Revisão de Probabilidade e Estatística

Aula 5

Aishameriane Schmidt

PPGECO/UFSC

Fevereiro de 2019.

Programa
Aula 5

1. Fazendo o cartãozinho da UFSC: *número de matrícula + documento com foto*

Programa

Aula 5

1. Fazendo o cartãozinho da UFSC: *número de matrícula + documento com foto*
2. Cadastro na biblioteca central: *número de matrícula (e acho que documento)*

Programa

Aula 5

1. Fazendo o cartãozinho da UFSC: *número de matrícula + documento com foto*
2. Cadastro na biblioteca central: *número de matrícula (e acho que documento)*
3. Cadastro no ID UFSC: *cpf + e-mail*

Programa

Aula 5

1. Fazendo o cartãozinho da UFSC: *número de matrícula + documento com foto*
2. Cadastro na biblioteca central: *número de matrícula (e acho que documento)*
3. Cadastro no ID UFSC: *cpf + e-mail*
4. CAGR e Moodle

Programa

Aula 5

1. Fazendo o cartãozinho da UFSC: *número de matrícula + documento com foto*
2. Cadastro na biblioteca central: *número de matrícula (e acho que documento)*
3. Cadastro no ID UFSC: *cpf + e-mail*
4. CAGR e Moodle
5. Distribuição normal multivariada

Programa

Aula 5

1. Fazendo o cartãozinho da UFSC: *número de matrícula + documento com foto*
2. Cadastro na biblioteca central: *número de matrícula (e acho que documento)*
3. Cadastro no ID UFSC: *cpf + e-mail*
4. CAGR e Moodle
5. Distribuição normal multivariada
6. Estatística
   - Distribuições amostrais
   - Função de Verossimilhança
   - Estimação pontual

Exemplos interessantes

Aula 4

Exemplo

- Prove que se $X \sim \text{Poisson}(\lambda_1)$ e $Y \sim \text{Poisson}(\lambda_2)$ são independentes e $Z = X + Y$, então

$$X|Z = n \sim Binomial\left(n, \frac{\lambda_1}{\lambda_1 + \lambda_2}\right)$$

Distribuições Multivariadas

Definição

Seja $\underset{\sim}{X} : (X_1, \ldots, X_j)$ um vetor aleatório. Então:

- Sua esperança é dada por

$$\mathbb{E}[\underset{\sim}{X}] = (\mathbb{E}[X_1], \ldots, \mathbb{E}[X_d])$$

Distribuições Multivariadas

Definição

Seja $\underset{\sim}{X} : (X_1, \ldots, X_j)$ um vetor aleatório. Então:

- Sua esperança é dada por

$$\mathbb{E}[\underset{\sim}{X}] = (\mathbb{E}[X_1], \ldots, \mathbb{E}[X_d])$$

- A matriz de variâncias e covariâncias de um vetor aleatório $\mathbf{X} = (X_1, \ldots, X_d)$, $\mathbb{V}[\underset{\sim}{X}]$, é a matriz $d \times d$ cujo componente (i, j) é dado por $Cov(X_i, X_j)$, isto é,

$$\mathbb{V}[\underset{\sim}{X}] = \begin{bmatrix} Var[X_1] & Cov[X_1, X_2] & \ldots & Cov[X_1, X_j] \\ Cov[X_2, X_1] & Var[X_2] & \ldots & Cov[X_2, X_j] \\ \vdots & \vdots & \ddots & \vdots \\ Cov[X_1, X_j] & Cov[X_2, X_j] & \ldots & Var[X_j] \end{bmatrix}.$$

- Vimos, utilizando a função geradora de momentos, que a soma de normais independentes segue também uma distribuição normal. Vamos continuar mexendo na distribuição normal, porém usando a versão multivariada da f.g.m.

- Vimos, utilizando a função geradora de momentos, que a soma de normais independentes segue também uma distribuição normal. Vamos continuar mexendo na distribuição normal, porém usando a versão multivariada da f.g.m.

