Notas de Probabilidades e Estatística

Giovani Loiola da Silva
Carlos Daniel Paulino
Departamento de Matemática - IST/UTL

Setembro 2012

Conteúdo

1 Introdução

Breve descrição do objecto da disciplina

Classificação de experiências segundo a (im)previsibilidade exacta dos seus resultados:

Experiências determinísticas ou causais - Exemplos:

1. Transformação de água pura em vapor quando aquecida a temperatura superior a 100°C sob pressão atmosférica de 760 mm Hg;
2. Distância percorrida por um móvel ao fim de um determinado tempo quando lançado verticalmente a uma dada velocidade;
3. Intensidade da corrente eléctrica que percorre um circuito com uma determinada resistência intercalada quando submetido a uma dada diferença de potencial nas condições de aplicabilidade da lei de Ohm.

Experiências aleatórias ou casuais - Exemplos:

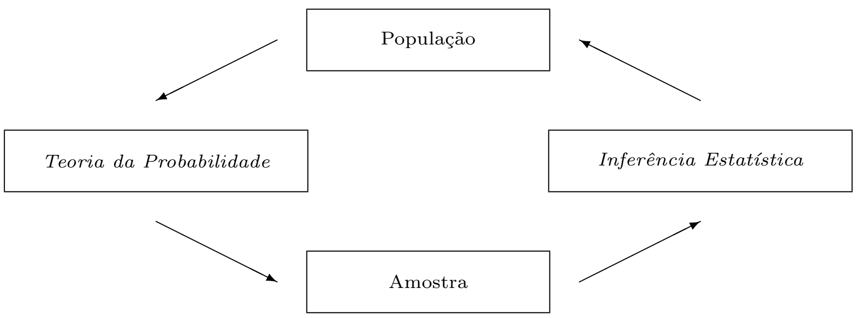
1. Desintegração radioactiva;

2. Repartição de defeitos estruturais em chapas metálicas usada na indústria;
3. Sexo do ser vivo resultante de um óvulo fecundado;
4. Extracção de prémios de uma lotaria;
5. Lançamento de um dado.

Embora esta dicotomia levante vários problemas, o que se pretende destacar aqui é se a descrição satisfatória da experiência/fenómeno em estudo requer uma análise de causa-efeito (as primeiras) ou carece de uma análise probabilístico-estatística (as segundas).

- *Teoria da Probabilidade*: estudo de modelos matemáticos adequados para a descrição das experiências aleatórias (modelos probabilísticos, estocásticos ou estatísticos).
- *Estatística*: Estudo de métodos para a selecção de modelos estocásticos apropriados para a descrição das experiências aleatórias com base em dados observados.
- *Estatística Descritiva/Análise Exploratória de Dados*: Descrição de operações numéricas e gráficas que visam patentear sumariamente a informação relevante contida nos dados obtidos.
- *Estatística Indutiva/Inferência Estatística*: Estudo de como inferir (características de) um modelo estocástico adequado para a descrição da experiência/fenómeno a partir dos dados, com medição probabilística do grau de incerteza associado.

Representação esquemática da diferenciação entre os raciocínios da Teoria da Probabilidade e da Inferência Estatística



2 Noções de probabilidade

Motivação: Num estudo científico, o objectivo centra-se usualmente na descrição de um fenómeno de interesse através de modelo teórico.

- O fenómeno pode ser observável e o processo de recolha das suas observações é uma experiência.
- Se a realização da experiência determina previamente qual o seu resultado, o modelo teórico é dito determinístico. Caso contrário, o modelo é não determinístico ou aleatório (estocástico).

Exemplo 2.1: Sob certas condições, a distância (S) percorrida em queda livre por um objecto ao fim de um tempo t é $S_t = -16t^2 + v_0 t$, onde v_0 é a velocidade inicial imprimida ao objecto.

Exemplo 2.2: O número de partículas alfa emitidas por um fragmento de material radioactivo durante um dado intervalo de tempo.

Experiências aleatórias. Espaço de resultados

Definição 2.1: Uma experiência diz-se *aleatória* se:

- todos os seus possíveis resultados são conhecidos à partida.
- o resultado em cada realização concreta da experiência não é de facto conhecido a priori.

Frequentemente, acrescenta-se ainda à definição de experiência aleatória que ela pode ser repetida muitas vezes, essencialmente sob as mesmas condições.

Exemplo 2.3: Lançamento de um dado (experiência E_1).

Definição 2.2: Espaço de resultados ou *espaço amostral* de uma experiência aleatória é o conjunto de todos os seus possíveis resultados, denotado por Ω.

Acontecimentos

Definição 2.3: Dada uma experiência aleatória E com espaço de resultados Ω, um *acontecimento* ou evento é um subconjunto de Ω.

Um acontecimento pode ser, por exemplo, elementar ($\{\omega\}$), certo (Ω) e impossível ($\emptyset$). Note-se que dois acontecimentos A e B tais que $A \cap B = \emptyset$ são ditos mutuamente exclusivos ou disjuntos.

Definição 2.4: Dada uma experiência aleatória E com espaço de resultados Ω, denota-se por $\mathcal{A}$ é o conjunto de todos os acontecimentos possíveis de Ω.

Exemplo 2.3a: Na experiência E_1, $\Omega = \{1, 2, 3, 4, 5, 6\}$ e $\mathcal{A} = \{\emptyset, \{1\}, \ldots, \Omega\}$ com $\#\mathcal{A} = 64$.

Noção de probabilidade

Interpretação de Laplace

Para uma experiência aleatória E com espaço de resultados finito $\Omega = \{1, \ldots, N\}$, supondo que os N resultados são igualmente prováveis, a probabilidade de qualquer acontecimento A é a proporção de resultados de Ω favoráveis a A.

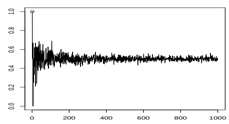
Exemplo 2.3b: Na experiência E_1, a probabilidade do acontecimento $A = \{\text{sair face par}\}$ é dada por $P(A) = 3/6 = 0.5$.

Interpretação frequencista

A probabilidade de um acontecimento A é o limite da frequência relativa da ocorrência de A numa longa sucessão de experiências realizadas sob as mesmas condições.

Exemplo 2.4: Num lançamento de uma moeda (E_2), a probabilidade de sair cara (acontecimento A) é dada por

$$P(A) = \lim_{n \to \infty} f_n(A) = \frac{1}{2}.$$



Interpretação subjectivista

A probabilidade de um acontecimento A é entendida como uma medida pessoal (entre 0 e 1) do grau de crença sobre a ocorrência de A.

Exemplo 2.5: Um geólogo afirma que uma dada região tem 60% de chance de haver petróleo, baseando-se quer nas características do terreno quer na sua semelhança com outras regiões com conhecida presença ou ausência de petróleo nos últimos anos.

Axiomática de probabilidade

Definição 2.5: Para cada evento A de uma experiência aleatória com espaço de resultados Ω, é suposto existir um número real, designado por *probabilidade de* A e denotado por $P(A)$, satisfazendo os axiomas:

A_1: $P(A) \geq 0$, $\forall A \in \mathcal{A}$ (acontecimentos possíveis de Ω).

A_2: $P(\Omega) = 1$.

A_3: Para qualquer sequência de acontecimentos disjuntos 2 a 2 $A_1, \ldots, A_n$ tem-se $P(\cup_{i=1}^n A_i) = \sum_{i=1}^n P(A_i)$, $n = 2, 3, \ldots$.

O conjunto $\mathcal{A}$ de Ω é dito ser uma σ-álgebra de acontecimentos se i) $\Omega \in \mathcal{A}$, ii) $A \in \mathcal{A} \Rightarrow \bar{A} \in \mathcal{A}$ e iii) $A_1, A_2, \ldots \in \mathcal{A} \Rightarrow \cup_{i=1}^{\infty} A_i \in \mathcal{A}$.

Definição 2.6: Sob a validade destes axiomas, $(\Omega, \mathcal{A}, P)$ é dito ser um *espaço de probabilidade*. O par $(\Omega, \mathcal{A})$ diz-se ser um espaço mensurável de acontecimentos.

Teoremas decorrentes

Sejam A e B acontecimentos de uma experiência aleatória com espaço de resultados Ω. Se $P(A)$ e $P(B)$ satisfazem os axiomas referidos anteriormente, então tem-se os seguintes teoremas decorrentes:

Teorema 2.1: $P(\bar{A}) = 1 - P(A)$;

Teorema 2.2: $P(\emptyset) = 0$;

Teorema 2.3: $A \subset B \Rightarrow P(A) \leq P(B)$ e $P(B - A) = P(B) - P(A)$;

Teorema 2.4: $P(A) \leq 1$;

Teorema 2.5: $P(B - A) = P(B) - P(A \cap B)$;

Teorema 2.6: $P(A \cup B) = P(A) + P(B) - P(A \cap B)$. Generalização: $P(\cup_{i=1}^n A_i) = \sum_{i=1}^n P(A_i) - \sum_{i=2}^n \sum_{j<i} P(A_i \cap A_j) + \cdots + (-1)^{n-1} P(\cap_{i=1}^n A_i)$.

Exemplo 2.6: Na experiência E_1, com $A_1 = \{\text{face par}\}$ e $A_2 = \{\text{face} \geq 4\}$, tem-se i) $P(A_1) = P(A_2) = \frac{1}{2}$; ii) $P(A_1 \cap A_2) = \frac{1}{3}$; iii) $P(A_1 \cup A_2) = \frac{2}{3}$.

Probabilidade condicional

Definição 2.7: Sejam A e B acontecimentos de um espaço de probabilidade $(\Omega, \mathcal{A}, P)$. Se $P(B) > 0$, a probabilidade do acontecimento A dado a ocorrência do acontecimento B (A dado B ou A se B ou A condicionado a B) é dada por

$$P(A|B) = \frac{P(A \cap B)}{P(B)}.$$

Analogamente, $P(B|A) = P(A \cap B)/P(A)$, se $P(A) > 0$.

- Em $P(A|B)$, B funciona como um espaço de resultados reduzido sobre o qual está avaliada a probabilidade de A.
- Se Ω é finito com resultados equiprováveis, pode-se calcular $P(A|B)$ directamente como $P(A|B) = \frac{\#\{A \cap B\}}{\#\{B\}}$.

No cenário vigente, a probabilidade condicional $P(A|B)$, com $P(B) > 0$, é uma probabilidade definida sobre o espaço de acontecimentos associado a B, verificando-se os axiomas:

A_1: $P(A|B) \geq 0$, $\forall$ acontecimento A.

A_2: $P(\Omega|B) = 1$.

A_3: Para acontecimentos disjuntos $A_1, \ldots, A_n$, $P(\cup_{i=1}^n A_i|B) = \sum_{i=1}^n P(A_i|B)$, $n = 1, 2, \ldots$.

E igualmente teoremas decorrentes tais como:

1. $P(\bar{A}|B) = 1 - P(A|B)$;
2. $P(\emptyset|B) = 0$;
3. $P(A|B) \leq 1$;
4. $A_1 \subset A_2 \Rightarrow P(A_1|B) \leq P(A_2|B), P(A_2 - A_1|B) = P(A_2|B) - P(A_1|B)$;
5. $P(A_2 - A_1|B) = P(A_2|B) - P(A_2 \cap A_1|B)$;
6. $P(A_1 \cup A_2|B) = P(A_1|B) + P(A_2|B) - P(A_1 \cap A_2|B)$.

Teorema da probabilidade composta

A partir da definição de probabilidade condicional obtém-se que

$$P(A \cap B) = P(A)P(B|A) \text{ ou } P(B)P(A|B),$$

bem como relações estendidas do tipo

$$P(A \cap B \cap C) = P(A)P(B \cap C|A) = P(A)P(B|A)P(C|A \cap B).$$

Teorema 2.7: Se $A_1, \ldots, A_n$ são acontecimentos de um espaço de resultados Ω, então

$$P(\cap_{i=1}^n A_i) = P(A_1)P(A_2|A_1)\ldots P(A_n|A_1 \cap A_2 \cap \ldots \cap A_{n-1}).$$

Exemplo 2.7: Num sorteio de 3 prémios, sem reposição, para 12 homens e 8 mulheres, a probabilidade de nenhum homem ganhar o sorteio (A) é $P(A) \simeq 0.049$.

Teorema da probabilidade total

Definição 2.8: Os acontecimentos $A_1, \ldots, A_n$ formam uma *partição* do espaço de resultados Ω quando

1. $A_i \cap A_j = \emptyset$, $\forall i \neq j = 1, \ldots, n$.
2. $\cup_{i=1}^n A_i = \Omega$.

Teorema 2.8: Se B é um acontecimento qualquer de um espaço de resultados Ω e $A_1, \ldots, A_n$ uma partição de Ω, então

$$P(B) = \sum_{i=1}^n P(A_i)P(B|A_i).$$

Exemplo 2.8: Numa caixa com 20 peças do tipo A e 80 do tipo B, sabe-se que 30% e 25% das peças do tipo A e B, respec., são defeituosas. Qual a probabilidade de uma peça, seleccionada ao acaso, ser defeituosa (D)? $P(D) = 0.26$.

Teorema de Bayes

Teorema 2.9: Se os acontecimentos $A_1, \ldots, A_n$ formam uma partição do espaço de resultados Ω e B é um acontecimento qualquer de Ω com $P(B) > 0$, então $\forall\, i = 1, \ldots, n$,

$$P(A_i|B) = \frac{P(A_i \cap B)}{P(B)} = \frac{P(A_i)P(B|A_i)}{\sum_{j=1}^{n} P(A_j)P(B|A_j)}.$$

Exemplo 2.8a: Na caixa com 100 peças dos tipos A e B (Exemplo 2.8), qual a probabilidade de uma peça seleccionada ao acaso ser do tipo A, sabendo que ela é defeituosa? E ser do tipo B, se defeituosa?

$$P(A|D) \simeq 0.231$$
$$P(B|D) \simeq 0.769$$

Acontecimentos independentes

Definição 2.9: Diz-se que dois acontecimentos A e B de um mesmo espaço de resultados Ω são *independentes* se

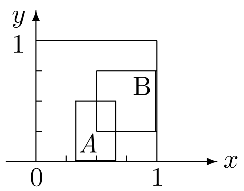
$$P(A \cap B) = P(A) \times P(B).$$

- Todo o acontecimento A é independente de $\emptyset$ e Ω.
- Se A e B são acontecimentos independentes, $P(A|B) = P(A)$ se $P(B) > 0$ e $P(B|A) = P(B)$ se $P(A) > 0$.
- Se A e B são acontecimentos independentes, também o são $\bar{A}$ e B, A e $\bar{B}$ e $\bar{A}$ e $\bar{B}$.
- Acontecimentos A e B são condicionalmente independentes ao acontecimento C, $P(C) > 0$, se $P(A \cap B|C) = P(A|C)P(B|C)$.
- Os acontecimentos A, B e C são completamente independentes, se $P(A \cap B) = P(A)P(B)$, $P(A \cap C) = P(A)P(C)$, $P(B \cap C) = P(B)P(C)$ e $P(A \cap B \cap C) = P(A)P(B)P(C)$.

  Nota: Independência 2 a 2 $\not\Leftrightarrow$ independência completa dos 3.
- Generalização: Os acontecimentos $A_1, \ldots, A_n$ dizem-se independentes se para todo o $k = 2, \ldots, n$ e todo o subconjunto $\{A_{i_j}, j = 1, \ldots, k\}$ de k desses acontecimentos, $P(\cap_{j=1}^{k} A_{i_j}) = \prod_{j=1}^{k} P(A_{i_j})$.

  Nota: O número de relações é dado por $2^n - (n+1)$.

Exemplo 2.9: Considere o espaço de resultados Ω como o quadrado de vértices (0,0), (0,1), (1,0) e (1,1). Suponha que a probabilidade de uma região (acontecimento) contida em Ω seja a área desta região. Os acontecimentos $A = \{(x, y) : 1/3 \leq x \leq 2/3, 0 \leq y \leq 1/2\}$ e $B = \{(x, y) : 1/2 \leq x \leq 1, 1/4 \leq y \leq 3/4\}$ são independentes?



$$P(A) = 1/6, \;\; P(B) = 1/4$$
$$P(A \cap B) = 1/24 = P(A) \times P(B)$$
$\therefore$ A e B são independentes.

3 Variáveis aleatórias

Numa experiência aleatória, independentemente de o seu espaço de resultados ser expresso numericamente, há interesse em considerar-se funções reais em Ω, denominadas por variáveis aleatórias.

Definição 3.1: Uma *variável aleatória* (v.a.) X é uma função que associa um número real a cada resultado possível de uma experiência aleatória.

Rigorosamente, dado um espaço de probabilidade $(\Omega, \mathcal{A}, P)$, uma variável aleatória X é uma função com domínio Ω e contradomínio na recta real ($X : \Omega \to I\!R$) tal que o conjunto $A_r \equiv \{w \in \Omega : X(\omega) \leq r\} \in \mathcal{A}$, $\forall\, r \in I\!R$.

As variáveis aleatórias podem assumir um número finito ou infinito (numerável ou não numerável) de valores possíveis.

O modelo probabilístico induzido em $I\!R$ pela v.a. X pode ser cabalmente definido de vários modos, *e.g.*, através da função de distribuição.

Função de distribuição

Definição 3.2: Dada uma variável aleatória X, a *função de distribuição* (cumulativa) de X é dada por

$$F_X(x) \equiv P(X \leq x), \forall\, x \in I\!R.$$

Por exemplo, $P(a < X \leq b) = F_X(b) - F_X(a),\ \forall\, a < b$.

Propriedades da função de distribuição.

A função de distribuição de X, $F_X(x)$, satisfaz as seguintes propriedades:

P_1: Se $x \leq y$, então $F_X(x) \leq F_X(y)$. Ou seja, F_X é uma função não decrescente.

P_2: Se $x_n \downarrow x$ $(n \to \infty)$, então $F_X(x_n) \downarrow F_X(x)$. Ou seja, F_X é uma função contínua à direita.

P_3: Se $x_n \downarrow -\infty$ $(n \to \infty)$, então $F_X(x_n) \downarrow 0 = F_X(-\infty)$.

P_4: Se $x_n \uparrow \infty$ $(n \to \infty)$, então $F_X(x_n) \uparrow 1 = F_X(\infty)$.

Variáveis aleatórias discretas

Se o conjunto dos possíveis valores de uma variável aleatória for finito ou infinito enumerável, a v.a. diz-se discreta. Nesse caso, outro modo de definir cabalmente o modelo probabilístico induzido em $I\!R$ é através da função massa de probabilidade.

Função (massa) de probabilidade

Definição 3.3: Diz-se que X é uma v.a. discreta, com os possíveis valores $x_1, x_2, \ldots$, se existir uma função ($I\!R \to [0,1]$) $f_X(x) = P(X = x)$, denotando a probabilidade de ocorrência de $\{x\}$, conhecida por função (massa) de probabilidade (*f.m.p.*), e satisfazendo as condições:

1. $f_X(x_i) > 0$, $\forall\, i = 1, 2, \ldots$;
2. $\sum_{i \geq 1} f_X(x_i) = 1$.

Observe-se que

1. $P(X \in B) = \sum\limits_{x_i \in B} f_X(x_i)$;

2. $F_X(x) = \sum\limits_{x_i \leq x} f_X(x_i)$;

3. $f_X(x) = F_X(x) - F_X(x^-)$, onde $F_X(x^-) \equiv P(X < x)$;

4. $P(a \leq X \leq b) = F_X(b) - F_X(a) + f_X(a)$;

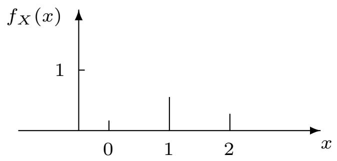
5. $P(a < X < b) = F_X(b) - F_X(a) - f_X(b)$:

6. $P(a \leq X < b) = F_X(b) - F_X(a) - f_X(b) + f_X(a)$;

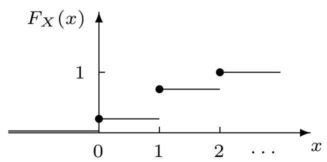
7. $P(a < X \leq b) = F_X(b) - F_X(a)$.

Exemplo 3.1: Na extracção, sem reposição, de 2 peças de uma urna com 5 peças defeituosas e 4 perfeitas, qual a f.m.p. de X (número de peças defeituosas nas 2 peças retiradas)? E a sua função de distribuição?

$$f_X(x) = \begin{cases} \frac{12}{72}, & x = 0; \\ \frac{40}{72}, & x = 1; \\ \frac{20}{72}, & x = 2; \\ 0, & \text{caso contrário} \end{cases}$$



$$F_X(x) = \begin{cases} 0, & x < 0; \\ \frac{12}{72}, & 0 \leq x < 1; \\ \frac{52}{72}, & 1 \leq x < 2; \\ 1, & x \geq 2. \end{cases}$$



Variáveis aleatórias contínuas

Se o conjunto dos possíveis valores de uma v.a. for infinito não numerável, a v.a. diz-se contínua se satisfizer certas condições adicionais.

Função densidade de probabilidade

Definição 3.4: Diz-se que X é uma v.a. contínua, se existir uma função f_X, denominada função densidade de probabilidade (*f.d.p.*) de X tal que:

1. $f_X(x) \geq 0, \forall\, x \in \mathbb{R}$;

2. $\int_{\mathbb{R}} f_X(x)dx = 1$;

3. A função de distribuição é contínua e dada por

$$F_X(x) \equiv P(X \leq x) = \int_{-\infty}^{x} f_X(u)du.$$

Definição 3.5: Dada uma v.a. contínua X com f.d.p. $f_X(x)$, a massa probabilística contida em qualquer acontecimento $B \subset I\!R$ é dada por

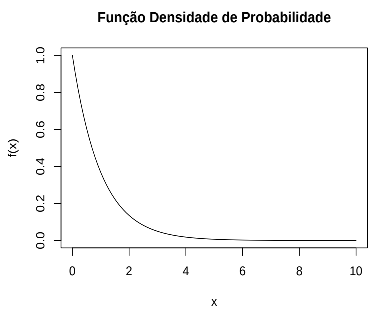
$$P(X \in B) = \int_B f_X(x)dx.$$

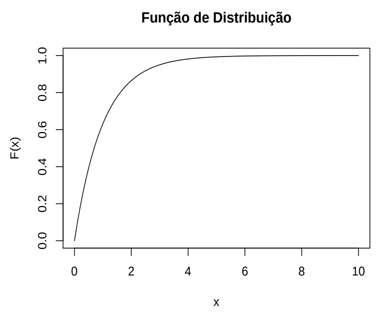
Se X é uma v.a. contínua com função de distribuição $F_X(x)$,

- $P(a < X \leq b) = F_X(b) - F_X(a)$: área sob $f_X(x)$ entre a e b.
- $\forall x \in I\!R,\ P(X = x) \leq F_X(x) - F_X(x-h),\ \forall h > 0$
  $\Rightarrow P(X = x) \leq \lim_{h \to 0^+} [F_X(x) - F_X(x-h)] = 0$
  $\Rightarrow P(a < X \leq b) = P(a \leq X \leq b) = P(a < X < b) = P(a \leq X < b)$.
- A f.d.p. de X pode ser obtida pela derivação de $F_X(x)$, *i.e.*, $f_X(x) = \frac{d}{dx}F_X(x)$, nos pontos de diferenciabilidade desta.
- $f_X(x)$ pode ser interpretada como uma massa de probabilidade por unidade de comprimento pois para Δx suficientemente pequeno $P(x - \frac{\Delta x}{2} \leq X \leq x + \frac{\Delta x}{2}) \approx \Delta x f_X(x)$, à luz do Teorema do Valor Médio do Cálculo Integral.

