



IND 1115

Inferência Estatística

Aula 9
Outubro 2004

Mônica Barros

monica@ele.puc-rio.br

1



Conteúdo

- Diferença entre Probabilidade e Estatística
- Amostra Aleatória
- Objetivos da Estatística
- Distribuição Amostral
- Estimação Pontual
- Estimação Bayesiana X Clássica
- Método de Máxima Verossimilhança

monica@ele.puc-rio.br

2

Diferença entre Probabilidade e Estatística



- A partir de agora começamos realmente a falar de Estatística.
- Os capítulos anteriores lidavam com Probabilidade. Qual a diferença?
- **Em Probabilidade, a densidade (ou função de probabilidade) era inteiramente conhecida.**
- Em Estatística, teremos uma amostra aleatória de uma distribuição com certos **parâmetros desconhecidos**, e procuraremos descobrir alguma coisa sobre estes parâmetros.

monica@ele.puc-rio.br

3

Amostra Aleatória (a.a.)



- É apenas um conjunto de variáveis aleatórias iid (independentes e identicamente distribuídas).
- Se X_1, X_2, \dots, X_n formam uma a.a. então, em particular, todas as variáveis têm a mesma densidade ou função de probabilidade, e portanto suas médias são todas iguais (o mesmo ocorre com suas variâncias).

monica@ele.puc-rio.br

4

Objetivos - Estatística



- A distribuição da amostra é conhecida exceto por alguns parâmetros que buscamos estimar.
- Objetivo: obter maneiras de **encontrar estimadores** ("chutes") destes parâmetros. Estes estimadores serão **pontuais** (e começaremos a estudar um importante método de estimação pontual hoje) ou **por intervalos** (nas próximas aulas).

Objetivos - Estatística



- Também é preciso ter uma idéia clara das **propriedades** desejáveis **destes estimadores**, e saber, segundo algum critério, se o estimador encontrado é bom ou ruim.
- Finalmente, em Estatística estamos interessados também em **testar hipóteses** sobre parâmetros desconhecidos.

Distribuição Amostral



- Uma "estatística" é qualquer função das observações numa amostra aleatória.
- Por exemplo, duas das estatísticas mais usadas são \bar{X} (a média amostral) e S^2 (a variância amostral).
- Já vimos que:

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

Distribuição Amostral



- Dada uma amostra aleatória X_1, X_2, \dots, X_n com uma densidade (ou função de probabilidade), podemos tentar encontrar a densidade da média e da variância amostral, e usá-las para inferir sobre a média e variância verdadeiras (e desconhecidas) de X_1, X_2, \dots, X_n .

Estimação Pontual



- ❑ Problemas de estimação de parâmetros surgem freqüentemente em Ciências e Engenharia. Por exemplo, muitas vezes desejamos estimar os seguintes parâmetros:
- ❑ a média de uma população,
- ❑ a variância ou desvio padrão de uma população,
- ❑ a proporção de itens numa população que pertencem a uma classe de interesse,
- ❑ a diferença entre as médias de duas populações.

Estimação Pontual



- ❑ Como estimar estas quantidades? Alguns estimadores razoáveis nestas situações são:
- ❑ a média amostral,
- ❑ a variância ou desvio padrão amostrais,
- ❑ a proporção de itens na amostra que pertencem à classe de interesse,
- ❑ a diferença entre as médias amostrais de duas amostras independentes, cada uma representando uma das populações.

Estimação Pontual



- ❑ X_1, X_2, \dots, X_n variáveis aleatórias.
- ❑ x_1, x_2, \dots, x_n valores observados das variáveis aleatórias.
- ❑ Seja X uma variável aleatória. com densidade $f(x, \theta)$, onde θ é um parâmetro, e $\theta \in \Omega$.
- ❑ O conjunto Ω é chamado de **espaço paramétrico**.
- ❑ Objetivo: estimar θ .

