Formulário de Probabilidades e Estatística

Probabilidades

$$n! = n \times (n-1) \times \cdots \times 2 \times 1, \qquad {}^nA_k = \frac{n!}{(n-k)!}, \qquad {}^nA'_k = n^k, \qquad {}^nC_k = \binom{n}{k} = \frac{n!}{(n-k)!k!}.$$

Probabilidade condicionada: $P(A|B) = \dfrac{P(A \cap B)}{P(B)}$, com $P(B) \neq 0$; A e B *independentes* se $P(A \cap B) = P(A) \times P(B)$.

Sendo A_i, $i = 1, \cdots, n$, uma partição de Ω, tem-se: $P(B) = \sum_{j=1}^{n} P(B|A_j)\, P(A_j)$, $P(A_i|B) = \dfrac{P(B|A_i)\, P(A_i)}{P(B)}$.

Variáveis aleatórias

Discretas: $f(x_i) \geq 0$, $\sum_{i=1}^{n} f(x_i) = 1$, $F(x) = \sum_{x_i \leq x} f(x_i)$, $E(X) = \sum_{i=1}^{n} x_i f(x_i)$, $E[g(X)] = \sum_{i=1}^{n} g(x_i) f(x_i)$.

Contínuas: $f(x) \geq 0$, $\int_{-\infty}^{+\infty} f(x)dx = 1$, $F(x) = \int_{-\infty}^{x} f(t)dt$, $E(X) = \int_{-\infty}^{+\infty} x f(x)dx$, $E[g(X)] = \int_{-\infty}^{+\infty} g(x) f(x)dx$.

$$Var(X) = \sigma_X^2 = E\left[(X - \mu_X)^2\right] = E(X^2) - \mu_X^2$$

Distribuição conjunta de duas variáveis

$Cov(X,Y) = E[(X - \mu_X)(Y - \mu_Y)] = E(XY) - E(X)E(Y)$, $\qquad Corr(X,Y) = \dfrac{Cov(X,Y)}{\sigma_X\, \sigma_Y}$.

$E[g(X,Y)] = \sum_i \sum_j g(x_i, y_j) f(x_i, y_j)$

$E(a + bX + cY) = a + bE(X) + cE(Y)$, $\qquad Var(a + bX + cY) = b^2 Var(X) + c^2 Var(Y) + 2bc\,\mathrm{Cov}(X,Y)$.

Distribuições teóricas

$X \sim B(n,p)$	$f(x) = \binom{n}{x} p^x (1-p)^{n-x}$, para $n \in \mathbb{N}$ e $x = 0, \cdots, n$; $E(X) = np$ e $Var(X) = np(1-p)$
$X \sim BN(r,p)$	$f(x) = \binom{x-1}{r-1} p^r (1-p)^{x-r}$, para $r \in \mathbb{N}$ e $x = r, r+1, \cdots$; $E(X) = \dfrac{r}{p}$ e $Var(X) = \dfrac{r \times (1-p)}{p^2}$
$X \sim H(N,n,r)$	$f(x) = \dfrac{\binom{r}{x}\binom{N-r}{n-x}}{\binom{N}{n}}$, $\max\{0, n-(N-r)\} \leq x \leq \min\{r,n\}$ e $x, N, n, r \in \mathbb{N}$; $E(X) = n \times \dfrac{r}{N}$ e $Var(X) = n \times \dfrac{r}{N} \times \dfrac{N-r}{N} \times \dfrac{N-n}{N-1}$
$X \sim P(\lambda)$	$f(x) = \dfrac{e^{-\lambda}\lambda^x}{x!}$, com $\lambda > 0$ e $x = 0, 1, \cdots$; $E(X) = \lambda$ e $Var(X) = \lambda$
$X \sim U(a,b)$	$f(x) = \begin{cases} \dfrac{1}{b-a} & a \leq x \leq b \\ 0 & x < a \ \vee \ x > b \end{cases}$, com $a < b$; $E(X) = \dfrac{a+b}{2}$ e $Var(X) = \dfrac{(b-a)^2}{12}$
$X \sim Exp(\lambda)$	$f(x) = \begin{cases} \lambda e^{-\lambda x} & x > 0 \\ 0 & x \leq 0 \end{cases}$, com $\lambda > 0$; $E(X) = \dfrac{1}{\lambda}$ e $Var(X) = \dfrac{1}{\lambda^2}$
$X \sim N(\mu,\sigma)$	para $\sigma > 0$ e $\mu, x \in \mathbb{R}$; $E(X) = \mu$, $Var(X) = \sigma^2$ e $Z = \dfrac{X-\mu}{\sigma} \sim N(0,1)$
$X \sim \chi^2_{(n)}$	para $x > 0$ e $n \in \mathbb{N}$; $E(X) = n$ e $Var(X) = 2n$
$X \sim t_{(n)}$	para $x \in \mathbb{R}$ e $n \in \mathbb{N}$; $E(X) = 0$ e $Var(X) = \dfrac{n}{n-2}$ (para $n \geq 3$)