Exemplo 3.2: Seja X o tempo de vida de uma componente electrónica (em determinadas unidades), suposto distribuído com f.d.p. $f_X(x) = \mathrm{e}^{-x}$, se $x > 0$, e $f_X(x) = 0$, caso contrário. Qual a função de distribuição de X?

$$F_X(x) = \begin{cases} \int_0^x \mathrm{e}^{-u}du = 1 - \mathrm{e}^{-x}, & x > 0; \\ 0, & x \leq 0. \end{cases}$$





Funções de variáveis aleatórias

Seja X uma variável aleatória definida no espaço de resultados Ω associado à experiência aleatória E. Se $y = g(x)$ é uma função real (mensurável) de x, com $x = X(\omega)$ para algum $\omega \in \Omega$, então $Y = g(X)$ é também uma variável aleatória definida no mesmo espaço de probabilidade.

Por exemplo, se X é uma v.a. discreta com f.m.p. $f_X(x)$ e contradomínio $D = \{x_1, x_2, \ldots\}$, então $Y = g(X)$ é também uma variável aleatória discreta com f.m.p.

$$f_Y(y) = P(Y = y) = P(X \in A_y) = \sum_{x_i \in A_y} f_X(x_i),\ y \in D^*$$

onde $A_y = \{x \in D : g(x) = y\}$ e $D^* = g(D)$ é o contradomínio de Y.

Se X é uma v.a. contínua, a continuidade de $Y = g(X)$ depende do tipo da função $g(\cdot)$. No caso de Y ser contínua, a sua distribuição é determinável da de X, como se ilustra em seguida.

Exemplo 3.3: Seja X uma v.a. contínua com f.d.p.

$$f_X(x) = \begin{cases} 1, & 0 < x < 1; \\ 0, & \text{caso contrário.} \end{cases}$$

Qual a f.d.p. de $Y = \mathrm{e}^X$ (exemplo de função real de X diferenciável em todos os pontos do respectivo domínio e estritamente monótona)?

$$\forall y > 0, F_Y(y) \equiv P(Y \leq y) = P(X \leq \log y) \equiv F_X(\log y).$$

$$\Longrightarrow f_Y(y) = \frac{d}{dy} F_Y(y) = \frac{d}{dx} F_X(x)\bigg|_{x=\log y} \frac{dx}{dy} = f_X(\log y)\frac{1}{y}$$

$$\therefore \quad f_Y(y) = \begin{cases} 1/y, & 1 < y < \mathrm{e}; \\ 0, & \text{caso contrário.} \end{cases}$$

Vectores aleatórios bidimensionais

Na maioria das situações, a consideração de uma única variável não é suficiente para explicar cabalmente um fenómeno aleatório, sendo necessário explicitar mais do que uma v.a. e, por conseguinte, definir a distribuição de probabilidade conjunta no espaço euclidiano multidimensional.

Exemplo 3.4: Sejam X e Y os números de cartas rei/dama e ás em 2 cartas retiradas (sem reposição) de um baralho com 52 cartas, respec. Quais as probabilidades conjuntas (não nulas) do par aleatório (X, Y)?

$X\backslash Y$	0	1	2	
0	0.589	0.121	0.004	0.714
1	0.241	0.024	0	0.265
2	0.021	0	0	0.021
	0.851	0.145	0.004	1

Função de distribuição conjunta

Um par aleatório (X, Y) é uma função (mensurável) $\Omega \to I\!R^2$.

Definição 3.6: Dado um par aleatório (X, Y), a sua função de distribuição *conjunta* é dada por

$$F_{X,Y}(x, y) \equiv P(X \leq x, Y \leq y).$$

Propriedades da função de distribuição de um par aleatório (X, Y):

P_1: $F_{X,Y}(x, y)$ é uma função não decrescente em cada uma das variáveis, *e.g.*, $\forall\, x, y_1 \leq y_2$, $F_{X,Y}(x, y_1) \leq F_{X,Y}(x, y_2)$.

P_2: $F_{X,Y}(x,y)$ é uma função contínua à direita em cada uma das variáveis, *e.g.*, se $x_n \downarrow x$ $(n \to \infty)$, então $F_{X,Y}(x_n,y) \downarrow F_{X,Y}(x,y)$.

P_3: $\lim\limits_{x,y\to-\infty} F_{X,Y}(x,y) = \lim\limits_{x\to-\infty} F_{X,Y}(x,y) = \lim\limits_{y\to-\infty} F_{X,Y}(x,y) = 0$.

P_4: $\lim\limits_{x,y\to\infty} F_{X,Y}(x,y) = 1$.

Funções de distribuição marginais

Definição 3.7: Dado um par aleatório (X,Y), a função de distribuição *marginal* de X é dada por

$$F_X(x) = \lim_{y\to\infty} F_{X,Y}(x,y) = P(X \le x),$$

enquanto a função de distribuição marginal de Y é

$$F_Y(y) = \lim_{x\to\infty} F_{X,Y}(x,y) = P(Y \le y).$$

Note-se que as funções de distribuição marginais $F_X(x)$ e $F_Y(y)$ de um par aleatório (X,Y) satisfazem as propriedades da função de distribuição (unidimensional) referidas previamente.

Distribuição conjunta de um par aleatório

Definição 3.8: Diz-se que (X,Y) é um par aleatório discreto (contínuo), quando existe uma função $f_{X,Y}(x,y)$, denominada função massa (densidade) de probabilidade *conjunta* de (X,Y), satisfazendo as seguintes condições:

1. $f_{X,Y}(x,y) \ge 0$, $\forall\,(x,y)$.
2. $\sum_{x_i}\sum_{y_j} f_{X,Y}(x_i,y_j) = 1$ (caso discreto),
   $\int_{I\!R}\int_{I\!R} f_{X,Y}(x,y)\,dxdy = 1$ (caso contínuo).
3. $P((X,Y)\in B) = \begin{cases} \sum\sum_{(x_i,y_j)\in B} f_{X,Y}(x_i,y_j) & \text{(caso discreto),} \\ \int\int_B f_{X,Y}(x,y)\,dxdy & \text{(caso contínuo).} \end{cases}$

Por conseguinte, a função de distribuição conjunta de (X,Y) é

$$F_{X,Y}(x,y) = \begin{cases} \sum_{x_i\le x}\sum_{y_j\le y} f_{X,Y}(x_i,y_j) & \text{(caso discreto),} \\ \int_{-\infty}^{x}\int_{-\infty}^{y} f_{X,Y}(u,v)\,dvdu & \text{(caso contínuo).} \end{cases}$$

Observe-se que a f.d.p. conjunta de (X,Y) (que representa a massa probabilística por unidade de área) pode ser obtida a partir da respectiva função de distribuição por diferenciação, nos pontos (x,y) de diferenciabilidade desta, *i.e.*,

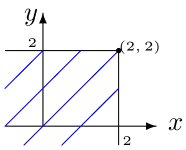
$$f_{X,Y}(x,y) = \frac{\partial^2}{\partial x \partial y} F_{X,Y}(x,y).$$

Exemplo 3.5: Num sistema com 2 componentes electrónicas, seja X (Y) a duração (em horas) da sua primeira (segunda) componente. Será $f_{X,Y}(x,y)$ abaixo uma f.d.p. conjunta do par aleatório (X,Y)?

$$f_{X,Y}(x,y) = \begin{cases} e^{-x-y}, & x > 0, y > 0; \\ 0, & \text{c.c.} \end{cases}$$

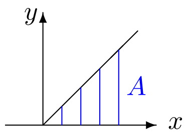
- Sim, $f_{X,Y}(x,y) \geq 0$, $\forall\,(x,y) \in I\!R^2$ e
- $\displaystyle\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{X,Y}(x,y)\,dxdy = \int_0^{\infty} e^{-y} \int_0^{\infty} e^{-x} dxdy = 1.$

Qual a probabilidade de as duas componentes durarem no máximo 2 horas cada uma?



$$\begin{aligned} P(X \leq 2, Y \leq 2) &= F_{X,Y}(2,2) \\ &= \textstyle\int_0^2 \int_0^2 e^{-x-y} dxdy \\ &\simeq 0.7477 \end{aligned}$$

Qual a probabilidade de a primeira componente durar mais do que a segunda?



$$\begin{aligned} P(X > Y) &= \textstyle\int\int_A f_{X,Y}(x,y)\,dxdy \\ &= \textstyle\int_0^{\infty} \int_y^{\infty} e^{-x-y} dxdy \\ &= 0.5 \end{aligned}$$

Distribuições marginais

Definição 3.9: Dado um par aleatório (X, Y) de v.a. discretas (contínuas) com função massa (densidade) de probabilidade conjunta $f_{X,Y}(x,y)$, as funções massa (densidade) de probabilidade *marginais* de X e de Y são, respectivamente, dadas por

$$f_X(x) = \sum_y f_{X,Y}(x,y) \quad \left(\int_{I\!R} f_{X,Y}(x,y)\,dy\right),$$
$$f_Y(y) = \sum_x f_{X,Y}(x,y) \quad \left(\int_{I\!R} f_{X,Y}(x,y)\,dx\right).$$

Note-se que as funções $f_X(x)$ e $f_Y(y)$ satisfazem as propriedades de f.m.p (f.d.p.), estando associadas igualmente a funções de distribuição (marginais). Por exemplo, se (X, Y) é contínuo: i) $f_X(x) \geq 0$, $\forall\, x \in I\!R$; ii) $\int_{-\infty}^{\infty} f_X(x)dx = 1$; iii) $F_X(x) = P(X \leq x) = \int_{-\infty}^{x} f_X(u)du$.

Exemplo 3.5a: No sistema com duas componentes electrónicas, qual a função de distribuição conjunta de (X, Y), sendo X e Y as durações das componentes?

$$\begin{aligned} F_{X,Y}(x,y) &= \textstyle\int_{-\infty}^{y}\int_{-\infty}^{x} f_{X,Y}(u,v)dudv \overset{x,y>0}{=} \int_0^y \int_0^x e^{-u-v} dudv \\ &= \textstyle\int_0^y e^{-v}(-e^{-u})|_0^x dv = (1 - e^{-x}) \int_0^y e^{-v} dv, \\ &= \begin{cases} (1-e^{-x})(1-e^{-y}), & x, y > 0; \\ 0, & \text{c.c.} \end{cases} \end{aligned}$$

E as funções densidade de probabilidade marginais de X e Y?

$$f_X(x) \overset{x \geq 0}{=} \int_0^{\infty} e^{-x-y} dy = e^{-x}(-e^{-y})|_0^{\infty} = \begin{cases} e^{-x}, & x > 0; \\ 0, & \text{c.c.} \end{cases}$$

$$f_Y(y) \overset{y \geq 0}{=} \int_0^{\infty} e^{-x-y} dx = e^{-y}(-e^{-x})|_0^{\infty} = \begin{cases} e^{-y}, & y > 0; \\ 0, & \text{c.c.} \end{cases}$$

Distribuições condicionais

Definição 3.10: Dado um par aleatório (X, Y) de v.a. discretas (contínuas) com função massa (densidade) de probabilidade conjunta $f_{X,Y}(x, y)$, a função massa (densidade) de probabilidade *condicional* de X dado $Y = y$ é expressa por

$$f_{X|Y=y}(x) = f_{X,Y}(x, y)/f_Y(y), \quad \text{se } f_Y(y) > 0.$$

Analogamente, a função massa (densidade) de probabilidade *condicional* de Y dado $X = x$ é

$$f_{Y|X=x}(y) = f_{X,Y}(x, y)/f_X(x), \quad \text{se } f_X(x) > 0.$$

Observe-se que, e.g., a função $f_{X|Y=y}(x)$ satisfaz as propriedades de f.m.p (f.d.p.) unidimensional, estando associada com a correspondente função de distribuição (condicional):

$$F_{X|Y=y}(x) = P(X \leq x | Y = y) = \begin{cases} \sum_{x_i \leq x} f_{X|Y=y}(x_i) & \text{(discreto)}, \\ \int_{-\infty}^{x} f_{X|Y=y}(u)\, du & \text{(contínuo)}. \end{cases}$$

Exemplo 3.4a: X e Y são os números de cartas rei/dama e ás em 2 cartas retiradas do baralho (sem reposição), respectivamente.

$X \backslash Y$	0	1	2	
0	0.589	0.121	0.004	0.714
1	0.241	0.024	0	0.265
2	0.021	0	0	0.021
	0.851	0.145	0.004	1

A função massa de probabilidade condicional de Y dado $X = 0$ é

$Y \mid X = 0$	0	1	2
$f_{Y\mid X=0}(y)$	$\frac{0.589}{0.714} = 0.825$	$\frac{0.121}{0.714} = 0.169$	$\frac{0.004}{0.714} = 0.006$

Note-se que

$$\sum_y f_{Y|X=0}(y) = \sum_y \frac{f_{X,Y}(0, y)}{f_X(0)} = \frac{1}{f_X(0)} \sum_y f_{X,Y}(0, y) = 1$$

Independência entre variáveis aleatórias

Definição 3.11: Duas v.a. X e Y são ditas *independentes*, se para todo A e B, os eventos $X \in A$ e $Y \in B$ são independentes, *i.e.*,

$$P(X \in A, Y \in B) = P(X \in A)P(Y \in B).$$

Teorema 3.1: Duas variáveis aleatórias X e Y são independentes, se e só se a função de distribuição conjunta de (X, Y) é dada por

$$F_{X,Y}(x, y) = F_X(x)F_Y(y),\ \forall\,(x, y),$$

onde $F_X(x)$ e $F_Y(y)$ são as funções de distribuição marginal de X e Y.

Teorema 3.2: Duas v.a. X e Y discretas (contínuas) são independentes, se e só se a f.m.p. (f.d.p.) conjunta de (X, Y) é dada por

$$f_{X,Y}(x, y) = f_X(x)f_Y(y),\ \forall\,(x, y),$$

onde $f_X(x)$ e $f_Y(y)$ são as f.m.p. (f.d.p.) marginal de X e Y.

Teorema 3.3: Duas v.a. X e Y são independentes, se e só se a f.m.p. (f.d.p.) condicional de X dado $Y = y$ é dada por

$$f_{X|Y=y}(x) = f_X(x),\ \forall\,(x, y), \text{tal que } f_Y(y) > 0.$$

onde $f_X(x)$ e $f_Y(y)$ são as f.m.p. (f.d.p.) marginal de X e Y. Analogamente, X e Y são independentes se e só se $f_{Y|X=x}(y) = f_Y(y)$, $\forall\,(x, y)$, tal que $f_X(x) > 0$.

Exemplo 3.5b: Serão X e Y, durações (em horas) das duas componentes electrónicas do sistema, v.a. independentes?

Sim, visto que

$$\begin{aligned} f_{X,Y}(x, y) &= \begin{cases} \mathrm{e}^{-(x+y)} = \mathrm{e}^{-x}\mathrm{e}^{-y}, & x, y > 0; \\ 0, & \text{c.c.} \end{cases} \\ &= f_X(x)f_Y(y), \forall\,(x, y). \end{aligned}$$

Vectores aleatórios discretos e contínuos

Definição 3.12: Seja $(X_1, \ldots, X_n) \in \mathbb{R}^n$ um vector aleatório, onde X_i, $1 \le i \le n$ são variáveis aleatórias discretas e/ou contínuas. $(X_1, \ldots, X_n)$ é dito ser um *vector aleatório discreto* ou *contínuo* com função de distribuição $F_{X_1,\ldots,X_n}(x_1, \ldots, x_n)$, quando existe uma função não negativa $f_{X_1,\ldots,X_n}(x_1, \ldots, x_n)$ verificando, respectivamente,

$$F_{X_1,\ldots,X_n}(x_1, \ldots, x_n) = \sum_{u_1 \le x_1} \cdots \sum_{u_n \le x_n} f_{X_1,\ldots,X_n}(u_1, \ldots, u_n)$$
$$F_{X_1,\ldots,X_n}(x_1, \ldots, x_n) = \int_{-\infty}^{x_1} \cdots \int_{-\infty}^{x_n} f_{X_1,\ldots,X_n}(u_1, \ldots, u_n)du_1 \ldots du_n$$

e consequentemente

$$\sum_{u_1 \le \infty} \cdots \sum_{u_n \le \infty} f_{X_1,\ldots,X_n}(u_1, \ldots, u_n) = 1$$
$$\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f_{X_1,\ldots,X_n}(u_1, \ldots, u_n)du_1 \ldots du_n = 1.$$

Por generalização óbvia a vectores aleatórios em $I\!R^n$,

Definição 3.13: $X_1, \ldots, X_n$ são v.a. *independentes*, se a função de distribuição de $(X_1, \ldots, X_n)$ é dada por

$$F_{X_1,\ldots,X_n}(x_1, \ldots, x_n) \equiv \; P(X_1 \le x_1, \ldots, X_n \le x_n) = \prod_{i=1}^{n} F_{X_i}(x_i),$$

onde $F_{X_i}(x_i)$ é a função de distribuição marginal de X_i, $i=1, \ldots, n$.

ou equivalentemente, se

$$f_{X_1,\ldots,X_n}(x_1, \ldots, x_n) = \prod_{i=1}^{n} f_{X_i}(x_i),$$

onde $f_{X_i}(x_i)$ é a f.m.p. (f.d.p.) marginal de X_i, $i=1, \ldots, n$.

Exemplo 3.6: Se $X_1, \ldots, X_n$ são v.a. independentes e identicamente distribuídas ($P(X_1 = 1) \equiv p = 1-P(X_1=0)$), qual a f.m.p conjunta de $(X_1, \ldots, X_n)$?

$$f_{X_1,\ldots,X_n}(x_1, \ldots, x_n) = \prod_{i=1}^{n} p^{x_i}(1-p)^{1-x_i} = p^{\sum_i x_i}(1-p)^{n-\sum_i x_i}.$$

4 Distribuições de probabilidade e características

Valor esperado de uma variável aleatória

Definição 4.1: Dada uma v.a. discreta (contínua) X com f.m.p. (f.d.p.) $f_X(x)$, o *valor esperado* (ou valor médio ou esperança matemática) de X, caso exista, é dado por

$$E(X) = \begin{cases} \sum_{x_i} x_i \; f_X(x_i) & \text{(caso discreto)} \\ \int_{I\!R} x \; f_X(x) dx & \text{(caso contínuo).} \end{cases}$$

Exemplo 4.1: Numa questão de escolha múltipla com 5 respostas das quais só uma está correcta, qual a nota esperada de um aluno que responde à questão ao acaso?

$$X = \begin{cases} 1, & \text{resposta correcta} \\ 0, & \text{resposta incorrecta} \end{cases} \qquad E(X) = P(X = 1) = 0.2.$$

Valor esperado de uma função de uma variável aleatória

Teorema 4.1: Seja X uma v.a. discreta (contínua) com f.m.p. (f.d.p.) $f_X(x)$ e $g(X)$ uma função de X. O valor esperado de $g(X)$, se existir, é

$$E(g(X)) = \begin{cases} \sum_{x_i} g(x_i) \; f_X(x_i) & (X \text{ discreta}) \\ \int_{I\!R} g(x) \; f_X(x) dx & (X \text{ e } g(\cdot) \text{ contínuas).} \end{cases}$$

Corolário 4.1: Seja X uma v.a. discreta (contínua) com f.m.p. (f.d.p.) $f_X(x)$ e $a \neq 0$ e b constantes reais. O valor esperado de $aX + b$ é

$$E(aX + b) = aE(X) + b.$$

Observe-se, por exemplo, que $E(\frac{1}{X}) \neq \frac{1}{E(X)}$ e $E(|X|) \neq |E(X)|$.

Momentos simples e centrais

Corolário 4.2: Seja X uma v.a. discreta (contínua) com f.m.p. (f.d.p.) $f_X(x)$ e k inteiro positivo. O valor esperado de X^k, conhecido por momento simples de ordem k de X, caso exista, é

$$E(X^k) = \begin{cases} \sum_{x_i} x_i^k \ f_X(x_i) & \text{(caso discreto)} \\ \int_{I\!R} x^k \ f_X(x)dx & \text{(caso contínuo).} \end{cases}$$

Corolário 4.3: Seja X uma v.a. discreta (contínua) com f.m.p. (f.d.p.) $f_X(x)$ e k inteiro positivo. O valor esperado de $(X - E(X))^k$, conhecido por momento central de ordem k de X, caso exista, é

$$E((X - E(X))^k) = \begin{cases} \sum_{x_i} (x_i - E(X))^k \ f_X(x_i) & \text{(caso discreto)} \\ \int_{I\!R} (x - E(X))^k \ f_X(x)dx & \text{(caso contínuo).} \end{cases}$$

Definição 4.2: Dada uma v.a. discreta (contínua) X com f.m.p. (f.d.p.) $f_X(x)$, a *variância* de X é o momento central de ordem 2 de X, i.e.,

$$Var(X) = E[(X - E(X))^2] = \begin{cases} \sum_{x_i} (x_i - E(X))^2 \ f_X(x_i) & \text{(discreto)} \\ \int_{I\!R} (x - E(X))^2 \ f_X(x)dx & \text{(contínuo).} \end{cases}$$

Se X é uma v.a. com valor esperado $E(X)$ e variância $Var(X)$, têm-se as seguintes propriedades:

P_1: $Var(aX + b) = a^2 Var(X)$, com $a \neq 0, b \in I\!R \Rightarrow Var(b) = 0$.

P_2: $Var(X) = E(X^2) - E(X)^2$.

P_3: $Var(X) = 0 \Rightarrow X$ é constante com probabilidade 1.

Definição 4.3: Se X é uma v.a. com variância $Var(X)$, o *desvio padrão* (outra medida de dispersão) de X é dado por

$$\sigma(X) = +\sqrt{Var(X)}.$$

Definição 4.4: Se X é uma v.a. com valor esperado $E(X) \neq 0$ e desvio padrão $\sigma(X)$, o *coeficiente de variação* (medida de dispersão relativa) de X é

$$CV(X) = \sigma(X)/|E(X)|.$$

Exemplo 4.2: Num lançamento de um dado, um jogador aposta 5 euros nas seguintes condições: i) Se sair face 6, ele ganha 4 vezes o montante apostado; ii) Se sair face 4 ou 5, ele ganha 5 euros; iii) caso contrário, ele nada ganha. Qual o lucro (X) esperado do jogador?

X	-5	0	15
$f_X(x)$	$1/2$	$1/3$	$1/6$

$$\begin{aligned} E(X) &= \sum_x x\, f_X(x) \\ &= 0\,\text{euros}. \end{aligned}$$

E a variância de X? $\quad Var(X) = \sum_x (x - E(X))^2 f_X(x) = 50\,\text{euros}^2$.

Outros parâmetros: Moda e quantis

Definição 4.5: Dada uma v.a. discreta (contínua) X com f.m.p. (f.d.p.) $f_X(x)$, a *moda* de X é dada por

$$m_o(X) = x_o : f_X(x_o) = \max_x f_X(x).$$

Se $X(\Omega) = I\!N$, pode-se encontrar x_o usando as relações

$$f_X(x_o)/f_X(x_o - 1) \geq 1 \quad \text{e} \quad f_X(x_o)/f_X(x_o + 1) \geq 1.$$

Definição 4.6: Dada uma v.a. X com função de distribuição $F_X(x)$, a *mediana* de X é

$$m_d(X) = x_d : F_X(x_d) \geq 0.5 \quad \text{e} \quad P(X \geq x_d) \geq 0.5,$$

ou equivalentemente $0.5 \leq F_X(x_d) \leq 0.5 + P(X = x_d)$.

Definição 4.7: Dado qualquer número p, $0 < p < 1$, o p-ésimo *quantil* de uma v.a. X com função de distribuição $F_X(x)$, denotado por q_p, é dado por

$$F_X(q_p) \geq p \quad \text{e} \quad P(X \geq q_p) \geq 1 - p,$$

ou equivalentemente

$$p \;\leq\; F_X(q_p) \;\leq\; p + P(X = q_p).$$

Observe-se que a mediana é o quantil $q_{0.5}$ e que, no caso contínuo, o p-ésimo quantil é obtido usando somente $F_X(q_p) = p$.

