Estimação Pontual

Elementos de Probabilidades e Estatística

Sumário

1. Inferência Estatística

Exemplo

Considere-se os estudantes da FCUL entre os quais há uma proporção θ que pratica desporto. Escolhem-se ao acaso, com reposição, n estudantes. Seja $n = 10$. Se θ fosse conhecido, por exemplo $\theta = 0.3$, a probabilidade de encontrar x praticantes $(0 \leq x \leq 10)$ nesse grupo de 10 pessoas é

$$\binom{10}{x}(0.3)^x(0.7)^{10-x} \quad \text{(problema de probabilidade)}$$

Na prática, quase sempre θ é desconhecido. Neste caso, olharemos para a proporção de praticantes na amostra retirada, $x/10$, para tirar conclusões sobre a proporção de praticantes na população donde a amostra foi retirada. (problema de inferência)

Conceitos básicos

Variável aleatória de interesse

Não passa de uma característica ou atributo crucial para o conhecimento do fenómeno aleatório em estudo.

Exemplo: A resistência de certo tipo de mola; idade do estudante de Física; a intenção de voto de certo eleitor.

População e unidade estatística

- Uma **população** é um conjunto de objetos ou indivíduos que têm em comum pelo menos uma característica de interesse.
- A cada elemento da população damos o nome de **unidade estatística**.

Exemplo: Todas as molas produzidas do referido tipo; idade de todos os estudantes de Física; intenção de voto dos eleitores de um país.

Conceitos básicos (cont.)

Amostra

Uma **amostra** é um subconjunto de valores da população. Tem que ser selecionada de forma aleatória e deve ser representativa da população de onde foi retirada.

$(x_1, \ldots, x_n)-$ amostra observada de dimensão n.

Amostragem

O processo de seleção de uma amostra da população de modo a estimar algum aspeto de interesse da mesma é designada **amostragem**.

Estatística descritiva

A estatística descritiva engloba um conjunto de métodos gráficos e numéricos que possibilitam a apresentação de forma concisa das informações essenciais contidas nos dados.

Inferência estatística

Conjunto de técnicas que visam a partir de um conjunto representativo de dados, fazer estimativas e previsões com vista a caracterizar um conjunto mais amplo de dados e retirar conclusões.

Conceitos básicos (cont.)

Nosso interesse

- Conhecer a população X corresponde a conhecer a sua função de distribuição $F_X(x)$.
- O que é equivalente a conhecer a f.d.p. caso X seja contínua, ou a f.m.p. caso X seja discreta.

Casos a considerar

- Desconhecemos completamente $F_X(x)$ sabendo-se apenas se X é discreta ou contínua;
- Admite-se que $F_X(x)$ pertence a determinada família de distribuições, mas com parâmetros desconhecidos.

Conceitos básicos (cont.)

Objetivos da Inferência Estatística

- Estimar $F_X(x)$ ou estimar os parâmetros de $F_X(x)$ conhecendo a sua forma;
- Fazer testes em relação aos parâmetros ou em relação à forma de $F_X(x)$.

Estimação de parâmetros:

- pontual;
- intervalar.

Testes de hipóteses:

- sobre parâmetros;
- sobre a forma de $F_X(x)$.

Conceitos básicos (cont.)

Amostragem Aleatória

Cada elemento da amostra é obtido totalmente ao acaso da população X e de forma independente dos outros elementos.

Amostra Aleatória

A sucessão $\underline{X} = (X_1, \ldots, X_n)$ constituem uma **amostra aleatória (a.a.)** de dimensão **n** da população X se forem **independentes** e **identicamente distribuídas** a X (**i.i.d.**).

Amostra

A uma observação particular da a.a. $\underline{X} = (X_1, \ldots, X_n)$ dá-se o nome de **amostra** e representa-se por $\underline{x} = (x_1, \ldots, x_n)$.

Ciências ULisboa

Conceitos básicos (cont.)

Caracterização da amostra aleatória

Pelo facto de a a.a. ser constituída por n v.a. i.i.d. a X, a caracterização probabilística da a.a. $\underline{X} = (X_1, \ldots, X_n)$ faz-se sem grande dificuldade. Com efeito, temos para os casos:

- discreto - f.p. conjunta de $\underline{X}$

$$\begin{aligned} P(\underline{X} = \underline{x}) &= P(X_1 = x_1, \ldots, X_n = x_n) \\ &\overset{X_i \text{ indep.}}{=} \prod_{i=1}^{n} P(X_i = x_i) \\ &\overset{X_i \sim X}{=} \prod_{i=1}^{n} P(X = x_i) \end{aligned}$$

- contínuo - f.d.p. conjunta de $\underline{X}$

$$\begin{aligned} f_{\underline{X}}(\underline{x}) &= f_{X_1, \ldots, X_n}(x_1, \ldots, x_n) \\ &\overset{X_i \text{ indep.}}{=} \prod_{i=1}^{n} f_{X_i}(x_i) \overset{X_i \sim X}{=} \prod_{i=1}^{n} f_X(x_i) \end{aligned}$$

Conceitos básicos (cont.)

