

Métodos Estatísticos de Apoio à Decisão

Aula 7

Mônica Barros, D.Sc.

Agosto de 2007

monica@ele.puc-rio.br

1

Aula 7

- ❑ Distribuições Derivadas da Normal
- ❑ Diferença entre Probabilidade e Estatística
- ❑ Amostra Aleatória
- ❑ Objetivos da Estatística
- ❑ Distribuição Amostral
- ❑ Estimação Pontual
- ❑ Estimação Bayesiana X Clássica
- ❑ Estimação por Máxima Verossimilhança
- ❑ Estimação por Método de Momentos

monica@ele.puc-rio.br

2

Distribuições Derivadas da Normal

- ❑ Densidade Qui-quadrado com k graus de liberdade

Seja X uma variável aleatória contínua e positiva com densidade dada por:

$$f(x) = \frac{1}{2^{k/2} \cdot \Gamma\left(\frac{k}{2}\right)} \cdot x^{\frac{k}{2}-1} \cdot e^{-x/2} \quad \text{onde } x > 0$$

- ❑ Então X tem densidade Qui-quadrado com k graus de liberdade, e escrevemos: $X \sim \chi^2_k$

monica@ele.puc-rio.br

3

Distribuições Derivadas da Normal

- ❑ A densidade Qui-quadrado é apenas um caso particular de uma outra densidade chamada densidade Gama, que também inclui a Exponencial como caso particular.

- ❑ Se X é Qui-quadrado com n graus de liberdade então:

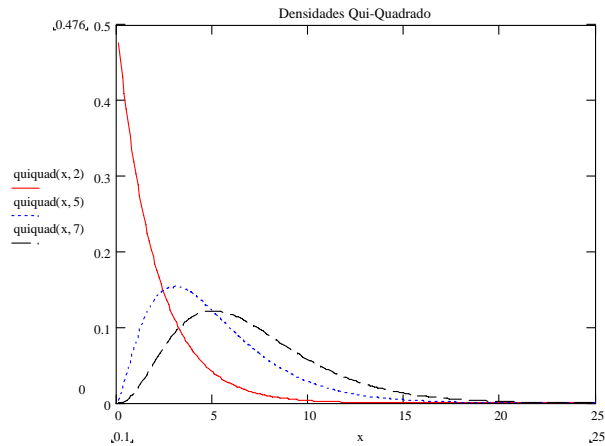
$$E(X) = \frac{n/2}{1/2} = n$$

$$VAR(X) = \frac{n/2}{(1/2)^2} = 2n$$

monica@ele.puc-rio.br

4

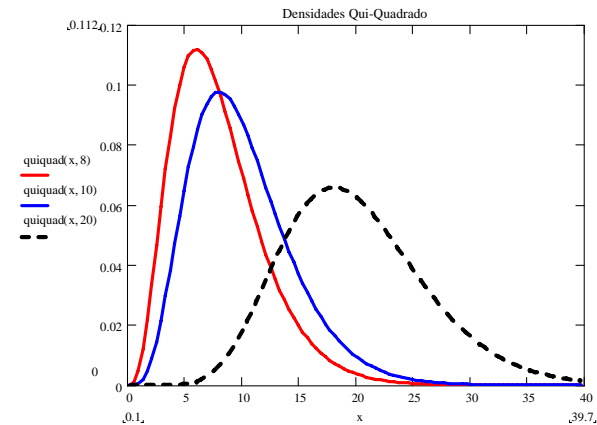
Distribuição Qui-Quadrado



monica@ele.puc-rio.br

5

Distribuição Qui-Quadrado



monica@ele.puc-rio.br

6

Distribuição Qui-quadrado



- Tabelas da função de distribuição Qui-quadrado
 - A densidade Qui-quadrado é tabelada para diversos graus de liberdade.
 - As tabelas geralmente fornecem o valor $x_{1-\alpha}$ tal que $\Pr(X < x_{1-\alpha}) = 1 - \alpha$ para $\alpha = 1\%, 5\%, 10\%$. Também existem tabelas que apresentam o valor x_α tais que $\Pr(X < x_\alpha) = \alpha$, isto é, $\Pr(X > x_\alpha) = 1 - \alpha$.