Exemplo 4.2a: Qual o lucro modal do jogador (Exemplo 4.2)? E o lucro mediano?

$$\begin{aligned} &m_o(X) = -5\,\text{euros} \\ &m_d(X) \in [-5, 0];\ \textit{e.g.}, -2.5\,\text{euros}. \end{aligned}$$

Exemplo 4.3: A percentagem de uma substância $(100X)$ em um certo composto químico é tal que X é uma v.a. descrita pela função

$$f_X(x) = \begin{cases} 6\,x\,(1-x), & 0 \leq x \leq 1; \\ 0, & \text{c.c.} \end{cases}$$

- Será $f_X(x)$ uma f.d.p.? Sim, $f_X(x) \geq 0$, $\forall\, x$, e $\int_0^1 6\,x\,(1-x)dx = 1$.
- Qual a percentagem média da substância no composto?
  $E(100X) = 100 \int_0^1 x\,6\,x\,(1-x)dx = 50\%$.
- E a moda de X? $m_o(X) = 0.5$ pois $f_X(0.5) = \max_x f_X(x)$
- E a mediana de X? $m_d(X) = 0.5$ pois, sendo $f_X(x)$ simétrica em torno de 0.5, $F_X(0.5) = \int_0^{0.5} 6\,x\,(1-x)dx = 0.5$.
- Qual a variância de X? $Var(X) = E(X^2) - E(X)^2 = \frac{3}{10} - \frac{1}{4} = \frac{1}{20}$.
- E o coeficiente de variação de X? $CV(X) = \frac{\sqrt{0.05}}{|0.5|} \approx 0.447$.

Distribuição uniforme discreta

Definição 4.8: Diz-se que uma v.a. X, com contradomínio finito, tem distribuição *uniforme discreta* se todos os seus valores $x_1, \ldots, x_k$ são igualmente prováveis, com f.m.p dada por

$$f_X(x) = \begin{cases} 1/k, & x = x_1, \ldots, x_k; \\ 0, & \text{caso contrário.} \end{cases}$$

O valor esperado e a variância de uma variável aleatória X com distribuição uniforme discreta $\{x_1, \ldots, x_k\}$ são, respectivamente,

$$E(X) = \tfrac{1}{k} \textstyle\sum_{i=1}^k x_i \quad \text{e} \quad Var(X) = \tfrac{1}{k} \textstyle\sum_{i=1}^k x_i^2 - \left(\tfrac{1}{k} \textstyle\sum_{i=1}^k x_i\right)^2.$$

Exemplo 4.4: Seja X uma v.a. com distribuição uniforme discreta $\{1, \ldots, k\}$. Qual a variância de X?

$$E(X) = \tfrac{k+1}{2},\ E(X^2) = \tfrac{(k+1)(2k+1)}{6} \ \Rightarrow\ Var(X) = \tfrac{k^2-1}{12}.$$

Distribuição Bernoulli

Definição 4.9: Uma experiência aleatória com somente dois resultados possíveis, sucesso (ocorrência de um acontecimento de interesse) e fracasso (caso contrário), é conhecida por ensaio (ou prova) de Bernoulli, e a v.a. subjacente definida por

$$X = \begin{cases} 1, & \text{se ocorre sucesso;} \\ 0, & \text{se ocorre fracasso} \end{cases}$$

possui f.m.p. (conhecida por distribuição *Bernoulli*)

$$f_X(x) = \begin{cases} p^x(1-p)^{1-x}, & x = 0, 1; \\ 0, & \text{c.c.,} \end{cases}$$

onde $p = P(X = 1)$ é a probabilidade de sucesso, $0 < p < 1$. Consequentemente, $E(X) = p$ e $Var(X) = p(1-p)$.

Distribuição binomial

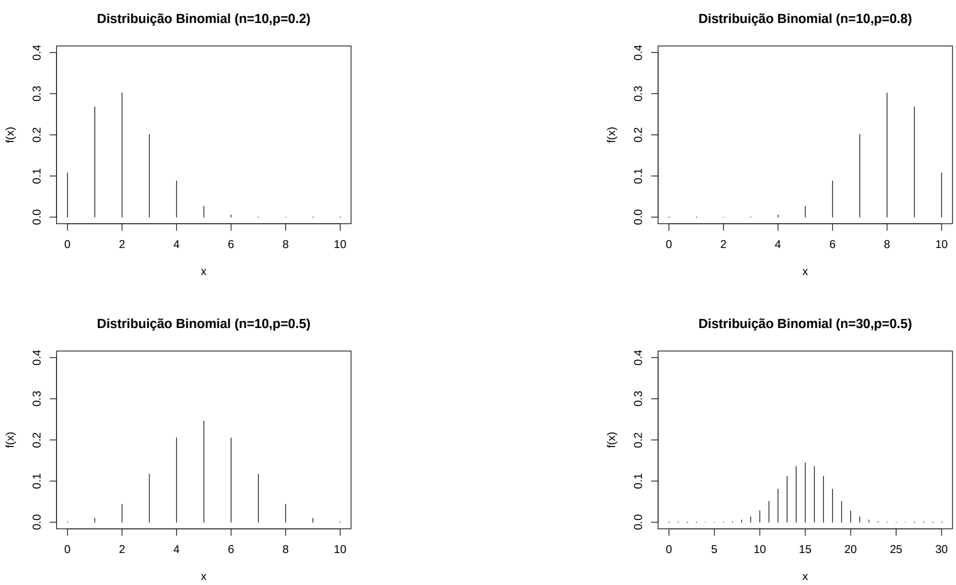
Definição 4.10: Considere uma experiência aleatória com n ensaios de Bernoulli independentes e todos com probabilidade de sucesso p. A v.a. correspondente ao número X de sucessos na experiência tem distribuição *binomial* com parâmetros n e p, com f.m.p.

$$f_X(x) = \begin{cases} \binom{n}{x} p^x (1-p)^{n-x}, & x = 0, 1, \ldots, n; \\ 0, & \text{c.c.} \end{cases}$$

- O valor esperado e a variância de $X \sim \text{Binomial}(n, p)$ são

$$E(X) = n\,p \quad \text{e} \quad Var(X) = n\,p(1-p).$$

- A moda de X satisfaz a relação $np+p-1 \leq m_o(X) \leq np+p$.
- Se $X_i \sim \text{Bernoulli}(p)$, $i=1, \ldots, n$, são v.a. independentes, então $X = \sum_{i=1}^{n} X_i \sim \text{Binomial}(n, p)$.
- Se $X \sim \text{Binomial}(n, p)$, então $n-X \sim \text{Binomial}(n, (1-p))$.



Exemplo 4.5: Considere um teste de múltipla escolha com 10 questões, onde somente uma das 5 alíneas de cada questão está correcta. Qual a probabilidade de um aluno acertar em pelo menos metade das questões fazendo o teste ao acaso?

Seja X o número de respostas correctas no teste do aluno. $X \sim \text{Binomial}(n=10, p=1/5)$.

$$P(X \geq 5) = 1 - F_X(4) = 0.0328$$

Qual a nota esperada desse aluno, se a cotação de cada questão é 1? E a nota modal e mediana? E o desvio padrão de X?

$$\begin{aligned}
E(X) &= 10 \times \tfrac{1}{5} = 2\,\text{valores} \\
m_o(X) &= 2\,\text{valores}\ (f_X(2)/f_X(1) \geq 1, f_X(2)/f_X(3) \geq 1) \\
m_d(X) &= 2\,\text{valores}\ (F_X(1) = 0.3758, F_X(2) = 0.6778) \\
\sigma(X) &= \sqrt{10 \times \tfrac{1}{5} \times \tfrac{4}{5}} = 1.265\,\text{valores}
\end{aligned}$$

Distribuição hipergeométrica

Definição 4.11: Considere uma população com N elementos dos quais M possuem uma certa característica (sucesso). Retira-se uma amostra, sem reposição, de dimensão n, anotando-se o número X de elementos com a característica na amostra. A distribuição de probabilidade da v.a. X é designada distribuição *hipergeométrica*, cuja f.m.p. é

$$f_X(x) = \begin{cases} \frac{\binom{M}{x}\binom{N-M}{n-x}}{\binom{N}{n}}, & \max(0, n+M-N) \leq x \leq \min(n, M); \\ 0, & \text{c.c.} \end{cases}$$

O valor esperado e a variância de uma variável aleatória X com distribuição hipergeométrica (N, M, n) são, respectivamente,

$$E(X) = n\,\frac{M}{N} \quad \text{e} \quad Var(X) = n\,\frac{M}{N}\,\frac{N-M}{N}\,\frac{N-n}{N-1}.$$

Exemplo 4.6: Numa turma de 10 estudantes dos quais 3 são mulheres, 2 estudantes foram sorteados para formar uma comissão. Qual a probabilidade de haver pelo menos uma mulher na comissão?

Seja X o número de mulheres na comissão de 2 estudantes. $X \sim$ Hipergeométrica$(N=10, M=3, n=2)$.

$$P(X \geq 1) \;=\; 1 - P(X = 0) \;\simeq\; 0.533$$

Se $X_i = 1$ ($X_i = 0$) denota a ocorrência de mulher (homem) no sorteio do estudante i, $i = 1, 2$, qual a probabilidade de ser mulher o primeiro estudante sorteado? E o segundo?

$$\begin{aligned}
&P(X_1=1) = 0.3 \\
&P(X_2=1) = P(X_1=0, X_2=1) + P(X_1=1, X_2=1) = 0.3
\end{aligned}$$

Note-se que $X = \sum_i X_i$, onde X_i são v.a. Bernoulli(p) dependentes.

Distribuição geométrica

Definição 4.12: Considere uma experiência aleatória envolvendo a realização de ensaios de Bernoulli independentes, com probabilidade de sucesso p, até à ocorrência do primeiro sucesso. A v.a. X número de ensaios realizados até à ocorrência do primeiro sucesso tem distribuição *geométrica* com parâmetro p, $0 < p < 1$, com f.m.p.

$$f_X(x) = \begin{cases} (1-p)^{x-1}p, & x = 1, 2, \ldots; \\ 0, & \text{c.c.,} \end{cases}$$

O valor esperado e a variância de $X \sim$ Geométrica(p) são, respectivamente,

$$E(X) = \frac{1}{p} \quad \text{e} \quad Var(X) = \frac{1-p}{p^2}.$$

Teorema 4.2: (Propriedade da falta de memória) Se uma v.a. $X \sim$ Geométrica(p), então

$$P(X > i + j | X > j) = P(X > i), \ \forall\, i, j = 1, 2, \ldots.$$

Exemplo 4.7: Seja X o número de lançamentos de um dado até ao surgimento da primeira face 6. Qual o número esperado de lançamentos do dado até sair a face 6?

Como $X \sim$ Geométrica$(p = \frac{1}{6})$, $E(X) = 6$ lançamentos.

Qual a probabilidade de serem necessários mais de 7 lançamentos, sabendo que já houve 3 lançamentos do dado sem que a face 6 saísse?

$$P(X > 7 | X > 3) = P(X > 4) = \sum_{x \geq 5} \left(\frac{5}{6}\right)^{x-1} \frac{1}{6} \simeq 0.4822$$

Distribuição Poisson

Definição 4.13: Em algumas experiências aleatórias, anota-se por vezes o número X de ocorrências de um evento de interesse num dado intervalo de tempo, superfície, volume, etc. A v.a. X tem distribuição de *Poisson* de parâmetro λ quando a sua f.m.p. é dada por

$$f_X(x) = \begin{cases} \frac{e^{-\lambda}\lambda^x}{x!}, & x = 0, 1, 2, \ldots; \\ 0, & \text{c.c.}, \end{cases}$$

onde λ é a taxa (esperada) de ocorrência do evento de interesse na base considerada. Se $X \sim$ Poisson(λ),

- X poderá representar, *e.g.*, o número de electrões emitidos por uma substância fotosensível, sob a acção da luz durante uma dada unidade de tempo.
- O valor esperado e a variância de X são iguais a $E(X) = Var(X) = \lambda$ e a moda de X satisfaz a relação $\lambda - 1 \leq m_o(X) \leq \lambda$.

Teorema 4.3: Seja X o número de ocorrências de um evento de interesse num dado período de tempo (região). Se $X \sim$ Poisson(λ), então

$$X = \sum_{i=1}^{t} X_i,$$

com $X_i \sim$ Poisson$(\frac{\lambda}{t})$, $i = 1, \ldots, t$ (independentes e identicamente distribuídas), sendo X_i o número de ocorrências do evento em cada uma das t fracções do período de tempo (região).

Exemplo 4.8: Suponha que X é o número de passas de um bolo-rei oriundo de uma padaria que se sabe ter uma distribuição de Poisson com taxa média de 5 passas por bolo. Qual a probabilidade de encontrar pelo menos 1 passa em meio bolo-rei dessa padaria?

Seja X^* o número de passas em meio bolo-rei produzido nessa padaria. $X^* \sim \text{Poisson}(\lambda^* = 2.5)$.

$$P(X^* \geq 1) = 1 - P(X^* = 0) = 1 - \mathrm{e}^{-2.5} \simeq 0.918.$$

Distribuição uniforme contínua

Definição 4.14: Diz-se que uma variável aleatória X tem distribuição *uniforme contínua* (ou rectangular) se, para qualquer ponto entre a e b ($a < b$), a sua f.d.p. é dada por

$$f_X(x) = \begin{cases} \frac{1}{b-a}, & a < x < b; \\ 0, & \text{c.c.} \end{cases}$$

O valor esperado e a variância de uma v.a. X com distribuição uniforme contínua (a, b) são, respectivamente,

$$E(X) = \frac{b+a}{2} \quad \text{e} \quad Var(X) = \frac{(b-a)^2}{12}.$$

Exemplo 4.9: Sabe-se que o tempo X gasto por um aluno no trajecto de casa para a escola pode ser qualquer valor entre 20 a 40 minutos (valores igualmente prováveis). Saindo de casa às 12:30 para assistir a aula das 13:00, qual a probabilidade de ele chegar atrasado?

Seja p a probabilidade de o aluno chegar atrasado à escola. Se a v.a. $X \sim \text{Uniforme}(20, 40)$,

$$p = P(X > 30) = \int_{30}^{40} \frac{1}{20} dx = 0.5.$$

Em 12 dias, qual o número esperado de dias em que ele chega atrasado?

Seja Y o número de dias entre os 12 em que o aluno chega atrasado à escola. Supondo independência entre os tempos gastos nos 12 dias e a mesma probabilidade de atraso p, $Y \sim \text{Bi}(n = 12, p = 0.5)$ e, por conseguinte,

$$E(Y) = 12 \times 0.5 = 6 \text{ dias}.$$

Distribuição exponencial

Definição 4.15: Diz-se que uma v.a. contínua X tem distribuição *exponencial*, com parâmetro $\lambda > 0$, se a sua f.d.p. é dada por

$$f_X(x) = \begin{cases} \lambda \mathrm{e}^{-\lambda x}, & x \geq 0; \\ 0, & \text{c.c.} \end{cases}$$

Esta distribuição é bastante utilizada para descrever tempos de vida de materiais ou seres vivos em estudos de Análise de Fiabilidade e Sobrevivência.

O valor esperado e a variância de uma v.a. X com distribuição exponencial (λ) são, respectivamente,

$$E(X) = \frac{1}{\lambda} \quad \text{e} \quad Var(X) = \frac{1}{\lambda^2}.$$

Exemplo 4.10: Suponha que o tempo X de falha de duas componentes electrónicas tem distribuição exponencial com média de 5 horas (componente C_1) e de 10 horas (C_2). Considere ainda que elas estão ligadas num sistema em paralelo e que o funcionamento de cada uma não depende do da outra. Qual a fiabilidade do sistema após 20 horas?

A fiabilidade do sistema com as duas componentes em paralelo é a probabilidade de pelo menos uma componente funcionar, denotada por

$$P(F_1 \cup F_2) = P(F_1) + P(F_2) - P(F_1 \cap F_2) = 0.1511,$$

uma vez que elas são independentes e a fiabilidade de cada uma é

$$\begin{aligned} P(F_1) &= P(X_1 > 20) = \int_{20}^{\infty} \tfrac{1}{5} \mathrm{e}^{-\frac{1}{5}\,x_1} dx_1 = \mathrm{e}^{-\frac{20}{5}} = 0.0183, \\ P(F_2) &= P(X_2 > 20) = \int_{20}^{\infty} \tfrac{1}{10} \mathrm{e}^{-\frac{1}{10}\,x_2} dx_2 = \mathrm{e}^{-\frac{20}{10}} = 0.1353. \end{aligned}$$

Teorema 4.4: (Propriedade da falta de memória) Se uma v.a. $X \sim$ Exponencial(λ), então $P(X > s+t | X > t) = P(X > s)$, $\forall\, s, t \geq 0$.

Processo de Poisson (*)

Definição 4.16: Suponha-se que acontecimentos (e.g., chegadas de clientes a um caixa de banco) ocorrem aleatoriamente num intervalo de tempo $[0, t]$, onde N_t é o número de ocorrências do acontecimento no intervalo de tempo de comprimento t. Diz-se que essas ocorrências constituem um *Processo de Poisson* com taxa unitária λ, $\lambda > 0$, se

- Os números de acontecimentos que ocorrem em intervalos não sobrepostos são independentes (independência).
- A distribuição do número de acontecimentos que ocorrem em um dado intervalo depende somente do comprimento do intervalo e não da sua localização (estacionariedade).
- A probabilidade de ocorrer exactamente um acontecimento em qualquer intervalo de comprimento Δt arbitrariamente pequeno é aproximadamente $\lambda \Delta t$ (i.e., $P(N_{\Delta t} = 1) \approx \lambda \Delta t$).
- A probabilidade de ocorrerem dois ou mais acontecimentos em qualquer intervalo de comprimento Δt arbitrariamente pequeno é aproximadamente igual a 0 (i.e., $P(N_{\Delta t} \geq 2) \approx 0$).

$\therefore \quad N_t \sim$ Poisson($\lambda\, t$).

Teorema 4.5: Seja N_t o número de ocorrências num intervalo de tempo de comprimento t, com $N_t \sim$ Poisson($\lambda\, t$). Considere ainda que X_1 é o tempo decorrido até à primeira ocorrência, enquanto X_i, $i > 1$, é o tempo decorrido entre as ocorrências $i-1$ e i. A sequência $X_1, X_2, \ldots$ é formada por v.a. i.i.d. com $X_i \sim$ Exponencial(λ), $i = 1, 2, \ldots$, onde λ é a taxa média de ocorrências por unidade de tempo. Nomeadamente,

- $P(X_1 > t) = P(N_t = 0) = \mathrm{e}^{-\lambda\, t} \Leftrightarrow X_1 \sim$ Exponencial(λ).
- $P(X_2 > t | X_1 = s) = P(N_{(s,s+t]} = 0 | N_{(0,s]} = 1) \overset{*}{=} P(N_{(s,s+t]} = 0) \overset{**}{=} P(N_t = 0) = \mathrm{e}^{-\lambda\, t} \equiv P(X_2 > t) \Leftrightarrow$ $X_2 \sim$ Exponencial(λ).

  * (**) pela suposição de independência (estacionariedade) das ocorrências.

Distribuição normal (ou de Gauss)

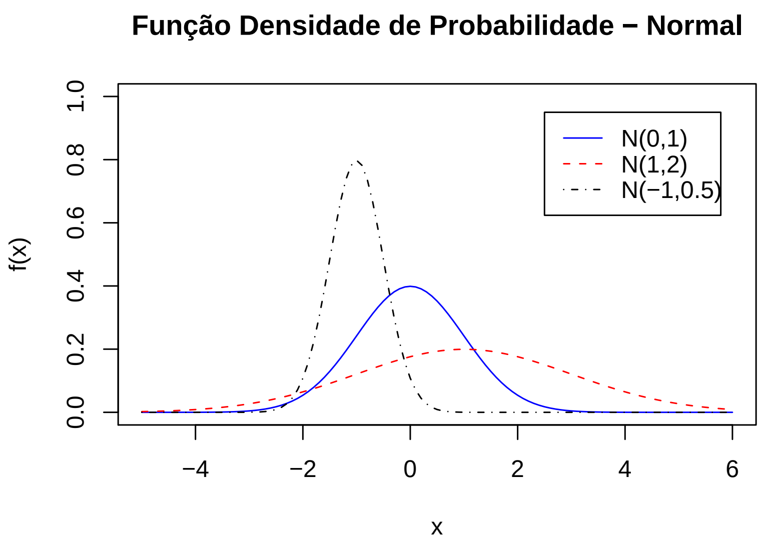
Definição 4.17: Diz-se que uma v.a. contínua X tem distribuição *normal* (ou gaussiana) com média μ e variância σ^2, denotada por $N(\mu,\sigma^2)$, se a sua f.d.p. é dada por

$$f_X(x) = \frac{1}{\sqrt{2\,\pi}\,\sigma}\exp\left[-\frac{1}{2\,\sigma^2}(x-\mu)^2\right], -\infty < x < \infty.$$

Propriedades da curva gaussiana $f_X(x)$:

- Como a função é simétrica em relação a μ, a mediana de X é μ.
- $f_X(x)$ atinge o ponto máximo em $x=\mu$ com valor $1/(\sqrt{2\,\pi}\,\sigma)$ e portanto a moda de X é μ.
- A curva gaussiana tem 2 pontos de inflexão em $x = \mu \pm \sigma$.

Neste cenário, X poderá ser, e.g., a velocidade numa dada direcção de uma molécula de gás de massa M à temperatura absoluta T, com distribuição $N(0, \frac{K\,T}{M})$, onde K é a constante de Boltzmann.



Teorema 4.6: Se uma v.a. $X \sim N(\mu,\sigma^2)$, $Y = aX+b \sim N(a\mu+b, a^2\,\sigma^2)$, com a e b constantes reais.

Corolário 4.4: Se uma v.a. $X \sim N(\mu,\sigma^2)$, então $Z = (X-\mu)/\sigma \sim N(0,1)$, conhecida por distribuição normal reduzida (ou padrão), cujas probabilidades são dadas em calculadoras ou tabelas.

Exemplo 4.11: Suponha que a altura X dos alunos de uma turma de PE tem distribuição normal com média $\mu = 160$ cm e desvio padrão $\sigma = 20$ cm. Qual a probabilidade de um aluno seleccionado ao acaso ter altura ente 150 e 170 cm?

$$
\begin{aligned}
P(150 < X < 170) &= P\left(\tfrac{150-160}{20} < \tfrac{X-\mu}{\sigma} < \tfrac{170-160}{20}\right) \\
&= P(-0.5 < Z < 0.5) = F_Z(0.5) - F_Z(-0.5) \\
&= 0.6915 - 0.3085 \\
&= 0.383.
\end{aligned}
$$

5 Complementos das distribuições de probabilidade

Valor esperado de uma função de um par aleatório discreto e contínuo

Definição 5.1: Dado um par aleatório (X,Y) com f.m.p. ou f.d.p. conjunta $f_{X,Y}(x,y)$, o *valor esperado* de uma função $g(X,Y)$ é dado por

$$
E(g(X,Y)) = \begin{cases} \sum_{x_i}\sum_{y_i} g(x_i,y_j) f_{X,Y}(x_i,y_j) & \text{(caso discreto)}, \\ \int_{I\!R}\int_{I\!R} g(x,y) f_{X,Y}(x,y)\, dxdy & \text{(caso contínuo)}. \end{cases}
$$

Exemplo 5.1: Seja (X,Y) um par aleatório com f.d.p. conjunta $f_{X,Y}(x,y)$ e marginais $f_X(x)$ e $f_Y(y)$. Qual o valor esperado de $X+Y$?

$$
\begin{aligned}
E(X+Y) &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(x+y) f_{X,Y}(x,y)dxdy \\
&= \int_{-\infty}^{\infty} x f_X(x)dx + \int_{-\infty}^{\infty} y f_Y(y)dy = E(X) + E(Y).
\end{aligned}
$$

E a variância de $X+Y$?

$$
\begin{aligned}
E((X+Y)^2) &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(x+y)^2 f_{X,Y}(x,y)dxdy \\
&= \int_{-\infty}^{\infty} x^2 f_X(x)dx + \int_{-\infty}^{\infty} y^2 f_Y(y)dy \\
&\quad +2\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xy f_{X,Y}(x,y)dxdy \\
&= E(X^2) + E(Y^2) + 2E(XY)
\end{aligned}
$$

$$
\begin{aligned}
\therefore\ Var(X+Y) &= E((X+Y)^2) - (E(X+Y))^2 \\
&= E(X^2) + 2E(XY) + E(Y^2) \\
&\quad -(E(X))^2 - 2E(X)E(Y) - (E(Y))^2 \\
&= Var(X) + Var(Y) + 2(E(XY) - E(X)E(Y)).
\end{aligned}
$$

Teorema 5.1: Se X e Y são variáveis aleatórias independentes com f.m.p. (f.d.p.) conjunta $f_{X,Y}(x,y)$ e marginais $f_X(x)$ e $f_Y(y)$, então $E(X\,Y) = E(X)E(Y)$.

