Notas de aula: Fundamentos Estatísticos para Ciência dos Dados

Ricardo Pagoto Marinho

22 de maio de 2018

- 13/03

  $P(\cup A_i) \leq \sum P(A_i) \rightarrow$ é igual quantos os $A_i s$ forem disjuntos.

- 15/03

  $P(A|B) = P(B) \rightarrow$ quando A ocorre e não tem nenhuma influência sobre B_0.

- 20/03

  Variável aleatória: Lista de valores possíveis e lista de probabilidades associadas

  ω dentro de um Ω. Exemplo: $\Omega =$ todos e-mails enviados.

  - $\omega_0 =$ é spam?
  - $\omega_1 =$ número de caracteres.
  - $\cdots$

  Elementos em uma mesma linha (ω_n), são correlacionados.

  - Atribuir valores de probabilidades a uma V.A. $\rightarrow$ contar quantos elementos no Ω possuem aquela característica.
    $P(X = 3) = P(A)$ onde A=$\{\omega \in \Omega / \omega \ tem \ 3 \ caras\}$ em $\Omega =$ lançamento de 6 moedas.
  - Esperança matemática $E(X)$
    $E(X) = \sum_i x_i p(xi) \approx \sum_i xi \times \frac{N_i}{N}$
  - Distribuição Binomial
    $P(X = 0) = (1 - \theta)^n$
    $[X = 0] = \{\omega \in \Omega : X(\omega) = 1\} = \{\omega \in \Omega : \omega \in \{(\neg c, \neg c, \neg c, \cdots, \neg c)\}\} =$
    $P(\neg c \ no \ 1^{o}) \times P(\neg \ c \ no \ 2^{o}) \times \cdots = (1 - \theta) \times (1 - \theta) \cdots = (1 - \theta)^n$

- 27/03

  $P(Y \in (y_0 \pm \frac{\delta}{2})) = \int_{y_0 - \frac{\delta}{2}}^{y_0 + \frac{\delta}{2}} f^*(y)dy \approx f^*(y_0) 2 \times \frac{\delta}{2} = f^*(y_0) \times \delta$

  Teste de Kolmogorov:

  $\sqrt{n}(D_n) \rightarrow K$, onde K é uma Variável Aleatória contínua.

  Se o modelo é o verdadeiro, quando comparado com os dados, a distância entre eles multiplicado por $\sqrt{n}$ vai cair dentro da densidade de K. Se não cair, provavelmente seu modelo não é adequado. Quanto maior o número de dados, mais confiável o resultado.

- 03/04

  Variáveis aleatórias: Lista de valores possíveis + probabilidades associadas

	Discretas	Contínuas
Valores	0,1,2,...	[0,1] ou [0,∞)
Probabilidades	$p_0, p_1, p_2, ...$	Densidade sob a curva

	$E(X)$
Discreta	$\sum_i x_i \times\ P(X = x_i)$
Contínua	$\int x \times\ f(x)dx$

Teste qui-quadrado: Compara o modelo com os dados. Serve para dados contínuos e discretos.

pchisq(15,4)-pchisq(10,4) $\rightarrow$ comando em R para saber o valor do teste qui-quadrado no intervalo [10,15] com 4 graus de liberdade.

pchisq(1.13,4) $\rightarrow$ probabilidade de uma distribuição qui-quadrado com 4 graus de liberdade ser menor do que 1.13.

1-pchisq(1.13,4)=pvalors

- 05/04

  As variáveis são i.i.d. (independentes e identicamente distribuídas) se:

  - elas forem todas independentes
  - possuírem todas a mesma distribuição

  - Transformação de V.A.s
    X é V.A.
    Y=g(X) é V.a.
    g(X) é função matemática.
    Distribuição de Y?
    * inverter g : Obter $F_y(y) = P(Y \leq\ y)$ e deriva para obter a densidade $f_y(Y) = F'(y)$
    * Y=g(X) e $X = g^{-1}(Y) = h(Y)$
      então $f_y(y) = f_x(h(y)).|h'(y)|$
      Exemplo:$f_x(x) = \begin{cases} 0, & \text{se } x \ni (0,1) \\ 1, & \text{se } x \in (0,1) \end{cases}$
      $Y = X^2 \rightarrow X = \sqrt{Y} = h(Y)$
      Então:
      $f_y(y) = f_x(\sqrt{y}) \times |\frac{d\sqrt{y}}{dy}|$
      Se quisermos $E(Y) = E(g(x)$
      1. $E(Y) = \int y\ f_Y(y)dy$ ($f_Y(y)$ é obtida de uma das duas maneiras anteriores)
      2. $= \int_{-\infty}^{\infty} g(x) \times\ f_x(x)dx$

