Revisão de Probabilidade e Estatística

Aula 4 - Parte 1

Aishameriane Schmidt

PPGECO/UFSC

Fevereiro de 2019.

Programa

Aula 4

1. Continuação das distribuições da aula passada

Programa

Aula 4

1. Continuação das distribuições da aula passada
2. Função geradora de momentos

Programa
Aula 4

1. Continuação das distribuições da aula passada
2. Função geradora de momentos
3. Distribuição normal multivariada

Programa

Aula 4

1. Continuação das distribuições da aula passada
2. Função geradora de momentos
3. Distribuição normal multivariada
4. Estatística
    - Distribuições amostrais
    - Função de Verossimilhança
    - Estimação pontual
5. Introdução à inferência bayesiana

Distribuição Exponencial

Densidade Acumulada

Teorema

Se $X \sim Exp(\lambda)$, então sua f.d.a. é dada por

$$F_X(x) = \mathbb{P}(X \leq x) = 1 - e^{-\frac{x}{\lambda}}$$

- **Observação:** Note que a f.d.a. nos diz a probabilidade de X ser menor ou igual que um valor qualquer, de maneira que seu complementar é a probabilidade de X ultrapassar um determinado valor;

Distribuição Exponencial

Densidade Acumulada

Teorema

Se $X \sim Exp(\lambda)$, então sua f.d.a. é dada por

$$F_X(x) = \mathbb{P}(X \leq x) = 1 - e^{-\frac{x}{\lambda}}$$

- **Observação:** Note que a f.d.a. nos diz a probabilidade de X ser menor ou igual que um valor qualquer, de maneira que seu complementar é a probabilidade de X ultrapassar um determinado valor;
- Se X é o tempo até um equipamento falhar:
  - Usamos a f.d.a. para calcular a probabilidade de que o equipamento queime em menos de 10 minutos;

Distribuição Exponencial

Densidade Acumulada

Teorema

Se $X \sim Exp(\lambda)$, então sua f.d.a. é dada por

$$F_X(x) = \mathbb{P}(X \leq x) = 1 - e^{-\frac{x}{\lambda}}$$

- **Observação:** Note que a f.d.a. nos diz a probabilidade de X ser menor ou igual que um valor qualquer, de maneira que seu complementar é a probabilidade de X ultrapassar um determinado valor;
- Se X é o tempo até um equipamento falhar:
  - Usamos a f.d.a. para calcular a probabilidade de que o equipamento queime em menos de 10 minutos;
  - E usamos o complementar para calcular a probabilidade de durar pelo menos 10 minutos.

Distribuição Exponencial
Falta de memória

Teorema

Se $X \sim Exp(\lambda)$, então

$$\mathbb{P}(X > t + s | X > t) = \mathbb{P}(X > s)$$

- Se estamos avaliando tempo entre clientes que chegam em uma agência bancária e queremos saber a probabilidade de um cliente chegar 30 minutos depois da agência estar aberta há 3 horas, isso é a mesma coisa que calcular a probabilidade de um cliente chegar meia hora depois do banco abrir.

Distribuição Exponencial

Exercício

Sejam $X_1 \sim \text{Exp}(\lambda_1)$ e $X_2 \sim \text{Exp}(\lambda_2)$ duas v.a. aleatórias independentes e seja $Z = \min\{X_1, X_2\}$.

Encontre a f.d.a. de Z. Que distribuição é essa?

Distribuição Exponencial

Exercício

Sejam $X_1 \sim \text{Exp}(\lambda_1)$ e $X_2 \sim \text{Exp}(\lambda_2)$ duas v.a. aleatórias independentes e seja $Z = \min\{X_1, X_2\}$.

Encontre a f.d.a. de Z. Que distribuição é essa?

Dica: Utilize a definição: Se X e Y são independentes, então $\mathbb{P}(X \cap Y) = \mathbb{P}(X)\mathbb{P}(Y)$.