Definição

$\underset{\sim}{X} : (X_1, \ldots, X_n)$ um vetor aleatório. Definimos a f.g.m. de $\underset{\sim}{X}$ por

$$M_{\underset{\sim}{X}} : \mathbb{R}^n \to \mathbb{R}$$

$$M_{\underset{\sim}{X}}(t_1, \ldots, t_n) = \mathbb{E}\left[e^{\sum_{i=1}^{n} t_i X_i}\right]$$

Exemplo

1. Sejam X e Y i.i.d. com distribuição $N(\mu, \sigma^2)$. Encontre a distribuição conjunta de $Z = X + Y$ e $W = X - Y$.

Distribuições Multivariadas

A normal multivariada

Exemplo

1. Sejam X e Y i.i.d. com distribuição $N(\mu, \sigma^2)$. Encontre a distribuição conjunta de $Z = X + Y$ e $W = X - Y$.
2. Se $\underset{\sim}{X} : (X_1, \ldots, X_d) \sim \text{Normal}(\underset{\sim}{\mu}, \Sigma)$, então,

$$M_{\underset{\sim}{X}}(t_1, \ldots, t_d) = exp\left\{\underset{\sim}{\mu}\underset{\sim}{t} + \frac{1}{2}\underset{\sim}{t}\ \Sigma\ \underset{\sim}{t}'\right\}, \qquad \underset{\sim}{t} = (t_1, \ldots, t_d)$$

Exemplo

1. Sejam X e Y i.i.d. com distribuição $N(\mu, \sigma^2)$. Encontre a distribuição conjunta de $Z = X + Y$ e $W = X - Y$.
2. Se $\underset{\sim}{X} : (X_1, \dots, X_d) \sim \text{Normal}(\underset{\sim}{\mu}, \Sigma)$, então,

$$M_{\underset{\sim}{X}}(t_1, \dots, t_d) = exp\left\{\underset{\sim}{\mu}\underset{\sim}{t} + \frac{1}{2}\underset{\sim}{t}\,\Sigma\,\underset{\sim}{t}'\right\}, \qquad \underset{\sim}{t} = (t_1, \dots, t_d)$$

- Mostre que se Σ é diagonal, as componentes de $\underset{\sim}{X}$ são independentes e diga qual a sua distribuição. O que isso significa?

Exemplo

1. Sejam X e Y i.i.d. com distribuição $N(\mu, \sigma^2)$. Encontre a distribuição conjunta de $Z = X + Y$ e $W = X - Y$.
2. Se $\underset{\sim}{X} : (X_1, \ldots, X_d) \sim \text{Normal}(\underset{\sim}{\mu}, \Sigma)$, então,

$$M_{\underset{\sim}{X}}(t_1, \ldots, t_d) = exp\left\{\underset{\sim}{\mu}\underset{\sim}{t} + \frac{1}{2}\underset{\sim}{t}\ \Sigma\ \underset{\sim}{t}'\right\}, \qquad \underset{\sim}{t} = (t_1, \ldots, t_d)$$

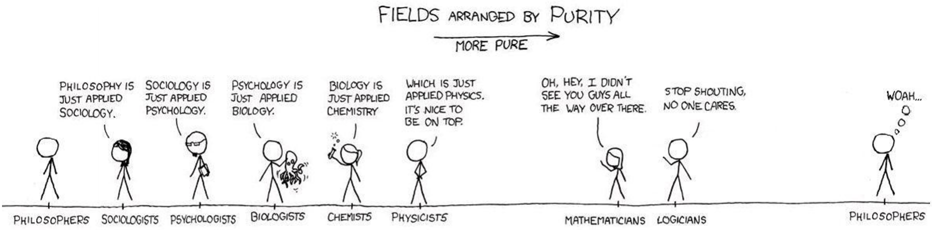
- Mostre que se Σ é diagonal, as componentes de $\underset{\sim}{X}$ são independentes e diga qual a sua distribuição. O que isso significa?
- A Normal é um dos poucos casos onde ausência de correlação implica independência!