monica@ele.puc-rio.br

7

Distribuição Qui-quadrado



probabilidade →	0.01	0.05	0.10	0.25	0.50	0.75	0.90	0.95	0.99
graus de liberdade ↓									
1	0.000	0.004	0.016	0.102	0.455	1.323	2.706	3.841	6.635
2	0.020	0.103	0.211	0.575	1.386	2.773	4.605	5.991	9.210
3	0.115	0.352	0.584	1.213	2.366	4.108	6.251	7.815	11.345
4	0.297	0.711	1.064	1.923	3.357	5.385	7.779	9.488	13.277
5	0.554	1.145	1.610	2.675	4.351	6.626	9.236	11.070	15.086
6	0.872	1.635	2.204	3.455	5.348	7.841	10.645	12.592	16.812
7	1.239	2.167	2.833	4.255	6.346	9.037	12.017	14.067	18.475
8	1.647	2.733	3.490	5.071	7.344	10.219	13.362	15.507	20.090
9	2.088	3.325	4.168	5.899	8.343	11.389	14.684	16.919	21.666
10	2.558	3.940	4.865	6.737	9.342	12.549	15.987	18.307	23.209
11	3.053	4.575	5.578	7.584	10.341	13.701	17.275	19.675	24.725
12	3.571	5.226	6.304	8.438	11.340	14.845	18.549	21.026	26.217
13	4.107	5.892	7.041	9.299	12.340	15.984	19.812	22.362	27.688
14	4.660	6.571	7.790	10.165	13.339	17.117	21.064	23.685	29.141
15	5.229	7.261	8.547	11.037	14.339	18.245	22.307	24.996	30.578
16	5.812	7.962	9.312	11.912	15.338	19.369	23.542	26.296	32.000
17	6.408	8.672	10.085	12.792	16.338	20.489	24.769	27.587	33.409
18	7.015	9.390	10.865	13.675	17.338	21.605	25.989	28.869	34.805
19	7.633	10.117	11.651	14.562	18.338	22.718	27.204	30.144	36.191
20	8.260	10.851	12.443	15.452	19.337	23.828	28.412	31.410	37.566

monica@ele.puc-rio.br

8

Distribuição Qui-quadrado



- ❑ Função de Distribuição Qui-quadrado no Excel
- ❑ Use as funções **DIST.QUI** e **INV.QUI**
- ❑ **A tabela anterior foi produzida usando INV.QUI** – dada uma probabilidade e o grau de liberdade, a função INV.QUI retorna o ponto correspondente da densidade tal que a probabilidade de estar **ACIMA** do ponto é a especificada como argumento da função.

monica@ele.puc-rio.br

9

Distribuição Qui-quadrado



- ❑ Função de Distribuição Qui-quadrado no Excel
- ❑ Por exemplo, para uma Qui-quadrado com 10 graus de liberdade:
- ❑ $INV.QUI(0.99, 10) = 2.558$
- ❑ $INV.QUI(0.01, 10) = 23.209$
- ❑ Ou seja, a probabilidade de uma v.a. Qui-quadrado com 10 graus de liberdade exceder 2.558 é 0.99, e a probabilidade da mesma variável exceder 23.209 é 0.01.

monica@ele.puc-rio.br

10

Distribuição Qui-Quadrado



Valor de x para o qual desejamos

$Pr(X > x)$

Graus de liberdade da Qui-quadrado

Microsoft Excel - Book1

File Edit View Insert Format Tools Data Window Help

Arial 10 B I U

CHIDIST X ✓ = =CHIDIST()

1 -CHIDIST()

2

3 CHIDIST

4 X = number

5 Deg_freedom = number

6

7 Returns the one-tailed probability of the chi-squared distribution.

8 X is the value at which you want to evaluate the distribution, a nonnegative number.

9

10

11 Formula result =

12

13

monica@ele.puc-rio.br

11

Distribuição Qui-Quadrado



Da figura segue que, a $Pr(X > 15)$ quando X é uma Qui-quadrado com 12 graus de liberdade é 0.2414

Microsoft Excel - Book1

File Edit View Insert Format Tools Data Window Help

Arial 10 B I U

CHIDIST X ✓ = =CHIDIST(15;12)

