para σ^2.

Exemplo

No caso do estimador $\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2$, vimos que $E(\hat{\sigma}^2) = \sigma^2(1 - \frac{1}{n})$, de forma que $\lim_{n\to\infty} E(\hat{\sigma}^2) = \sigma^2$, temos que:

$$Var(\hat{\sigma}^2) = \left(\frac{n-1}{n}\right)^2 Var(S^2) = \frac{n-1}{n^2}(2\sigma^4)$$

o que mostra que $Var(\hat{\sigma}^2) \to 0$ quando $n \to \infty$. Logo, $\hat{\sigma}^2 = \hat{\sigma}_n^2$ também é consistente para σ^2

obtemos que

$$Var(\hat{\sigma}^2) < \frac{2\sigma^4}{n-1} = Var(S^2)$$

Usando o critério da variância menor, $\hat{\sigma}^2$ é melhor estimador para σ^2

Exemplo

No caso do estimador $\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2$, vimos que $E(\hat{\sigma}^2) = \sigma^2(1 - \frac{1}{n})$, de forma que $\lim_{n\to\infty} E(\hat{\sigma}^2) = \sigma^2$, temos que:

$$Var(\hat{\sigma}^2) = \left(\frac{n-1}{n}\right)^2 Var(S^2) = \frac{n-1}{n^2}(2\sigma^4)$$

o que mostra que $Var(\hat{\sigma}^2) \to 0$ quando $n \to \infty$. Logo, $\hat{\sigma}^2 = \hat{\sigma}^2_n$ também é consistente para σ^2
obtemos que

$$Var(\hat{\sigma}^2) < \frac{2\sigma^4}{n-1} = Var(S^2)$$

Usando o critério da variância menor, $\hat{\sigma}^2$ é melhor estimador para σ^2

Estimadores Eficientes

Definição:
Se $\hat{\theta}$ e $\hat{\theta}^{`}$ são dois estimadores não viesados de um mesmo parâmetro θ, e ainda

$$var(\hat{\theta}) < var(\hat{\theta}^{`})$$

Então diz-se que $\hat{\theta}$ é mais eficiente que $\hat{\theta}^{`}$

Erro Quadrático Médio (EQM)

Seja $e = T - \theta$ o erro amostral cometido ao estimar o parâmetro θ da distribuição X por meio de $T = g(X_1, ..., X_n)$ baseado na amostra $X_1, ..., X_n$.

o EQM é definido por:

$$EQM(T;\theta) = E(e^2) = E(T - \theta^2)$$

Erro Quadrático Médio (EQM)

Seja $e = T - \theta$ o erro amostral cometido ao estimar o parâmetro θ da distribuição X por meio de $T = g(X_1, ..., X_n)$ baseado na amostra $X_1, ..., X_n$.

o EQM é definido por:

$$EQM(T;\theta) = E(e^2) = E(T - \theta^2)$$

Erro Quadrático Médio (EQM)

Temos que :

$$
\begin{aligned}
EQM(T;\theta) &= E(T - E(T) + E(T) - \theta)^2 \\
&= E(T - E(T))^2 + 2E[(T - E(T))(E(T) - \theta)] + \\
&\quad E(E(T) - \theta)^2 \\
&= E(T - E(T))^2 + E(E(T) - \theta)^2
\end{aligned}
$$

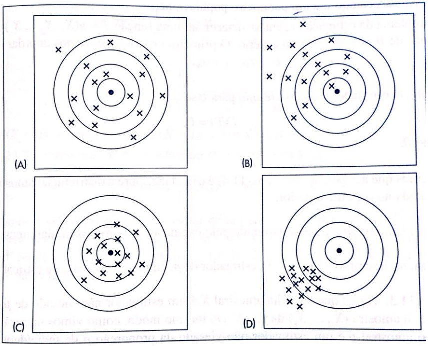
como $E(T) - \theta$ é uma constante, e $E(T - E(T)) = 0$ podemos escrever:

$$EQM(T;\theta) = Var(T) + V^2$$

onde $V = V(T) = E(T) - \theta$ (viês de T)

Representação das propriedades dos estimadores

Figura 11.1: Resultados de 15 tiros dados por 4 rifles.



Desse modo, podemos descrever cada arma da seguinte maneira:

Arma A: não-viesada, pouco acurada e baixa precisão.
Arma B: viesada, pouco acurada e baixa precisão.
Arma C: não-viesada, muito acurada e boa precisão.
Arma D: viesada, pouco acurada e alta precisão.

Representação do EQM

Vemos, portanto, que um estimador preciso tem variância pequena, mas pode ter EQM grande.

Figura 11.2: Representação gráfica para o *EQM*.

