Revisão de probabilidade

Densidade de probabilidade e outras funções e medidas para distrib. contínua

André Gustavo dos Santos

Departamento de Informática
Universidade Federal de Viçosa

INF222 - 2022/2

Tópicos da aula

1. Variáveis contínuas
2. Densidade de probabilidade
3. Densidades conjunta e marginal
4. Esperança e variância

UFV

Variáveis contínuas

- Para toda variável contínua, a função massa de probabilidade é zero

$$f(x) = 0, \quad \forall x$$

  pois uma variável contínua, mesmo restrita a um intervalo, tem infinitos valores
- De fato, como $\sum_x f(x) = 1$, a fmp pode ter probabilidade positiva apenas para um conjunto finito ou contável de valores
  - Apenas 2 valores de x podem ter $f(x) \geq 1/2$
  - Apenas 4 valores de x podem ter $f(x) \geq 1/4$
  - Continuando, podemos sempre listar os valores com $f(x) > 0$
  - O conjunto com $f(x) > 0$ é contável; não pode ser um intervalo com infinitos valores
- Portanto, a fmp não traz nenhuma informação sobre uma var. aleatória contínua
- Mas podemos usar a fda, função de distribuição acumulada, $F(x) = \mathbf{P}\{X \leq x\}$

Variáveis contínuas – função de distribuição

- Assim como para variáveis discretas, a função de distribuição $F(x) = \mathbf{P}\{X \leq x\}$ é não-decrescente e tem valores entre 0 e 1
- Porém não dá "saltos" de tamanho $f(x)$, pois $f(x) = 0$
- Ou seja, a função de distribuição $F(x)$ é uma função contínua
- Note ainda que, para variáveis contínuas, $F(x) = \mathbf{P}\{X \leq x\} = \mathbf{P}\{X < x\}$, pois a diferença entre elas é $f(x)$, que vale 0

Função densidade

- Vimos que a fda, $F(x)$, é contínua para variáveis aleatórias contínuas
- Suponha, adicionalmente, que seja diferenciável[1]

Definição

A função densidade de probabilidade (fdp), ou densidade, é a derivada da fda

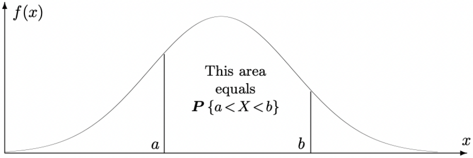
$$f(x) = F'(x)$$

- $F(x)$ é então a primitiva (antiderivada) da densidade $f(x)$
- Assim, $\mathbf{P}\{a < X \leq b\} = F(b) - F(a) = \int_a^b f(x)dx$
- Note que, como $\mathbf{P}\{X \leq x\} = \mathbf{P}\{X < x\}$, temos que $\mathbf{P}\{a < X \leq b\}$ é o mesmo que $\mathbf{P}\{a \leq X \leq b\}$, $\mathbf{P}\{a \leq X < b\}$ e $\mathbf{P}\{a < X < b\}$

[1] não é verdade no caso geral, mas ocorre em todas as funções de distribuição comumente usadas

Função densidade

- A função densidade é uma curva contínua
- A área sob a curva entre os pontos *a* e *b* é a probabilidade do intervalo
- A área total sob a curva é 1



Função densidade

- $f(x) = F'(x)$
- $\mathbf{P}\{a \leq X \leq b\} = \int_a^b f(x)dx$
- $\int_{-\infty}^{b} f(x)dx = \mathbf{P}\{-\infty \leq X \leq b\} = \mathbf{P}\{X \leq b\}$
- $\int_{-\infty}^{+\infty} f(x)dx = \mathbf{P}\{-\infty \leq X \leq +\infty\} = 1$

- Note mais uma vez por que $f(x) = 0$: temos $f(x) = \mathbf{P}\{x \leq X \leq x\} = \int_x^x f = 0$ (área sob a curva de intervalo de largura 0)

Função densidade

Vida útil

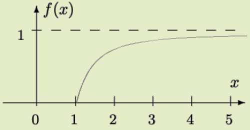
A vida útil de certo componente eletrônico, em anos, é uma variável aleatória contínua com densidade $f(x)$ ao lado:

$$f(x) = \begin{cases} \frac{k}{x^3} & \text{se } x > 1 \\ 0 & \text{se } x \leq 1 \end{cases}$$

- Encontre k e desenhe o gráfico da fda $F(x)$
- Calcule a probabilidade da vida útil exceder 5 anos

Solução:

- Temos que $\int_{-\infty}^{+\infty} f(x)dx = 1$,
  então: $\int_{-\infty}^{+\infty} f(x)dx = \int_{1}^{+\infty} \frac{k}{x^3} dx = -\frac{k}{2x^2}\Big|_{x=1}^{+\infty} = \frac{k}{2} = 1 \therefore k = 2$
- A fda é a integral da densidade:
  $F(x) = \int_{-\infty}^{x} f(y)dy = \int_{1}^{x} \frac{2}{y^3} dy = -\frac{1}{y^2}\Big|_{y=1}^{x} = 1 - \frac{1}{x^2}$ para $x > 1$
- Gráfico de $F(x)$:



- Probabilidade de exceder 5 anos:
  $\mathbf{P}\{X > 5\} = 1 - F(5) = 1 - (1 - \frac{1}{5^2}) = 0.04$

  ou, integrando a densidade:
  $\mathbf{P}\{X > 5\} = \int_{5}^{+\infty} f(x)dx = \int_{5}^{+\infty} \frac{2}{x^3} dx = -\frac{1}{x^2}\Big|_{x=5}^{+\infty} = \frac{1}{25} = 0.04$

