Probabilidade de acontecimentos

$P(A-B)=P(A\cap\overline{B})=P(A)-P(A\cap B)$

$P(A\cup B\cup C)=P(A)+P(B)+P(C)-P(A\cap B)$
$\qquad -P(A\cap C)-P(B\cap C)+P(A\cap B\cap C)$

$P(A\cap B\cap C)=P(ABC)=P(A)P(B|A)P(C|AB)$

$P(\overline{A}|B)=1-P(A|B),\quad$ se $P(B)\neq 0$

X v.a. contínua com f. densidade $f_X(x)$, $Y=\varphi(X)$, estritamente monótona e derivável, então

$$f_Y(y)=f_X(x)\left|\frac{dx}{dy}\right| \text{ com } x=\varphi^{-1}(y)$$

Parâmetros (de funções) de uma v.a. X

Quantil de probabilidade p, χ_p, de uma v.a. X é o menor valor x tal que $F(x)\geq p$ (se X contínua $F(\chi_p)=p$)

$$E[\varphi(X)]=\begin{cases}\sum_i\varphi(x_i)p_i, & \text{se } X \text{ discreta}\\ \int_{-\infty}^{+\infty}\varphi(x)f(x)dx, & \text{se } X \text{ contínua}\end{cases}$$

$\sigma^2=Var(X)=E[(X-\mu)^2]=E[X^2]-\mu^2\ (\mu=E[X])$

$Var(aX+b)=a^2Var(X)$

Função geradora de momentos

$M_X(t)=E[e^{tX}]$

$M_{aX+b}(t)=e^{bt}M_X(at)$

Se X,Y v.a.'s independentes, $M_{X+Y}(t)=M_X(t)M_Y(t)$

Se $M_X(t)$ definida numa vizinhança de zero,

$$\frac{d^{(r)}M_X(t)}{dt^r}\Big|_{t=0}=E[X^r],\qquad r=1,2,\ldots$$

Pares aleatórios (X,Y) discretos com f. massa de probabilidade conjunta $P[X=x_i,Y=y_j]=p_{ij}$

$p_{i\cdot}=\sum_j p_{ij}\qquad p_{\cdot j}=\sum_i p_{ij}$

$P[X=x_i|Y=y_j]=\dfrac{p_{ij}}{p_{\cdot j}}$

Pares aleatórios (X,Y) contínuos com f. densidade conjunta $f(x,y)$

$f_X(x)=\int_{-\infty}^{+\infty}f(x,y)dy\qquad f_Y(y)=\int_{-\infty}^{+\infty}f(x,y)dx$

$f_{X|Y=y}(x|y)=\dfrac{f(x,y)}{f_Y(y)}$, $f_Y(y)\neq 0$

Parâmetros de funções de um par aleatório (X,Y)

$$E[g(X,Y)]=\begin{cases}\sum_i\sum_j g(x_i,y_j)p_{ij}, & (X,Y) \text{ discreto}\\ \iint_{\mathbb{R}^2}g(x,y)f(x,y)dxdy, & (X,Y) \text{ contínuo}\end{cases}$$

$\sigma_{X,Y}=Cov(X,Y)=E[(X-\mu_X)(Y-\mu_Y)]$
$\qquad =E[XY]-E[X]E[Y]\ (\mu_X=E[X],\ \mu_Y=E[Y])$

$Var(X\pm Y)=Var(X)+Var(Y)\pm 2Cov(X,Y)$

$Cov(aX+b,cY+d)=acCov(X,Y)$

$\rho=\rho_{X,Y}=\dfrac{\sigma_{X,Y}}{\sigma_X\sigma_Y}$, $\sigma_X\neq 0$, $\sigma_Y\neq 0$

$\rho_{aX+b,cY+d}=\rho_{X,Y}$ se $ac>0$

Distribuição binomial

$X\cap B(n,p)\Leftrightarrow n-X\cap B(n,q)$, com $q=1-p$

Aproximações das distribuições

$X\cap H(N,n,k)$ e $\dfrac{N}{n}>10\Rightarrow X\sim B(n,p)$, com $p=\dfrac{k}{N}$

$X\cap B(n,p)$, $n\geq 20$ e $p\leq 0.05\Rightarrow X\sim P(\lambda)$, com $\lambda=np$

$X\cap B(n,p)$, $np>5$ e $nq>5\Rightarrow X\sim\mathcal{N}(\mu,\sigma)$, com
$\mu=np$ e $\sigma=\sqrt{npq}$