Distribuição Gama

Definição

Função Gama

$\Gamma : \mathbb{R}^+ \rightarrow \mathbb{R}^+$ é chamada de função Gama e é tal que:

$$\Gamma(z) = \int_0^{+\infty} t^{z-1} e^{-t} dt$$

Distribuição Gama

Definição

Função Gama

$\Gamma : \mathbb{R}^+ \to \mathbb{R}^+$ é chamada de função Gama e é tal que:

$$\Gamma(z) = \int_0^{+\infty} t^{z-1} e^{-t} dt$$

Propriedades: (demonstração em [Stern and Izbicki, 2016])

1. Para $a \geq 1$, temos que $\Gamma(a) = (a-1)\Gamma(a-1)$
2. Se $n \in \mathbb{N}$, $\Gamma(n) = (n-1)!$

Distribuição Gama

Definição

Função Gama

Dizemos que uma variável aleatória X tem distribuição Gama com parâmetros(k, λ) e denotamos por $X \sim \text{Gama}(k, \lambda)$ se a densidade de X é

$$f_X(x) = \begin{cases} \frac{1}{\Gamma(k)\lambda^k} x^{k-1} e^{-\frac{x}{\lambda}} & \text{, if } x > 0 \\ 0 & \text{, caso contrário} \end{cases}$$

- Em est. bayesiana, a distribuição Gama é utilizada como priori para modelar variâncias (ou precisões);

Distribuição Gama

Definição

Função Gama

Dizemos que uma variável aleatória X tem distribuição Gama com parâmetros(k, λ) e denotamos por $X \sim \text{Gama}(k, \lambda)$ se a densidade de X é

$$f_X(x) = \begin{cases} \frac{1}{\Gamma(k)\lambda^k} x^{k-1} e^{-\frac{x}{\lambda}} & \text{, if } x > 0 \\ 0 & \text{, caso contrário} \end{cases}$$

- Em est. bayesiana, a distribuição Gama é utilizada como priori para modelar variâncias (ou precisões);
  - No modelo de regressão linear múltiplo, a priori Gama para a precisão dos erros com uma priori Normal para os coeficientes junto com verossimilhança normal resulta em uma posteriori conjunta Normal-Gama.

Distribuição Gama

Definição

Função Gama

Dizemos que uma variável aleatória X tem distribuição Gama com parâmetros(k, λ) e denotamos por $X \sim \text{Gama}(k, \lambda)$ se a densidade de X é

$$f_X(x) = \begin{cases} \frac{1}{\Gamma(k)\lambda^k} x^{k-1} e^{-\frac{x}{\lambda}} & \text{, if } x > 0 \\ 0 & \text{, caso contrário} \end{cases}$$

- Em est. bayesiana, a distribuição Gama é utilizada como priori para modelar variâncias (ou precisões);
  - No modelo de regressão linear múltiplo, a priori Gama para a precisão dos erros com uma priori Normal para os coeficientes junto com verossimilhança normal resulta em uma posteriori conjunta Normal-Gama.
- Se $X \sim \text{Exp}(\lambda)$, então $X \sim Gama(1, \lambda)$
  - Dizemos que a distribuição exponencial é um *caso particular* da distribuição Gama.

Distribuição Gama

Lema

- *A densidade Gama de fato é uma densidade de probabilidade (sempre positiva e integra 1);*
- *Se $X \sim Gama(k, \lambda)$, então:*
  - $\mathbb{E}[X] = k\lambda$
  - $Var[X] = k\lambda^2$

Distribuição Beta

Definição

Dizemos que X tem distribuição Beta com parâmetros (α, β) e denotamos por $X \sim \text{Beta}(\alpha, \beta)$ se a densidade de X é

$$f_X(x) = \begin{cases} \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \cdot x^{\alpha-1}(1-x)^{\beta-1} & \text{, if } 0 < x < 1 \\ 0 & \text{, caso contrário} \end{cases}$$

- Em bayesiana, a densidade Beta é comumente utilizada como priori para parâmetros que signifiquem elasticidades (em módulo), probabilidades e/ou o coeficiente de um processo *AR* que seja estacionário;

Distribuição Beta

Definição

Dizemos que X tem distribuição Beta com parâmetros (α, β) e denotamos por $X \sim \text{Beta}(\alpha, \beta)$ se a densidade de X é

$$f_X(x) = \begin{cases} \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \cdot x^{\alpha-1}(1-x)^{\beta-1} & \text{, if } 0 < x < 1 \\ 0 & \text{, caso contrário} \end{cases}$$

- Em bayesiana, a densidade Beta é comumente utilizada como priori para parâmetros que signifiquem elasticidades (em módulo), probabilidades e/ou o coeficiente de um processo *AR* que seja estacionário;
  - A densidade Beta é conjugada da Binomial, isso significa que uma priori Beta com uma verossimilhança Binomial resulta em uma posteriori Beta.