Estatística

Uma **estatística** é uma variável aleatória que é função unicamente da amostra aleatória e que não depende de parâmetros desconhecidos. Denota-se usualmente por $T = T(\underline{X}) = T(X_1, \ldots, X_n)$.

Exemplos	Estatística	Valor observado da estatística
média amostral	$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$	$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$
var. amostral corrigida	$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$	$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$
var. amostral não corrig.	$(S^{'})^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2$	$(s^{'})^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2$
mínimo da a.a.	$X_{(1)} = \min(X_1, \ldots, X_n)$	$x_{(1)} = \min(x_1, \ldots, x_n)$
máximo da a.a.	$X_{(n)} = \max(X_1, \ldots, X_n)$	$x_{(n)} = \min(x_1, \ldots, x_n)$

Conceitos básicos (cont.)

Parâmetro

O termo parâmetro refere-se a uma quantidade da qual depende a distribuição de uma estatística ou de uma v.a. (e.g. μ, σ, etc.).

Espaço paramétrico

Corresponde ao conjunto de todos os valores possíveis para o parâmetro desconhecido θ e é frequentemente representado por Θ.

Modelo paramétrico

Seja X uma variável aleatória com função massa/densidade de probabilidade $f_X(x;\theta)$ pertencente a uma família $\mathcal{F} = \{f_X(x;\theta), \theta \in \Theta\}$, onde o valor do parâmetro θ é desconhecido.

Exemplo (Modelo paramétrico):

$$\{\text{Poisson}(\theta) : \theta \in \mathbb{R}^+\} \quad \text{ou} \quad \{e^{-\theta}\theta^x/x!, x = 0, 1, 2, \cdots : \theta \in \mathbb{R}^+\}$$

$$\{\text{exponencial}(\theta) : \theta \in \Theta\} \quad \text{ou} \quad \{\theta e^{-\theta x}, x \geq 0 : \theta \in \Theta\}$$

Conceitos básicos (cont.)

Estimador e Estimativa

Uma estatística $T = T(X_1, \ldots, X_n)$ diz-se um **estimador** do parâmetro desconhecido θ caso tome valores exclusivamente no espaço paramétrico Θ.

O valor concreto do estimador numa amostra particular $(x_1, \ldots, x_n)$, isto é, $t = T(x_1, \ldots, x_n)$, chama-se **estimativa** de θ.

Exemplo: Seja $X_1, \ldots, X_n$ uma amostra aleatória de uma população com distribuição $N(\mu, \sigma)$.

Alguns estimadores possíveis para μ são

1. $\bar{X}$ - média amostral
2. mediana amostral
3. moda amostral
4. $\frac{X_{(1)}+X_{(n)}}{2}$ - centro do intervalo de variação amostral
5. ...

Ciências ULisboa

Exemplo: Admita que vai inquirir *n* condutores/as quanto à sua preferência (ou não) por motores elétricos e que as respostas possíveis (admissíveis) neste inquérito são:

- Sim (1), prefiro moto elétrico;
- Não (2), não prefiro motor elétrico.

Procure identificar: a v.a. de interesse, a respetiva distribuição; o parâmetro desconhecido; o modelo e o espaço paramétrico; uma estimativa e um estimador do parâmetro desconhecido.

V.a. de interesse e sua distribuição

$$\begin{aligned} X &= \text{preferência do/a condutor/a inquirido/a} \\ &= \begin{cases} 1 \text{ (prefere motor elétrico), com prob. } \theta \\ 0 \text{ (não prefere motor elétrico), com prob. } (1-\theta) \end{cases} \end{aligned}$$

$X \sim \text{Bernoulli}(\theta)$

Conceitos básicos (cont.)