Estimação Pontual



- ❑ A densidade de X , $f(x, \theta)$, tem uma forma conhecida, **exceto pelo parâmetro θ** que varia no conjunto Ω .
- ❑ Assim, não temos apenas uma densidade, mas uma família de densidades. A cada valor de θ em Ω . corresponde um membro da família.
- ❑ Aqui adotaremos o **enfoque "clássico"** de estimação, no qual θ é um **parâmetro desconhecido**, suposto constante, e não uma variável aleatória.

Estimação Bayesiana X Clássica



- Na **estimação Bayesiana**, θ será encarado como uma **variável aleatória**, e a ele associaremos uma distribuição de probabilidade.
- A distribuição de probabilidade de θ antes de observarmos os dados será chamada de distribuição a priori, e muitas vezes representa o nosso conhecimento subjetivo sobre o parâmetro θ .
- A distribuição de θ após observarmos a amostra é conhecida como distribuição a posteriori de θ .

monica@ele.puc-rio.br

13

Estimação Bayesiana versus Clássica



- Em estatística Bayesiana a verossimilhança (que iremos definir em breve) "carrega" a informação sobre θ contida na amostra, e resulta na atualização da densidade de θ , passando de uma priori para uma posteriori.
- A densidade a posteriori combina a "informação" subjetiva trazida pela priori com a "informação" proveniente da amostra.
- Os dois enfoques, Clássico e Bayesiano, concordam se o tamanho da amostra é grande.

monica@ele.puc-rio.br

14

Definição do Problema de Estimação Pontual



- *O problema geral aqui é ...*
- A partir dos dados observados x_1, x_2, \dots, x_n precisamos **escolher um membro** de uma família de densidades para representar estes dados.
- Ou seja, precisamos de um **estimador pontual** de θ (um "chute educado" para o valor desconhecido de θ).

monica@ele.puc-rio.br

15

Definição do Problema de Estimação Pontual



- Seja X_1, X_2, \dots, X_n uma amostra aleatória da densidade $f(x, \theta)$.
- O **objetivo** agora é **definir uma estatística** $T = T(X_1, X_2, \dots, X_n)$ de tal modo que, após observarmos $X_1 = x_1, X_2 = x_2, \dots, X_n = x_n$ $t = t(x_1, x_2, \dots, x_n)$ seja uma boa estimativa pontual de θ .
- Na verdade, a cada amostra obtida, encontraremos um valor para a estatística usada para "chutar" θ , pois esta estatística é também uma variável aleatória

monica@ele.puc-rio.br

16

Exemplo



- No próximo exemplo exibimos a média amostral de 5 amostras de tamanho 50 geradas a partir da densidade $N(0,1)$ no Excel. A média amostral serve para estimar a média da distribuição (zero, neste caso) e portanto deve ser, para todas as amostras, um valor próximo de zero.
- Os resultados para as 5 amostras geradas estão a seguir.

	Amostra 1	Amostra 2	Amostra 3	Amostra 4	Amostra 5
Média	0,076	0,150	0,180	-0,199	0,055
Desvio Padrão	1,108	1,060	1,020	1,017	0,923
Mediana	0,168	0,179	0,241	-0,206	0,072

monica@ele.puc-rio.br

17

Exemplo



- Note que os valores estimados da média em cada amostra são todos diferentes entre si, e diferentes do valor real da média da população, que é $= 0$.
- Da mesma maneira, as estimativas do desvio padrão (cujo valor real é 1) são todas diferentes do valor real, e diferentes entre si. Note que, na prática, os valores de μ e σ são desconhecidos, o que não acontece neste exemplo, onde geramos amostras de uma distribuição conhecida.