Estatística

- Enquanto a probabilidade é uma *área* da matemática, a estatística é considerada uma **ciência**;

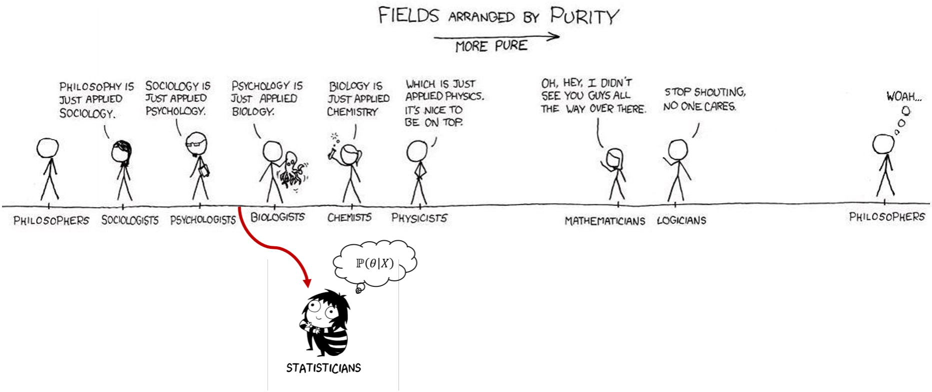
Estatística

- Enquanto a probabilidade é uma *área* da matemática, a estatística é considerada uma **ciência**;



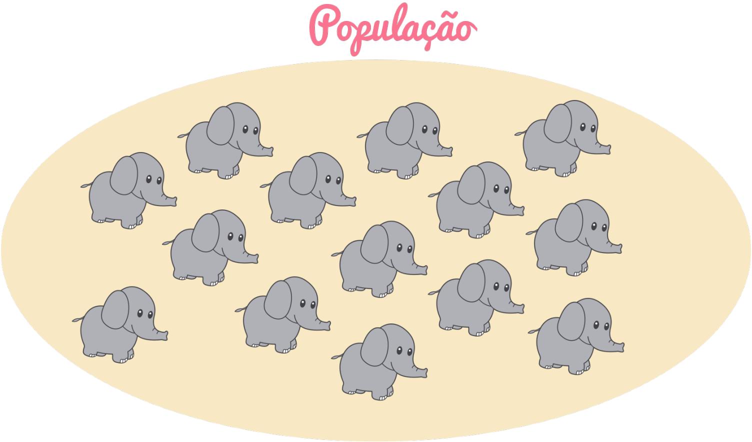
Estatística

- Enquanto a probabilidade é uma *área* da matemática, a estatística é considerada uma **ciência**;



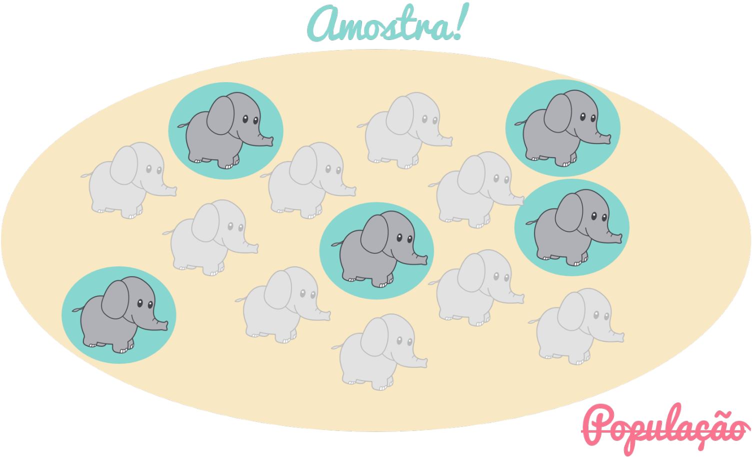
Estatística

O problema fundamental da estatística



Estatística

O problema fundamental da estatística



Usamos os dados da *amostra*

↓

Em conjunto com *estimadores*

↓

Para obter *estimativas*

↓

Dos *parâmetros populacionais*

Estatística

O problema fundamental da estatística

Usamos os dados da *amostra*

Em conjunto com *estimadores*

Precisamos de métodos de encontrar estimadores como GMM, MLE, MQO, etc

Para obter *estimativas*

Dos *parâmetros populacionais*

Estatística

O problema fundamental da estatística

Usamos os dados da *amostra*

Em conjunto com *estimadores*

Para obter *estimativas*

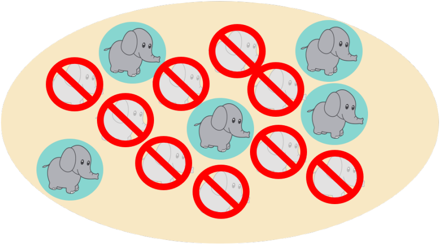
Nossos estimadores devem produzir estimativas com propriedades boas

Dos *parâmetros populacionais*

Estatística

O problema fundamental da estatística

No "mundo real", nós geralmente **não temos como coletar mais dados** para comparar estimativas a partir de várias amostras.