Parâmetro desconhecido e espaço paramétrico

$\theta = P(X = 1) = P(\text{prefere motor elétrico})\ \Theta = [0, 1]$

Modelo paramétrico

$\{\text{Bernoulli}(\theta) : \theta \in \Theta\}$ ou $\{\theta^x(1-\theta)^{1-x}, x = 0, 1 : \theta \in \Theta\}$

A.a. e amostra
$\underline{X} = (X_1, \ldots, X_n)$ a.a. de dimensão n proveniente da população X
$\underline{x} = (x_1, \ldots, x_n)$ onde $\underline{x} \in \{0, 1\}^n$

Estimativa de θ

Candidata: um valor razoável para θ é a proporção observada de condutores/as que preferem motor elétrico, i.e., $T(\underline{x}) = \frac{1}{n}\sum_{i=1}^n x_i = \overline{x}$.

Estimador de θ

Candidato: $T(\underline{X}) = \frac{1}{n}\sum_{i=1}^n X_i = \overline{X}$.
Verificações

- $T(\underline{X})$ só depende de $\underline{X}$;
- $T(\underline{X})$ toma valores em $\{0, \frac{1}{n}, \frac{2}{n}, \ldots, \frac{n-1}{n}, 1\} \subset [0, 1]$

Portanto $T(\underline{X})$ é um estimador de θ.

Conceitos básicos (cont.)

Exemplo: Imagina que estás a estudar o tempo de reação de uma pessoa ao ver uma luz acender-se (por exemplo, num laboratório com um cronómetro). Queres estimar o tempo médio de reação μ da população (ou seja, de todas as pessoas).

Parâmetro desconhecido:
O tempo médio de reação μ da população — não o conheces, é o que queres estimar.

Estimador (fórmula geral):
Vais usar a média amostral $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$, onde $X_1, X_2, \ldots, X_n$ são os tempos de reação medidos numa amostra de n pessoas.

$\bar{X}$ é estimador do tempo médio de reação μ. Ela é uma função dos dados amostrais e varia de amostra para amostra.

Conceitos básicos (cont.)

Estimativa (valor numérico):
Suponhamos que mediste os tempos de reação de 11 pessoas e obtiveste os seguintes valores (em segundos):

$$0.22, 0.25, 0.21, 0.24, 0.23, 0.26, 0.22, 0.23, 0.25, 0.20, 0.45$$

A média (estimador aplicado aos dados) é:

$$\bar{X} = \frac{2.76}{11} \approx 0.2509 \text{ segundos.}$$

Esse 0.2509 segundos é a tua estimativa do tempo médio de reação μ.

Como avaliar estimadores?

Como avaliar estimadores?

- Há um grande número de critérios para avaliar e comparar estimadores de acordo com diferentes propriedades desejáveis.
- Podemos avaliar a **exatidão** de um estimador recorrendo ao seu valor esperado. A exatidão de um estimador reporta-se à proximidade do seu valor esperado do verdadeiro valor do parâmetro desconhecido θ.
- Podemos avaliar a **precisão** de um estimador recorrendo à sua variância.
- Podemos avaliar em conjunto as duas caraterísticas usando o **erro quadrático médio**.

Estimador centrado; estimador enviesado; enviesamento de um estimador

Um estimador T de um parâmetro θ diz-se **centrado** se e só se

$$E(T) = \theta, \ \forall \theta \in \Theta.$$

Um estimador T de um parâmetro θ diz-se **não centrado ou enviesado** se

$$\exists \theta \in \Theta : E(T) \neq \theta.$$

O estimador T de um parâmetro θ, possui **enviesamento** dado por

$$\text{bias}_\theta(T) = E(T) - \theta.$$

O **erro quadrático médio** de um estimador T de um parâmetro θ é definido por

$$\begin{aligned} EQM_\theta(T) &= E[(T-\theta)^2] \\ &= V(T) + [E(T) - \theta]^2 \\ &= V(T) + [\text{bias}_\theta(T)]^2. \end{aligned}$$

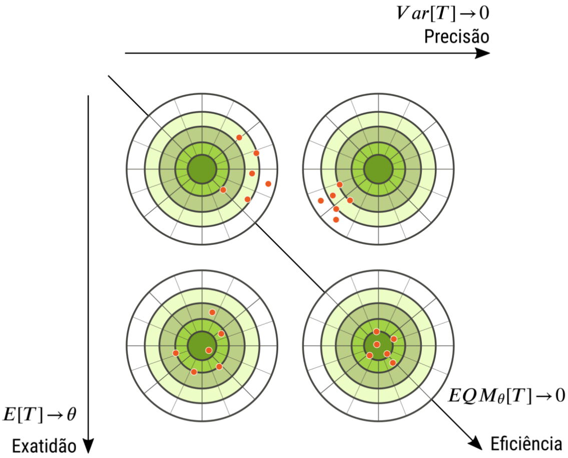

Figura: Imagem retirada de
`https://web.tecnico.ulisboa.pt/paulo.soares/pe/`

Conceitos básicos (cont.)