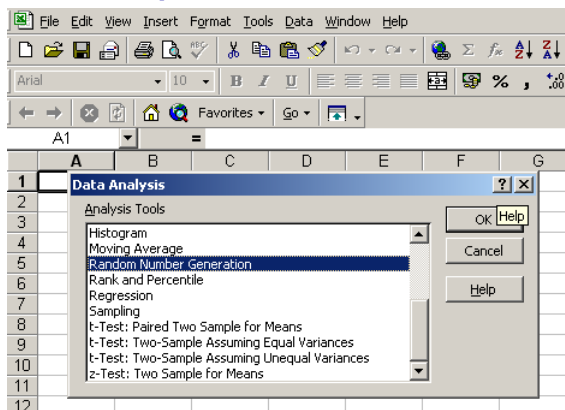
monica@ele.puc-rio.br

18

Exemplo



- Como fazer a geração destas variáveis Normais no Excel? Lembre-se que o suplemento de análise de dados deve estar previamente instalado.



19

Exemplo



5 variáveis

50 valores por variável

Pasta onde armazenar resultados (opcional)

Semente do gerador (opcional)

monica@ele.puc-rio.br

20

O que é um bom estimador?



- ❑ Existem potencialmente milhares de estimadores para um certo parâmetro.
- ❑ Por exemplo, para estimar a média de uma população poderíamos usar a média amostral, a mediana amostral, a média entre a menor e a maior observação na amostra e uma infinidade de outros estimadores "razoáveis".

O que é um bom estimador?



- ❑ **Como escolher dentre eles? Quais serão os critérios usados para comparar estimadores e caracterizar os bons estimadores?**
- ❑ Por enquanto não responderemos a esta questão, mas começaremos a estudar o (talvez) mais tradicional método de estimação pontual.

Método da Máx. Verossimilhança



- ❑ A função de verossimilhança (likelihood function)
- ❑ Esta é uma função relativamente simples com um nome indigesto!
- ❑ "Likelihood" em inglês é uma palavra de uso corrente, que indica "plausibilidade". Ao contrário, "verossimilhança" é uma coisa meio obscura.
- ❑ Seja X_1, X_2, \dots, X_n uma amostra aleatória da densidade $f(x, \theta)$.

Método da Máx. Verossimilhança



- ❑ A **função de verossimilhança** é a densidade conjunta encarada como função do parâmetro θ . Isto é:

$$L(\theta) = f(x_1, x_2, \dots, x_n) = \prod_{i=1}^n f(x_i, \theta)$$

- ❑ A partir da verossimilhança podemos encontrar um estimador, o estimador de máxima verossimilhança (MLE = maximum likelihood estimator).
- ❑ O MLE é obtido a partir da maximização da verossimilhança, geralmente feita através da equação $dL(\theta)/d\theta = 0$.

Método da Máx. Verossimilhança



- É equivalente maximizar $L(\theta)$ ou seu logaritmo natural, $l(\theta) = \log L(\theta)$ onde $\log(\cdot)$ indica o logaritmo na base e .
- Esta última função é chamada **log-verossimilhança** e é freqüentemente mais fácil de maximizar do que $L(\theta)$, pois as verossimilhanças muitas vezes podem ser escritas como $\exp\{ \dots \}$.
- A equivalência da maximização de $L(\theta)$ e $l(\theta)$ decorre do fato de $L(\theta)$ ser sempre maior que 0 (pois é o produto de densidades) e do logaritmo ser uma função bijetora.

monica@ele.puc-rio.br

25

Método da Máx. Verossimilhança



- *Por que maximizar a verossimilhança?*
- Suponha que temos uma amostra aleatória X_1, X_2, \dots, X_n de uma densidade qualquer, completamente conhecida exceto pelo parâmetro θ .
- Ao observarmos cada x_i , a densidade conjunta fica completamente especificada exceto pelo valor de θ . Então, por que não "chutar" para θ o valor que torna esta função um máximo?
- Este "chute" para θ é o valor que **mais concorda com os dados** observados.

monica@ele.puc-rio.br

26

Exemplo 1 - MLE (Poisson)