1

2 -CHIDIST

3 X = 15 = 15

4 Deg_freedom = 12 = 12

5

6 = 0.241436451

7 Returns the one-tailed probability of the chi-squared distribution.

8 X is the value at which you want to evaluate the distribution, a nonnegative number.

9

10 Formula result = 0.241436451

11

12

monica@ele.puc-rio.br

12

Distribuição Qui-Quadrado



- Por exemplo:
- Supondo que X seja uma variável aleatória com densidade qui-quadrado com 6 graus de liberdade, a probabilidade de X exceder 0.87 é 99%.
- Analogamente, a probabilidade de X exceder 12.59 é 5% e a probabilidade de X estar acima de 16.81 é apenas 1%.

monica@ele.puc-rio.br

13

Distribuição Qui-Quadrado



- Podemos estar interessados na pergunta “ao contrário”. Dada uma Qui-Quadrado com k graus de liberdade e uma probabilidade α , qual é o ponto tal que a probabilidade de estar ACIMA dele é α ?
- O **Excel** também nos dá esta resposta, através da função **INV.CHI**.

monica@ele.puc-rio.br

14

Distribuição Qui-Quadrado



Da figura segue que, a $\Pr(X > 31.4104)$ quando X é uma Qui-quadrado com 20 graus de liberdade é 0.05

monica@ele.puc-rio.br

15

Distribuição Qui-quadrado



- A densidade Qui-quadrado é importante no contexto de amostras aleatórias Normais, na estimação da variância.
- Também pode-se provar que **o quadrado de uma variável Normal padrão (que estudaremos a seguir) tem densidade Qui-quadrado com um grau de liberdade.**

monica@ele.puc-rio.br

16

Distribuições Derivadas da Normal



- Uma propriedade muito importante da densidade Qui-quadrado é a **preservação da mesma família** de densidades **quando somamos** variáveis independentes.

- Ou seja, se X_1, X_2, \dots, X_n são variáveis independentes, cada uma com distribuição Qui-quadrado, a soma de X_1, X_2, \dots, X_n também é uma variável aleatória qui-quadrado.

Distribuições Derivadas da Normal



- **Teorema (aditividade da densidade Qui-quadrado)**
- Sejam X_1, X_2, \dots, X_n v.a. aleatórias independentes, e suponha que X_i tem densidade qui-quadrado com k_i graus de liberdade. Seja $Y = X_1 + X_2 + \dots + X_n$.
- Então Y tem também uma densidade Qui-quadrado, mas com $k = k_1 + k_2 + \dots + k_n$ graus de liberdade.
- O próximo teorema exhibe a relação existente entre as densidades $N(0,1)$ e Qui-quadrado.

Distribuições Derivadas da Normal



Teorema

Seja $Z \sim N(0,1)$. Então $V = Z^2$ tem densidade Qui-quadrado com 1 grau de liberdade.

- **A combinação dos dois últimos teoremas leva a um resultado importante.**
- Sejam Z_1, Z_2, \dots, Z_n v.a. independentes e identicamente distribuídas com densidade $N(0,1)$. Então:

Distribuições Derivadas da Normal



$$V = \sum_{i=1}^n Z_i^2 = Z_1^2 + Z_2^2 + \dots + Z_n^2$$

- tem densidade Qui-quadrado com n graus de liberdade.
- Este resultado segue trivialmente dos dois últimos, se lembrarmos que cada Z_i^2 tem densidade qui-quadrado com 1 grau de liberdade (e são todos independentes).

Distribuições Derivadas da Normal



- ❑ **Por que a densidade Qui-quadrado é importante?**
- ❑ Porque está relacionada com a distribuição da variância amostral de uma amostra aleatória Normal, como indicado no próximo teorema.
- ❑ Por exemplo, se desejarmos encontrar um intervalo baseado na variância amostral que contenha, com alta probabilidade, a variância (desconhecida) da distribuição Normal, este intervalo será construído a partir da distribuição Qui-quadrado.

Distribuições Derivadas da Normal



- ❑ Teorema
- ❑ Sejam X_1, X_2, \dots, X_n uma amostra aleatória da distribuição $N(\mu, \sigma^2)$. Seja S^2 a variância amostral:

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

- ❑ Então:

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

- ❑ **tem distribuição Qui-quadrado com (n-1) graus de liberdade.**

Distribuições Derivadas da Normal



- ❑ A partir deste teorema podemos deduzir facilmente a média e variância de S^2 .
- ❑ Teorema
- ❑ Sejam X_1, X_2, \dots, X_n uma amostra aleatória da distribuição $N(\mu, \sigma^2)$. Seja S^2 a variância amostral. Então:

$$E(S^2) = \sigma^2$$

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

Distribuições Derivadas da Normal



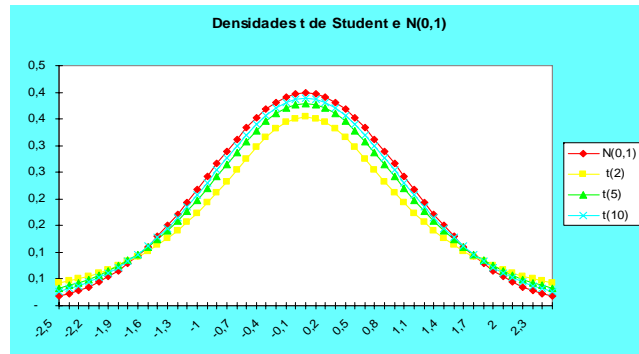
- ❑ A distribuição t de Student
- ❑ Tem apenas um parâmetro k, o número de graus de liberdade, e é definida como:

$$T = \frac{Z}{\sqrt{V/k}}$$

- ❑ Onde Z é $N(0,1)$ e V é Qui-Quadrado com k graus de liberdade, e ambos são independentes.
- ❑ Esta distribuição é simétrica em torno de zero, também tem forma de sino e, à medida que o número de graus de liberdade cresce, se aproxima da $N(0,1)$.

Distribuições Derivadas da Normal

- Quando n (número de graus de liberdade) cresce, a densidade t de Student se torna cada vez mais parecida com uma $N(0,1)$



monica@ele.puc-rio.br

25

Distribuições Derivadas da Normal

- Exemplo (uso de uma tabela t)

graus de liberdade	0.9	0.95	0.975	0.99	0.995
1	3.078	6.314	12.706	31.821	63.657
5	1.476	2.015	2.571	3.365	4.032
10	1.372	1.812	2.228	2.764	3.169
15	1.341	1.753	2.131	2.602	2.947
20	1.325	1.725	2.086	2.528	2.845

- Por exemplo, se T tem 10 graus de liberdade, a probabilidade de T ser menor que 1.372 é 90%. Se o número de graus de liberdade passa a 15, o valor tal que a probabilidade de T ser menor que ele é 90% passa a ser 1.341.

monica@ele.puc-rio.br

26

Distribuições Derivadas da Normal

- Função do Excel para a distribuição t

Função	Descrição
<code>invt(p; gl)</code>	Para a distribuição t de Student, calcula o valor t para $p = 2.\alpha$, com gl graus de liberdade

- Por exemplo, $INVT(0.05, 20) = 2.086$ é o valor da distribuição t com 20 graus de liberdade tal que $Pr(T > 2.086) = 0.05/2 = 0.025$.

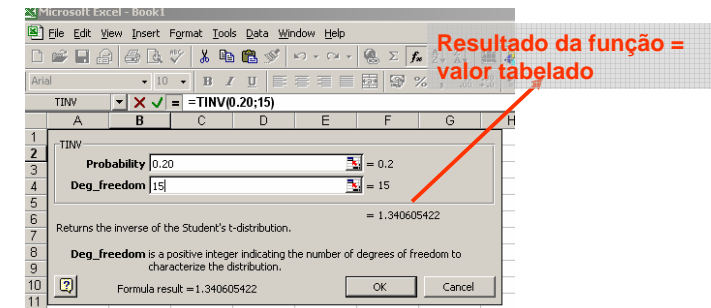
CUIDADO com a especificação da probabilidade para esta função, a função `INVT` fornece as probabilidades “bi-laterais”.

monica@ele.puc-rio.br

27

Distribuições Derivadas da Normal

- Refazemos a seguir o exemplo anterior com a função `INVT` do Excel. Note a especificação da probabilidade como $0.20 = 2.\alpha$, enquanto na nossa tabela as colunas referem-se a $1 - \alpha$.



monica@ele.puc-rio.br

28

Distribuições Derivadas da Normal



- ❑ **Por que a densidade t é importante?**
- ❑ Ela é essencial no contexto de intervalos de confiança e testes de hipóteses, como veremos posteriormente. A justificativa vem, em parte, do próximo resultado.
- ❑ Teorema
- ❑ Sejam X_1, X_2, \dots, X_n uma amostra aleatória da distribuição $N(\mu, \sigma^2)$. Sejam \bar{X} e S^2 a média e variância amostrais. Então:

$$T = \frac{\bar{X} - \mu}{S/\sqrt{n}} = \frac{\sqrt{n}(\bar{X} - \mu)}{S} \sim t_{n-1}$$

monica@ele.puc-rio.br

29

Probabilidade e Estatística – Qual a Diferença?