- Não basta que um estimador de θ seja centrado (com exatidão elevada) para garantir estimativas rigorosas. Estas serão tanto mais rigorosas quanto menos o estimador se dispersar em torno do verdadeiro valor do parâmetro desconhecido θ.
- $EQM_\theta(T)$ quantifica a dispersão esperada do estimador T em torno do verdadeiro valor do parâmetro desconhecido θ e combina duas parcelas:
  - $V(T)$ que diz respeito à variabilidade do estimador e dá ideia da precisão do mesmo;
  - $[\text{bias}_\theta]^2$ que se reporta ao quadrado do viés do estimador e reflete a exatidão do mesmo.
- Um estimador será tanto *melhor* quanto menor for o seu EQM.

Conceitos básicos (cont.)

Exemplo: Considere que X é uma v.a. de interesse com distribuição arbitrária, valor esperado μ e variância σ^2.

(a) Prove que a média da a.a. $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$, e a variância corrigida da a.a., $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2 = \frac{1}{n-1}\left[\sum_{i=1}^{n} X_i^2 - n(\bar{X})^2\right]$, são estimadores centrados de μ e σ^2 (respetivamente).

(b) Demonstre que a variância não corrigida da a.a., $(S^{'})^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2 = \frac{1}{n}\left[\sum_{i=1}^{n} X_i^2 - n(\bar{X})^2\right] = \frac{n-1}{n}S^2$, é um estimador enviesado de σ^2. Calcule e comente o respetivo enviesamento.

Conceitos básicos (cont.)

Eficiência relativa de estimadores; estimador mais eficiente

Sejam $T_1 = T_1(\underline{X})$ e $T_2 = T_2(\underline{X})$ dois estimadores do parâmetro desconhecido θ. Então, a eficiência de T_1 - com respeito a T_2 na estimação de θ - é dada por

$$e_\theta(T_1, T_2) = \frac{EQM_\theta(T_2)}{EQM_\theta(T_1)}.$$

Assim sendo, se

$$e_\theta(T_1, T_2) > 1 \Leftrightarrow EQM_\theta(T_2) > EQM_\theta(T_1),$$

diremos que o estimador T_1 é mais eficiente que T_2 na estimação de θ.

Ciências ULisboa

Conceitos básicos (cont.)

Exemplo: Num estudo prévio ao lançamento no mercado de uma nova pilha de *pacemaker* foram colocadas algumas questões acerca da sua duração (em milhares de dias) a um engenheiro. Estudos anteriores (embora com outros tipos de pilhas) levam a crer que tal v.a. possui distribuição uniforme $(0, \theta)$, onde o parâmetro θ é positivo, desconhecido e representa a idade máxima da pilha.

Calcule a eficiência relativa de $X_{(n)} = \max\{(X_1, \ldots, X_n)\}$ com respeito a $2\bar{X}$ no que se refere à estimação do parâmetro θ. Para o efeito, atente que $E(X_{(n)}) = \frac{n}{n+1}\theta$ e $V(X_{(n)}) = \frac{n}{(n+2)(n+1)^2}\theta^2$. Qual dos dois estimadores de θ é mais eficiente?

V.a. de interesse

$X =$ duração da pilha

Distribuição

$X \sim \text{uniforme}(0, \theta)$

Parâmetro desconhecido e espaço paramétrico

θ

$\Theta = \mathbb{R}^+$

A.a.

$\underline{X} = (X_1 \ldots, X_n)$ a.a. de dimensão n proveniente da população X

Estimador de θ

$X_{(n)} = \max\{(X_1, \ldots, X_n)\}$

Erro quadrático médio de $X_{(n)}$

$$\begin{aligned}
EQM_\theta(X_{(n)}) &= V(X_{(n)}) + [bias_\theta[X_{(n)}]]^2 \\
&= V(X_{(n)}) + [E[X_{(n)}] - \theta]^2 \\
&= \frac{n}{(n+2)(n+1)^2}\theta^2 + \left(\frac{n}{n+1}\theta - \theta\right)^2 \\
&= \frac{2}{(n+2)(n+1)}\theta^2
\end{aligned}$$