Convergência assimptótica

$X \sim H(N, n, r)$	$\Rightarrow$	$X \overset{\bullet}{\sim} B(n, p)$ se $\dfrac{n}{N} \leq 0.05$, sendo $p = \dfrac{r}{N}$
$X \sim B(n, p)$	$\Rightarrow$	$X \overset{\bullet}{\sim} P(\lambda)$ se $n \geq 20$ e $p \leq 0.05$, sendo $\lambda = np$
$X \sim B(n, p)$	$\Rightarrow$	$X' \overset{\bullet}{\sim} N(\mu, \sigma)$ se $n \times p \geq 15$ e $n \times (1-p) \geq 15$, sendo $\mu = np$ e $\sigma^2 = n \times p \times (1-p)$
$X \sim P(\lambda)$	$\Rightarrow$	$X' \overset{\bullet}{\sim} N(\mu, \sigma)$ se $\lambda > 20$, sendo $\mu = \lambda$ e $\sigma^2 = \lambda$
$X \sim \chi^2_{(n)}$	$\Rightarrow$	$\left(\sqrt{2X} - \sqrt{2n}\right) \overset{\bullet}{\sim} N(0, 1)$ se $n \geq 30$
$X \sim \chi^2_{(n)}$	$\Rightarrow$	$X \overset{\bullet}{\sim} N(\mu, \sigma)$ se $n \geq 100$, sendo $\mu = n$ e $\sigma^2 = 2n$
$X \sim t_{(n)}$	$\Rightarrow$	$X \overset{\bullet}{\sim} N(0, 1)$ se $n \geq 30$
X_i *i. i. d.*	$\Rightarrow$	$\sum_{i=1}^{n} X_i \overset{\bullet}{\sim} N(n\mu, \sqrt{n}\,\sigma)$ se $n \geq 30$, sendo $\mu = E(X_i)$ e $\sigma^2 = Var(X_i)$

Correcção de continuidade $P(X = k) \simeq P(k - 0.5 < X' \leq k + 0.5)$

$X_i \sim P(\lambda_i)$	$\Rightarrow$	$\sum_{i=1}^{n} X_i \sim P\left(\sum_{i=1}^{n} \lambda_i\right)$ se X_i independentes para $i = 1, \cdots, n$
$X_i \sim N(\mu_i, \sigma_i)$	$\Rightarrow$	$a + \sum_{i=1}^{n} b_i X_i \sim N\left(a + \sum_{i=1}^{n} b_i \mu_i, \sqrt{\sum_{i=1}^{n} b_i^2 \sigma_i^2}\right)$ se X_i independentes para $i = 1, \cdots, n$
$X_i \sim \chi^2_{(n_i)}$	$\Rightarrow$	$\sum_{i=1}^{n} X_i \sim \chi^2_{\left(\sum_{i=1}^{n} n_i\right)}$ se X_i independentes para $i = 1, \cdots, n$
$Z_i \sim N(0, 1)$	$\Rightarrow$	$\sum_{i=1}^{n} Z_i^2 \sim \chi^2_{(n)}$ se Z_i independentes para $i = 1, \cdots, n$
$Z \sim N(0, 1)$ e $X \sim \chi^2_{(n)}$	$\Rightarrow$	$\dfrac{Z}{\sqrt{\dfrac{X}{n}}} \sim t_{(n)}$ se Z e X independentes