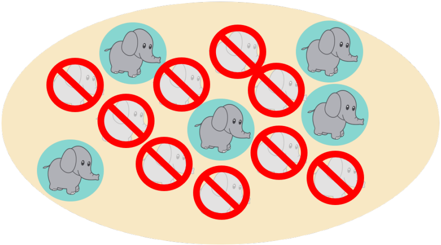


Estatística

O problema fundamental da estatística

No "mundo real", nós geralmente **não temos como coletar mais dados** para comparar estimativas a partir de várias amostras.

Nosso estimador irá produzir um número* (ou intervalo) e teremos que confiar que ele está sendo informativo a respeito da população de interesse.

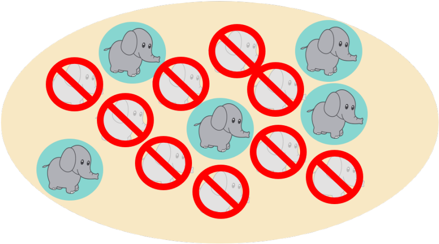


Estatística

O problema fundamental da estatística

No "mundo real", nós geralmente **não temos como coletar mais dados** para comparar estimativas a partir de várias amostras.

Nosso estimador irá produzir um número* (ou intervalo) e teremos que confiar que ele está sendo informativo a respeito da população de interesse.



Uma forma de "dormirmos tranquilos" é utilizar estimadores com boas propriedades teóricas
(mesmo que isso seja válido apenas sob certas condições de regularidade).

Amostras aleatórias

Oi, você por aqui?

Definição

(Amostra Aleatória) Sejam as variáveis aleatórias $X_1, X_2, \ldots, X_n$ com densidade conjunta dada por $f_{X_1,X_2,\cdots,X_n}(\cdot, \ldots, \cdot)$ que pode ser escrita da seguinte forma:

$$f_{X_1,X_2,\cdots,X_n}(\cdot, \ldots, \cdot) = f_{X_1}(x_1) f_{X_2}(x_2) \cdots f_{X_n}(x_n)$$

onde $f(\cdot)$ é a densidade de cada X_i. Então, $X_1, X_2, \ldots, X_n$ é definida como sendo *uma amostra aleatória* de tamanho n de uma população com densidade $f(\cdot)$.

Amostras aleatórias

Oi, você por aqui?

Definição

(Amostra Aleatória) Sejam as variáveis aleatórias $X_1, X_2, \ldots, X_n$ com densidade conjunta dada por $f_{X_1, X_2, \cdots, X_n}(\cdot, \ldots, \cdot)$ que pode ser escrita da seguinte forma:

$$f_{X_1, X_2, \cdots, X_n}(\cdot, \ldots, \cdot) = f_{X_1}(x_1) f_{X_2}(x_2) \cdots f_{X_n}(x_n)$$

onde $f(\cdot)$ é a densidade de cada X_i. Então, $X_1, X_2, \ldots, X_n$ é definida como sendo *uma amostra aleatória* de tamanho n de uma população com densidade $f(\cdot)$.

Observação 1: Note que necessariamente os X_i precisarão ser amostrados ***COM*** reposição.

Observação 2: Para ser uma a.a., precisa ser i.i.d..

Amostras aleatórias

Oi, você por aqui?

Exemplo

a.a. com $n = 2$ **de uma Bernoulli**

Suponha que X só pode assumir dois valores, 0 e 1, com probabilidades p e $q = 1 - p$, respectivamente. Isto é, X é uma variável aleatória discreta com distribuição de Bernoulli:

$$p_X(x) = p^x q^{1-x} \mathbb{I}_{\{0,1\}}(x)$$

Onde $\mathbb{I}$ é a função *indicadora*, que será igual a 1 se $x = 0$ ou $x = 1$ e será igual a 0 em todos os outros casos.