Outro estimador de θ

$2\bar{X}$

Erro quadrático médio de $2\bar{X}$

Dado que $E(\bar{X}) = E(X)$, $V(\bar{X}) = \frac{V(X)}{n}$, $E(X) \overset{form.}{=} \frac{\theta}{2}$ e $V(X) \overset{form.}{=} \frac{\theta^2}{12}$ temos

$$\begin{aligned}
EQM_\theta(2\bar{X}) &= V(2\bar{X}) + [bias_\theta(2\bar{X})]^2 \\
&= V(2\bar{X}) + [E(2\bar{X}) - \theta]^2 \\
&= \frac{2^2}{n} V(X) + [2E(X) - \theta]^2 \\
&= \frac{2^2}{n} \frac{\theta^2}{12} + \left(2 \times \frac{\theta}{2} - \theta\right)^2 \\
&= \frac{1}{3n} \theta^2
\end{aligned}$$

Conceitos básicos (cont.)

Eficiência relativa de $X_{(n)}$ com respeito a $2\bar{X}$

$$
\begin{aligned}
e_\theta[X_{(n)}, 2\bar{X}] &= \frac{EQM_\theta(2\bar{X})}{EQM_\theta(X_{(n)})} \\
&= \frac{\frac{1}{3n}\theta^2}{\frac{2}{(n+2)(n+1)}\theta^2} \\
&= \frac{(n+2)(n+1)}{6n},
\end{aligned}
$$

que constitui o termo geral de uma sucessão monótona não decrescente, cujos dois primeiros termos são iguais a 1.

Comentário

Tendo em conta a expressão de $e_\theta[X_{(n)}, 2\bar{X}]$ podemos afirmar que

- $X_{(n)}$ e $2\bar{X}$ são igualmente eficientes, para $n = 1, 2$,
- $X_{(n)}$ é mais eficiente que $2\bar{X}$, para $n > 2$.

Curiosamente, $X_{(n)}$ não é estimador centrado de θ, ao contrário de $2\bar{X}$.

Como obter estimadores ou estimativas?

Conceitos básicos (cont.)

Como obter estimadores ou estimativas?

Diversos métodos de estimação estão disponíveis:

- Método da máxima verosimilhança
- Método dos momentos
- Método dos mínimos quadrados
- ...

3. Método da Máxima Verosimilhança

Método da Máxima Verosimilhança

O método da máxima verosimilhança é uma técnica usada para estimar parâmetros desconhecidos de um modelo estatístico, a partir de uma amostra observada.

Imagine que você tem dados, mas não sabe exatamente qual é a "melhor" distribuição que os gerou. Suponha que desconfia que os dados vêm de uma distribuição conhecida (como a normal, binomial, Poisson, etc.), mas não conhece os valores dos parâmetros (por exemplo, a média e o desvio padrão no caso da normal).

O que o método da máxima verosimilhança faz é o seguinte:

Escolher os valores dos parâmetros que tornam os dados observados mais prováveis.

Ou seja, dado que já tens os dados, o método pergunta:

"Qual valor do parâmetro mais provavelmente teria gerado exatamente esses dados?"

Método da Máxima Verosimilhança

Ideia

O parâmetro θ é desconhecido, mas $\underline{x} = (x_1, , \ldots, x_n)$ são conhecidos. Então, qual é a melhor aproximação para θ? O número que maximiza a probabilidade/densidade de obter a amostra observada. Ou seja, esse valor de θ é o mais compatível com os dados observados.

Exemplo: Seja $X \sim$ Bernoulli(p), $P(X=1) = p$ e $P(X=0) = 1-p$. O parâmetro p é desconhecido.

A.a.: $(X_1, \ldots, X_{10})$, $\quad X_i \overset{iid}{\sim}$ Bern(p)

Amostra particular:

$$x_1 = x_4 = x_7 = x_8 = x_9 = 0$$
$$x_2 = x_3 = x_5 = x_6 = x_{10} = 1$$

Função de verosimilhança desta amostra:

$$\begin{aligned} L(p|x_1, \ldots, x_{10}) &= f_X(x_1, p) \times \cdots \times f_X(x_{10}, p) \\ &= p^5 \times (1-p)^5, \quad 0 \leq p \leq 1 \end{aligned}$$

Qual o valor de p que maximiza L?

$$\begin{aligned} \frac{dL}{dp} = \frac{d(p^5(1-p)^5)}{dp} &= 5p^4(1-p)^5 - 5p^5(1-p)^4 \\ &= 5p^4(1-p)^4(1-2p) = 0 \end{aligned}$$

$\Rightarrow p = 0 \quad \vee \quad p = 1 \quad \vee \quad p = 1/2$

p	0		1/2		1
L'	0	$+$	0	-	0
L	0	$\nearrow$	máx	$\searrow$	0

Portanto a estimativa de máxima verosimilhança (m.v.) de p com base nesta amostra é $\hat{p} = 1/2$.