- ❑ Até agora tivemos estivessemos interessados em Probabilidade, ou seja, nosso objetivo era:
 - ❑ apresentar alguns dos modelos probabilísticos mais usuais e as situações em que eles surgem na prática.

monica@ele.puc-rio.br

30

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

31

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

32

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 frequentemente 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 versus 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

41

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

42

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

43

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

44

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

45

Exemplo

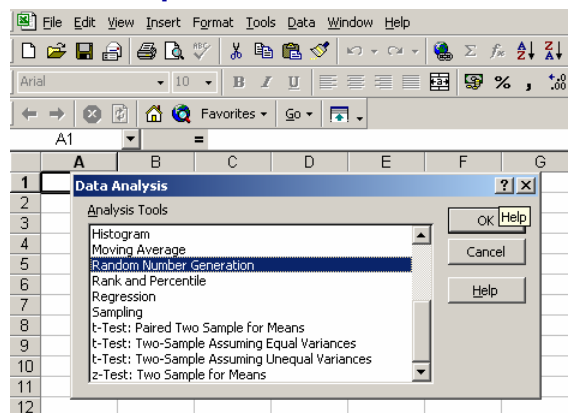
- ❑ Note que os valores estimados da média em cada amostra são todos diferentes entre si, e diferentes do valor real da média da população, que é = 0.
- ❑ Da mesma maneira, as estimativas do desvio padrão (cujo valor real é 1) são todas diferentes do valor real, e diferentes entre si. Note que , na prática, 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

46

Exemplo

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



47

Exemplo

5 variáveis

50 valores por variável

Pasta onde armazenar resultados (opcional)

Semente do gerador (opcional)

monica@ele.puc-rio.br

48

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.

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.

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:

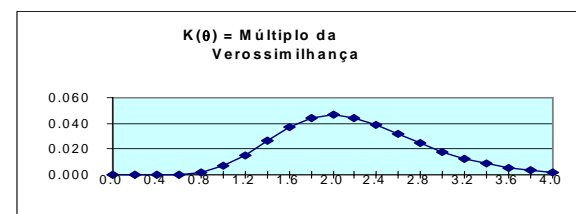
$$L(\theta) = \prod_{i=1}^n f(x_i, \theta) = \prod_{i=1}^n \frac{\theta^{x_i} e^{-\theta}}{x_i!} = \frac{e^{-5\theta} e^{\sum x_i}}{\prod_{i=1}^5 x_i!}$$

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

Exemplo 1 - MLE (Poisson)



- Seja $K(\theta) = 1440 \cdot L(\theta) = \theta^{10} e^{-5\theta}$
- 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 $\theta = 2$.

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 com o exemplo 1. Este resultado não é mera coincidência.

Exemplo 3 (Bernoulli)

- Sejam X_1, X_2, \dots, X_n iid Bernoulli(θ).
- A função de probabilidade de cada X_i é:

$$f(x_i, \theta) = \theta^{x_i} (1 - \theta)^{1 - x_i} \quad \theta \in (0, 1), x_i = 0, 1$$

- A função de verossimilhança o produto das funções de probabilidade individuais, isto é:

$$\begin{aligned} L(\theta) &= f(x_1, x_2, \dots, x_n, \theta) = \prod_{i=1}^n \theta^{x_i} (1 - \theta)^{1 - x_i} = \theta^{\sum x_i} (1 - \theta)^{n - \sum x_i} = \\ &= \theta^{n\bar{x}} (1 - \theta)^{n - n\bar{x}} = \exp\{n\bar{x} \cdot \log \theta + (n - n\bar{x}) \log(1 - \theta)\} \end{aligned}$$

Exemplo 3 (Bernoulli)

- A log verossimilhança é:

$$\begin{aligned} l(\theta) &= \log(L(\theta)) = (\sum x_i) \log \theta + (n - \sum x_i) \log(1 - \theta) = \\ &= n\bar{x} \cdot \log \theta + (n - n\bar{x}) \log(1 - \theta) \end{aligned}$$

- Resolvendo a equação de verossimilhança leva a:

$$\begin{aligned} \frac{dl}{d\theta} = 0 &\Rightarrow \frac{n\bar{x}}{\theta} - \frac{(n - n\bar{x})}{1 - \theta} = 0 \\ \Leftrightarrow (1 - \theta)n\bar{x} &= n\theta - n\bar{x}\theta \Leftrightarrow n\theta = n\bar{x} \end{aligned}$$

- E então o MLE para θ é a média amostral.
- Verifique que $\left. \frac{d^2 l}{d\theta^2} \right|_{\theta = \bar{X}} < 0$
- de tal forma que a média amostral realmente **MAXIMIZA** a verossimilhança.