- 10/04

$\sum_{i=1}^{n}(Y_i-\bar{Y})^2 = \sum_i(Y_i^2-2Y_i\bar{Y}+\bar{Y}^2) = \sum_i Y_i^2-2\sum_i(Y_i\bar{Y})+\sum_i(\bar{Y}^2) = \sum_i Y_i^2 - 2\bar{Y}\sum_i(Y_i) + n\bar{Y}^2 = \sum_i Y_i^2 - 2\bar{Y}.n\bar{Y} + n\bar{Y}^2 = \sum_i Y_i^2 - n(\bar{Y}^2)$

$\mathbb{P}(X_1 = i, X_2 = j, X_4 = k|X_3 = 2) = \frac{\mathbb{P}(X_1=i,X_2=j,X_4=k,X_3=2)}{\mathbb{P}(X_3=2)}$

Exemplo:

$\mathbb{P}(X_1 = 0, X_2 = 1, X_4 = 1|X_3 = 2) = \frac{\mathbb{P}(X_1=0,X_2=1,X_4=1,X_3=2)}{\mathbb{P})X_3=2)} = \frac{\frac{0.2}{100}}{\frac{32.4}{100}}$

Para comparar resultados na distribuição condicional, basta olhar na tabela a condição não condicional.

- 12/04

  $\int(x^2y + x^3y^4)dx$

  $Y+(Y_1, \ldots, Y_p)$ $(Y_1|Y_2 = a_2, \ldots, Y_p = a_p) = \frac{f_Y(y_1,a_2,\ldots,a+p)}{f_{Y_2\ldots Y_p}} \propto f(y_2, a_2, \ldots, a_p)$

  Desvios padronizados:

  Pearson sugeriu multiplicar eles. O interesse é olhar a esperança desse valor: $\mathbb{E}(Z_1 * Z_2)$, que será a correlação entre as variáveis aleatórias.

- 17/04

  $\rho = \mathbb{E}[\frac{X-\mu_x}{\sigma_x} \times \frac{Y-\mu_y}{\sigma_y}]$

  Os valores da matriz de correlação variam entre (-1,1).

  A distância de desvio padronizado tira o problema de escala entre as duas variáveis calculadas.

  $\sigma y = \lambda y$. $\lambda = auto-valor, y = auto-vetor$

  $\sigma^{-1}(\sigma y) = \sigma^{-1}(\lambda y)$. $y = \lambda\sigma^{-1}y$ $\frac{1}{\lambda}y = \sigma^{-1}y$

- 19/04

  A covariância é uma medida confusa, pois não tem um limite definido para variar, diferente da correlação.

  $$y'Ay = \sum_{ij} A_{ij}y_iy_j$$

  A pode ser considerada simétrica, pois se não for, pode-se criar uma simétrica que tal que a forma quadrática dê a mesma coisa.

- 03/05 Considerando que para um indivíduo, FV é alta (próximo de 2) e FQ é baixa (próximo de 0) $X_{mat} = \mu_{mat} + 0FV + 15FQ + \xi_{mat} + \sim N(,\psi_{mat}) \approx \mu_{mat}$, já que FQ é baixa.

  $X_{gram} = \mu_{gram} + 10FV + 0.1FQ + \xi_{gram} + \sim N(,\psi_{gram}) \gg \mu_{mat}$, já que FV é alta.

- 10/05

  Definir se um indivíduo vem de uma população ou outra, pode-se fazer calculando as distâncias estatísticas do indivíduo para as populações ou fazer o produto $\frac{f_1(x)}{f_2(x)}$ das densidades. Se > 1, é da população 1, se < 1, da população 2.

- 22/05

  Para resolver o sistema linear $Y \approx Xb$, com X com muito mais linhas do que colunas (o que impede de achar a inversa), fazemos (com o exemplo dos apartamentos)

$$\begin{aligned} Y &\approx Xb \\ X^tY &\approx X^tXb \end{aligned}$$

Onde X^t é uma matriz 31x1500, Y é 31x1, X é 1500x31. Logo, X^tY é 31x1, X^tX é 31x31, o que é resolvível.

Espaço vetorial V=R^1500.

Subespaço vetorial fechado para soma, multiplicação por escalar e possui o 0:

$V_1 + V_2 \in W \in V$

$cV_1 \in W$

$0 \in W$

$$\begin{aligned} v_1 = Xb_1, v_2 = Xb_2 \rightarrow v_1 + v_2 = Xb_1 + Xb_2 &= X(b_1 + b_2) \\ 3v_1 = 3Xb_1 &= X(3b_1) \\ b &= \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix} \end{aligned}$$

linearmente independentes se NÃO existir uma combinação linear tal que:

$$b_0[x_0] + b_1[x_1] + b_2[x_2] + b_3[x_3] = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}$$

Queremos um vetor de coeficientes (b) que torne a soma dos erros de predição o menor possível:

$$\min_b \sum_{i=1}^{1500} (Y_i - \hat{Y}_i)^2 = \min_b \sum_{i=1}^{1500} (Y_i - (linha\ i\ de\ x \times b))^2$$