Definição: Função de verosimilhança

Seja X uma v.a. com função massa/densidade de probabilidade $f(x,\theta)$, onde θ é um parâmetro desconhecido. Sejam $(x_1,\ldots,x_n)$ os valores observados de uma a.a. de dimensão n.

A **função de verosimilhança da amostra** é:

$$L(\theta|x_1,\ldots,x_n) = f(x_1,\theta)\times\cdots\times f(x_n,\theta) = \prod_{i=1}^{n} f(x_i,\theta), \quad \theta\in\Theta.$$

A **estimativa de máxima verosimilhança** de θ é o valor $\hat{\theta}$ que maximiza L. i.e., tal que

$$\hat{\theta}:\ L(\hat{\theta}|x_1,\ldots,x_n) = \max_{\theta\in\Theta} L(\theta|x_1,\ldots,x_n).$$

A função de verosimilhança dá ideia de quão plausível é o valor θ para o parâmetro desconhecido, caso se tenha recolhido a amostra $\underline{x}$.

Exemplo: Seja $X \sim \text{Bern}(p)$. Sabemos que

$$f_X(x) = P(X = x) = p^x(1-p)^{1-x}, \quad 0 < p < 1.$$

Dada uma amostra $(x_1, \dots, x_n)$, temos que

$$\begin{aligned} L(p) &\equiv L(p|x_1, \dots, x_n) \\ &= f_X(x_1) \times \cdots \times f_X(x_n) \\ &= \prod_{i=1}^{n} p^{x_i}(1-p)^{1-x_i} \\ &= p^{\sum_{i=1}^{n} x_i}(1-p)^{\sum_{i=1}^{n}(1-x_i)} \\ &= p^k(1-p)^{n-k}, \end{aligned}$$

com $k = \sum_{i=1}^{n} x_i$ é número de sucessos na amostra e $n-k$ é o número de insucessos.

Observação importante

Para uma função $f > 0$ qualquer, f e $\log f$ têm máximo e mínimo nos mesmos pontos (pois log é estritamente crescente em $\mathbb{R}^+$). Assim, em vez de determinar p que maximiza L pode-se determinar p que maximiza $\log L$, o que envolve cálculos mais simples. A função $\log L$ é designada por log-verosimilhança.

$$\log L(\theta|x_1 \ldots, x_n) \quad - \text{(função de log-verosimilhança)}$$

$$\hat{\theta}: \; \log L(\hat{\theta}|x_1, \ldots, x_n) = \max_{\theta \in \Theta} \log L(\theta|x_1, \ldots, x_n) \quad - \text{(estimativa de MV)}$$

Exemplo (cont.):

- $\log L(p) = \log\left(p^k(1-p)^{n-k}\right) = k\log p + (n-k)\log(1-p)$

$\Rightarrow \frac{d\log L(p)}{dp} = \frac{k}{p} - \frac{n-k}{1-p} = \frac{k(1-p)-(n-k)p}{p(1-p)} = 0$

$\stackrel{p\neq 0,1}{\Rightarrow} k - np = 0 \Rightarrow p = \frac{k}{n} = \frac{\sum_{i=1}^n x_i}{n} = \bar{x}$

- $\frac{d^2\log L(p)}{dp^2} = -\frac{k}{p^2} - \frac{n-k}{(1-p)^2} < 0, \quad \forall\, 0 < p < 1$

A estimativa de m.v. é $\hat{p} = \frac{\sum_{i=1}^n x_i}{n} = \bar{x}$ (valor numérico).

O estimador de m.v. é $EMV(p) = \hat{P} = \frac{\sum_{i=1}^n X_i}{n} = \bar{X}$ (variável aleatória).

Método da Máxima Verosimilhança (cont.)

Observações

- O estimador de MV de θ, denotado por $EMV(\theta)$, obtém-se substituindo $x_1, \ldots, x_n$ por $X_1, \ldots, X_n$, na expressão geral de $\hat{\theta}$ = estimativa de MV.
- Maximizar $L \Leftrightarrow$ Maximizar $\log L$.
- Cálculos com $\log L$ são normalmente mais simples do que os cálculos com L.