Exemplo 4 (Normal)

- Sejam X_1, X_2, \dots, X_n iid Normal($\mu, 1$), ou seja, uma amostra Normal com média desconhecida e variância conhecida (e suposta igual a um sem perda de generalidade).
- Mostre que o MLE de μ é a média amostral.

Exemplo 4 (Normal)



- A verossimilhança é:

$$L(\mu) = \prod_{i=1}^n \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{1}{2}(X_i - \mu)^2\right] = (2\pi)^{-n/2} \exp\left\{-\frac{1}{2} \sum_{i=1}^n (X_i - \mu)^2\right\}$$

- Note que a verossimilhança é máxima quando $Q(\mu) = \sum (X_i - \mu)^2$ é mínimo.

- Então é equivalente maximizar $L(\mu)$ ou minimizar $Q(\mu)$.

Exemplo 4 (Normal)



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

- Derivando $Q(\mu)$ em relação a μ e igualando a zero nos leva a um ponto crítico:

$$\frac{dQ(\mu)}{d\mu} = 0 \rightarrow -2 \sum_{i=1}^n X_i + 2n\mu = 0 \Leftrightarrow \hat{\mu} = \frac{1}{n} \sum_{i=1}^n X_i = \bar{X}$$

- Logo, o MLE de μ é a média amostral.

Método dos Momentos



- É uma alternativa ao método de máxima verossimilhança para encontrar estimadores de maneira simples.

Idéia Geral

- Igualar os momentos da distribuição aos momentos da amostra. Isso é razoável pois a distribuição empírica converge estocasticamente para a função de distribuição $F(X)$.

Método dos Momentos



- Considere uma amostra aleatória da distribuição com densidade $f(x, \theta_1, \theta_2, \dots, \theta_p)$ onde $(\theta_1, \theta_2, \dots, \theta_p) \in \Omega$.
- Seja $E(X^k)$ o k -ésimo momento da distribuição ($k = 1, 2, \dots$).
- Seja $M_k = \frac{1}{n} \sum_{i=1}^n X_i^k$ o k -ésimo momento amostral
- Faça $E(X^k) = M_k$ para $k = 1, 2, \dots$
- Faça isto para quantos k 's forem necessários até obter soluções únicas para os parâmetros desconhecidos

Método dos Momentos



- ❑ Nota
- ❑ Uma formulação equivalente do método dos momentos usa os momentos centrais, igualando os momentos centrais da distribuição e da amostra até obter soluções únicas para todos os parâmetros desconhecidos.

- ❑ Ou seja, a outra formulação do método dos momentos é igualar:

$$E(X - \mu)^k = m_k = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^k \text{ para } k = 1, 2, \dots$$

- ❑ Até encontrar uma solução única. Muitas vezes a resposta pelas duas formulações será igual.

Método dos Momentos



- ❑ Na maneira dos casos práticos, os estimadores pelo método dos momentos são consistentes.
- ❑ **O grande problema com este método é que ele não fornece estimadores únicos para um certo parâmetro.**
- ❑ Por exemplo, se procuramos estimar a média e variância de uma distribuição é possível encontrar estimadores diferentes pelo método de momentos dependendo de quais momentos amostrais são igualados aos momentos da distribuição.

Exemplo 5



- ❑ Vide a situação do exemplo 4. Sejam X_1, X_2, \dots, X_n iid $\text{Normal}(\mu, 1)$, ou seja, uma amostra Normal com média desconhecida e variância conhecida (e suposta igual a um sem perda de generalidade).
- ❑ Mostre que o estimador por método de momentos de μ é também a média amostral.
- ❑ Neste caso, há apenas um parâmetro desconhecido e basta igualar o 1º. Momento da amostra ao 1º. Momento da população (μ).