Demonstração (caso discreto):

$$
\begin{aligned}
E(X\,Y) &= \sum_{x_i}\sum_{y_j} x_i\, y_j f_{X,Y}(x_i,y_j) = \sum_{x_i}\sum_{y_j} x_i\, y_j f_X(x_i) f_Y(y_j) \\
&= \sum_{x_i} x_i f_X(x_i) \sum_{y_j} y_j f_Y(y_j) \\
&= E(X)E(Y).
\end{aligned}
$$

Exemplo 5.2: Sejam X e Y v.a. com f.d.p. conjunta

$f_{X,Y}(x,y) = \mathrm{e}^{-x-y}$ (Exemplo 3.5) e f.d.p. marginais $f_X(x) = \mathrm{e}^{-x}$ e $f_Y(y) = \mathrm{e}^{-y}$ (Exemplo 3.5a). Encontre $E(X,Y)$. Como X e Y são v.a. independentes com distribuição Exponencial ($\lambda = 1$),

$$E(XY) = E(X)E(Y) = 1.$$

Note-se que $E(X)=1=E(Y)$ e $E(XY)=\int_0^\infty\int_0^\infty xy\,\mathrm{e}^{-(x+y)}dxdy=1$.

Covariância

Definição 5.2: Dadas duas v.a. X e Y, a *covariância* de X e Y é o valor esperado do produto dos desvios médios de X e Y, *i.e.*,

$$Cov(X,Y) = E[(X - E(X))(Y - E(Y))].$$

Propriedades da covariância: Dado um par aleatório (X,Y) com f.m.p. (f.d.p.) conjunta $f_{X,Y}(x,y)$,

1. $Cov(X,Y) = \sum_{x_i}\sum_{y_j}(x_i - E(X))(y_j - E(Y))f_{X,Y}(x_i,y_j)$ (caso discreto).
2. $Cov(X,Y) = \int_{\mathbb{R}}\int_{\mathbb{R}}(x - E(X))(y - E(Y))f_{X,Y}(x,y)dxdy$ (caso contínuo).
3. $Cov(X,X) = Var(X)$.
4. $Cov(X,Y) = E(XY) - E(X)E(Y)$.

Teorema 5.2: Se X e Y são v.a. independentes, então a covariância de X e Y é nula.

$$Cov(X,Y) = E(XY) - E(X)E(Y) = 0,$$

visto que $E(XY) = E(X)E(Y)$, quando X e Y são independentes.

Exemplo 5.3: Sejam X e Y duas v.a. contínuas com f.d.p. conjunta $f_{X,Y}(x,y) = 1$, se $0 \le x,y \le 1$, e 0, no caso contrário.

- $E(XY) = \int_0^1\int_0^1 xy \times 1\,dxdy = \frac{1}{4}$.
- $E(X) = \int_0^1\int_0^1 x \times 1\,dxdy = \frac{1}{2}$ e $E(Y) = \int_0^1\int_0^1 y \times 1\,dxdy = \frac{1}{2}$.

  $\therefore \quad Cov(X,Y) = \frac{1}{4} - \frac{1}{2}\frac{1}{2} = 0.$

Resultado previsível pois X e Y são v.a. independentes ($f_{X,Y}(x,y) = f_X(x) \times f_Y(y)$, $\forall\,(x,y) \in \mathbb{R}^2$).

Exemplo 5.4: Sejam X, Y e Z variáveis aleatórias. Qual a covariância de $X + Z$ e Y?

$$\begin{aligned} Cov(X+Z,Y) &= E((X+Z)Y) - E(X+Z)E(Y) \\ &= E(XY) + E(ZY) - E(X)E(Y) - E(Z)E(Y) \\ &= Cov(X,Y) + Cov(Z,Y) \end{aligned}$$

E de $a\,X$ e $Y+b$, onde $a \neq 0$ e b são constantes reais?

$$\begin{aligned} Cov(a\,X, Y+b) &= E((a\,X - E(a\,X))(Y+b-E(Y+b))) \\ &= a\,E((X-E(X))(Y-E(Y))) = a\,Cov(X,Y) \end{aligned}$$

E a variância de $X - Y$?

$$\begin{aligned} Var(X-Y) &= E((X-Y)^2) - (E(X-Y))^2 \\ &= E(X^2) - 2\,E(XY) + E(Y^2) \\ &\quad -(E(X))^2 + 2\,E(X)E(Y) - (E(Y))^2 \\ &= Var(X) + Var(Y) - 2\,Cov(X,Y) \end{aligned}$$

Correlação

Definição 5.3: Dado um par aleatório (X,Y), o *coeficiente de correlação* (linear) de X e Y é um parâmetro adimensional dado por

$$Corr(X,Y) = \frac{Cov(X,Y)}{\sqrt{Var(X)\,Var(Y)}}.$$

Propriedades do coeficiente de correlação: Sejam (X,Y) um par aleatório e $a \neq 0$ e b constantes reais.

1. $-1 \leq Corr(X,Y) \leq 1 \Leftrightarrow [Cov(X,Y)]^2 \leq Var(X)Var(Y)$.
2. $Y = a\,X + b \ \Leftrightarrow Corr(X,Y) = \pm 1$ (correlação linear perfeita).
3. $Corr(a\,X, Y+b) = \frac{a}{|a|} Corr(X,Y)$.

Note-se que a não correlação entre X e Y, $Corr(X,Y) = 0$ (*i.e.*, $Cov(X,Y) = 0$), não implica independência entre X e Y.

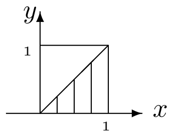
Exemplo 5.5: Sejam X e Y os números de cartas rei/dama e ás em 2 cartas retiradas do baralho (sem reposição), respectivamente (Exemplo 3.4). Qual o coeficiente de correlação de X e Y?

$X \backslash Y$	0	1	2	
0	0.589	0.121	0.004	0.714
1	0.241	0.024	0	0.265
2	0.021	0	0	0.021
	0.851	0.145	0.004	1

- $E(X) = 0.307$, $E(X^2) = 0.349$ e $Var(X) = 0.2547$.
- $E(Y) = 0.153$, $E(Y^2) = 0.161$ e $Var(Y) = 0.1376$.
- $E(XY) = 0.024$ e $Cov(X,Y) = -0.023$.
- $Corr(X,Y) = \frac{-0.023}{\sqrt{0.2547 \times 0.1376}} = \frac{-0.023}{0.1872} = -0.123$.
- X e Y estão pouco correlacionadas linearmente e de uma forma negativa (quando uma variável cresce, a outra decresce).

Exemplo 5.6: Seja (X, Y) um par aleatório contínuo com a seguinte f.d.p. conjunta $f_{X,Y}(x, y)$. Qual o coeficiente de correlação de X e Y?

$$f_{X,Y}(x,y) = \begin{cases} 2, & 0 < y < x < 1; \\ 0, & \text{c.c.} \end{cases}$$



- $E(X) = 2/3$, $E(X^2) = 1/2$ e $Var(X) = 1/18$.
- $E(Y) = 1/3$, $E(Y^2) = 1/6$ e $Var(Y) = 1/18$.
- $E(XY) = 1/4$ e $Cov(X, Y) = 1/36$.
- $Corr(X, Y) = \frac{1/36}{\sqrt{1/18 \times 1/18}} = 0.5$.
- X e Y estão moderadamente correlacionadas de uma forma linear e positiva (quando uma variável cresce, a outra também cresce).
- Note-se que $Corr(X, Y) \neq 0 \Rightarrow X$ e Y não são independentes.

Definição 5.4: Dadas n v.a. $X_1, \ldots, X_n$ e n constantes reais $c_1, \ldots, c_n$, uma combinação linear das variáveis aleatórias é uma v.a. Y tal que

$$Y = \sum_{i=1}^{n} c_i X_i.$$

Por generalização da Definição 5.1,

- $E(Y) = E(\sum_{i=1}^{n} c_i X_i) = \sum_{i=1}^{n} c_i E(X_i)$.
- $Var(Y) = \sum_{i=1}^{n} c_i^2 Var(X_i) + 2 \sum_{i<j} \sum_{j=2}^{n} c_i c_j Cov(X_i, X_j)$. Se $X_1, \ldots, X_n$ são v.a. independentes, $Var(Y) = \sum_{i=1}^{n} c_i^2 Var(X_i)$.

Teorema 5.3: Se $X_1, \ldots, X_n$ são v.a. independentes tais que $X_i \sim N(\mu_i, \sigma_i^2)$, $i = 1, \ldots, n$, então para $c_1, \ldots, c_n$ constantes reais

$$Y = \sum_{i=1}^{n} c_i X_i \sim N\left(\sum_{i=1}^{n} c_i \mu_i, \sum_{i=1}^{n} c_i^2 \sigma_i^2\right).$$

Proposição 5.1: Sejam $X_1, \ldots, X_n$ e $Y_1, \ldots, Y_m$ variáveis aleatórias.

$$Cov\left(\sum_{i=1}^{n} X_i, \sum_{j=1}^{m} Y_j\right) = \sum_{i=1}^{n} \sum_{j=1}^{m} Cov(X_i, Y_j).$$

Corolário 5.1: Sejam $X_1, \ldots, X_n$ variáveis aleatórias.

$$Var\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} Var(X_i) + \sum_{i=1}^{n} \sum_{j \neq i=1}^{n} Cov(X_i, X_j).$$

Corolário 5.2: Se $X_1, \ldots, X_n$ são variáveis aleatórias independentes,

$$Var\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} Var(X_i).$$

Teorema 5.4: Se $X_1, \ldots, X_n$ são v.a. independentes tais que $X_i \sim$ Exponencial(λ_i), $i=1, \ldots, n$, então

$$Y = \min(X_1, \ldots, X_n) \sim \text{Exponencial}(\lambda = \sum_{i=1}^{n} \lambda_i).$$

Teorema 5.5: Se $X_1, \ldots, X_n$ são v.a. independentes tais que $X_i \sim$ Bernoulli(p), $i=1, \ldots, n$, então

$$Y = \sum_{i=1}^{n} X_i \sim \text{Binomial}(n, p).$$

- $E(Y) = \sum\limits_{i=1}^{n} E(X_i) = np$ e $Var(Y) = \sum\limits_{i=1}^{n} Var(X_i) = np(1-p)$.

Teorema 5.6: Se $X_1, \ldots, X_n$ são v.a. dependentes associadas com n extracções sem reposição de uma população de dimensão N com M sucessos, tais que $X_i \sim$ Bernoulli$(p = \frac{M}{N})$, $i=1, \ldots, n$, então

$$Y = \sum_{i=1}^{n} X_i \sim \text{Hipergeométrica}(N, M, n).$$

- $E(Y) = n\frac{M}{N}$ e $Var(Y) = n\frac{M}{N}\frac{N-M}{N}\frac{N-n}{N-1}$.

Teorema 5.7: Se $X_1, \ldots, X_n$ são v.a. independentes tais que $X_i \sim$ Poisson(λ_i), $i=1, \ldots, n$, então

$$Y = \sum_{i=1}^{n} X_i \sim \text{Poisson}(\lambda = \sum_{i=1}^{n} \lambda_i).$$

Exemplo 5.7: Suponha que X_1 e X_2 são v.a. independentes com distribuição de Poisson com parâmetros λ_1 e λ_2, respectivamente. Qual a distribuição de $X_1 + X_2$?

Seja $Z = X_1 + X_2$ uma v.a. com f.m.p. $f_Z(z)$, *i.e.*,

$$\begin{aligned}
f_Z(z) &\equiv P(Z = z) = \sum_{x_1} P(X_1 = x_1, Z = z) \\
&= \sum_{x_1} P(X_1 = x_1) P(Z = z | X_1 = x_1) \\
&= \sum_{x_1} P(X_1 = x_1) P(X_2 = z - x_1 | X_1 = x_1) \\
&= \sum_{x_1} P(X_1 = x_1) P(X_2 = z - x_1) \\
&= \sum_{x_1=0}^{z} \frac{\mathrm{e}^{-\lambda_1} {\lambda_1}^{x_1}}{x_1!} \frac{\mathrm{e}^{-\lambda_2} {\lambda_2}^{z-x_1}}{(z-x_1)!} = \frac{\mathrm{e}^{-(\lambda_1+\lambda_2)}}{z!} \sum_{x_1=0}^{z} \frac{z!}{x_1!\,(z-x_1)!} \lambda_1^{x_1} \lambda_2^{z-x_1} \\
&= \frac{\mathrm{e}^{-(\lambda_1+\lambda_2)}}{z!} (\lambda_1 + \lambda_2)^z
\end{aligned}$$

Consequentemente, $Z = X_1 + X_2 \sim$ Poisson$(\lambda = \lambda_1 + \lambda_2)$.

Valor esperado e matriz de covariâncias de um par aleatório

Seja $\mathbf{X} = (X_1, X_2)$ um vector aleatório em $I\!R^2$, onde X_1 e X_2 são v.a. com $E(X_i) = \mu_i$, $Var(X_i) = \sigma_i^2$, $i=1,2$, e $Cov(X_1, X_2) = \sigma_{12}$. O valor esperado de $\mathbf{X}$ é entendido como

$$E(\mathbf{X}) = \boldsymbol{\mu} \equiv \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}$$

enquanto a matriz de covariâncias de $\mathbf{X}$ é

$$Var(\mathbf{X}) = E((\mathbf{X}-\boldsymbol{\mu})(\mathbf{X}-\boldsymbol{\mu})^T) \equiv \begin{pmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{21} & \sigma_2^2 \end{pmatrix}, \quad \sigma_{21}=\sigma_{12}$$

Se X_1 e X_2 são v.a. independentes e $Var(X_i)=\sigma^2$, $i=1,2$, então $Var(\mathbf{X}) = \sigma^2\mathbf{I}_2$, onde $\mathbf{I}_2$ é a matriz identidade de ordem 2.

Valor esperado condicional e propriedades

Definição 5.5: Dado um par aleatório (X, Y) discreto (contínuo) com f.m.p. (f.d.p.) condicional de X dado $Y = y$ denotada por $f_{X|Y=y}(x)$, o *valor esperado condicional* de X dado $Y = y$ é dado por

$$E(X|Y = y) = \begin{cases} \sum_{x_i} x_i f_{X|Y=y}(x_i) & \text{(caso discreto)}, \\ \int_{I\!R} x f_{X|Y=y}(x)\, dx & \text{(caso contínuo)} \end{cases}$$

enquanto a variância condicional de X dado $Y = y$ é dada por

$$\begin{aligned} Var(X|Y=y) &\equiv E\left[(X - E(X|Y = y))^2 | Y=y\right] \\ &= \begin{cases} \sum_{x_i}(x_i - E(X|Y=y))^2 f_{X|Y=y}(x_i) & \text{(discreto)}, \\ \int_{I\!R}(x - E(X|Y=y))^2 f_{X|Y=y}(x)dx & \text{(contínuo)}. \end{cases} \end{aligned}$$

Nota: Se X e Y são independentes, $E(X|Y=y)=E(X)$ e $Var(X|Y=y)=Var(X)$, $\forall\, y$.

Propriedades do valor esperado condicional: Se (X, Y) um par aleatório com f.m.p. (f.d.p.) conjunta $f_{X,Y}(x, y)$ e marginal de Y $f_Y(y)$,

$$E(E(X|Y)) = E(X), \quad \text{caso } E(X) < \infty.$$

Demonstração (par aleatório contínuo): Denotando $g(y)$ como o valor esperado condicional de X dado $Y=y$ com $f_Y(y) > 0$, *i.e.*,

$$g(y) = E(X|Y = y) = \int_{I\!R} x f_{X|Y=y}(x)dx = \int_{I\!R} x \frac{f_{X,Y}(x, y)}{f_Y(y)} dx,$$

$$\begin{aligned} E(E(X|Y)) &= E(g(Y)) = \textstyle\int_{I\!R} g(y) f_Y(y)dy \\ &= \textstyle\int_{I\!R} \left[\int_{I\!R} x \frac{f_{X,Y}(x,y)}{f_Y(y)} dx \right] f_Y(y)dy \\ &= \textstyle\int_{I\!R} x \left[\int_{I\!R} f_{X,Y}(x, y)dy \right] dx \\ &= \textstyle\int_{I\!R} x f_X(x)dx \\ &= E(X) \end{aligned}$$

Analogamente $E(E(Y|X)) = E(Y)$, caso $E(Y) < \infty$.

Distribuição multinomial*

Definição 5.6: Considere uma experiência aleatória constituída de n ensaios independentes em cada um dos quais pode ocorrer um de k acontecimentos mutuamente exclusivos tal que p_i é a probabilidade do acontecimento i em cada ensaio com $\sum_{i=1}^{k} p_i = 1$. Seja X_i a v.a. que designa o número de ensaios em que o acontecimento i ocorre na experiência, $i=1,\ldots,k$, pelo que $\sum_{i=1}^{k} X_i = n$. O vector aleatório $\mathbf{X} = (X_1,\ldots,X_k)$ tem distribuição *multinomial* (de dimensão $k-1$) com parâmetros n e $\mathbf{p} = (p_1,\ldots,p_k)$, definida pela f.m.p. conjunta

$$f_{\mathbf{X}}(\mathbf{x}) = \begin{cases} \frac{n!}{x_1!x_2!\ldots x_k!} {p_1}^{x_1} {p_2}^{x_2} \ldots {p_k}^{x_k}, & x_i \in \{0,1,\ldots,n\}; \\ 0, & \text{caso contrário} \end{cases}$$

com $p_k = 1-p_1 \ldots -p_{k-1}$ e $x_k = n-x_1 \ldots -x_{k-1}$.

Pode-se provar que $X_i \sim$ Binomial(n, p_i), $E(X_i) = np_i$, $Var(X_i) = np_i(1-p_i)$ e $Cov(X_i, X_j) = -np_ip_j$, $j \neq i = 1,\ldots,k$.

Relações entre distribuições

Teorema 5.8: Se X é uma v.a. com distribuição binomial de parâmetros n e p, X tem aproximadamente distribuição de Poisson com parâmetro $\lambda = np$, quando $n \to \infty$ e $p \to 0$.

$$\begin{aligned} f_X(x) &= \tbinom{n}{x} p^x (1-p)^{n-x} \\ &= \tfrac{n!}{x!(n-x)!} (\tfrac{\lambda}{n})^x (1-\tfrac{\lambda}{n})^{n-x}, \quad \text{onde } p = \lambda/n \\ &= \tfrac{n(n-1)\ldots(n-x+1)}{n^x} \times \tfrac{\lambda^x}{x!} \times \tfrac{(1-\lambda/n)^n}{(1-\lambda/n)^x} \\ &\to 1 \times \tfrac{\lambda^x}{x!} \times \tfrac{\mathrm{e}^{-\lambda}}{1}, \quad \text{quando } n \to \infty, p \to 0 \\ &= \tfrac{\mathrm{e}^{-\lambda}\lambda^x}{x!}. \end{aligned}$$

Ou seja, $X \overset{a}{\sim}$ Poisson$(\lambda = np)$, para n grande e p pequeno. Esta relação determina a característica fundamental do Processo de Poisson.

Teorema 5.9: Se X é uma v.a. com distribuição hipergeométrica de parâmetros N, M e n, X tem aproximadamente distribuição binomial com parâmetros n (fixo) e $p = M/N$, quando $N, M \to \infty$ tal que p constante.

$$\begin{aligned} f_X(x) &= [\tbinom{M}{x}\tbinom{N-M}{n-x}]/\tbinom{N}{n} \\ &= \tfrac{M!}{x!(M-x)!} \times \tfrac{(N-M)!}{(n-x)!(N-M-n+x)!} \times \tfrac{n!(N-n)!}{N!} \\ &= \tfrac{n!}{x!(n-x)!} \times \tfrac{M!}{(M-x)!} \times \tfrac{(N-M)!}{(N-M-(n-x))!} \times \tfrac{(N-n)!}{N!} \\ &= \tbinom{n}{x} \times \tfrac{M(M-1)\ldots(M-x+1)}{N(N-1)\ldots(N-x+1)} \times \tfrac{(N-M)(N-M-1)\ldots(N-M-(n-x)+1)}{(N-x)(N-x-1)\ldots(N-n+1)} \\ &\to \tbinom{n}{x} \times [\tfrac{M}{N}\,\tfrac{M}{N}\cdots\tfrac{M}{N}] \times [\tfrac{N-M}{N}\,\tfrac{N-M}{N}\cdots\tfrac{N-M}{N}], M, N \to \infty \\ &= \tbinom{n}{x} (\tfrac{M}{N})^x (\tfrac{N-M}{N})^{n-x} \\ &= \tbinom{n}{x} p^x (1-p)^{n-x}, \quad x = 0,1,\ldots,n \;\text{ e }\; p = M/N. \end{aligned}$$

Ou seja, $X \overset{a}{\sim}$ Binomial$(n, p = \frac{M}{N})$, para N muito maior do que n.

Convergência em distribuição

Os Teoremas 5.8

e 5.9

são aplicação do conceito de convergência em função de probabilidade. Outro conceito de convergência é:

Definição 5.7: Sejam $X, X_1, X_2 \ldots$ v.a. com respectivas funções de distribuição $F_X, F_{X_1}, F_{X_2}, \ldots$. Diz-se que a sucessão $\{X_n\}$ converge em distribuição para X ($X_n \stackrel{\mathcal{D}}{\to} X$), se

$$F_{X_n}(x) \to F_X(x), \text{quando } n \to \infty,$$

$\forall\, x$ ponto de continuidade de F_X. Ou seja,

$$\forall x, \lim_{n\to\infty} F_{X_n}(x) = F_X(x)$$

$\Leftrightarrow \forall \delta > 0, \exists n_1(\delta) : n > n_1(\delta) \Rightarrow |F_{X_n}(x) - F_X(x)| < \delta, \forall x.$

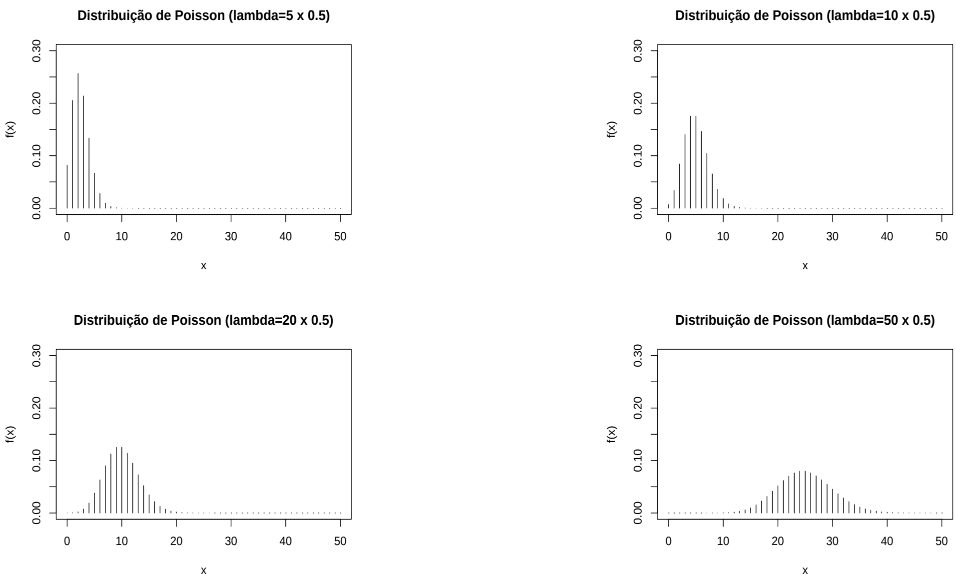
Teorema Limite Central

Teorema 5.10 (T.L.C.): Seja $X_1, X_2 \ldots$ uma sucessão de v.a. independentes e identicamente distribuídas com valor esperado μ e variância σ^2, ambos finitos. Para $S_n = \sum_{i=1}^{n} X_i$, tem-se

$$\frac{S_n - E(S_n)}{\sqrt{Var(S_n)}} \;=\; \frac{S_n - n\mu}{\sigma\sqrt{n}} \quad \stackrel{\mathcal{D}}{\to} \quad N(0,1).$$

Ou seja, para n razoavelmente grande, $P\big(\frac{S_n - n\mu}{\sigma\sqrt{n}} \leq x\big) \approx \Phi(x)$, onde $\Phi(\cdot)$ é a função de distribuição da normal reduzida, *i.e.*, $N(0,1)$. Assim, $S_n \stackrel{a}{\sim} N(n\mu, n\sigma^2)$ para n suficientemente grande.