$X\cap P(\lambda)$ e $\lambda>12\Rightarrow X\sim\mathcal{N}(\mu,\sigma)$, com $\mu=\lambda$ e $\sigma=\sqrt{\lambda}$

Teorema Limite Central

Sejam $X_1,\ldots,X_n$ v.a.'s independentes e identicamente distribuidas com valor médio μ e variância σ^2 (finita), $S_n=\sum_{i=1}^n X_i$ e $\bar{X}_n=\sum_{i=1}^n\dfrac{X_i}{n}$, então

$$\frac{S_n-n\mu}{\sigma\sqrt{n}}\sim\mathcal{N}(0,1)\quad\text{e}\quad\frac{\bar{X}_n-\mu}{\sigma/\sqrt{n}}\sim\mathcal{N}(0,1)$$

Construção de v.a.'s

$Z_1,\ldots,Z_n$ normais reduzidas independentes
$\Rightarrow\sum_{i=1}^n Z_i^2\cap\chi^2_{(n)}$

$Z\cap\mathcal{N}(0,1)$, $Y\cap\chi^2_{(n)}$, independentes $\Rightarrow\dfrac{Z}{\sqrt{Y/n}}\cap t_{(n)}$

$U\cap\chi^2_{(m)}$, $V\cap\chi^2_{(n)}$, independentes $\Rightarrow\dfrac{U/m}{V/n}\cap F_{(m,n)}$

$X\cap F_{(m,n)}\Rightarrow\dfrac{1}{X}\cap F_{(n,m)}$

Expressões úteis

Combinações de n elementos k a k, $n,k\in\mathbb{N}_0$, $k\leq n$

$C_k^n=\dbinom{n}{k}=\dfrac{n!}{k!(n-k)!}$

$\int_0^{+\infty}x^ne^{-x}dx=n!$, $n\in\mathbb{N}_0$

$e^x=1+x+\frac{x^2}{2!}+\frac{x^3}{3!}+\cdots=\sum_{n=0}^{\infty}\frac{x^n}{n!}$

$\sum_{k=0}^n ar^k=a\dfrac{1-r^{n+1}}{1-r}\qquad\sum_{n=0}^{\infty}ar^n=\dfrac{a}{1-r}$ se $|r|<1$

Algumas regras de primitivas

Uma primitiva de xe^{-x} é $-e^{-x}(x+1)$

$P(fg)=Fg-P(Fg')$, com $F=Pf$

$P(f'f^\alpha)=\frac{f^{\alpha+1}}{\alpha+1}+C$, com $\alpha\in\mathbb{R}-\{-1\}$

$P(f'e^f)=e^f+C,\qquad P\frac{f'}{f}=\log|f|+C$

Indicadores para dados univariados

$x_1, x_2, \cdots, x_n$

f.d. empírica $\quad F^*(x) = \frac{\text{n. de } x_i \leq x}{n}$

amostra ordenada $\quad x_{(1)} \leq x_{(2)} \leq \cdots \leq x_{(n)}$

mediana $\quad \tilde{x} = \begin{cases} x_{(\frac{n+1}{2})} & n \text{ ímpar} \\ \frac{x_{(n/2)} + x_{(n/2+1)}}{2} & n \text{ par} \end{cases}$

Quantil de ordem θ $(0 < \theta < 1)$

$Q^*_\theta = \begin{cases} \frac{1}{2}\left(x_{(n\theta)} + x_{(n\theta+1)}\right) & \text{se } n\theta \text{ inteiro} \\ x_{([n\theta]+1)} & \text{se } n\theta \text{ não inteiro} \end{cases}$

$[n\theta]$ a parte inteira de $n\theta$

barreira inferior $\quad BI = Q_1 - 1.5(Q_3 - Q_1)$

barreira superior $\quad BS = Q_3 + 1.5(Q_3 - Q_1)$

$$s_x^2 = \frac{\sum_{i=1}^n (x_i - \overline{x})^2}{n-1} = \frac{n\sum_{i=1}^n x_i^2 - \left(\sum_{i=1}^n x_i\right)^2}{n(n-1)}$$

se $x_i' = a + bx_i$ então $\overline{x'} = a + b\overline{x}$, $\quad s_{x'}^2 = b^2 s_x^2$

regra de Sturges (divisão em classes)

o n. de classes próximo de $\left[1 + \frac{\ln n}{\ln 2}\right]$

Dados agrupados em c classes

n_i, x_i' frequência absoluta e ponto médio da classe i

$\tilde{\overline{x}} = \frac{1}{n}\sum_{i=1}^c n_i x_i' = \sum_{i=1}^c f_i x_i'$

$$\tilde{s}_x^2 = \frac{\sum_{i=1}^c (x_i' - \tilde{\overline{x}})^2\ n_i}{n} = \frac{\sum_{i=1}^c x_i'^2\ n_i}{n} - \tilde{\overline{x}}^2$$