Exemplo: Estimação por MV do parâmetro λ da exponencial

Considere-se $X \sim \text{Exp}(\lambda)$, $\lambda > 0$, $f_X(x) = \lambda e^{-\lambda x}$, $x > 0$. Dada uma amostra $(x_1, \ldots, x_n)$ a função de verosimilhança é:

$$L(\lambda) \equiv L(\lambda | x_1, \ldots, x_n) = f(x_1) \times \cdots \times f(x_n) = \prod_{i=1}^{n} (\lambda e^{-\lambda x_i})$$

$$= \lambda^n e^{-\lambda \sum_{i=1}^{n} x_i}, \ \lambda > 0, \ x_i > 0$$

$$\log L(\lambda) = n \log \lambda - \lambda \sum_{i=1}^{n} x_i$$

Determinação do ponto de máximo:

$$\begin{cases} \frac{d \log L(\lambda)}{d\lambda} = 0 \\ \frac{d^2 \log L(\lambda)}{d\lambda^2} < 0 \end{cases} \Leftrightarrow \begin{cases} \frac{n}{\lambda} - \sum_{i=1}^{n} x_i = 0 \Leftrightarrow n - \lambda \sum_{i=1}^{n} x_i = 0 \\ -\frac{n}{\lambda^2} < 0, \quad \forall \lambda > 0 \end{cases}$$

A estimativa de m.v. de λ é $\hat{\lambda} = \frac{n}{\sum_{i=1}^{n} x_i} = \frac{1}{\bar{x}}$

O estimador de m.v. de λ é $\Lambda = EMV(\lambda) = \frac{n}{\sum_{i=1}^{n} X_i} = \frac{1}{\bar{X}}$

Método da Máxima Verosimilhança (cont.)

Propriedade de invariância dos estimadores de MV

Se $EMV(\theta)$ (ou $\hat{\theta}$) é estimador de MV do parâmetro θ, então $g(EMV(\theta))$ é o estimador de MV de $g(\theta)$, isto é,

$$EMV(g(\theta)) = g(EMV(\theta)).$$

Ciências ULisboa

Exemplo: Seja $X_1, \ldots, X_n$ uma a.a. de uma população X com distribuição Exp(λ). Determine o estimador de MV de $P(X > 2)$.

$$P(X > 2) = \int_2^{+\infty} f_X(x)\,dx = \int_2^{+\infty} \lambda e^{-\lambda x}\,dx = \left[-e^{-\lambda x}\right]_2^{+\infty} = -e^{-\infty} + e^{-2\lambda} = e^{-2\lambda}$$

$$\Rightarrow EMV(P(X > 2)) = EMV(e^{-2\lambda}) = e^{-2EMV(\lambda)} = e^{-2/\bar{X}}.$$

Método da Máxima Verosimilhança (cont.)

Exemplo: Considere uma urna com bolas brancas e pretas na proporção de 3/1 desconhecendo-se, no entanto, qual a cor dominante. Seja p a probabilidade de sair uma bola preta numa extração.

Qual é a estimativa de MV de p se, ao extrairmos com reposição três bolas da urna, encontrássemos

(a) uma bola preta?

(b) duas bolas pretas?

V.a. de interesse

$$X = \text{indicador de extração de bola preta}$$
$$= \begin{cases} 1, & \text{se sair bola preta} \\ 0, & \text{se sair bola branca} \end{cases}$$

Distribuição de X

$X \sim \text{Bernoulli}(p)$

F.m.p de X

$P(X = x) = p^x (1 - p)^{1-x}, \quad x = 0, 1$

Parâmetro desconhecido e espaço paramétrico

$p = P(X = 1) = P(\text{sair bola preta})$

Uma vez que a urna possui bolas brancas e pretas na proporção de 3 para 1, o conjunto de valores possíveis de p é $\Theta = \{\frac{1}{4}, \frac{3}{4}\}$.

Com efeito: $p = \frac{1}{4}$, se houver 3 vezes mais bolas brancas que pretas; $p = \frac{3}{4}$, se houver 4 vezes mais bolas pretas que brancas.

Método da Máxima Verosimilhança (cont.)

Amostra

Lidamos com $\underline{x} = (x_1, \ldots, x_n)$, uma amostra de dimensão n proveniente da população X. Em particular,

$\underline{x} \quad : \quad n = 3$

$\sum_{i=1}^{3} x_i =$ número de bolas pretas nas 3 extrações $= \begin{cases} (a) & 1 \text{ bola preta} \\ (b) & 2 \text{ bolas pretas} \end{cases}$

Obtenção da estimativa de MV de p

Será representada por $\hat{p}$ e $L(\hat{p}|\underline{x}) = \max_{p \in \Theta} L(p|\underline{x})$, onde $L(p|\underline{x})$ representa a função de verosimilhança. Assim, $\hat{p}$ é o valor mais plausível/verosímil para o parâmetro desconhecido p, tendo em conta a amostra $\underline{x}$.