Distribuição Beta

Lema

A densidade na Definição 6 é uma fdp válida. Em particular,

$$\int_0^1 \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(1-x)^{\beta-1}dx = 1$$

Distribuição Beta

Lema

A densidade na Definição 6 é uma fdp válida. Em particular,

$$\int_0^1 \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(1-x)^{\beta-1} dx = 1$$

Lema

Se $X \sim Beta(\alpha, \beta)$, então para todo $c, d \geq 0$ $\mathbb{E}[X^c(1-X)^d] = \frac{\Gamma(\alpha+c)\Gamma(\beta+d)}{\Gamma(\alpha+\beta+c+d)}$. Portanto, $\mathbb{E}[X] = \frac{\alpha}{\alpha+\beta}$ e $Var[X] = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$.

Demonstração: Fica como exercício :) A média e variância estão nas notas de aula. A densidade se alguém for tentar e quiser ajuda, pode pedir pra Aisha.

Distribuição Normal

Definição

Dizemos que X tem distribuição Normal com parâmetros (μ, σ^2) e denotamos por $X \sim \mathrm{N}(\mu, \sigma^2)$ se a densidade de X é

$$f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

Se $\mu = 0$, e $\sigma = 1$, dizemos que X tem distribuição normal padrão.

Distribuição Normal

Definição

Dizemos que X tem distribuição Normal com parâmetros (μ, σ^2) e denotamos por $X \sim \mathrm{N}(\mu, \sigma^2)$ se a densidade de X é

$$f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

Se $\mu = 0$, e $\sigma = 1$, dizemos que X tem distribuição normal padrão.

Definição

Seja $Z \sim \mathrm{N}(0, 1)$.

$$\phi(z) = \mathbb{P}(Z \leq z) = F_Z(z)$$

ϕ não tem solução analítica.

Distribuição Normal

Lema

Se $X \sim N(\mu, \sigma^2)$, *então* $\frac{(X-\mu)}{\sigma} \sim N(0, 1)$.

Distribuição Normal

Lema

Se $X \sim N(\mu, \sigma^2)$*, então* $\frac{(X-\mu)}{\sigma} \sim N(0, 1)$*.*

Lema

Se $X \sim N(\mu, \sigma^2)$*, então* $\mathbb{E}[X] = \mu$ *e* $Var[X] = \sigma^2$*.*

Densidades Conjuntas

Motivação

Considere duas v.a. discretas, X e Y, associadas ao mesmo experimento. As probabilidades que X e Y podem assumir são descritas pela função densidade de probabilidade conjunta de X e Y:

$$f_{X,Y}(x, y) = \mathbb{P}(\{X = x\} \cap \{Y = y\}) = \mathbb{P}(X = x, Y = y)$$

Densidades Conjuntas

Motivação

Considere duas v.a. discretas, X e Y, associadas ao mesmo experimento. As probabilidades que X e Y podem assumir são descritas pela função densidade de probabilidade conjunta de X e Y:

$$f_{X,Y}(x,y) = \mathbb{P}(\{X = x\} \cap \{Y = y\}) = \mathbb{P}(X = x, Y = y)$$

A f.d.p. conjunta de X e Y determina a probabilidade de qualquer evento que possa ser especificado em termos de X e Y:

$$\mathbb{P}((X,Y) \in A) = \sum_{(x,y)\in A} f_{X,Y}(x,y)$$

Densidades Conjuntas

Motivação

Considere duas v.a. discretas, X e Y, associadas ao mesmo experimento. As probabilidades que X e Y podem assumir são descritas pela função densidade de probabilidade conjunta de X e Y:

$$f_{X,Y}(x, y) = \mathbb{P}(\{X = x\} \cap \{Y = y\}) = \mathbb{P}(X = x, Y = y)$$