Amostras aleatórias

Oi, você por aqui?

Exemplo

a.a. com $n = 2$ **de uma Bernoulli**

Suponha que X só pode assumir dois valores, 0 e 1, com probabilidades p e $q = 1 - p$, respectivamente. Isto é, X é uma variável aleatória discreta com distribuição de Bernoulli:

$$p_X(x) = p^x q^{1-x} \mathbb{I}_{\{0,1\}}(x)$$

Onde $\mathbb{I}$ é a função *indicadora*, que será igual a 1 se $x = 0$ ou $x = 1$ e será igual a 0 em todos os outros casos.

A função densidade conjunta para uma amostra aleatória da $f(\cdot)$ que tenha 2 valores é:

$$f_{X_1,X_2}(x_1, x_2) = f(x_1)f(x_2) = p^{x_1+x_2} q^{2-x_1-x_2} \mathbb{I}_{\{0,1\}}(x_1)\mathbb{I}_{\{0,1\}}(x_2)$$

Amostras aleatórias

Oi, você por aqui?

Exemplo

Distribuição amostral da exponencial (Retirado de [Casella and Berger, 2002])

Seja $X_1, \ldots, X_n$ uma a.a. de uma população com distribuição exponencial de parâmetro λ. Então, a densidade conjunta é dada por:

Amostras aleatórias

Oi, você por aqui?

Exemplo

Distribuição amostral da exponencial (Retirado de [Casella and Berger, 2002])

Seja $X_1, \ldots, X_n$ uma a.a. de uma população com distribuição exponencial de parâmetro λ. Então, a densidade conjunta é dada por:

$$\begin{aligned} f(x_1, \ldots, x_n|\lambda) &= \prod_{i=1}^{n} f(x_i|\lambda) \\ &= \prod_{i=1}^{n} \frac{1}{\lambda} e^{-\frac{x_i}{\lambda}} \\ &= \frac{1}{\lambda^n} e^{-\frac{x_1+x_2+\ldots+x_n}{\lambda}} = \frac{1}{\lambda^n} e^{-\frac{1}{\lambda}\sum_{i=1}^{n} x_i} \end{aligned}$$

Amostras aleatórias

Oi, você por aqui?

Exemplo

Distribuição amostral da exponencial (Retirado de [Casella and Berger, 2002])
Seja $X_1, \ldots, X_n$ uma a.a. de uma população com distribuição exponencial de parâmetro λ. Então, a densidade conjunta é dada por:

$$
\begin{aligned}
f(x_1, \ldots, x_n|\lambda) &= \prod_{i=1}^{n} f(x_i|\lambda) \\
&= \prod_{i=1}^{n} \frac{1}{\lambda} e^{-\frac{x_i}{\lambda}} \\
&= \frac{1}{\lambda^n} e^{-\frac{x_1 + x_2 + \ldots + x_n}{\lambda}} = \frac{1}{\lambda^n} e^{-\frac{1}{\lambda} \sum_{i=1}^{n} x_i}
\end{aligned}
$$

Podemos usar a f.g.m. para calcular densidades de a.a.!

Estatística

Definição

Estatística
Seja $X_1, \ldots, X_n$ uma a.a. de tamanho n de uma população e seja $T(X_1, \ldots, X_n)$ uma função real (ou um vetor de funções reais) cujo domínio inclui o espaço amostral de $(X_1, \ldots, X_n)$. Então a v.a. ou o vetor aleatório $Y = T(X_1, \ldots, X_n)$ é chamado de *estatística*. A função densidade de probabilidade de uma estatística Y é chamada de *distribuição amostral de Y*.

Observação: Note que a definição de estatística é bastante abrangente e não necessariamente Y irá ser uma função do parâmetro populacional θ.