Analogia fmp x fdp

- A função densidade para distribuição contínua tem papel semelhante ao da função massa de probabilidade para distribuições discretas
- Troca-se fmp por fdp e somatório por integração

Distribuição	Discreta	Contínua
Definição	$f(x) = \mathbf{P}\{X = x\}$ (fmp)	$f(x) = F'(x)$ (fdp)
Cálculo de probabilidade	$\mathbf{P}\{X \in A\} = \sum_{x \in A} f(x)$	$\mathbf{P}\{X \in A\} = \int_A f(x)dx$
Função distribuição acumulada	$F(x) = \mathbf{P}\{X \leq x\} = \sum_{y \leq x} f(y)$	$F(x) = \mathbf{P}\{X \leq x\} = \int_{-\infty}^{x} f(y)dy$
Probabilidade total	$\sum_x f(x) = 1$	$\int_{-\infty}^{+\infty} f(x)dx = 1$

UFV

Densidade conjunta

Definição

Para vetores aleatórios temos a distribuição de probabilidade conjunta, definida por

$$F_{XY}(x, y) = \mathbf{P}\{X \leq x \cap Y \leq y\}$$

e a densidade conjunta é sua derivada mista

$$f_{XY}xy = \frac{\partial^2}{\partial x \partial y} F_{XY}(x, y)$$

Densidade marginal

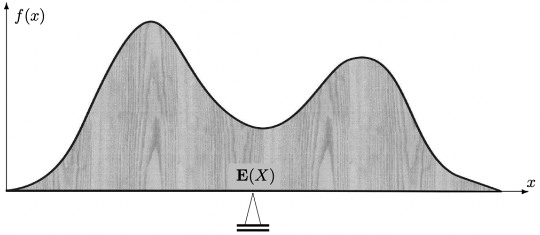
- Densidades marginais das variáveis de um vetor aleatório podem ser calculadas de forma semelhante ao cálculo da distribuição marginal de variáveis discretas
- Troca-se somatório por integração
- Variáveis são independentes se a densidade conjunta é o produto das densidades marginais

Distribuição	Discreta	Contínua
Distribuições marginais	$f(x) = \sum_y f(x,y)$ $f(y) = \sum_x f(x,y)$	$f(x) = \int f(x,y)dy$ $f(y) = \int f(x,y)dx$
Independência	$f(x,y) = f(x)f(y)$ (fmp)	$f(x,y) = f(x)f(y)$ (fdp)
Cálculo de probabilidades	$\mathbf{P}\{(X,Y) \in A\}$ $= \sum\sum_{(x,y)\in A} f(x,y)$	$\mathbf{P}\{(X,Y) \in A\}$ $= \int\int_{(x,y)\in A} f(x,y)dxdy$

Esperança matemática

- Continuando a analogia com o caso discreto, a esperança matemática também é o centro de gravidade no caso contínuo

$$\mu = \mathbf{E}(X) = \int x\, f(x)dx$$



Variância e outras

- Variância, desvio padrão, covariância e correlação também são definidos de forma simular ao caso discreto
- Troca-se fmp por fdp e somatório por integração

Discreta	Contínua
$\mathbf{E}(X) = \sum_x x\, f(x)$	$\mathbf{E}(X) = \int x\, f(x) dx$
$\begin{aligned} \mathrm{Var}(X) &= \mathbf{E}((X-\mu)^2) \\ &= \sum_x (x-\mu)^2 f(x) \\ &= \sum_x x^2 f(x) - \mu^2 \end{aligned}$	$\begin{aligned} \mathrm{Var}(X) &= \mathbf{E}((X-\mu)^2) \\ &= \int (x-\mu)^2 f(x) dx \\ &= \int x^2 f(x) dx - \mu^2 \end{aligned}$
$\begin{aligned} \mathrm{Cov}(X,Y) &= \mathbf{E}((X-\mu_X)(Y-\mu_Y)) \\ &= \sum_x \sum_y (x-\mu_X)(y-\mu_Y) f(x,y) \\ &= \sum_x \sum_y (xy) f(x,y) - \mu_X \mu_Y \end{aligned}$	$\begin{aligned} \mathrm{Cov}(X,Y) &= \mathbf{E}((X-\mu_X)(Y-\mu_Y)) \\ &= \int \int (x-\mu_X)(y-\mu_Y) f(x,y) dx dy \\ &= \int \int (xy) f(x,y) dx dy - \mu_X \mu_Y \end{aligned}$

Exemplo

Vida útil (continuação)

A vida útil, variável aleatória do exemplo anterior, tem densidade

$$f(x) = 2x^{-3}, \quad \text{para } x \geq 1$$

- Valor esperado: $\mu = \mathbf{E}(X) = \int x\, f(x) dx = \int_1^{\infty} 2x^{-2} dx = -2x^{-1}\Big|_1^{\infty} = 2$
- Variância:
  $\sigma^2 = \mathrm{Var}(X) = \int x^2 f(x) dx - \mu^2 = \int_1^{\infty} 2x^{-1} dx - 4 = 2\ln x\Big|_1^{\infty} - 4 = +\infty$

uma surpresa aqui... a variável tem um valor esperado, 2, mas sua variância é infinita!