Método da Máxima Verosimilhança (cont.)

Função de verosimilhança

$$
\begin{aligned}
L(p|\underline{x}) &= P(\underline{X} = \underline{x}) \\
&\overset{X_i \text{ indep}}{=} \prod_{i=1}^{n} P(X_i = x_i) \\
&\overset{X_i \sim X}{=} \prod_{i=1}^{n} P(X = x_i) \\
&= \prod_{i=1}^{n} [p^{x_i}(1-p)^{1-x_i}] \\
&\overset{n=3}{=} p^{\sum_{i=1}^{3} x_i}(1-p)^{3-\sum_{i=1}^{3} x_i}, \quad p \in \Theta = \{\frac{1}{4}, \frac{3}{4}\}
\end{aligned}
$$

Maximização e concretização

Dado que Θ é um conjunto discreto, não devemos recorrer à técnica usual de maximização que passa por calcular a função de log-verosimilhança, a 1a e 2a derivadas de $\ln L(p|\underline{x})$, resolver a equação de verosimilhança, etc.

De facto, a estimativa de MV de p obtém-se calculando os vários valores de $L(p|\underline{x})$, para $p \in \Theta$, e identificando o ponto de máximo - ou seja, faz-se por pesquisa ponto por ponto.

Ciências ULisboa

Método da Máxima Verosimilhança (cont.)

	$\sum_{i=1}^{3} x_i$	$\sum_{i=1}^{3} x_i$
$L(p \mid \underline{x})$	1	2
$p = \frac{1}{4}$	$(1/4)^1(1-1/4)^{3-1} = \frac{9}{64}$	$(1/4)^2(1-1/4)^{3-2} = \frac{3}{64}$
$p = \frac{3}{4}$	$(3/4)^1(1-3/4)^{3-1} = \frac{3}{64}$	$(3/4)^2(1-3/4)^{3-2} = \frac{9}{64}$

A inspeção da tabela anterior leva a concluir que

$$\hat{p} = \begin{cases} \text{(a) } \frac{1}{4}, & \text{se } \sum_{i=1}^{3} x_i = 1 \text{ (bola preta), pois } L\left(\frac{1}{4} \mid \underline{x}\right) > L\left(\frac{3}{4} \mid \underline{x}\right) \\ \text{(b) } \frac{3}{4}, & \text{se } \sum_{i=1}^{3} x_i = 2 \text{ (bolas pretas), já que } L\left(\frac{3}{4} \mid \underline{x}\right) > L\left(\frac{1}{4} \mid \underline{x}\right) \end{cases}$$

Método da Máxima Verosimilhança (cont.)

O estimador de MV pode não ser único...

Suponha que $X_1, \ldots, X_n$ é uma amostra aleatória de uma distribuição uniforme no intervalo $(\theta, \theta + 1)$, e o valor do parâmetro θ é desconhecido $(-\infty < \theta < +\infty)$. Qual o estimador de MV de θ?

A f.d.p de X é $f_X(x) = \begin{cases} 1, & \theta \leq x \leq \theta + 1 \\ 0, & \text{caso contrário.} \end{cases}$ A função de verosimilhança é:

$$L(\theta) = \begin{cases} 1, & \theta \leq x_i \leq \theta + 1 \quad (i = 1, \ldots, n) \\ 0, & \text{caso contrário.} \end{cases}$$

A condição que $\theta \leq x_i$ para $i = 1, \ldots, n$ é equivalente à condição que $\theta \leq \min(x_1, \ldots, x_n)$. Analogamente, a condição que $x_i \leq \theta + 1$ para $i = 1, \ldots, n$ é equivalente à condição que $\theta \geq \max(x_1, \ldots, x_n) - 1$. Então, podemos reescrever a função de verosimilhança da forma:

$$L(\theta) = \begin{cases} 1, & \max(x_1, \ldots, x_n) - 1 \leq \theta \leq \min(x_1, \ldots, x_n) \\ 0, & \text{caso contrário.} \end{cases}$$

Assim, podemos selecionar qualquer valor no intervalo $[\max(x_1, \ldots, x_n) - 1, \min(x_1, \ldots, x_n)]$ como estimativa de MV para θ. Então, o estimador de MV não é unicamente especificado neste exemplo.

Referências

- Manuel Cabral Morais (2020): Probabilidade e Estatística: Teoria, Exemplos & Exercícios. IST Press, 1a edição.