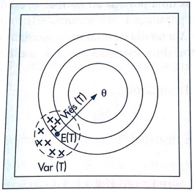


Figura: EQM

Definição

Seja $X_1, X_2, ..., X_n$ uma a.a. de un f.m.p ou f.d.p $f(x)$. Para $k = 1, 2, ...$, o k-ésimo momento populacional , ou o k-ésimo momento da distribuição $f(x)$ é $E(X^k)$
O k-ésimo momento é:

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

Estimadores dos momentos

Seja $X_1, X_2, ..., X_n$ uma a.a. de uma distribuição com f.m.p ou f.d.p. $f(x; \theta_1, ..., \theta_m)$ onde $\theta_1, ..., \theta_m$ são parâmetros cujos valores são desconhecidos. Então os estimadores dos momentos $\theta_1, ..., \theta_m$ são obtidos igualando os primeiros m momentos amostrais aos correspondentes primeiros m momentos populacionais e resolvemos para $\theta_1, ..., \theta_m$.

Método dos momentos

A média populacional é o primeiro momento.
Dizemos que $\hat{\theta}_1, ..., \hat{\theta}_n$ são estimadores obtidos pelo método dos momentos se eles forem as soluções das equações

$$m_k = \mu_k, \quad k = 1, 2, ..., r$$

Método dos momentos

Se $X \sim N(\mu, \sigma^2)$ teremos as seguintes relações para os dois primeiros momentos populacionais:

$$E(X) = \mu, \quad E(X^2) = \sigma^2 + \mu^2$$

do que se obtém:

$$\mu = E(X), \quad \sigma^2 = E(X^2) - (E(X))^2$$

no caso dos momentos amostrais temos:

$$m_1 = \frac{1}{n}\sum_{i=1}^{n} X_i = \bar{X}$$

$$m_2 = \frac{1}{n}\sum_{i=1}^{n} X_i^2$$

Método dos momentos

Se $X \sim N(\mu, \sigma^2)$ teremos as seguintes relações para os dois primeiros momentos populacionais:

$$E(X) = \mu, \quad E(X^2) = \sigma^2 + \mu^2$$

do que se obtém:

$$\mu = E(X), \quad \sigma^2 = E(X^2) - (E(X))^2$$

no caso dos momentos amostrais temos:

$$m_1 = \frac{1}{n} \sum_{i=1}^{n} X_i = \bar{X}$$

$$m_2 = \frac{1}{n} \sum_{i=1}^{n} X_i^2$$

Método dos momentos

OS estimadores obtidos pelo método dos momentos são:

$$\hat{\mu}_M = m_1 = \bar{X}$$

$$\hat{\sigma}^2_M = m_2 - m_1^2 = \frac{1}{n}\sum_{i=1}^{n} X_i^2 - \bar{X}^2 = \hat{\sigma}^2$$

Verossimilhança

Seja $X_1, X_2, ..., X_n$ uma a.a. com f.m.p conjunta ou f.d.p. conjunta

$$f(x_1, ..., x_n; \theta_1, ..., \theta_m)$$

onde os parâmetros $\theta_1, ..., \theta_m$ tem valores desconhecidos.
quando $x_1, ..., x_n$ são valores amostrais observados e f é considerada como uma função de $\theta_1, ..., \theta_m$, ela é chamada de "função de verossimilhança".

EMV

Os estimadores de máxima verossimilhança (EMV) $\hat{\theta}_1, ..., \hat{\theta}_m$ são aqueles valores dos θ_i's que maximizam a função de verossimilhança, tal que:

$$f(x_1, ..., x_n; \hat{\theta}_1, ..., \hat{\theta}_m) \geq f(x_1, ..., x_n; \theta_1, ..., \theta_m)$$

para todo $(\theta_1, ..., \theta_m)$
quando os X_i's são substituidos no lugar dos x_i's, resultam nos estimadores de máxima verossimilhança.

Exemplo

Suponha que temos n provas de Bernoulli com probabilidade de sucesso p $0 < p < 1$, e $X =$número de sucessos.

tomaremos como estimador, aquele valor de p que torna a amostra observada a máis provável de ocorrer

Exemplo

Suponha que $n = 3$ e se obtiveram 2 sucessos e um fracasso. A função de verossimilhança é:

$$L(p) = P(2 \text{ sucessos e } 1 \text{ fracasso}) = p^2(1-p)$$

maximizamos essa função em relação a p:

$$L`(p) = 2p(1-p) - p^2 = 0 \Rightarrow p(2-3p) = 0$$

de onde tem-se $p = 0$ ou $p = 2/3$.

Exemplo

Suponha que $n = 3$ e se obtiveram 2 sucessos e um fracasso. A função de verossimilhança é:

$$L(p) = P(2 \text{ sucessos e } 1 \text{ fracasso}) = p^2(1-p)$$

maximizamos essa função em relação a p:

$$L`(p) = 2p(1-p) - p^2 = 0 \Rightarrow p(2-3p) = 0$$

de onde tem-se $p = 0$ ou $p = 2/3$. Vemos que o ponto máximo é $\hat{p} = 2/3$ que é o estimador de máxima verossimilhança de p.