A f.d.p. conjunta de X e Y determina a probabilidade de qualquer evento que possa ser especificado em termos de X e Y:

$$\mathbb{P}((X, Y) \in A) = \sum_{(x,y)\in A} f_{X,Y}(x, y)$$

Pode-se calcular a f.d.p. somente de X ou de Y (*densidade marginal*) através da conjunta:

$$f_X(x) = \sum_y f_{X,Y}(x, y) \qquad f_Y(y) = \sum_x f_{X,Y}(x, y)$$

Definição

Sejam X e Y duas variáveis aleatórias discretas. A fmp conjunta de X e Y é $p_{X,Y}(x,y) := \mathbb{P}(X = x, Y = y)$. De forma similar, se X e Y são duas variáveis aleatórias com distribuições contínuas, a fdp conjunta de X e Y é $f_{X,Y}(x,y)$. Seja A um subconjunto de $\mathbb{R}^2$,

$$\mathbb{P}((X,Y) \in A) = \mathbb{P}(\{\omega \in \Omega : (X(\omega), Y(\omega)) \in A\}) =$$
$$\begin{cases} \sum_{(x,y)\in A} p_{X,Y}(x,y) & \text{, se } X \text{ e } Y \text{ são discretas.} \\ \int_{(x,y)\in A} f_{X,Y}(x,y)d(x,y) & \text{, se } X \text{ e } Y \text{ são contínuas.} \end{cases}$$

Exemplo

Considere que X e Y are são variáveis aleatórias discretas com a fmp conjunta dada na Tabela abaixo.

X/Y	0	1
0	0.2	0.4
1	0.3	0.1

Exemplo

Sejam X e Y variáveis aleatórias contínuas, com a seguinte fdp

$$f_{X,Y}(x,y) = \begin{cases} cxy & \text{, if } x > 0,\ y > 0,\ x + y < 1 \\ 0 & \text{, caso contrário} \end{cases}$$

Qual é o valor de c?

Independência de v.a.'s

Definição

Duas variáveis aleatórias contínuas X e Y são independentes se e somente se $f_{X,Y}(x, y) = f_X(x)f_Y(y)$.

Independência de v.a.'s

Definição

Duas variáveis aleatórias contínuas X e Y são independentes se e somente se $f_{X,Y}(x,y) = f_X(x) f_Y(y)$.

Lema

Sejam X e Y duas variáveis aleatórias. X e Y são independentes se e somente se para todos $A, B \subset \mathfrak{R}$, $\mathbb{P}(X \in A, Y \in B) = \mathbb{P}(X \in A)\mathbb{P}(Y \in B)$.

Independência de v.a.'s

Definição

Duas variáveis aleatórias contínuas X e Y são independentes se e somente se $f_{X,Y}(x,y) = f_X(x)f_Y(y)$.

Lema

Sejam X e Y duas variáveis aleatórias. X e Y são independentes se e somente se para todos $A, B \subset \mathfrak{R}$, $\mathbb{P}(X \in A, Y \in B) = \mathbb{P}(X \in A)\mathbb{P}(Y \in B)$.

Lema

Duas variáveis contínuas, X e Y, são independentes se e somente se a distribuição conjunta pode ser escrita como $f_{X,Y}(x,y) = g(x)h(y)$ para alguma função g e h.

Densidades Condicionais

Caso discreto

Definição

Se X e Y são variáveis aleatórias discretas, definimos a *massa de probabilidade condicional* de X dado que $Y = y$ por

$$p_{X|Y}(x|y) = \mathbb{P}(X = x|Y = y) = \frac{\mathbb{P}(X = x, Y = y)}{\mathbb{P}(Y = y)},$$

para todos os valores de y tais que $\mathbb{P}(Y = y) > 0$.

Densidades Condicionais

Caso discreto

Definição

Se X e Y são variáveis aleatórias discretas, definimos a *massa de probabilidade condicional* de X dado que $Y = y$ por

$$p_{X|Y}(x|y) = \mathbb{P}(X = x|Y = y) = \frac{\mathbb{P}(X = x, Y = y)}{\mathbb{P}(Y = y)},$$

para todos os valores de y tais que $\mathbb{P}(Y = y) > 0$.

Note que, para um dado y, $g(x) = p_{X|Y}(x|y)$ é uma fmp no sentido em que a estudamos antes.

Densidades Condicionais

Caso contínuo

Definição

Se X e Y são variáveis aleatórias contínuas com densidade conjunta $f(x, y)$, definimos a *densidade de probabilidade condicional* de X dado $Y = y$ por

$$f_{X|Y}(x|y) = \frac{f(x, y)}{f_Y(y)},$$

para todos os valores de y tais que $f_Y(y) > 0$.