- Suponha que obtemos uma amostra aleatória de tamanho 5 da distribuição Poisson com média θ .
- Os valores observados na amostra são: 0, 6, 1, 2 e 1.
- Então a função de probabilidade conjunta é:

$$L(\theta) = \prod_{i=1}^n f(x_i, \theta) = \prod_{i=1}^n \frac{\theta^{x_i} e^{-\theta}}{x_i!} = \frac{e^{-5\theta} e^{\sum x_i}}{\prod_{i=1}^5 x_i!}$$

$$L(\theta) = \frac{e^{-5\theta} \theta^{10}}{0!6!1!2!1!} = \frac{\theta^{10} e^{-5\theta}}{1440}$$

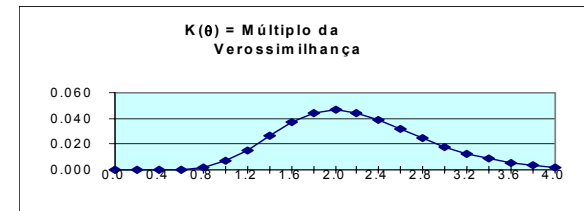
monica@ele.puc-rio.br

27

Exemplo 1 - MLE (Poisson)



- Seja $K(\theta) = 1440.L(\theta) = \theta^{10}e^{-5\theta}$
- Podemos fazer um gráfico de $K(\theta)$ e ver qual o valor que aparentemente maximiza esta função, ou, alternativamente, fazer um gráfico de $L(\theta)$ ou $l(\theta)$. O gráfico de $K(\theta)$ é:



- O máximo aparente ocorre em $\theta = 2$.

monica@ele.puc-rio.br

28

Método da Máx. Verossimilhança



- Podemos confirmar se este valor realmente corresponde ao máximo através de técnicas simples do Cálculo.
- Lembre-se que uma **condição necessária** (mas não suficiente) para a existência de um **máximo local** é que a **primeira derivada** da função de interesse seja **zero**.
- Isso nos leva à idéia de "equação de máxima verossimilhança", discutida a seguir.

Método da Máx. Verossimilhança



- Para maximizar $L(\theta)$, uma condição necessária é que sua primeira derivada seja igual a zero.
- Assim, a **equação de máxima verossimilhança** é:
$$\frac{dL(\theta)}{d\theta} = 0$$
- e esta equação deve ser resolvida, por métodos analíticos ou numéricos para θ .
- Para assegurar que a solução de $dL/d\theta = 0$ seja realmente um máximo da verossimilhança, precisamos garantir que a segunda derivada seja ≤ 0 .

Método da Máx. Verossimilhança



- A equação de máxima verossimilhança pode ser reescrita em termos da log-verossimilhança. Assim, é equivalente resolver:

$$\frac{d(\log L(\theta))}{d\theta} = 0 \Leftrightarrow \frac{dl(\theta)}{d\theta} = 0 \text{ para } \theta.$$

- O estimador obtido pela maximização da função de verossimilhança é chamado de **estimador de máxima verossimilhança (MLE)**.
- Notação:
- Geralmente denotaremos o MLE por $\hat{\theta} = T(X_1, X_2, \dots, X_n)$

Método da Máx. Verossimilhança



- Atenção**
- Em muitos casos o estimador de máxima verossimilhança é único e pode ser obtido por métodos analíticos.
- Em outros casos, a equação de máxima verossimilhança $dL/d\theta = 0$ (ou $dl/d\theta = 0$) não nos dá o resultado correto e precisaremos encontrar o máximo da verossimilhança por outros métodos (por exemplo, graficamente)

Exemplo 2 - MLE (Poisson)



- Considere o exemplo 1. A log verossimilhança é:

$$l(\theta) = -5\theta + 10 \cdot \log \theta - \log(1440)$$

- Derivando esta última expressão com relação a θ e igualando a zero leva a:

$$\frac{dl}{d\theta} = 0 \rightarrow -5 + \frac{10}{\theta} = 0 \rightarrow -5\theta = -10 \rightarrow \hat{\theta} = 2$$

- é o estimador de máxima verossimilhança para θ .
- Compare este resultado como exemplo 1. Este resultado não é mera coincidência.