Aplicação à distribuição Poisson: $X_i \sim$ Poisson(λ), $i = 1, 2, \ldots \Rightarrow S_n \sim$ Poisson($n\lambda$) $\stackrel{a}{\sim} N(n\lambda, n\lambda)$ (a seguir).



Exemplo 5.8: Suponha que X_i é o tempo de atendimento (em minutos) do cliente i num caixa de banco. Considere ainda que X_i, $i=1,\ldots,n$, são v.a. independentes com distribuição Uniforme$(0,5)$. Havendo 60 clientes no momento da abertura do banco às 9 horas, qual a probabilidade de o caixa do banco atender todos os clientes até às 12 horas?

Se $X_i \sim$ Uniforme$(0,5)$, $i=1,\ldots,60$, e $S_{60} = \sum_{i=1}^{60} X_i$, então

- $E(X_i) = \frac{(5+0)}{2} = 2.5 \;\Rightarrow\; E(S_{60}) = 60 \times 2.5 = 150m.$
- $Var(X_i) = \frac{(5-0)^2}{12} = \frac{25}{12} \;\Rightarrow\; Var(S_{60}) = 60 \times \frac{25}{12} = 125m^2.$

Como $n = 60$ (grande) e X_i são v.a. independentes e identicamente distribuídas, pode-se usar o T.L.C. ($S_{60} \approx N(150,125)$), *i.e.*,

$$\begin{aligned} P(S_{60} \leq 180) &\approx P\left(\frac{S_{60}-E(S_{60})}{\sqrt{Var(S_{60})}} \leq \frac{180-150}{\sqrt{125}}\right) \\ &= F_{N(0,1)}(2.68) \\ &= 0.9963. \end{aligned}$$

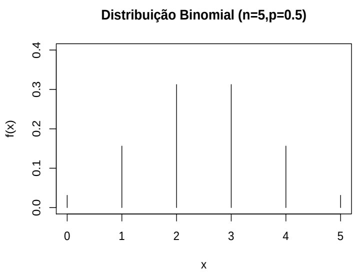
Aplicação à distribuição binomial

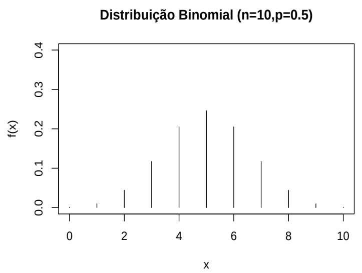
Corolário 5.3 (Teorema de DeMoivre-Laplace): Seja $X_1, X_2 \ldots$ uma sucessão de v.a. $X_i \sim$ Bernoulli(p), $i=1,\ldots,n$, independentes com valor esperado $\mu = E(X_i) = p$ e variância $\sigma^2 = Var(X_i) = p(1-p)$, onde $p = P(X_i = 1) \in (0,1)$. Para $S_n = \sum_{i=1}^{n} X_i$,

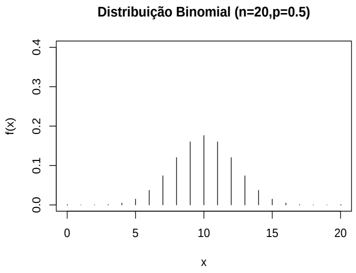
$$\frac{S_n - np}{\sqrt{np(1-p)}} \quad \overset{\mathcal{D}}{\to} \quad N(0,1).$$

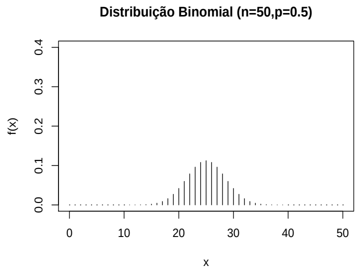
Consequentemente, $P(a < S_n \leq b) = F_{Bi(n,p)}(b) - F_{Bi(n,p)}(a) \approx$

$$\begin{cases} F_{N(0,1)}\left(\frac{b-np}{\sqrt{np(1-p)}}\right) - F_{N(0,1)}\left(\frac{a-np}{\sqrt{np(1-p)}}\right), & n \text{ bastande grande} \\ F_{N(0,1)}\left(\frac{b-np+0.5}{\sqrt{np(1-p)}}\right) - F_{N(0,1)}\left(\frac{a-np+0.5}{\sqrt{np(1-p)}}\right), & n \text{ moderadamente grande.} \end{cases}$$









Exemplo 5.9: Se $X = \sum_{i=1}^{n} X_i \sim$ Binomial$(n = 12, p = 0.5)$, qual a probabilidade de X ser pelo menos 7?

$$P(X \geq 7) = \sum_{x=7}^{12} \binom{12}{x} 0.5^x (1 - 0.5)^{12-x} = 0.3872.$$

$$\begin{aligned} P(X \geq 7) & \approx 1 - F_{N(0,1)}\left(\tfrac{7-12\times 0.5}{\sqrt{12\times 0.5\times 0.5}}\right) = 1 - F_{N(0,1)}(0.58) \\ & = 1-0.719 = 0.281 \text{ (sem correcção de continuidade)} \end{aligned}$$

$$\begin{aligned} P(X \geq 7) & \approx 1 - F_{N(0,1)}\left(\tfrac{7-12\times 0.5-1/2}{\sqrt{12\times 0.5\times 0.5}}\right) = 1 - F_{N(0,1)}(0.29) \\ & = 1-0.6141 = 0.3859 \text{ (com correcção de continuidade)} \end{aligned}$$

A aproximação $N(np, np(1-p))$ para a distribuição Binomial é tanto melhor quanto maior for n e tanto mais intermédios forem os valores de p $(0.1 < p < 0.9)$. Para valores de p pequenos, a melhor aproximação é a distribuição Poisson (np) que, por sua vez, é aproximável pela distribuição $N(np, np)$ para valores bastante grandes de n.

6 Amostragem e estimação pontual

Diferenciação entre Teoria da Probabilidade e Estatística

Exemplo 6.1: Uma escola está situada junto a uma rodovia com grande intensidade de tráfego. Num processo de monitorização de efeitos perniciosos da poluição ambiental na população escolar mediu-se a concentração de chumbo (Pb), expressa em $ng = 10^{-9}\, g/ml$, na corrente sanguínea de 50 crianças seleccionadas ao acaso.

- População: conjunto Ω de crianças da referida escola.

Fases de um trabalho estatístico:

1. Recolha dos dados estatísticos (Amostragem)

- Amostra: Subconjunto de crianças seleccionadas de Ω por algum processo de amostragem que envolva aleatoriedade, $(\omega_1, \ldots, \omega_n)$, $n = 50$.

- Variável aleatória associada à população (ou abreviadamente população): X= concentração de Pb no sangue de uma criança em ng/ml.
- Amostra numérica de interesse: $(x_1, \ldots, x_n)$ em que $x_i = X(\omega_i) \equiv X_i(\omega_1, \ldots, \omega_n)$, $i = 1, \ldots, n$ com X_i traduzindo X nos i-ésimos elementos de todas as amostras susceptíveis de serem obtidas.

2. Descrição da amostra (Estatística Descritiva)

- Classificação da amostra: traçado de gráficos (e.g., histograma de frequências relativas/acumuladas e correspondente polígono).
- Condensação da amostra: Cálculo de momentos ou outras quantidades empíricas (e.g., quartis) com eventual representação gráfica.

Média da amostra:

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i \equiv \frac{1}{n}\sum_{i=1}^{k} n_i x_i = 10.12\, ng/ml.$$

Variância corrigida da amostra:

$$s_{n-1}^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2 \equiv \frac{1}{n-1}\sum_{i=1}^{k} n_i (x_i - \bar{x})^2 = 0.41\, ng^2/ml$$

$$\to s_{n-1} = 0.64\, ng/ml.$$

3. Inferência Estatística (Estatística Indutiva)

- Ponto de vista probabilístico: conhecimento completo da distribuição de X, e.g., $X \sim N(\mu, \sigma^2)$ com μ e σ especificados

  $\Rightarrow$ $P(X > x_c) = 1 - \Phi(\frac{x_c - \mu}{\sigma})$

- Ponto de vista estatístico: Distribuição de X total ou parcialmente desconhecida

  $\Rightarrow$ Necessidade de induzir os aspectos desconhecidos do modelo a partir dos dados, com medição do respectivo grau de incerteza (feita probabilisticamente).

a) Forma distribucional de X conhecida (e.g., Normal) mas parâmetros desconhecidos

  $\Rightarrow$ Estimação pontual desses parâmetros (e.g., pelos momentos empíricos correspondentes) e estimação por intervalos para medição concomitante da precisão da estimação.

  $\Rightarrow$ Indagar a evidência nos dados a favor ou contra certas conjecturas prévias sobre os parâmetros através de testes de hipóteses paramétricas.

b) Distribuição de X completamente desconhecida

  $\Rightarrow$ Inspecção do histograma para formulação de alguma conjectura distribucional (caso ainda não haja nenhuma) a ser testada via testes de ajustamento.

4. Predição

- Com a monitorização ao longo do tempo do tráfego e perante uma eventual intensificação deste, que consequências incidem sobre $f_X(x)$? A forma desta mantém-se? Que medidas devem ser tomadas em face de um potencial aumento de $E(X)$ ou de $P(X > x_c)$? Deslocalização da escola ou construção de desvios rodoviários?

Amostra aleatória (acepção restrita)

A definição de cada amostra numérica de dimensão n remete para um vector aleatório $(X_1, \ldots, X_n)$ cujas características distribucionais dependem do processo de amostragem casual adoptado.

Definição 6.1: Dada uma população a que está associada uma variável aleatória X com uma certa distribuição de probabilidade, uma *amostra aleatória* (a.a.) de tamanho n dessa população é uma sequência de n variáveis aleatórias $X_1 \ldots, X_n$ independentes e identicamente distribuídas (i.i.d.) com a mesma distribuição de X.

Definição 6.2: Dada uma amostra aleatória $(X_1 \ldots, X_n)$ de uma população X com f.m.p. (f.d.p.) $f_X(x)$, a *distribuição de probabilidade amostral* (f.m.p. ou f.d.p. conjunta) é dada por

$$f_{X_1,\ldots,X_n}(x_1, \ldots, x_n) = \prod_{i=1}^{n} f_{X_i}(x_i) = \prod_{i=1}^{n} f_X(x_i).$$

Exemplo 6.2: Uma a.a. de dimensão n de uma população em que se inquire intenções de voto num dado partido reporta-se a n v.a. $X_1 \ldots, X_n$ i.i.d., tal que

$$X_i = \begin{cases} 1, & \text{se o eleitor } i \text{ tenciona votar no partido;} \\ 0, & \text{c.c.} \end{cases}$$

Sendo $p = P(X_i\!=\!1) = 1 - P(X_i\!=\!0)$, $i\!=\!1, \ldots, n$, a respectiva distribuição de probabilidade amostral é dada por

$$f_{X_1,\ldots,X_n}(x_1, \ldots, x_n) = \prod_{i=1}^{n} p^{x_i}(1-p)^{1-x_i} = p^{\sum_i x_i}(1-p)^{n-\sum_i x_i}.$$

Estatísticas

Definição 6.3: Dada uma amostra $(X_1 \ldots, X_n)$ de uma população X, uma *estatística* T é uma variável aleatória (vector aleatório) função da amostra, *i.e.*,

$$T = T(X_1, \ldots, X_n).$$

Exemplos de estatísticas:

- Média amostral: $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$.
- Variância amostral (corrigida): $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$.

- Mínimo amostral: $X_{(1)} = \min(X_1, \ldots, X_n)$.
- Máximo amostral: $X_{(n)} = \max(X_1, \ldots, X_n)$.
- Amplitude amostral: $R = X_{(n)} - X_{(1)}$.

Definição 6.4: Um *parâmetro* é uma medida usada para descrever uma característica da população.

Notação usual de parâmetros, estatísticas e valores observados destas:

		Amostra	
Característica	População	aleatória	concreta
média	μ	$\bar{X}$	$\bar{x}$
variância	σ^2	S^2	s^2
número de elementos	N	n	n
proporção	p	$\bar{X}$	$\bar{x}$

Se $(X_1 \ldots, X_n)$ é uma a.a. de uma população X, então

- média populacional: $\mu = E(X)$,
- média amostral: $\bar{X} = (X_1 + \cdots + X_n)/n$.

Estimação pontual: estimador e estimativa

Definição 6.5: Seja $(X_1 \ldots, X_n)$ uma amostra aleatória de uma população X indexada pelo parâmetro θ. Um *estimador* de θ é uma estatística $T = T(X_1, \ldots, X_n)$ usada para estimar θ.

Definição 6.6: O valor observado de um estimador em cada amostra concreta $t = T(x_1, \ldots, x_n)$ é conhecido por *estimativa*.

Exemplo 6.2a: Numa amostra aleatória de $n = 100000$ eleitores, observaram-se 38900 eleitores com intenção de voto no partido em causa. Neste cenário, $X_1, \ldots, X_n$ são v.a. i.i.d. com distribuição de Bernoulli (p), onde p é a proporção (populacional) de votantes no partido. O parâmetro p pode ser estimado pela média amostral $\bar{X}$, *i.e.*, a proporção amostral de votantes no partido, cujo estimativa é

$$\bar{x} = 38900/100000 = 0.389 \text{ ou } 38.9\%.$$

Propriedades dos estimadores

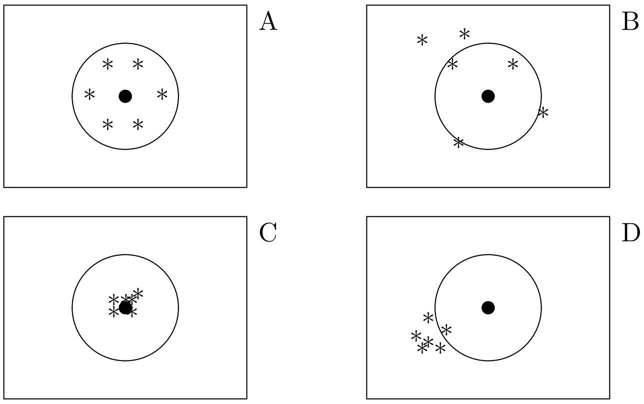
As propriedades básicas dos estimadores estão relacionadas com noções de exactidão e precisão à semelhança da caracterização dos métodos experimentais de medição de uma quantidade desconhecida em termos da concordância das medidas repetidas obtidas, em que se considera

$$\boxed{\textit{Exactidão} = \text{concordância das observações com o valor visado.}}$$

$$\boxed{\textit{Precisão} = \text{concordância das observações entre si.}}$$

A exactidão (*accuracy*) está associada aos erros sistemáticos, *e.g.*, deficiências de instrumentos de medição, enquanto a precisão (*precision*) se reporta aos erros aleatórios que são responsáveis por pequenas variações imprevisíveis nas medições realizadas, cujas causas não são completamente conhecidas.

Exemplo 6.3: Ilustração (informal) de jogadores de tiro ao alvo ("estimadores") com boa exactidão (A,C) e boa precisão (C,D).



Definição 6.7: Seja $(X_1 \ldots, X_n)$ uma a.a. de X com distribuição indexada pelo parâmetro θ. O estimador $T = T(X_1, \ldots, X_n)$ é dito ser um estimador *centrado* (não enviesado) de θ se $E(T) = \theta$.

Exemplo 6.4: Seja $(X_1 \ldots, X_n)$ uma a.a. de X com $E(X) = \mu$ e $Var(X) = \sigma^2$. Será $\sum_{i=1}^{n}(X_i - \bar{X})^2$ um estimador centrado de σ^2?

Se $X_1, \ldots, X_n$ são v.a. i.i.d. com $E(X_i) = \mu$ e $Var(X_i) = \sigma^2$, $i = 1, \ldots, n$, então $E(\bar{X}) = \mu$ e $Var(\bar{X}) = \sigma^2/n$. Logo,

$$\begin{aligned} E(\textstyle\sum_{i=1}^{n}(X_i - \bar{X})^2) &= E(\textstyle\sum_i X_i^2 - 2\bar{X}\sum_i X_i + n\bar{X}^2) \\ &= n\left[E(X^2) - E(\bar{X}^2)\right] \\ &= n\left[(\sigma^2 + \mu^2) - (\sigma^2/n + \mu^2)\right] \\ &= (n-1)\sigma^2 \end{aligned}$$

$\therefore$ Não, mas $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$ é um estimador centrado de σ^2.

Definição 6.8: Seja $T = T(X_1, \ldots, X_n)$ um estimador do parâmetro θ. Chama-se *viés* (enviesamento) de T como estimador de θ ao valor médio do erro de estimação, $E(T - \theta) = E(T) - \theta$. Note-se que o viés é nulo se e somente se T é um estimador centrado de θ.

Definição 6.9: Seja $T = T(X_1, \ldots, X_n)$ um estimador do parâmetro θ. Uma medida da variabilidade do estimador T é o *erro quadrático médio* (EQM), dado por

$$EQM(T) \equiv E((T - \theta)^2) = Var(T) + (E(T) - \theta)^2.$$

Definição 6.10: Sejam $T = T(X_1, \ldots, X_n)$ e $U = U(X_1, \ldots, X_n)$ dois estimadores do parâmetro θ. Diz-se que T é mais *eficiente* do que U, se

$$EQM(T) \leq EQM(U), \ \forall\ \theta$$

com desigualdade estrita para algum θ.

Se T e U são estimadores centrados do parâmetro θ, então T é mais eficiente do que U se $Var(T) \leq Var(U)$, $\forall\ \theta$ com desigualdade estrita para algum θ.

Exemplo 6.5: Seja $(X_1 \ldots, X_n)$ uma a.a. de $X \sim$ Bernoulli(p). Considere ainda X_1 e $\bar{X}$ como dois estimadores de p. Qual dos dois é o estimador mais eficiente?

Sendo X_i's v.a. i.i.d. Bernoulli (p), $\sum_{i=1}^n X_i \sim$ Binomial (n, p),

- $E(X_1) = p$ e
  $E(\bar{X}) = n^{-1}E(\sum_{i=1}^n X_i) = n^{-1}n\,p = p$.
  $\therefore\ X_1$ e $\bar{X}$ são estimadores centrados de p.
- $Var(X_1) = p(1-p)$ e
  $Var(\bar{X}) = n^{-2}\,Var(\sum_{i=1}^n X_i) = n^{-1}\,p\,(1-p)$
  $\Rightarrow\ \frac{Var(\bar{X})}{Var(X_1)} = \frac{1}{n} < 1,\ \forall\ n > 1.$
  $\therefore\ \bar{X}$ é mais eficiente do que X_1 na estimação de p.

Exemplo 6.6: Seja $(X_1 \ldots, X_n)$ uma a.a. de uma população X Normal com $E(X) = \mu$ e $Var(X) = \sigma^2$. Será a variância amostral (corrigida) $S^2 = (n-1)^{-1}\sum_{i=1}^n (X_i - \bar{X})^2$ mais eficiente do que $\hat{\sigma}^2 = n^{-1}\sum_{i=1}^n (X_i - \bar{X})^2$ na estimação de σ^2?

Como

- $E(\sum_{i=1}^n (X_i - \bar{X})^2) = (n-1)\sigma^2$,
  $\Rightarrow\ E(S^2) = \sigma^2$ e $E(\hat{\sigma}^2) = \frac{n-1}{n}\sigma^2$.
- $Var(\sum_{i=1}^n (X_i - \bar{X})^2) = 2(n-1)\sigma^4$,
- $EQM(S^2) = Var(S^2) + (E(S^2) - \sigma^2)^2 = 2(n-1)^{-1}\sigma^4$,
- $EQM(\hat{\sigma}^2) = Var(\hat{\sigma}^2) + (E(\hat{\sigma}^2) - \sigma^2)^2 = (2n-1)n^{-2}\sigma^4$,
  $\Rightarrow\ \frac{EQM(S^2)}{EQM(\hat{\sigma}^2)} = \frac{2n^2}{(n-1)(2n-1)} > 1,\ \forall\ n > 1.$
  $\therefore\ \hat{\sigma}^2$ é mais eficiente do que S^2 $(n > 1)$ na estimação de σ^2.

Definição 6.11: Seja $(X_1 \ldots, X_n)$ uma a.a. de uma população X indexada pelo parâmetro θ. Uma sucessão $\{T_n\}$ de estimadores de θ é *consistente* se $\lim_{n\to\infty} P(|T_n - \theta| > \epsilon) = 0$, $\forall\,\epsilon > 0$, o que é garantido por

$$\text{i) } \lim_{n\to\infty} E(T_n) = \theta, \qquad \text{ii) } \lim_{n\to\infty} Var(T_n) = 0.$$

Exemplo 6.7: Seja $(X_1 \ldots, X_n)$ uma a.a. de $X \sim$ Bernoulli(p). Será $\bar{X}$ um estimador consistente de p?

Sendo X_i's v.a. i.i.d. Bernoulli (p), $\sum_{i=1}^n X_i \sim$ Binomial (n, p),

- $E(\bar{X}) = E(\sum_{i=1}^n X_i)/n = p$. $\bar{X}$ é um estimador centrado de p. Condição i) logicamente satisfeita.

- $Var(\bar{X}) = Var(\sum_{i=1}^{n} X_i)/n^2 = p\,(1-p)/n$. Por conseguinte, $\lim_{n\to\infty} Var(\bar{X}) = \lim_{n\to\infty} \frac{p(1-p)}{n} = 0$. Condição ii) satisfeita.

Portanto, $\bar{X}$ é um estimador consistente de p.

Método da máxima verosimilhança

Definição 6.12: Dada uma a.a. $(X_1 \ldots, X_n)$ de uma população X com f.m.p. ou f.d.p. $f_X(x|\theta)$ indexada pelo parâmetro (desconhecido) θ, a função de *verosimilhança* de θ relativa à amostra $(x_1, \ldots, x_n)$, denotada por $L(\theta|x_1, \ldots, x_n)$, é a função de θ que é numericamente idêntica à distribuição de probabilidade amostral avaliada em $(x_1, \ldots, x_n)$, *i.e.*,

$$L(\theta|x_1, \ldots, x_n) \equiv f_{X_1,\ldots,X_n}(x_1, \ldots, x_n|\theta) = \prod_{i=1}^{n} f_X(x_i|\theta).$$

O método de máxima verosimilhança consiste em maximizar a função de verosimilhança para obter o valor dito mais verosímil de θ, denominado estimativa de máxima verosimilhança de θ.

Ao determinar o valor que maximiza $L(\theta|x_1, \ldots, x_n)$, usa-se frequentemente o facto de que $L(\theta|x_1, \ldots, x_n)$ e $\log L(\theta|x_1, \ldots, x_n)$ têm o seu máximo no mesmo valor de θ.

Exemplo 6.8: Seja $(X_1 \ldots, X_n)$ uma a.a. de uma população $X \sim$ Poisson(λ). Qual o estimador de máxima verosimilhança (EMV) de λ?