Densidades Condicionais

Caso contínuo

Exemplo

Seja densidade conjunta de X e Y definida por

$$f(x,y) = \begin{cases} \frac{15}{2}x(2-x-y) & \text{se } 0 < x, y < 1 \\ 0 & \text{caso contrário} \end{cases}$$

Calcule $X|Y = y$ quando $0 < y < 1$.

Densidades Condicionais

Esperança condicional - caso discreto

Definição

Esperança condicional

A esperança condicional de X dado $Y = y$, no caso discreto, é dada por:

$$\begin{aligned}
\mathbb{E}[X|Y = y] &= \sum_{\omega \in \Omega} X(\omega)\mathbb{P}(\{\omega\}|Y = y) \\
&= \sum_{x \in \mathcal{X}} \sum_{\{\omega : X(\omega) = x\}} X(\omega)\mathbb{P}(\{\omega\}|Y = y) \\
&= \sum_{x \in \mathcal{X}} x \sum_{\{\omega : X(\omega) = x\}} \mathbb{P}(\{\omega\}|Y = y) \\
&= \sum_{x \in \mathcal{X}} x \cdot \mathbb{P}(X = x|Y = y) \\
&= \sum_{x \in \mathcal{X}} x \cdot p_{X|Y}(x|y)
\end{aligned}$$

Densidades Condicionais

Esperança condicional - caso contínuo

Definição

Esperança condicional

A esperança condicional de X dado $Y = y$, no caso contínuo, é dada por:

$$\mathbb{E}[X|Y = y] = \int_{\mathbb{R}} x \cdot f_{X|Y}(x|y)dx$$

Densidades Condicionais

Esperança condicional - caso contínuo

Definição

Esperança condicional

A esperança condicional de X dado $Y = y$, no caso contínuo, é dada por:

$$\mathbb{E}[X|Y = y] = \int_{\mathbb{R}} x \cdot f_{X|Y}(x|y)dx$$

Atenção!

- $\mathbb{E}[X|Y]$ é uma variável aleatória,
- $\mathbb{E}[X|Y = y]$ é um número.

Exemplos interessantes

Aula 4

Exemplo

- Se $f_{X,Y}(x,y) = \frac{e^{-\frac{x}{y}} e^{-y}}{y}$ para $0 < x, y < +\infty$, qual é a distribuição de $X|Y = y$?

Exemplos interessantes

Aula 4

Exemplo

- Se $f_{X,Y}(x,y) = \frac{e^{-\frac{x}{y}} e^{-y}}{y}$ para $0 < x, y < +\infty$, qual é a distribuição de $X|Y = y$?
- Prove que se $X \sim \text{Poisson}(\lambda_1)$ e $Y \sim \text{Poisson}(\lambda_2)$ são independentes e $Z = X + Y$, então

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

Função Geradora de Momentos

Definição

(Função Geradora de Momentos) A função geradora de momentos de uma variável aleatória X é uma função $M_X : \mathbb{R} \to \mathbb{R}$ definida por:

$$M_X(t) = \mathbb{E}[e^{tX}] \tag{1}$$

Isto é, a função geradora de momentos é calculada através da esperança da função e^{tX}.

Função Geradora de Momentos

Lema

- *Se $X \sim$ Bernoulli(p), então a f.g.m. de X é dada por* $M_x(t) = (1 - p) + e^t \cdot p$;

Função Geradora de Momentos

Lema

- *Se $X \sim \text{Bernoulli}(p)$, então a f.g.m. de X é dada por* $M_x(t) = (1 - p) + e^t \cdot p$;
- *Se $X \sim \text{Binomial}(n, p)$, então a f.g.m. de X é dada por* $M_x(t) = \left[(1 - p) + e^t \cdot p\right]^n$;
  - *É possível usar a f.g.m. para provar que soma de Bernoulli é Binomial, mas aqui vamos fazer o caminho inverso e assumir que isso é verdade para encontrar a forma da f.g.m. de forma mais fácil.*

Função Geradora de Momentos

Lema

- *Se* $X \sim Normal(0,1)$, *então a f.g.m. de* X *é dada por* $M_x(t) = e^{\frac{t^2}{2}}$;