Desigualdade de Tchebycheff: $P\left(\lvert X - \mu \rvert < k\sigma\right) \geq 1 - \dfrac{1}{k^2}$ ou $P\left(\lvert X - \mu \rvert \geq k\sigma\right) \leq \dfrac{1}{k^2}, \ \forall k > 0$

Inferência Estatística

$\overline{X} = \dfrac{1}{n} \sum_{i=1}^{n} X_i$	$S^2 = \dfrac{1}{n} \sum_{i=1}^{n} \left(X_i - \overline{X}\right)^2 = \dfrac{1}{n} \sum_{i=1}^{n} X_i^2 - \overline{X}^2$	$S_C^2 = \dfrac{n}{n-1} S^2$

Estimador centrado: $E(\hat{\theta}) = \theta$; $\widehat{\theta_1}$ é *mais eficiente* que $\widehat{\theta_2}$ se $Var\left(\widehat{\theta_1}\right) < Var\left(\widehat{\theta_2}\right)$, com $E(\hat{\theta}_1) = E(\hat{\theta}_2) = \theta$

Condições suficientes para que $\hat{\theta}$ seja um *estimador consistente*: $\lim_{n \to +\infty} E\left(\hat{\theta}\right) = \theta$ e $\lim_{n \to +\infty} Var\left(\hat{\theta}\right) = 0$

$\overline{X}$ - com σ conhecido e população Normal: $\overline{X} \sim N\left(\mu, \dfrac{\sigma}{\sqrt{n}}\right)$; $\overline{X}$ - com σ desconhecido e população Normal: $\dfrac{\overline{X} - \mu}{\dfrac{S_C}{\sqrt{n}}} \sim t_{(n-1)}$	$\overline{X}$ - com σ conhecido e $n \geq 30$: $\overline{X} \overset{\bullet}{\sim} N\left(\mu, \dfrac{\sigma}{\sqrt{n}}\right)$
$\widehat{p}$ - População Bernoulli e $n \geq 30$: $\widehat{p} \overset{\bullet}{\sim} N\left(p, \sqrt{\dfrac{p(1-p)}{n}}\right)$	S_C^2 - População Normal: $\dfrac{(n-1) S_C^2}{\sigma^2} \sim \chi^2_{(n-1)}$

$\alpha = P(\text{rejeitar } H_0 \mid H_0 \text{ verdadeira})$, $\beta = P(\text{não rejeitar } H_0 \mid H_0 \text{ falsa})$, *Potência do teste*: $1 - \beta$

p-value		$H_1 : \theta > \theta_0$	$H_1 : \theta < \theta_0$	$H_1 : \theta \neq \theta_0$
	Normal	$P(Z > z_{obs.})$	$P(Z < z_{obs.})$	$2P(Z > \lvert z_{obs.} \rvert)$
	t-Student	$P(T > t_{obs.})$	$P(T < t_{obs.})$	$2P(T > \lvert t_{obs.} \rvert)$
	Qui-Quadrado	$P\left(\chi^2 > \chi^2_{obs.}\right)$	$P\left(\chi^2 < \chi^2_{obs.}\right)$	$2 \min\left\{P\left(\chi^2 > \chi^2_{obs.}\right), P\left(\chi^2 < \chi^2_{obs.}\right)\right\}$