A função de verosimilhança de λ, dado $(x_1, \ldots, x_n)$, é

$$L(\lambda|x_1, \ldots, x_n) = \prod_{i=1}^{n} \frac{\mathrm{e}^{-\lambda} \lambda^{x_i}}{x_i!}.$$

Seja $L_\lambda \equiv \log L(\lambda|x_1, \ldots, x_n) = -n\,\lambda + \log\lambda \sum_{i=1}^{n} x_i - \log \prod_{i=1}^{n} x_i!$.

- $\frac{dL_\lambda}{d\lambda} = -n + \lambda^{-1} \sum_{i=1}^{n} x_i = 0 \quad \Rightarrow \; \lambda = \frac{1}{n} \sum_{i=1}^{n} x_i = \bar{x}$
- $\frac{d^2 L_\lambda}{d\lambda^2} = -\lambda^{-2} \sum_{i=1}^{n} x_i < 0, \forall\lambda$.

$\therefore$ $\bar{x}$ é a estimativa de máxima verosimilhança de λ e o EMV de λ é

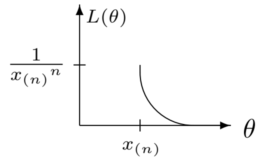
$$\hat{\lambda} = \bar{X} = \frac{1}{n} \sum_{i=1}^{n} X_i.$$

Teorema 6.1: Se $\hat{\theta}$ é o estimador de máxima verosimilhança de um parâmetro θ, então $g(\hat{\theta})$ é o estimador de máxima verosimilhança de $g(\theta)$ (*propriedade de invariância*).

Exemplo 6.9: Seja $(X_1 \ldots, X_n)$ uma a.a. de $X \sim$ Uniforme$(0, \theta]$. Qual o estimador de máxima verosimilhança de $\log\theta$?

A função de verosimilhança de θ, dado $x_1, \ldots, x_n$, é

$$\begin{aligned} L(\theta|x_1,\ldots,x_n) &= \prod_{i=1}^{n} \tfrac{1}{\theta}\; I_{(0,\theta]}(x_i) \\ &= \tfrac{1}{\theta^n}\; I_{[x_{(n)},\infty)}(\theta) \end{aligned}$$



$\Rightarrow$ $X_{(n)} = \max(X_1,\ldots,X_n)$ é o EMV de θ.

$\therefore$ Pela propriedade de invariância dos estimadores de máxima verosimilhança, $\log X_{(n)}$ é o EMV de $\log\theta$.

Distribuições amostrais da média e variância

Para melhor avaliar a estimação de um parâmetro θ a partir de uma estatística $T = T(X_1,\ldots,X_n)$, deve-se conhecer a distribuição de T.

A distribuição da estatística T, conhecida como distribuição amostral de T, tem em conta todos os valores possíveis da amostra $(X_1\ldots,X_n)$.

Exemplo 6.10: Suponha uma população com v.a. X de distribuição uniforme em $\{2,4,6\}$ da qual se retira (com reposição) uma amostra de tamanho 2. Qual o valor esperado da média e da variância amostrais?

Como os elementos da população X são equiprováveis,

- $E(X) = \sum_x x f_X(x) = \frac{1}{3}(2+4+6) = 4.$
- $E(X^2) = \sum_x x^2 f_X(x) = \frac{1}{3}(4+16+36) = 56/3$

  $\Rightarrow$ $Var(X) = E(X^2) - (E(X))^2 = 56/3 - 16 = 8/3.$

Seja X_i o resultado da extracção i, $i=1,\ldots,n$ $(n=2)$. Recorde-se que a média e a variância corrigida amostrais são, respectivamente,

$$\bar{X} = n^{-1}\textstyle\sum_{i=1}^{n} X_i \quad \text{e} \quad S^2 = (n-1)^{-1}\textstyle\sum_{i=1}^{n}(X_i-\bar{X})^2.$$

A distribuição de probabilidade conjunta de (X_1,X_2) é dada por

$X_1\backslash X_2$	2	4	6
2	1/9	1/9	1/9
4	1/9	1/9	1/9
6	1/9	1/9	1/9

Distribuição amostral da estatística $\bar{X} = \frac{1}{2}(X_1+X_2)$:

$\bar{X}$	2	3	4	5	6
$P(\bar{X}=\bar{x})$	1/9	2/9	3/9	2/9	1/9

$E(\bar{X}) = \sum_u u\,P(\bar{X}=u) = 2\times\frac{1}{9}+\cdots+6\times\frac{1}{9} = \frac{36}{9} = 4$
$\Rightarrow E(\bar{X}) = E(X).$

- $E(\bar{X}^2) = \sum_u u^2\, P(\bar{X}=u) = 4 \times \frac{1}{9} + \cdots + 36 \times \frac{1}{9} = \frac{156}{9}$
- $Var(\bar{X}) = E(\bar{X}^2) - (E(\bar{X})^2 = \frac{156}{9} - 16 = \frac{12}{9} = \frac{4}{3}$

  $\Rightarrow \; Var(\bar{X}) = Var(X)/n.$

Distribuição amostral da estatística $S^2 = \sum_{i=1}^{2}(X_i - \bar{X})^2 = \frac{(X_1 - X_2)^2}{2}$:

S^2	0	2	8
$P(S^2 = s^2)$	3/9	4/9	2/9

$E(S^2) = \sum_v v\, P(S^2 = v) = 0 \times \frac{3}{9} + 2 \times \frac{4}{9} + 8 \times \frac{2}{9} = \frac{24}{9} = \frac{8}{3}$
$\Rightarrow \; E(S^2) = Var(X).$

Teorema 6.2: Se $(X_1 \ldots, X_n)$ é uma a.a. de uma população X com $E(X) = \mu$ e $Var(X) = \sigma^2$, então o valor esperado e variância da média amostral $\bar{X}$ são, respectivamente,

- $E(\bar{X}) = n^{-1} \sum_i E(X_i) = n^{-1}\, n\, \mu = \mu;$
- $Var(\bar{X}) = n^{-2} \sum_i Var(X_i) = n^{-2}\, n\, \sigma^2 = \sigma^2/n.$

Teorema 6.3: Se $(X_1 \ldots, X_n)$ é uma a.a. de uma população $X \sim N(\mu, \sigma^2)$, $0 < \sigma^2 < \infty$, então

$$Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \sim N(0, 1).$$

Teorema 6.4: Seja $(X_1 \ldots, X_n)$ uma a.a. de uma população X com $E(X) = \mu$ e $Var(X) = \sigma^2$, $0 < \sigma^2 < \infty$. Pelo Teorema do Limite Central, a distribuição amostral de $\bar{X}$ é aproximada pela distribuição Normal com média μ e variância σ^2/n, para n suficientemente grande.

Exemplo 6.11: Seja $(X_1 \ldots, X_n)$ uma a.a. de $X \sim$ Bernoulli(p). Qual a distribuição aproximada da proporção amostral $\bar{X} = n^{-1} \sum_{i=1}^{n} X_i$?

Sabendo que $E(X) = p$ e $Var(X) = p(1-p)$, pelo Teorema 6.3

$$\frac{\bar{X} - p}{\sqrt{p(1-p)/n}} \overset{a}{\sim} N(0,1) \quad \Rightarrow \quad \bar{X} \overset{a}{\sim} N\left(p, \frac{p(1-p)}{n}\right).$$

Distribuição qui-quadrado

Definição 6.13: Se $X_1, \ldots, X_k$ são v.a. i.i.d. com distribuição $N(0, 1)$,

$$Q = X_1^2 + \cdots + X_k^2$$

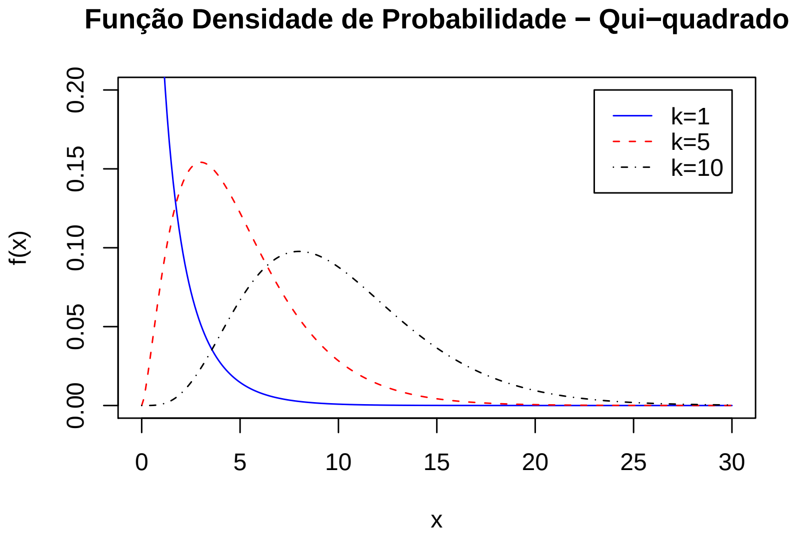
é dito ter uma distribuição *qui-quadrado* com k graus de liberdade, denotada por $\chi^2_{(k)}$, cuja f.d.p. é dada por

$$f_Q(q) = \frac{1}{2^{\frac{k}{2}}\, \Gamma(\frac{k}{2})} q^{\frac{k}{2}-1}\, \mathrm{e}^{-\frac{q}{2}}, \; q > 0,$$

onde $\Gamma(a) = \int_0^\infty x^{a-1} \mathrm{e}^{-x} dx$, $a > 0$.

O valor esperado e a variância de uma v.a. $Q \sim \chi^2_{(k)}$ são:

- $E(Q) = k$;
- $Var(Q) = 2\,k$.



Distribuição t-Student

Definição 6.14: Se Z e Q são v.a. independentes com $Z \sim N(0,1)$ e $Q \sim \chi^2_{(k)}$, então

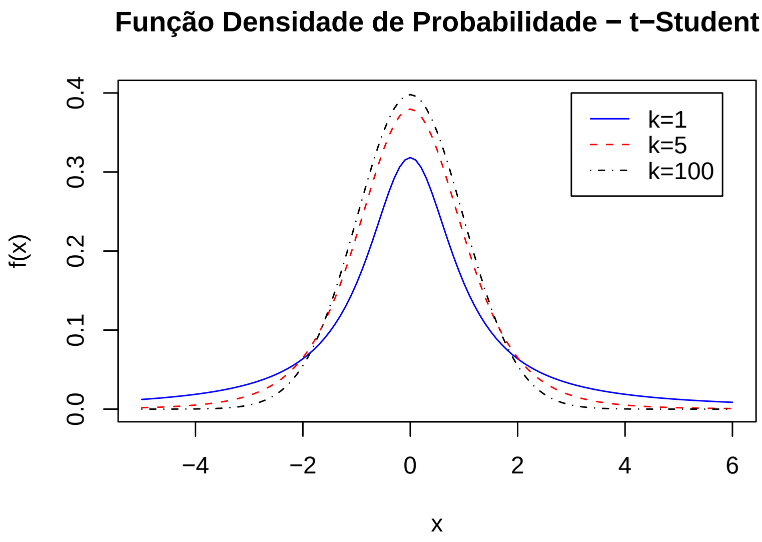
$$T = \frac{Z}{\sqrt{Q/k}}$$

é dito ter uma distribuição *t-Student* com k graus de liberdade, denotada por $t_{(k)}$, cuja f.d.p. é dada por

$$f_T(t) = \frac{1}{\sqrt{k\,\pi}}\,\frac{\Gamma(\frac{k+1}{2})}{\Gamma(\frac{k}{2})}\left(1+\frac{t^2}{k}\right)^{-(\frac{k+1}{2})}, \quad -\infty < t < \infty.$$

O valor esperado e a variância de uma v.a. $T \sim t_{(k)}$ são:

- $E(T) = 0$, $k > 1$.
- $Var(T) = k/(k-2)$, $k > 2$.



Teorema 6.5: Se $(X_1 \ldots, X_n)$ é uma a.a. de uma população $X \sim N(\mu, \sigma^2)$, então

$$\frac{\sum_{i=1}^{n}(X_i - \mu)^2}{\sigma^2} = \sum_{i=1}^{n}\left(\frac{X_i - \mu}{\sigma}\right)^2 \sim \chi^2_{(n)}$$

e

$$\frac{\sum_{i=1}^{n}(X_i - \bar{X})^2}{\sigma^2} = \frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{(n-1)}.$$

Teorema 6.6: Se $(X_1 \ldots, X_n)$ é uma a.a. de uma população $X \sim N(\mu, \sigma^2)$, então

$$\frac{(\bar{X} - \mu)/(\sigma/\sqrt{n})}{\sqrt{[(n-1)S^2/\sigma^2]/(n-1)}} = \frac{\bar{X} - \mu}{S/\sqrt{n}} \sim t_{(n-1)}.$$

7 Estimação por intervalos

Seja $(X_1, \ldots, X_n)$ uma amostra aleatória de uma população X indexada por um parâmetro θ. Por vezes, torna-se mais valioso especificar um intervalo que contém o verdadeiro valor de θ com um certo grau de confiança do que apenas estimar θ pontualmente.

Noções básicas

Definição 7.1: Seja $(X_1, \ldots, X_n)$ uma amostra aleatória de uma população X indexada por um parâmetro $\theta \in \Theta$. Se $T_i = T_i(X_1, \ldots, X_n)$, $i = 1, 2$, são duas estatísticas tais que $T_1 < T_2$ e

$$P(T_1 < \theta < T_2) = \gamma,$$

onde γ é um valor fixado entre 0 e 1, diz-se que (T_1, T_2) é um *intervalo aleatório de confiança* (IAC) para θ com grau de confiança γ.

Exemplo 7.1: Seja $(X_1, \ldots, X_n)$ uma a.a. de $X \sim N(\mu, \sigma^2 = 4)$. Qual o IAC para μ baseado no EMV com grau de confiança de 95%?

Sabe-se que $\bar{X}$ é o estimador de máxima verosimilhança de μ e que

$$\bar{X} \sim N(\mu, \sigma^2/n) \quad \Rightarrow \quad Z = \frac{\bar{X} - \mu}{\sqrt{4/n}} \sim N(0,1)$$

Por outro lado, $P(-1.96 < Z < 1.96) = 0.95$ e consequentemente

$$P(\bar{X} - 1.96\sqrt{4/n} < \mu < \bar{X} + 1.96\sqrt{4/n}) = 0.95,$$

indicando que o intervalo aleatório de confiança a 95% pretendido para μ é expresso por (T_1, T_2), em que

$$T_1 = \bar{X} - 1.96\sqrt{4/n} \quad \text{e} \quad T_2 = \bar{X} + 1.96\sqrt{4/n}.$$

Método pivotal

Definição 7.2: Seja $(X_1, \ldots, X_n)$ uma a.a. de uma população X indexada pelo parâmetro $\theta \in \Theta$. Diz-se que a função da a.a. e de θ,

$$W = W(X_1, \ldots, X_n; \theta)$$

é uma *variável fulcral* ou *quantidade pivotal* (ou simplesmente, *pivô*) usada na construção de um intervalo de confiança para θ quando a sua distribuição (f.m.p. ou f.d.p.) amostral não depende de θ.

Os intervalos de confiança são obtidos aqui pelo método da variável fulcral ou método pivotal. Isto é, dada uma variável fulcral $W = W(X_1, \ldots, X_n; \theta)$, os intervalos de confiança para θ são obtidos da seguinte forma:

1. Fixado um grau de confiança γ, obtêm-se a_γ e b_γ tais que

$$P(a_\gamma < W < b_\gamma) = \gamma, \ \forall \theta \in \Theta.$$

2. Se a partir de $a_\gamma < W < b_\gamma$, for possível obter uma desigualdade equivalente $T_1 < \theta < T_2$, onde $T_i = T_i(X_1, \ldots, X_n)$, $i = 1, 2$, tem-se

$$P(a_\gamma < W < b_\gamma) = P(T_1 < \theta < T_2) = \gamma, \ \forall \theta \in \Theta.$$

3. Dado uma amostra particular $(x_1, \ldots, x_n)$, a concretização do intervalo aleatório de confiança para θ com grau de confiança γ é designada por intervalo de confiança a $100\gamma\%$ para θ, dado por

$$(t_1, t_2) = (t_1(x_1, \ldots, x_n), t_2(x_1, \ldots, x_n)).$$

Nota: O intervalo acima é bilateral, sendo os respectivos intervalos de confiança unilaterais do tipo $(-\infty, u_2)$ ou (u_1, ∞), $u_1 < u_2$.

Em suma, o intervalo (t_1, t_2) é um intervalo de confiança para θ a $100\gamma\%$, denotado por

$$IC(\theta, \gamma) = (t_1, t_2),$$

sendo a concretização do intervalo aleatório de confiança, denotado por

$$IAC(\theta, \gamma) = (T_1, T_2).$$

A probabilidade γ é interpretada como a frequência relativa de todos os intervalos (t_1, t_2) que contêm θ obtidos numa sequência infinitamente grande de observações repetidas de $(X_1, \ldots, X_n)$ (perspectiva frequencista). Entretanto,

$$\gamma \neq P(t_1 < \theta < t_2) = \begin{cases} 1, & \text{se } \theta \in (t_1, t_2), \\ 0, & \text{c.c.} \end{cases}$$

Intervalos de confiança para parâmetros de populações normais

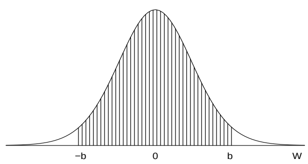
Seja $(X_1, \ldots, X_n)$ uma a.a. de uma população $X \sim N(\mu, \sigma^2)$.
Poder-se-á considerar $Z = \frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \sim N(0,1)$ como variável fulcral para obter um intervalo de confiança para μ? Não, se σ^2 for desconhecido.
Entretanto, sabe-se que $W = \frac{\bar{X}-\mu}{S/\sqrt{n}} \sim t_{(n-1)}$, onde S^2 é a variância amostral corrigida. Portanto,

$$P(a < W < b) = \gamma = 1-\alpha$$

$$P(\bar{X} - b\tfrac{S}{\sqrt{n}} < \mu < \bar{X} + b\tfrac{S}{\sqrt{n}}) = 1-\alpha$$

$$IAC(\mu, \gamma = 1-\alpha) = \left(\bar{X} \pm b\,\tfrac{S}{\sqrt{n}}\right)$$

$$IC(\mu, 1-\alpha) = \left(\bar{x} \ \pm \ b\,\tfrac{s}{\sqrt{n}}\right)$$



$$b = -a = F^{-1}_{t_{(n-1)}}(1 - \tfrac{\alpha}{2})$$

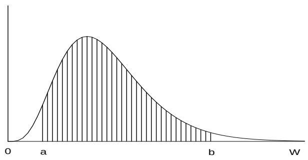
Exemplo 7.2: Suponha que o tempo X (em minutos) de reparação de uma máquina segue uma distribuição Normal com média μ e variância σ^2. Estime o tempo médio μ de reparação com um grau de confiança de 99%, baseando-se nos seguintes dados para uma amostra aleatória concretizada: $n = 10$, $\sum_{i=1}^n x_i = 846$ e $\sum_{i=1}^n x_i^2 = 71607$.
Nesse cenário, a variável fulcral é $W = \frac{\bar{X}-\mu}{S/\sqrt{n}} \sim t_{(9)}$ e as quantidades $\bar{x} = \frac{1}{n}\sum_{i=1}^n x_i = 84.6$ e $s^2 = \frac{1}{n-1}(\sum_{i=1}^n x_i^2 - n\,\bar{x}^2) = \frac{35.4}{9} = 3.933$.

- $P(-3.25 < W < 3.25) = 0.99$.
- $P(\bar{X} - 3.25\frac{S}{\sqrt{n}} < \mu < \bar{X} + 3.25\frac{S}{\sqrt{n}}) = 0.99$
- $IAC(\mu, \gamma = 0.99) = (\bar{X} \pm 3.25\frac{S}{\sqrt{n}})$
- $IC(\mu, 0.99) = (84.6 \pm 3.25\sqrt{\frac{3.933}{10}}) = (84.6 \pm 2.038)$
  $= (82.562, 86.638)$.

Seja $(X_1,\ldots,X_n)$ uma a.a. de uma população $X \sim N(\mu,\sigma^2)$. Na construção do intervalo de confiança para σ^2 com grau de confiança $\gamma = 1-\alpha$, tem-se como variável fulcral

$$W = \frac{(n-1)S^2}{\sigma^2} \;\sim\; \chi^2_{(n-1)}.$$

$P(a < W < b) = \gamma = 1-\alpha$, onde
$a = F^{-1}_{\chi^2_{(n-1)}}(\frac{\alpha}{2})$ e $b = F^{-1}_{\chi^2_{(n-1)}}(1-\frac{\alpha}{2})$
$P(\frac{(n-1)S^2}{b} < \sigma^2 < \frac{(n-1)S^2}{a}) = 1-\alpha$
$IAC(\sigma^2, 1-\alpha) = \ (\frac{(n-1)S^2}{b}, \frac{(n-1)S^2}{a})$



$$IC(\sigma^2, 1-\alpha) = \left(\frac{(n-1)s^2}{b}, \frac{(n-1)s^2}{a}\right)$$

Duas populações normais

Sejam $(X_{11},\ldots,X_{1n_1})$ e $(X_{21},\ldots,X_{2n_2})$ a.a. de duas populações independentes $X_1 \sim N(\mu_1,\sigma_1^2)$ e $X_2 \sim N(\mu_2,\sigma_2^2)$, respectivamente.

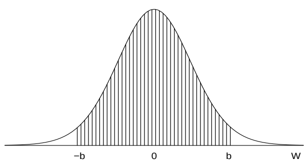
Se σ_1^2 e σ_2^2 são conhecidas, a variável fulcral para estimar $\mu_1-\mu_2$ com grau de confiança $\gamma = 1-\alpha$ tem como base os seguintes resultados:

- $\bar{X}_i \sim N(\mu_i, \sigma_i^2/n_i)$, $i = 1,2$ (independentes),
- $\bar{X}_1 - \bar{X}_2 \sim N(\mu_1-\mu_2, \sigma_1^2/n_1 + \sigma_2^2/n_2)$,

onde $\bar{X}_i = \frac{1}{n_i}\sum_{j=1}^{n_i} X_{ij}$, $i = 1,2$. Por conseguinte, a variável fulcral é dada por

$$W = \frac{\bar{X}_1 - \bar{X}_2 - (\mu_1-\mu_2)}{\sqrt{\sigma_1^2/n_1 + \sigma_2^2/n_2}} \sim N(0,1).$$

$P(a < W < b) = \gamma = 1-\alpha$,
$b = -a = F^{-1}_{N(0,1)}(1-\frac{\alpha}{2})$



- $P(\bar{X}_1 - \bar{X}_2 - b\sqrt{\sigma_1^2/n_1 + \sigma_2^2/n_2} < \mu_1 - \mu_2 <$

  $< \bar{X}_1 - \bar{X}_2 + b\sqrt{\sigma_1^2/n_1 + \sigma_2^2/n_2}) = 1-\alpha$
- $IAC(\mu_1-\mu_2, \gamma = 1-\alpha) = \ (\bar{X}_1 - \bar{X}_2 \pm b\sqrt{\sigma_1^2/n_1 + \sigma_2^2/n_2})$

Consequentemente,

$$IC(\mu_1-\mu_2, 1-\alpha) = \ (\bar{x}_1 - \bar{x}_2 \pm b\sqrt{\sigma_1^2/n_1 + \sigma_2^2/n_2})$$

Exemplo 7.3: Sejam $(X_{11},\ldots,X_{1n_1})$ e $(X_{21},\ldots,X_{2n_2})$ a.a. de duas populações independentes X_1 e X_2 com $E(X_j)=\mu_j$ e $\mathrm{Var}(X_j)=\sigma_j^2$, $j=1,2$, respectivamente. Deduza um intervalo de confiança para $\mu_1-\mu_2$ com grau de confiança $\gamma=1-\alpha$ para grandes amostras ($n_1,n_2\to\infty$).