Exemplo

Suponha que $n = 3$ e se obtiveram 2 sucessos e um fracasso. A função de verossimilhança é:

$$L(p) = P(2 \text{ sucessos e } 1 \text{ fracasso}) = p^2(1-p)$$

maximizamos essa função em relação a p:

$$L`(p) = 2p(1-p) - p^2 = 0 \Rightarrow p(2-3p) = 0$$

de onde tem-se $p = 0$ ou $p = 2/3$. Vemos que o ponto máximo é $\hat{p} = 2/3$ que é o estimador de máxima verossimilhança de p.

Exemplo

De modo geral, o EMV(p) de uma binomial é

$$\hat{p}_{MV} = \frac{X}{n}$$

Observe que no caso geral a função de verossimilhança é

$$L(p) = p^x (1-p)^{n-x}$$

que é a probabilidade de se obter x sucessos e $n-x$ fracassos.

Exemplo

De modo geral, o EMV(p) de uma binomial é

$$\hat{p}_{MV} = \frac{X}{n}$$

Observe que no caso geral a função de verossimilhança é

$$L(p) = p^x(1-p)^{n-x}$$

que é a probabilidade de se obter x sucessos e $n-x$ fracassos. O máximo dessa função ocorre quando $l(p) = ln(L(p))$

$$l(p) = xlog(p) + (n-x)log(1-p)$$

derivando e igualando a zero temos:

$$\hat{p}_{MV} = \frac{x}{n}$$

Exemplo

De modo geral, o EMV(p) de uma binomial é

$$\hat{p}_{MV} = \frac{X}{n}$$

Observe que no caso geral a função de verossimilhança é

$$L(p) = p^x(1-p)^{n-x}$$

que é a probabilidade de se obter x sucessos e $n-x$ fracassos. O máximo dessa função ocorre quando $l(p) = ln(L(p))$

$$l(p) = xlog(p) + (n-x)log(1-p)$$

derivando e igualando a zero temos:

$$\hat{p}_{MV} = \frac{x}{n}$$

Procedimento

O procedimento consiste em

1. obter a função de verossimilhança $L(\theta; X_1, ..., X_n)$, que depende dos parâmetros desconhecidos e dos valores amostrais
2. maximizar a função de verossimilhança (FV) ou o logaritmo da FV $l(\theta; X_1, ..., X_n) = Ln\left(L(\theta; X_1, ..., X_n)\right)$. Dependendo da conveniência da situação

Procedimento

A função de verossimilhança (FV) é:

$$L(\theta; x_1, ..., x_n) = f(x_1; \theta)...f(x_n; \theta)$$

que é uma função de θ. O EMV de θ é o valor $\hat{\theta}_{MV}$ que maximiza $L(\theta; x_1, ..., x_n)$

O princípio da invariância

Sejam $\hat{\theta}_1, ..., \hat{\theta}_m$ os EMV dos parâmetros $\theta_1, ..., \theta_m$, então o EMV de qualquer função $h(\theta_1, ..., \theta_m)$ destes parâmetros, é a função $h(\hat{\theta}_1, ..., \hat{\theta}_m)$ dos EMV.

Exemplo

Suponha que a v.a. X tenha distribuição exponencial com parâmetro $\alpha > 0$ desconhecido. Queremos obter o EMV de α .

$$f(x;\alpha) = \begin{cases} \frac{1}{\alpha} e^{-x/\alpha}, & \text{se } x \geq 0 \\ 0 & \text{se} x < 0 \end{cases}$$

a verossimilhança é dada por:

$$L(\alpha|x) = \left(\frac{1}{\alpha}\right)^n e^{-\sum_{i=1}^n x_i/\alpha}$$

Exemplo

A função de log-verossimilhança é:

$$l(\alpha|x) = -nlog(\alpha) - \frac{\sum_{i=1}^{n} x_i}{\alpha}$$

Derivando e igualando a zero obtemos:

$$\hat{\alpha}_{MV} = \frac{\sum_{i=1}^{n} x_i}{n}$$

lembre que $E(X) = \alpha$ e o estimador é a média amostral.

Exemplo

A função de log-verossimilhança é:

$$l(\alpha|x) = -nlog(\alpha) - \frac{\sum_{i=1}^{n} x_i}{\alpha}$$

Derivando e igualando a zero obtemos:

$$\hat{\alpha}_{MV} = \frac{\sum_{i=1}^{n} x_i}{n}$$

lembre que $E(X) = \alpha$ e o estimador é a média amostral.

Propriedades

Sob condições gerais na distribuição conjunta da amostra, quando o tamanho amostral n é grande, o EMV de qualquer parâmetro θ é aproximadamente não viesado:

$$\left[E(\hat{\theta}) \approx \theta\right]$$

e tem variância que é tão pequena como a que pode ser obtida por qualquer estimador:

$$\text{EMV}\hat{\theta} \approx \text{ENVVM de}\theta$$