Seja k a primeira classe tal que $F_k \geq \theta$

$$Q^*_\theta \simeq x_k^{\min} + \left(x_k^{\max} - x_k^{\min}\right)\frac{\theta - F_{k-1}}{f_k}$$

Indicadores para dados bivariados

$(x_1, y_1), (x_2, y_2), \cdots, (x_n, y_n)$

$$cov(x,y) = \frac{\sum_{i=1}^n (x_i - \overline{x})(y_i - \overline{y})}{n-1} = \frac{n\sum_{i=1}^n x_i y_i - \sum_{i=1}^n x_i \sum_{i=1}^n y_i}{n(n-1)}$$

$|cov(x,y)| \leq s_x s_y$

$r = r_{xy} = \frac{cov(x,y)}{s_x s_y}$ se $s_x \neq 0, s_y \neq 0$

se $x_i' = a + bx_i$ e $y_i' = c + dy_i$

$cov(x', y') = bd\ cov(x,y)$

$r_{x'y'} = \begin{cases} r_{xy} & \text{se } bd > 0 \\ -r_{xy} & \text{se } bd < 0 \end{cases}$

Regressão linear simples

$\widehat{y}_i = a + bx_i \quad \widehat{y}_i - \overline{y} = b(x_i - \overline{x})$

$y_i = a + bx_i + e_i$

$$\begin{cases} b & = \frac{cov(x,y)}{s_x^2} = r\frac{s_y}{s_x}, \qquad s_x \neq 0 \\ a & = \overline{y} - b\overline{x} \end{cases}$$

$\sum_{i=1}^n (y_i - \overline{y})^2 = \sum_{i=1}^n (y_i - \hat{y}_i)^2 + \sum_{i=1}^n (\hat{y}_i - \overline{y})^2$

$\Leftrightarrow SQ_T = SQ_E + SQ_R$

coeficiente de determinação $\quad R^2 = \frac{SQ_R}{SQ_T} = \frac{cov^2(x,y)}{s_x^2\ s_y^2} = r^2$

Estimação

Erro quadrático médio de $\widehat{\Theta}$, estimador de um parâmetro θ

$EQM = \left(E\left[\widehat{\Theta}\right] - \theta\right)^2 + Var(\widehat{\Theta})$

$(X_1, X_2, ..., X_n)$ amostra aleatória retirada de uma população $\mathcal{N}(\mu, \sigma)$

$\overline{X} = \frac{\sum_{i=1}^n X_i}{n}$ e $\frac{\overline{X} - \mu}{\sigma/\sqrt{n}} \cap \mathcal{N}(0,1)$

$S^2 = \frac{\sum_{i=1}^n \left(X_i - \overline{X}\right)^2}{n-1}$ e $\frac{(n-1)S^2}{\sigma^2} \cap \chi^2_{(n-1)}$

$\frac{\overline{X} - \mu}{S/\sqrt{n}} \cap t_{(n-1)}$

$X_1 \cap \mathcal{N}(\mu_1, \sigma_1)$ e $X_2 \cap \mathcal{N}(\mu_2, \sigma_2)$

e tendo duas amostras aleatórias independentes, uma de cada população, com dimensões n_1 e n_2, respectivamente.

$\overline{X_1} - \overline{X_2} \cap \mathcal{N}\left(\mu_1 - \mu_2, \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}\right)$

$\frac{\sigma_2^2\ S_1^2}{\sigma_1^2\ S_2^2} \cap F_{(n_1-1, n_2-1)}$

Testes de hipóteses

H_0 hipótese nula

H_1 hipótese alternativa

P(erro de 1^a espécie)=P(Rejeitar H_0| H_0 verdadeira)= α

P(erro de 2^a espécie)=P(Não rejeitar H_0| H_0 falsa) $=\beta$

regra de decisão

rejeitar H_0 se *p-value* $\leq \alpha$

O Teste de Shapiro Wilk

H_0 : X tem distribuição normal

H_1: X não tem distribuição normal