Para grandes amostras, pode-se substituir as variâncias σ_1^2 e σ_2^2 pelas suas variâncias amostrais S_1^2 e S_2^2 (estimadores consistentes). Portanto, na construção de um intervalo de confiança (aproximado) para $\mu_1-\mu_2$ com grau de confiança γ, tem-se como variável fulcral

$$W=\frac{\bar{X}_1-\bar{X}_2-(\mu_1-\mu_2)}{\sqrt{S_1^2/n_1+S_2^2/n_2}}\ \overset{a}{\sim}\ N(0,1).$$

Consequentemente,

- $P(a<W<b)=\gamma=1-\alpha$, onde $b=-a\simeq F^{-1}_{N(0,1)}(1-\frac{\alpha}{2})$.
- $IAC(\mu_1-\mu_2,1-\alpha)\approx\ (\bar{X}_1-\bar{X}_2\pm b\sqrt{S_1^2/n_1+S_2^2/n_2})$.
- $IC(\mu_1-\mu_2,1-\alpha)\approx\ (\bar{x}_1-\bar{x}_2\pm b\sqrt{s_1^2/n_1+s_2^2/n_2})$.

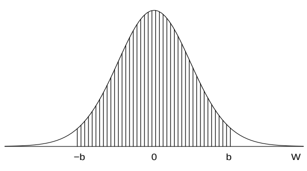
Sejam $(X_{11},\ldots,X_{1n_1})$ e $(X_{21},\ldots,X_{2n_2})$ a.a. de duas populações independentes $X_1\sim N(\mu_1,\sigma_1^2)$ e $X_2\sim N(\mu_2,\sigma_2^2)$, respectivamente.

Se $\sigma_1^2=\sigma_2^2=\sigma^2$ (desconhecida), a variável fulcral para estimar $\mu_1-\mu_2$ com grau de confiança $\gamma=1-\alpha$ tem como base os seguintes resultados:

- $Z=(\bar{X}_1-\bar{X}_2-(\mu_1-\mu_2))/\sqrt{\sigma^2(1/n_1+1/n_2)}\ \sim\ N(0,1)$.
- $((n_1-1)S_1^2+(n_2-1)S_2^2)/\sigma^2\equiv(n_1+n_2-2)S_c^2/\sigma^2\ \sim\ \chi^2_{(n_1+n_2-2)}$, onde $S_i^2=\frac{1}{n_i-1}\sum_{j=1}^{n_i}(X_{ij}-\bar{X}_i)^2$, $i=1,2$.
- Os pares $(\bar{X}_1,\bar{X}_2)$ e (S_1^2,S_2^2) são independentes.

$$\Rightarrow\quad W=\frac{\bar{X}_1-\bar{X}_2-(\mu_1-\mu_2)}{\sqrt{\frac{(n_1-1)S_1^2+(n_2-1)S_2^2}{n_1+n_2-2}(\frac{1}{n_1}+\frac{1}{n_2})}}\sim t_{(n_1+n_2-2)}.$$

$P(a<W<b)=\gamma=1-\alpha$,
$b=-a=F^{-1}_{t_{(n_1+n_2-2)}}(1-\frac{\alpha}{2})$



- $P(\bar{X}_1-\bar{X}_2-b\sqrt{\frac{(n_1-1)S_1^2+(n_2-1)S_2^2}{n_1+n_2-2}(\frac{1}{n_1}+\frac{1}{n_2})}<\mu_1-\mu_2<$

  $<\bar{X}_1-\bar{X}_2+b\sqrt{\frac{(n_1-1)S_1^2+(n_2-1)S_2^2}{n_1+n_2-2}(\frac{1}{n_1}+\frac{1}{n_2})})=1-\alpha$
- $IAC(\mu_1-\mu_2,1-\alpha)=\ \left(\bar{X}_1-\bar{X}_2\pm b\sqrt{\frac{(n_1-1)S_1^2+(n_2-1)S_2^2}{n_1+n_2-2}(\frac{1}{n_1}+\frac{1}{n_2})}\right)$

Consequentemente,

$$IC(\mu_1 – \mu_2, 1 – \alpha) = \left(\bar{x}_1 – \bar{x}_2 \pm b\sqrt{\frac{(n_1 – 1)s_1^2 + (n_2 – 1)s_2^2}{n_1 + n_2 - 2}\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}\right).$$

Exemplo 7.4: Sejam X_i o tempo de vida de uma bactéria do tipo i, $i = 1, 2$, independentes. Considere $(X_{11}, \ldots, X_{1n_1})$ e $(X_{21}, \ldots, X_{2n_2})$ duas a.a. de $X_1 \sim N(\mu_1, \sigma^2)$ e $X_2 \sim N(\mu_2, \sigma^2)$, respectivamente. Estime a diferença dos tempos de vida médios dos dois tipos de bactérias com 95% de grau de confiança, sabendo que $n_1 = 10$, $n_2 = 13$, $\sum_{j=1}^{n_1} x_{1j} = 300$, $\sum_{j=1}^{n_2} x_{2j} = 260$, $\sum_{j=1}^{n_1} x_{1j}^2 = 10000$ e $\sum_{j=1}^{n_2} x_{2j}^2 = 7000$.

Variável fulcral: $W = \frac{\bar{X}_1 – \bar{X}_2 – (\mu_1 - \mu_2)}{\sqrt{\frac{(n_1 – 1)S_1^2 + (n_2 – 1)S_2^2}{n_1 + n_2 - 2}(\frac{1}{n_1} + \frac{1}{n_2})}} \sim t_{(21)}$.

- $P(-2.08 < W < 2.08) = 0.95$, com $F^{-1}_{t_{(21)}}(0.975) = 2.08$.
- $S_c^2 = \frac{(n_1 – 1)S_1^2 + (n_2 – 1)S_2^2}{n_1 + n_2 - 2} \Rightarrow s_c^2 = \frac{9 \times 111.11 + 12 \times 150}{21} = 133.33$.
- $IAC(\mu_1 - \mu_2, 0.95) = (\bar{X}_1 – \bar{X}_2 \pm 2.08\sqrt{S_c^2(\frac{1}{n_1} + \frac{1}{n_2})})$
- $IC(\mu_1 – \mu_2, 0.95) = (10 \pm 2.08\sqrt{133.33(\frac{1}{10} + \frac{1}{13})}) = (-0.1, 20.1)$.

Intervalos de confiança para parâmetros de populações não normais uniparamétricas

Exemplo 7.5: Seja $(X_1, \ldots, X_n)$ uma a.a. de uma população $X \sim$ Exponencial com $E(X) = \lambda$. Encontre um intervalo de confiança aleatório a $100(1 – \alpha)\%$ para o logaritmo da média populacional λ.

Nesse cenário, sabe-se que

- O estimador MV de λ é $\bar{X}$ e $W = \frac{2}{\lambda}\sum_{i=1}^{n} X_i \sim \chi^2_{(2n)}$.
- $P(a < W < b) = 1 – \alpha$, onde $a = F^{-1}_{\chi^2_{(2n)}}(\frac{\alpha}{2})$ e $b = F^{-1}_{\chi^2_{(2n)}}(1 – \frac{\alpha}{2})$.
- $P\big(\frac{2\sum_{i=1}^{n} X_i}{b} < \lambda < \frac{2\sum_{i=1}^{n} X_i}{a}\big) = 1 – \alpha$.

Portanto, $IAC(\log \lambda, 1 – \alpha) = \left(\log(\frac{2\sum_{i=1}^{n} X_i}{b}), \log(\frac{2\sum_{i=1}^{n} X_i}{a})\right)$.

Para n grande, usa-se o pivô $Z = \frac{W - 2n}{2\sqrt{n}} \equiv \frac{\bar{X} - \lambda}{\sqrt{\lambda^2/n}} \overset{a}{\sim} N(0, 1)$.

Exemplo 7.6: Seja $(X_1, \ldots, X_n)$ uma a.a. de uma população $X \sim$ Bernoulli(p). Encontre um intervalo de confiança (aproximado) a $100(1 – \alpha)\%$ para a proporção populacional p.

Nesse cenário,

$$X_i = \begin{cases} 1, & \text{se ocorrer sucesso,} \\ 0, & \text{c.c.} \end{cases}, i = 1, \ldots, n.$$

- O estimador de máxima verosimilhança de p é $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$.

- $\sum_{i=1}^{n} X_i \sim$ Binomial(n,p) e, para grandes amostras, tem-se pelo T.L.C. que

$$W = \frac{\bar{X} - p}{\sqrt{p(1-p)/n}} \overset{a}{\sim} N(0,1).$$

- $P(-b < W < b) = 1-\alpha$, onde $a = F^{-1}_{N(0,1)}(\frac{\alpha}{2})$ e $b = F^{-1}_{N(0,1)}(1-\frac{\alpha}{2})$.

- Para encontrar os limites do intervalo de confiança de p, deve-se isolar o valor de p, encontrando as raízes do respectivo polinómio de 2º grau.
- Uma alternativa a este procedimento para n bem grande é usar o facto de que $\bar{X}$ é um estimador consistente de p e, para grandes amostras, pode-se substituir p no denominador de Z por $\bar{X}$.
  - $P\left(\bar{X} - b\sqrt{\frac{\bar{X}(1-\bar{X})}{n}} < p < \bar{X} + b\sqrt{\frac{\bar{X}(1-\bar{X})}{n}}\right) = 1-\alpha$.
  - $IAC(p, 1-\alpha) \approx \left(\bar{X} \pm b\sqrt{\frac{\bar{X}(1-\bar{X})}{n}}\right)$
  - Portanto, um intervalo de confiança (aproximado) a $100(1-\alpha)\%$ para p é dado por

$$IC(p, 1-\alpha) \approx \left(\bar{x} \pm b\sqrt{\frac{\bar{x}(1-\bar{x})}{n}}\right).$$

8 Testes de hipóteses

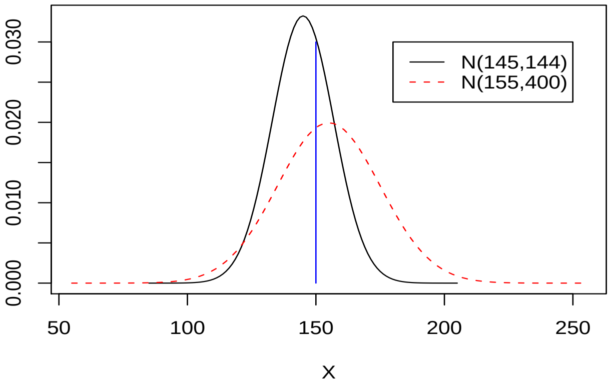
Uma outra forma de inferir sobre características de uma população é testar hipóteses previamente formuladas sobre os seus parâmetros, tendo em conta uma amostra aleatória da população e o valor tolerável para a probabilidade de rejeição incorrecta dessas hipóteses.

Exemplo 8.1: Uma empresa portuguesa compra usualmente parafusos americanos e japoneses devido às suas boas condições de resistência à tracção (X). Os americanos afirmam que a resistência à tracção dos seus parafusos tem média 145 kg e desvio padrão 12 kg, enquanto os japoneses dizem ter 155 kg de média e 20 kg de desvio padrão.

Um lote de parafusos será inspeccionado, desconhecendo-se a sua proveniência (americana ou japonesa). Com base numa amostra aleatória de 25 parafusos, calcula-se a resistência média à tracção ($\bar{x}$) a fim de investigar a origem dos mesmos.

Supondo distribuição normal para as duas populações, $N(145, 144)$ (americana) e $N(155, 400)$ (japonesa), pode-se considerar a seguinte regra de decisão:

- Se $\bar{x} \leq 150$, diz-se que os parafusos são de origem americana; caso contrário, são de procedência japonesa (Regra 1).



Na decisão do Exemplo 8.1, pode-se cometer dois tipos de erro:

- *Erro do tipo I:* Afirmar que os parafusos não são americanos (japoneses) quando na realidade o são.
- *Erro do tipo II:* Afirmar que os parafusos são americanos quando na realidade são japoneses.

As probabilidades destes dois tipos de erro (Exemplo 8.1) são:

$$\begin{aligned} \alpha &= P(\text{Erro tipo I}) = P(\bar{X} > 150 \mid \text{parafusos americanos}) \\ &= P(\bar{X} > 150 \mid X \sim N(145, 144)) = P(Z > 2.08) \\ &= 0.019 \end{aligned}$$

$$\begin{aligned} \beta &= P(\text{Erro tipo II}) = P(\bar{X} \leq 150 \mid \text{parafusos japoneses}) \\ &= P(\bar{X} \leq 150 \mid X \sim N(155, 400)) = P(Z \leq -1.25) \\ &= 0.106 \end{aligned}$$

Ao usar a regra de decisão do Exemplo 8.1 (Regra 1), a probabilidade do erro de tipo I é inferior à do erro de tipo II ($\alpha = 0.019 < \beta = 0.106$), "favorecendo" assim os parafusos americanos.

Cada regra de decisão deste tipo define um valor limítrofe para $\bar{x}$ (denotado aqui por $\bar{x}_c$). Por conseguinte, os valores de α e β variam consoante o valor fixado de $\bar{x}_c$.

- Se $\bar{x}_c < 150$, α aumenta e β diminui.
- Se $\bar{x}_c > 150$, α diminui e β aumenta.
- Se $\bar{x}_c = 148.75$, $\alpha = \beta = 0.059$ (ponto de equilíbrio).

Em suma, dada uma regra de decisão (*e.g.*, um valor para $\bar{x}_c$), pode-se calcular as duas probabilidades de erros para avaliar o teste. Outro procedimento possível é fixar a probabilidade de um tipo de erro e encontrar a correspondente regra de decisão. Por exemplo, $\alpha = 0.05$ implica $\bar{x}_c = 148.95$ e $\beta = 0.0651$, favorecendo a opção pelos parafusos americanos. Este é o esquema mais usado.

Noções básicas

Um teste de hipóteses paramétricas usualmente visa comparar diferentes valores para parâmetros de uma dada população X. Por exemplo, para o parâmetro desconhecido μ de $X \sim N(\mu, \sigma^2)$.

Procedimento geral de um teste de hipóteses paramétricas:

1. Hipóteses de interesse:
   - Hipótese nula H_0 (*e.g.*, $\mu = \mu_0$; $\mu \geq \mu_0$ ou $\mu \leq \mu_0$).
   - Hipótese alternativa H_1 (*e.g.*, $\mu \neq \mu_0 \rightarrow$ teste bilateral; $\mu < \mu_0$ ou $\mu > \mu_0 \rightarrow$ testes unilaterais).
2. Erros associados à regra de teste, cujas correspondentes probabilidades são dadas por
   - $\alpha = P(\text{Erro do tipo I}) = P(\text{Rejeitar } H_0 | H_0 \text{ verdadeira})$.
   - $\beta = P(\text{Erro do tipo II}) = P(\text{Aceitar } H_0 | H_0 \text{ falsa})$.

3. Região crítica (RC):
   - Região que conduz à rejeição da hipótese nula H_0 pela regra do teste. Construída com base numa estatística apropriada $T = T(X_1, \ldots, X_n)$ denominada estatística do teste.
   - A RC é construída tal que $P(T \in RC | H_0 \text{ verdadeira}) = \alpha$, com α (nível de significância) fixado previamente nos valores usuais de 1%, 5% e 10%. Esta RC será denotada por RC_α.
4. Regra do teste de hipótese:
   - Se $T \in RC_\alpha$, rejeita-se H_0 ao nível de significância de $100\alpha\%$. Caso contrário, não se rejeita H_0 a $100\alpha\%$.
   - Quanto menor for o nível de significância do teste, tanto maior será a precaução contra o risco de rejeição incorrecta de H_0.

A determinação de β exige a especificação de cada valor alternativo para o parâmetro em teste, dado que H_1 é geralmente composta (*e.g.*, $\beta(\mu) = P(T \notin RC | \mu \neq \mu_0)$). Identicamente, o nível α corresponde à probabilidade máxima do erro de tipo I, quando H_0 é composta.

A função $1-\beta(\mu)$ é conhecida por potência do teste para H_1 verdadeiro. Ou seja, para um dado valor de μ, a potência do teste é a probabilidade de rejeição de H_0 quando μ é o verdadeiro valor do parâmetro.

$$P(\text{Rejeitar } H_0 | \mu) = \begin{cases} \alpha(\mu), & H_0 \text{ verdadeiro} \\ 1 - \beta(\mu), & H_1 \text{ verdadeiro.} \end{cases}$$

Exemplo 8.1a: Considere o cenário do Exemplo 8.1 com $H_0 : \mu \leq 145$ (parafusos americanos) contra $H_1 : \mu > 145$ (parafusos não americanos) e regra de decisão 1 ($\bar{x}_c = 150$). A função potência do teste será

$$P(\text{Rejeitar } H_0 | \mu) = P(\bar{X} > 150 | \mu) = 1 - F_{N(0,1)}\left(\frac{150 - \mu}{12/\sqrt{25}}\right)$$

μ	143	145	147	150	153
$P(\text{Rejeitar}\, H_0 \vert \mu)$	0.002	0.019	0.106	0.5	0.894
	$\alpha(\mu)$			$1-\beta(\mu)$	

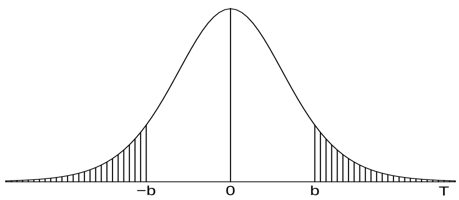
Testes de hipóteses para parâmetros de populações normais

Seja $(X_1 \ldots, X_n)$ uma a.a. de uma população $X \sim N(\mu, \sigma^2)$. Sabe-se que o estimador MV de μ é $\bar{X} \sim N(\mu, \sigma^2/n)$.

Teste de hipóteses para a média, supondo σ^2 conhecido:

1. Hipóteses:
   - $H_0 : \mu = \mu_0$
   - $H_1 : \mu \neq \mu_0$ (ou $H_1 : \mu > \mu_0$ ou $H_1 : \mu < \mu_0$)
2. Estatística do teste:
$$T = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} \overset{H_0}{\sim} N(0,1),$$
   cujo valor observado é denotado por t_0.

3. Região crítica bilateral: Fixado um valor para α,
$$RC_\alpha = \{t \in I\!R : (t < -b) \cup (t > b)\},$$
   onde $b = F^{-1}_{N(0,1)}(1 - \frac{\alpha}{2})$.



4. Conclusão:
   - Se $t_0 \in RC_\alpha$, os dados apontam contra H_0 ao nível de significância de $100\alpha\%$, pelo que esta deve ser rejeitada.
   - Caso contrário, não há evidência para rejeitar H_0 ao nível α.

Exemplo 8.2: Uma máquina foi regulada para encher pacotes de café de $500g$. Seja $x_1, \ldots, x_{16}$ concretização de uma a.a. de X (quantidade de café por pacote), cuja média é $492g$. Considerando que X segue uma distribuição normal com desvio padrão $20g$, teste a regulação da máquina ao nível de 1% de significância.

Teste de hipóteses:

1. Hipóteses: $H_0 : \mu = 500$ versus $H_1 : \mu \neq 500$

2. Estatística do teste: $T = \frac{\bar{X}-500}{\sigma/\sqrt{n}} \overset{H_0}{\sim} N(0,1)$, cujo valor observado é $t_0 = \frac{492-500}{\sqrt{400/16}} = -1.6$.

3. Região crítica bilateral: Fixado $\alpha = 0.01$, $F^{-1}_{N(0,1)}(0.995) = 2.58$ e $RC_{1\%} = (-\infty, -2.58) \cup (2.58, \infty)$.

4. Conclusão: Como $t_0 \notin RC_{1\%}$, não se rejeita H_0 ao nível de significância de 1%, *i.e.*, não há evidência contra a regulação da máquina a esse nível.

Exemplo 8.2a: Seja $X_1, \ldots, X_{16}$ uma a.a. de X (quantidade de café por pacote) em que a média e variância empíricas de uma sua concretização são $480g$ e $800g^2$, respectivamente. Considerando uma distribuição normal para X, teste se a máquina está a encher pacotes de café com pelo menos $500g$, ao nível de 5% de significância.

Teste de hipóteses:

1. Hipóteses: $H_0 : \mu \geq 500$ versus $H_1 : \mu < 500$.

2. Estatística do teste: $T = \frac{\bar{X}-500}{S/\sqrt{n}} \overset{\mu=500}{\sim} t_{(15)}$, cujo valor observado é $t_0 = (480-500)/\sqrt{800/16} = -2.83$.

3. Região crítica unilateral: Fixado $\alpha = 0.05$, $F^{-1}_{t_{(15)}}(0.95) = 1.753$ e $RC_{5\%} = (-\infty, -1.753)$, dado que valores decrescentes de T tendem a reflectir valores mais pequenos de μ.

4. Conclusão: Como $t_0 \in RC_{5\%}$, rejeita-se H_0 ao nível de significância de 5%, *i.e.*, há evidência contra a hipótese de enchimento de pacotes de café com pelo menos $500g$.

Exemplo 8.2b: No teste de hipóteses $H_0 : \mu \geq 500 \equiv \mu_0$ versus $H_1 : \mu < 500$, a decisão do teste varia com a escolha de α, *i.e.*,

α	RC_α	decisão do teste
0.05	$(-\infty, -1.753)$	rejeita-se H_0
0.01	$(-\infty, -2.602)$	rejeita-se H_0
0.006	$(-\infty, -2.857)$	não se rejeita H_0

Note-se que o menor valor do nível de significância α que conduz à rejeição de H_0 é $P = P(T < -2.83|H_0) = 0.0063$.

Valor-P do teste

Definição 8.1: O *valor-P* de um teste de hipóteses é a probabilidade sob H_0 de a estatística do teste tomar valores tão ou mais desfavoráveis a H_0 do que o seu valor observado. Deste modo, H_0 será rejeitado a todo nível de significância α tal que $P < \alpha$ e aceite no caso contrário.

Testes de hipóteses para a variância de uma população normal

Seja $(X_1 \ldots, X_n)$ uma a.a. de uma população $X \sim N(\mu, \sigma^2)$. Sabe-se que $\frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{(n-1)}$.

Teste de hipóteses para a variância:

1. Hipóteses:

   - $H_0 : \sigma^2 = \sigma_0^2$

- $H_1 : \sigma^2 \neq \sigma_0^2$ (ou $H_1 : \sigma^2 > \sigma_0^2$ ou $H_1 : \sigma^2 < \sigma_0^2$).

2. Estatística do teste:

$$T = \frac{(n-1)S^2}{\sigma_0^2} \overset{H_0}{\sim} \chi^2_{(n-1)},$$

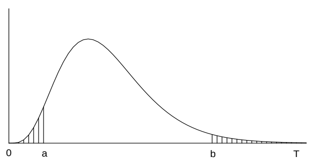
com valor observado t_0 e $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$.

Nota: Se μ for conhecido, $T = \frac{\sum_{i=1}^{n}(X_i-\mu)^2}{\sigma_0^2} \overset{H_0}{\sim} \chi^2_{(n)}$.

3. Região crítica bilateral: Fixado um valor para α,

$$RC_\alpha = \{t \in I\!R^+ : (t < a) \cup (t > b)\},$$

onde $a = F^{-1}_{\chi^2_{(n-1)}}(\frac{\alpha}{2})$ e $b = F^{-1}_{\chi^2_{(n-1)}}(1-\frac{\alpha}{2})$.



4. Conclusão:

- Se $t_0 \in RC_\alpha$, i.e. $P \equiv 2min\{P(T \geq t_0|H_0), P(T \leq t_0|H_0)\} < \alpha$, os dados tendem a desmentir H_0 ao n.s. de $100\alpha\%$, pelo que esta deve ser rejeitada.
- Caso contrário, não há evidência contra H_0 ao nível α.

Exemplo 8.3: Seja $(X_1, \ldots, X_{10})$ uma a.a. de X (tensão de ruptura de um material), resumida em $\sum_{i=1}^{10} x_i = 900$ e $\sum_{i=1}^{10} x_i^2 = 81108$. Considerando $X \sim N(\mu, \sigma^2)$, teste se $\sigma^2 = 10$.