Estatística

Exemplo

Distribuição amostral de $\bar{X}$ para a distribuição Normal

Considere uma a.a. de tamanho n de uma população com distribuição Normal de média μ e variância σ^2, isto é, $X_1, \ldots, X_n$ são i.i.d. com $X_i \sim \mathcal{N}(\mu, \sigma^2)$. Defina a estatística $\bar{X}$ como sendo:

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

Estatística

Exemplo

Distribuição amostral de $\bar{X}$ para a distribuição Normal

Considere uma a.a. de tamanho n de uma população com distribuição Normal de média μ e variância σ^2, isto é, $X_1, \ldots, X_n$ são i.i.d. com $X_i \sim \mathcal{N}(\mu, \sigma^2)$. Defina a estatística $\bar{X}$ como sendo:

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

Utilizando a função geradora de momentos, podemos encontrar qual a densidade de $\bar{X}$!

Função de Verossimilhança

Definição

Função de Verossimilhança
Seja $f(\boldsymbol{x}|\theta)$ a densidade conjunta de uma amostra $\boldsymbol{X} = (X_1, \ldots, X_n)$. Então, dado que $\boldsymbol{X} = \boldsymbol{x}$ foi observada, a função de $\boldsymbol{\theta}$ definida como

$$L(\theta|\boldsymbol{x}) = f(\boldsymbol{x}|\theta)$$

é chamada de *função de verossimilhança*.
Em particular, se $X_1, \ldots, X_n$ é uma amostra aleatória, então:

$$L(\theta|\boldsymbol{x}) = f(\boldsymbol{x}|\theta) = \prod_{i=1}^{n} f(x_i|\boldsymbol{\theta})$$

Função de Verossimilhança

- A função de verossimilhança contém toda a informação que a amostra tem sobre θ.
  - Se houver um escalar α tal que, para duas observações de uma distribuição com parâmetro θ, denotadas por X_1 e X_2, seja possível escrever $\ell(\theta|X_1) = \alpha \cdot \ell(\theta|X_2)$ para todo θ, então X_1 e X_2 levam às mesmas conclusões no processo de inferência.

Função de Verossimilhança

- A função de verossimilhança contém toda a informação que a amostra tem sobre θ.
  - Se houver um escalar α tal que, para duas observações de uma distribuição com parâmetro θ, denotadas por X_1 e X_2, seja possível escrever $\ell(\theta|X_1) = \alpha \cdot \ell(\theta|X_2)$ para todo θ, então X_1 e X_2 levam às mesmas conclusões no processo de inferência.
  - Ela **NÃO** é uma densidade!

Estimadores

- Estimadores são estatísticas e nos auxiliam no processo de inferência

Estimadores

- Estimadores são estatísticas e nos auxiliam no processo de inferência
- Existem métodos diferentes para encontrar estimadores: EMV, MM, MGM, MQO, etc

Estimadores

- Estimadores são estatísticas e nos auxiliam no processo de inferência
- Existem métodos diferentes para encontrar estimadores: EMV, MM, MGM, MQO, etc
- Iremos avaliar estimadores de acordo com suas propriedades: média, variância, EQM, etc

Estimadores

- Estimadores são estatísticas e nos auxiliam no processo de inferência
- Existem métodos diferentes para encontrar estimadores: EMV, MM, MGM, MQO, etc
- Iremos avaliar estimadores de acordo com suas propriedades: média, variância, EQM, etc
- Podemos avaliar estimadores de acordo com propriedades assintóticas também.

O estimador de máxima verossimilhança

Definição

Estimador de máxima verossimilhança

Para cada ponto da amostra $\boldsymbol{x}$, seja $\hat{\theta}(\boldsymbol{x})$ o valor do parâmetro no qual $L(\theta|\boldsymbol{x})$ atinge seu máximo como função de θ (e $\boldsymbol{x}$ fixo). O estimador de máxima verossimilhança (EMV) do parâmetro θ baseado na amostra $\boldsymbol{X}$ é $\hat{\theta}(\boldsymbol{X})$.

Observação: Se a função de verossimilhança é diferenciável em θ_i, os possíveis candidatos a EMV são os valores de $(\theta_1, \ldots, \theta_k)$ que são solução de

$$\frac{\partial}{\partial \theta_i} L(\boldsymbol{\theta}|\boldsymbol{x}) = 0, \quad i = 1, \ldots, k$$

Referências I

Casella, G. and Berger, R. (2002).
Statistical inference.
Duxbury, 2nd edition.