Função Geradora de Momentos

Lema

- *Se* $X \sim Normal(0, 1)$, *então a f.g.m. de* X *é dada por* $M_x(t) = e^{\frac{t^2}{2}}$;
- *Se* $X \sim Normal(\mu, \sigma^2)$, *então a f.g.m. de* X *é dada por* $M_x(t) = e^{t\mu + \frac{(t\sigma)^2}{2}}$;

Função Geradora de Momentos

Lema

- *Se* $X \sim Normal(0,1)$, *então a f.g.m. de* X *é dada por* $M_x(t) = e^{\frac{t^2}{2}}$;
- *Se* $X \sim Normal(\mu, \sigma^2)$, *então a f.g.m. de* X *é dada por* $M_x(t) = e^{t\mu + \frac{(t\sigma)^2}{2}}$;
- *Se* $X \sim Normal(\mu_Y, \sigma_Y^2)$ *e* $Y \sim Normal(\mu_Y, \sigma_Y^2)$ *(independentes), então a f.g.m. de* $X + Y$ *é dada por* $M_x(t) = e^{t(\mu_X+\mu_Y)+\frac{\left(t\sqrt{\sigma_X^2+\sigma_Y^2}\right)^2}{2}}$;

Função Geradora de momentos

Lema

- *Se* $X \sim Poisson(\lambda)$, *então a f.g.m. de* X *é dada por* $M_x(t) = exp\left\{\lambda\left(e^t - 1\right)\right\}$;

Função Geradora de momentos

Lema

- *Se* $X \sim Poisson(\lambda)$*, então a f.g.m. de X é dada por* $M_x(t) = exp\left\{\lambda\left(e^t - 1\right)\right\}$;
- *Se* $X_i \sim Poisson(\lambda_i)$ *para* $i = 1, 2, \ldots, n$ *e* X_i *é independente de* X_j *(para* $i \neq j$*), então a f.g.m. de* $\sum_{i=1}^{n} X_i$ *é dada por* $M_x(t) = exp\left\{\sum_{i=1}^{n} \lambda_i\left(e^t - 1\right)\right\}$;

Função Geradora de momentos

Lema

1. *A f.g.m. (quando pode ser definida) caracteriza unicamente uma distribuição;*

Função Geradora de momentos

Lema

1. *A f.g.m. (quando pode ser definida) caracteriza unicamente uma distribuição;*
2. *Se $X_1, \dots, X_n$ são i.i.d.,*

$$M_{\sum X_i}(t) = (M_{X_i}(t))^n$$

(Vamos usar esse resultado hoje)

Função Geradora de momentos

Lema

1. *A f.g.m. (quando pode ser definida) caracteriza unicamente uma distribuição;*
2. *Se $X_1, \ldots, X_n$ são i.i.d.,*

$$M_{\sum X_i}(t) = \left(M_{X_i}(t)\right)^n$$

   (Vamos usar esse resultado hoje)

3. *Se $M_X(t)$ é a f.g.m. de X, então,*

$$\left.\frac{\partial^n M_X(t)}{\partial t^n}\right|_{t=0} = \mathbb{E}\left[X^n\right] \quad \forall n \in \{1, 2, \ldots, \}.$$

A demonstração do item 2 é sugerida como exercício.

Função Geradora de momentos

Exemplo

Vimos que se $X \sim$ Normal(μ, σ^2), então a f.g.m. de X é dada por $M_x(t) = e^{t\mu + \frac{(t\sigma)^2}{2}}$. Vamos calcular o primeiro e o segundo momento de X usando a f.g.m..

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.

Distribuições Multivariadas

A normal multivariada

- 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

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

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

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 $W = X - Y$.

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,

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

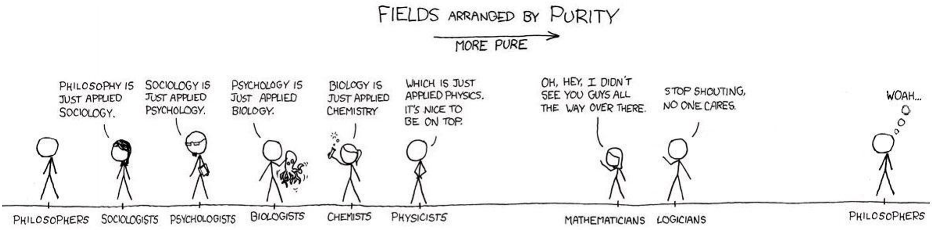
- Mostre que se Σ é diagonal, as componentes de $\underset{\sim}{X}$ são independentes e diga qual a sua distribuição. O que isso significa?
- A Normal é um dos poucos casos onde ausência de correlação implica independência!

Estatística

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

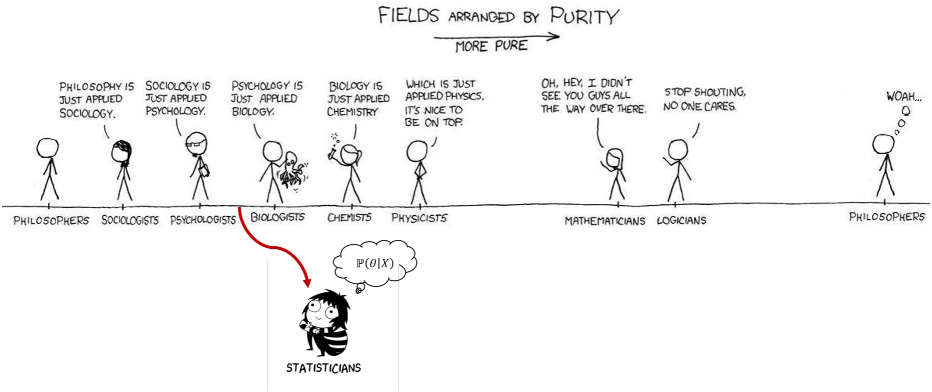
Estatística

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



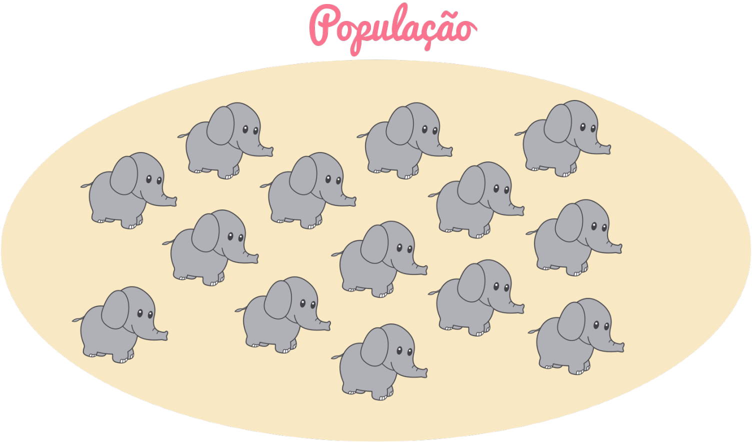
Estatística

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



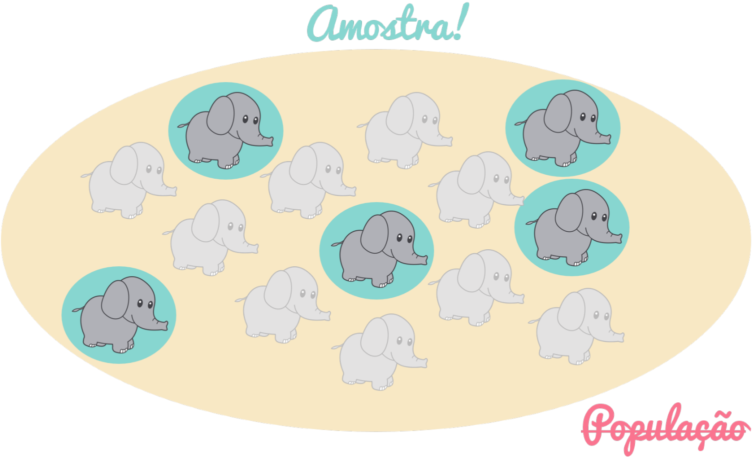
Estatística

O problema fundamental da estatística



Estatística

O problema fundamental da estatística



Estatística

O problema fundamental da estatística

Usamos os dados da *amostra*

Em conjunto com *estimadores*

Para obter *estimativas*

Dos *parâmetros populacionais*

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.

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:

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

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!

Referências I

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

Stern, R. and Izbicki, R. (2016).
Introducao à Teoria das Probabilidades e Processos Aleatorios.
UFSCAR.