Teste de hipóteses:

1. Hipóteses: $H_0 : \sigma^2 = 10$ versus $H_1 : \sigma^2 \neq 10$.
2. Estatística do teste: $T = \frac{(n-1)S^2}{10} \overset{\sigma^2=10}{\sim} \chi^2_{(9)}$, cujo valor observado é $t_0 = \frac{9\times 12}{10} = 10.8$.
3. Valor-P: $P = 2\ \min(P(T \geq 10.8|H_0), P(T < 10.8|H_0)) = 0.58$. Note-se que $F^{-1}_{\chi^2_{(9)}}(0.71) = 10.8$, $F^{-1}_{\chi^2_{(9)}}(0.7) = 10.66$ e $F^{-1}_{\chi^2_{(9)}}(0.8) = 12.24$.
4. Conclusão: Rejeita-se H_0 para $\alpha \geq 0.58$ e não se rejeita H_0 para $\alpha < 0.58$. Ou seja, há forte evidência a favor da hipótese $\sigma^2 = 10$ contra a alternativa bilateral aos níveis de significância usuais $\alpha = 1\%, 5\%, 10\%$.

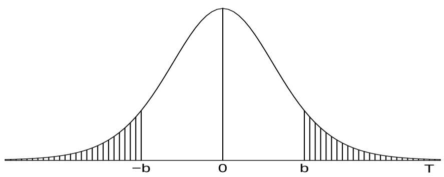
Duas populações normais

Sejam $(X_{11},\ldots,X_{1n_1})$ e $(X_{21},\ldots,X_{2n_2})$ amostras aleatórias de duas populações independentes $X_1 \sim N(\mu_1,\sigma_1^2)$ e $X_2 \sim N(\mu_2,\sigma_2^2)$, respectivamente. Sabe-se que o estimador MV de $\mu_1-\mu_2$ é $\bar{X}_1-\bar{X}_2 \sim N(\mu_1-\mu_2, \sigma_1^2/n_1+\sigma_2^2/n_2)$.

Teste de hipóteses para a diferença de médias, com σ_1^2 e σ_2^2 conhecidos:

1. Hipóteses:
   - $H_0 : \mu_1 = \mu_2 \;\Leftrightarrow\; \mu_1-\mu_2 = 0$
   - $H_1 : \mu_1 \neq \mu_2$ (ou $H_1 : \mu_1 > \mu_2$ ou $H_1 : \mu_1 < \mu_2$)
2. Estatística do teste:
$$T = \frac{\bar{X}_1-\bar{X}_2}{\sqrt{\sigma_1^2/n_1+\sigma_2^2/n_2}} \overset{\mu_1=\mu_2}{\sim} N(0,1),$$
   com valor observado que se denota por t_0.

3. Região crítica bilateral: Fixado um valor para α,
$$RC_\alpha = \{t \in I\!R : (t < -b) \cup (t > b)\},$$
   onde $b = F^{-1}_{N(0,1)}(1-\frac{\alpha}{2})$.



4. Conclusão:
   - Se $t_0 \in RC_\alpha$, rejeita-se H_0 ao nível de significância de $100\alpha\%$. Caso contrário, aceita-se H_0 ao nível α.
   - Alternativamente, calcula-se $P = P(|T| \geq |t_0|\,|H_0) = 2P(T \geq |t_0||H_0)$ e confronta-se com α do modo indicado.

Exemplo 8.4: Para testar a resistência de dois tipos de viga de aço (A e B), observou-se a resistência de algumas dessas vigas de aço, obtendo os seguintes resultados de duas amostras, uma de cada tipo:

tipo A	$n_1 = 15$	$\bar{x}_1 = 70.5$	$s_1^2 = 81.6$
tipo B	$n_2 = 10$	$\bar{x}_2 = 84.3$	$s_2^2 = 85.7$

Supondo que as amostras (aleatórias) são provenientes de duas populações normais independentes $X_1 \sim N(\mu_1,\sigma_1^2)$ e $X_2 \sim N(\mu_2,\sigma_2^2)$, teste a igualdade das resistências médias (populacionais) dos dois tipos de viga de aço.

Teste de hipóteses, supondo $\sigma_1^2 = \sigma_2^2 = \sigma^2$ desconhecido:

1. Hipóteses:

   - $H_0 : \mu_1 = \mu_2 \;\Leftrightarrow\; \mu_1 - \mu_2 = 0$
   - $H_1 : \mu_1 \neq \mu_2$.

2. Estatística do teste:

$$T = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{(n_1-1)S_1^2+(n_2-1)S_2^2}{n_1+n_2-2}\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}} \overset{H_0}{\sim} t_{(23)},$$

   cujo valor observado é $t_0 = \frac{70.5-84.3}{\sqrt{(83.204)(1/15+1/10)}} = -3.71$.

3. Valor-P:

$$P = P(|T| \geq 3.71|H_0) = 2(1-F_{t_{(23)}}(3.71)) = 0.0012.$$

4. Conclusão:

   - Rejeita-se H_0 para $\alpha \geq 0.0012$.
   - Aceita-se H_0 para $\alpha < 0.0012$.

   Há forte evidência contra a hipótese de igualdade entre as resistências médias dos dois tipos de viga de aço.

Testes de hipóteses para parâmetros de populações não normais uniparamétricas

Seja $(X_1 \ldots, X_n)$ uma a.a. de uma população X tal que, por exemplo, $X \sim \text{Poisson}(\lambda)$. Pelo T.L.C. ($n$ grande), $\bar{X} \overset{a}{\sim} N(\lambda, \lambda/n)$.

Teste de hipóteses (em grandes amostras):

1. Hipóteses:

   - $H_0 : \lambda = \lambda_0$
   - $H_1 : \lambda \neq \lambda_0$ (ou $H_1 : \lambda > \lambda_0$ ou $H_1 : \lambda < \lambda_0$).

2. Estatística do teste:

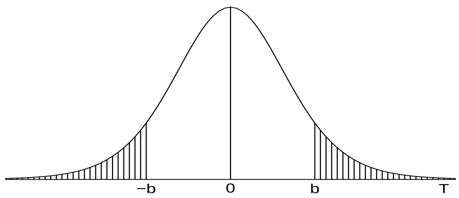
$$T = \frac{\bar{X} - \lambda_0}{\sqrt{\lambda_0/n}} \underset{a}{\overset{H_0}{\sim}} N(0,1),$$

   cujo valor observado é t_0.

3. Região crítica bilateral: Fixado um valor para α,

$$RC_\alpha = \{t \in I\!R : (t < a) \cup (t > b)\},$$

   onde $b = -a = F^{-1}_{N(0,1)}(1-\frac{\alpha}{2})$.



4. Conclusão:
   - Se $t_0 \in RC_\alpha$, rejeita-se H_0 ao nível de significância de $100\alpha\%$.
   - Caso contrário, não há evidência contra H_0 ao nível α.

Exemplo 8.5: Uma estação de TV afirma que no mínimo 60% dos telespectadores devem assistir ao seu programa especial de passagem de ano. A fim de avaliar esta afirmação, 200 famílias foram inquiridas constituindo as suas respostas uma suposta a.a. $X_1, \ldots, X_{200}$ de uma população Bernoulli(p), tendo-se verificado 104 respostas afirmativas.

Teste de hipóteses para uma proporção (em grandes amostras):

1. Hipóteses: $H_0 : p \geq 0.6$ versus $H_1 : p < 0.6$.
2. Estatística do teste: Pelo T.L.C., $T = \frac{\bar{X}-0.6}{\sqrt{0.6(1-0.6)/n}} \overset{p=0.6}{\underset{a}{\sim}} N(0,1)$, cujo valor observado é $t_0 = (0.52-0.6)/\sqrt{0.6\times 0.4/200} = -2.31$.
3. Valor-P: $P = P(T \leq -2.31|H_0) = 0.0104$, pela associação entre valores pequenos de $\bar{X}$ e de p.
4. Conclusão:
   - Rejeita-se H_0 para $\alpha \geq 0.0104$ e aceita-se H_0 de outro modo.

   Ou seja, há alguma evidência contra a afirmação da estação de TV.

Teste de ajustamento do qui-quadrado de Pearson

Até ao momento os testes estatísticos têm-se baseado na suposição de uma distribuição populacional conhecida (exacta ou aproximadamente) que raramente é legítima. Interessa, pois, saber como se podem testar hipóteses sobre a própria forma distribucional de uma dada população, objecto dos chamados testes de ajustamento.

Construção da estatística do teste do qui-quadrado de Pearson:

1. Considere uma amostra aleatória de n elementos sobre os quais se observa uma característica X, sendo as respectivas observações classificadas numa partição da recta real, $B_1, \ldots, B_k$, de modo que O_i denota o número de elementos da amostra agrupados em B_i, $i=1, \ldots, k$, tal que $\sum_{i=1}^k O_i = n$.
2. Seja $p_i = P(X \in B_i)$ a probabilidade de obter uma observação na i-ésima parte da partição, $i=1, \ldots, k$, tal que $\sum_{i=1}^k p_i = 1$.

3. O vector aleatório $\mathbf{O} = (O_1, \ldots, O_k)$ tem f.m.p. dada por

$$f_{\mathbf{O}}(o_1, \ldots, o_k) = \frac{n!}{o_1! \ldots o_k!}\, p_1^{o_1}\, p_2^{o_2} \cdots p_k^{o_k},$$

conhecida por distribuição Multinomial $(n, \mathbf{p} = (p_1, \ldots, p_k))$, podendo-se provar que $O_i \sim$ Binomial(n, p_i), $i = 1, \ldots, k$.

4. Hipóteses:
   - $H_0 : X \sim F_X(\cdot) \;\Rightarrow\; p_i = p_i^0,\ \forall\, i = 1, \ldots, k.$
   - $H_1 : X \nsim F_X(\cdot) \;\Rightarrow\; p_i \neq p_i^0$, para algum $i = 1, \ldots, k$.

5. Estatística do teste:

$$Q = \sum_{i=1}^{k} \frac{(O_i - E_i)^2}{E_i} \quad \underset{a}{\overset{H_0}{\sim}} \quad \chi^2_{(k-m-1)},$$

onde $E_i = E(O_i|H_0) = n\, p_i^0$ e m é o total de parâmetros estimados de $F_X(\cdot)$ sob H_0. Se $m > 0$, $\{p_i^0\}$ são ainda desconhecidos implicando que E_i sejam estimadores (apropriados) das frequências esperadas.

Exemplo 8.6: Acredita-se que o número X de acidentes por semana numa dada estrada segue uma distribuição Poisson. Para testar esta crença observou-se o número de acidentes nessa estrada durante 30 semanas, cujos resultados encontram-se a seguir. Teste a suposição de uma lei Poisson para X ao nível de significância de 5%.

8	4	1	1	3	0	0	0	8	9	2	4	7	1	3
3	1	0	2	0	3	4	2	1	12	5	0	5	4	2

Teste de hipóteses:

1. Hipóteses:
   - $H_0 : X \sim$ Poisson(λ) versus $H_1 : X \nsim$ Poisson(λ).

2. Estatística do teste:

$$Q = \sum_{i=1}^{k} \frac{(O_i - E_i)^2}{E_i} \quad \underset{a}{\overset{H_0}{\sim}} \quad \chi^2_{(k-m-1)},$$

onde $E_i = 30\ p_i^0$ com λ estimado segundo MV por $\bar{x} = 95/30 = 3.167$ $(m = 1)$. Se $X \sim$ Poisson(3.167), $P(X=0) = 0.0421$, $P(X=1) = 0.1334$, $P(X=2) = 0.2113$, $P(X=3) = 0.223$, $P(X=4) = 0.1766$, $P(X=5) = 0.1118$ e $P(X \geq 6) = 0.1018$.

i	B_i	$\hat{p}_i^0 = P(X \in B_i \vert H_0)$	E_i
1	$[0, 1]$	$P(X = 0) + P(X = 1) = 0.1755$	5.265
2	$(1, 2]$	$P(X = 2) = 0.2113$	6.339
3	$(2, 3]$	$P(X = 3) = 0.2230$	6.690
4	$(3, 4]$	$P(X = 4) = 0.1766$	5.298
5	$(4, \infty)$	$P(X = 5) + P(X \geq 6) = 0.2136$	6.408

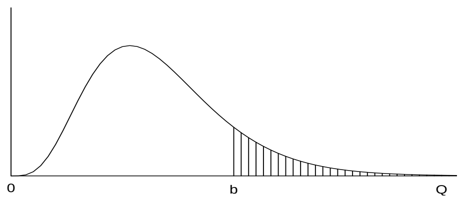
O valor observado da estatística do teste é

$$\begin{aligned} q_0 &= \tfrac{(11-5.265)^2}{5.265}+\tfrac{(4-6.339)^2}{6.339}+\tfrac{(4-6.69)^2}{6.69}+\tfrac{(4-5.298)^2}{5.298}+\tfrac{(7-6.408)^2}{6.408} \\ &= 6.247+0.863+1.081+0.318+0.055=8.564 \end{aligned}$$

3. Região crítica: Fixado $\alpha = 0.05$,

$$RC_{5\%} = \{q \in I\!R^+ : (q > b)\},$$

onde $b = F^{-1}_{\chi^2_{(3)}}(0.95) = 7.815$.



4. Conclusão:

   - Como $q_0 = 8.56 \in RC_{5\%}$, rejeita-se H_0 ao nível de significância de 5%. Ou seja, não há evidência a favor da hipótese de $X \sim$ Poisson(λ).

Notas:

1. O valor-P do teste de ajustamento no Exemplo 8.6 é

$$P = P(Q \geq 8.56|H_0) = 0.0357.$$

   Note-se que $F^{-1}_{\chi^2_{(3)}}(0.95) = 7.815$ e $F^{-1}_{\chi^2_{(3)}}(0.975) = 9.348$.

   Consequentemente, rejeita-se H_0 para $\alpha \geq 0.0357$ e aceita-se H_0 para $\alpha < 0.0357$. Ou seja, não há forte evidência nem contra nem a favor H_0.

2. Os valores E_i na estatística do teste do qui-quadrado devem ser

   - Todos maiores ou iguais a 1.
   - Pelo menos 80% deles devem ser no mínimo 5.

   Caso contrário, deve-se fazer reagrupamento de classes.

3. Caso seja necessário estimar m parâmetros no cálculo dos E_i, deve-se retirar m graus de liberdade da distribuição à estatística do teste do qui-quadrado de Pearson ($\chi^2_{(k-m-1)}$).

Teste de independência do qui-quadrado de Pearson em tabelas de contingência

Suponha que cada um dos n elementos amostrados de uma população pode ser classificado de acordo com duas características X e Y, com r e s categorias, respectivamente.

Seja $p_{ij} = P(X=i, Y=j)$ a probabilidade (conjunta) de um elemento da população pertencer a categoria (i,j) de (X,Y), $i=1,\ldots,r$, $j=1,\ldots,s$.

Consequentemente, as probabilidades (marginais) das duas características são dadas por

$$\begin{aligned} p_{i\bullet} &= P(X=i) = \textstyle\sum_{j=1}^{s} P(X=i, Y=j) \\ p_{\bullet j} &= P(Y=j) = \textstyle\sum_{i=1}^{r} P(X=i, Y=j). \end{aligned}$$

Para avaliar a independência das duas características pode-se construir um teste de hipótese com base na estatística do qui-quadrado de Pearson, efectuando as seguintes adaptações:

- A partição B_i, $i=1,\ldots,rs$ passa a ser denotada como a partição B_{ij}, $i=1,\ldots,r$, $j=1,\ldots,s$ das celas da dita tabela de contingência.
- $p_{ij} = P(B_{ij})$ é a probabilidade de obter uma observação na (i,j)-ésima parte da partição, tal que $\sum_{i=1}^{r}\sum_{j=1}^{s} p_{ij} = 1$.
- O_{ij} denota o número de elementos da amostra que pertence a B_{ij}, tal que $\sum_{i=1}^{r}\sum_{j=1}^{s} O_{ij} = N$.
- A estatística do teste é expressável por

$$Q = \sum_{i=1}^{r}\sum_{j=1}^{s} \frac{(O_{ij} - E_{ij})^2}{E_{ij}} \quad \overset{H_0}{\underset{a}{\sim}} \quad \chi^2_{(r-1)(s-1)},$$

  onde $E_{ij} = E(O_{ij}|H_0) = N\, p_{ij}^0$, como se indica a seguir.

- A hipótese de independência entre X e Y é

$$H_0 : p_{ij} = p_{i\,\bullet} \times p_{\bullet\, j} \equiv p_{ij}^0, \forall\, i, j.$$

- Sob H_0, as frequências esperadas das celas são

$$n\, p_{ij}^0 = n\, p_{i\bullet} \times p_{\bullet j}.$$

  Como $p_{i\bullet}$ e $p_{\bullet j}$ não são especificadas sob H_0, a sua estimação por MV conduz a $\hat{p}_{i\bullet} = O_{i\bullet}/n$ e $\hat{p}_{\bullet j} = O_{\bullet j}/n$, sendo então

$$n\hat{p}_{i,j}^0 = E_{ij} \;=\; (O_{i\bullet} \times O_{\bullet j})/n,$$

  onde $O_{i\bullet} = \sum_{j=1}^{s} O_{ij}$ e $O_{\bullet j} = \sum_{i=1}^{r} O_{ij}$, as frequências esperadas estimadas.

- Note-se que o vector aleatório $\mathbf{O} = (O_{11},\ldots,O_{rs})$ segue uma distribuição Multinomial $(n, \mathbf{p} = (p_{11},\ldots,p_{rs}))$ e que cada $O_{ij} \sim$ Binomial(n, p_{ij}), $i=1,\ldots,r$, $j=1,\ldots,s$.

Exemplo 8.7: Um fabricante de automóveis suspeita que a venda dos seus três últimos modelos está relacionada com o género dos seus compradores. Com base na seguinte tabela de contingência envolvendo 500 compradores, teste a hipótese de independência entre o tipo dos modelos de automóveis e o género dos compradores.

género	modelo 1	modelo 2	modelo 3	total
masculino	160	140	40	340
feminino	40	60	60	160
total	200	200	100	500

Teste de hipóteses:

1. Hipóteses:

   - $H_0 : p_{ij} = p_{i\bullet} \times p_{\bullet j}, \forall\, i, j$
   - $H_1 : p_{ij} \neq p_{i\bullet} \times p_{\bullet j}$, para algum i, j.

2. Estatística do teste:

$$Q = \sum_{i=1}^{2}\sum_{j=1}^{3} \frac{(O_{ij} - E_{ij})^2}{E_{ij}} \quad \underset{a}{\overset{H_0}{\sim}} \quad \chi^2_{(2)},$$

   onde $E_{ij} = E(O_{ij}|H_0) = n\,p^0_{ij}$ e $n\hat{p}^0_{ij} = E_{ij} = O_{i\bullet} \times O_{\bullet j}/n$. Por exemplo, $E_{11} = 200 \times \frac{340}{500} = 136$ e $E_{23} = 100 \times \frac{160}{500} = 32$.

   O valor observado da estatística (qui-quadrado de Pearson) é

$$q_0 = \frac{(160 - 136)^2}{136} + \cdots + \frac{(60 - 32)^2}{32} = 49.63.$$

3. Valor-P: $P = P(Q \geq 49.63|H_0) = 1.67 \times 10^{-11}$. Note-se que $F^{-1}_{\chi^2_{(2)}}(0.9995) = 15.2$.

4. Conclusão:

   - Rejeita-se H_0 para $\alpha \geq 1.67 \times 10^{-11}$ e aceita-se H_0 no caso contrário, *i.e.*, há forte evidência a favor da dependência entre o tipo dos modelos de automóveis e género dos compradores.

9 Introdução à regressão linear simples

Modelos de regressão

Há variáveis aleatórias que podem ser explicadas por acção conjunta de factores determinísticos e aleatórios. Por exemplo, o rendimento de um processo químico depende de um modo previsível da temperatura a que se realiza e da quantidade de catalisador usada, bem como de factores desconhecidos responsáveis pela variabilidade imprevisível dos resultados obtidos.

Ou seja, uma variável de interesse Y passa a ter uma componente determinística e outra aleatória. Supondo uma estrutura aditiva entre elas,

$$Y = g(x) + \epsilon,$$

onde $g(x)$ é a parte determinística de Y, formada por uma ou mais variáveis auxiliares x observáveis e ϵ é a sua parte aleatória.

A parte determinística de Y é considerada fixa, mesmo que dependa de parâmetros desconhecidos, enquanto a parte aleatória admite naturalmente uma distribuição de probabilidade. Nesse cenário, o conjunto de dados é formado por n pares (y_i, x_i), $i = 1, \ldots, n$, com os x_i supostamente especificados sem erro.

Considerando uma amostra casual (Y_i, x_i), $i = 1, \ldots, n$, um modelo estatístico para relacionar Y e x é dado por

$$Y_i = \beta_0 + \beta_1 \, x_i + \epsilon_i,$$

onde Y_i é a variável resposta do i-ésimo elemento da amostra, x_i é o correspondente valor da variável explicativa (fixa), β_0 e β_1 são parâmetros (desconhecidos) e ϵ_i é o erro aleatório do elemento i da amostra.

O modelo acima é conhecido por modelo de regressão linear simples, com parte determinística $g(x) = \beta_0 + \beta_1 \, x$ e parte aleatória ϵ, cuja distribuição de probabilidade se supõe habitualmente ser Normal.

Suposições usuais para os erros aleatórios ϵ_i, $i = 1, \ldots, n$:

- $E(\epsilon_i) = 0$. Isso implica que para um dado valor de x,

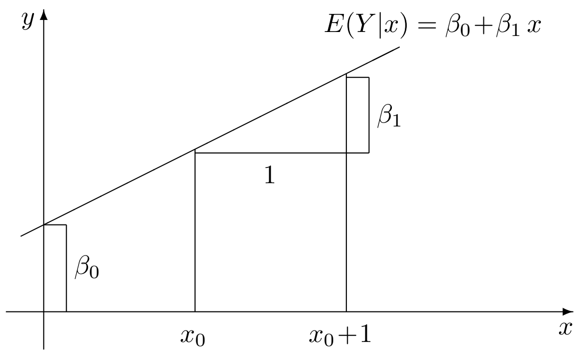
$$E(Y|x) = \beta_0 + \beta_1 \, x,$$

  conhecida por equação ou recta de regressão do modelo.

- $Var(\epsilon_i) = \sigma^2, \forall \, i$ (variância constante).
- $\epsilon_1, \ldots, \epsilon_n$ são não correlacionados (ou independentes).
- ϵ_i segue uma distribuição Normal, $i = 1, \ldots, n$.

Interpretação dos parâmetros de regressão:

- A ordenada na origem β_0 é o valor esperado de Y para um valor nulo da variável explicativa x.
- O declive da recta de regressão β_1 é a variação do valor esperado de Y por cada incremento unitário em x.



Parâmetros de regressão:

- $\beta_0 = E(Y|x{=}0)$.
- $\beta_1 = E(Y|x_0{+}1) - E(Y|x_0)$, $\forall\, x_0$.

Método dos mínimos quadrados em regressão linear simples

Um método de estimação dos coeficientes de regressão é o método de mínimos quadrados que consiste em minimizar a soma de quadrados dos erros aleatórios. Ou seja, o valor que minimiza a função

$$SQ(\beta_0,\beta_1) = \sum_{i=1}^{n} \epsilon_i^2 = \sum_{i=1}^{n} (Y_i - \beta_0 - \beta_1\, x_i)^2,$$

denotado por $(\hat\beta_0, \hat\beta_1)$, é denominado o estimador de mínimos quadrados do vector dos coeficientes de regressão.

Para a determinação da estimativa associada a $(\hat\beta_0, \hat\beta_1)$, deve-se encontrar as derivadas parciais da função $SQ(\beta_0,\beta_1)$ avaliada em $\{(y_i, x_i)\}$ em relação aos parâmetros β_0 e β_1.

$$\begin{array}{l} \frac{\partial}{\partial\beta_0} SQ(\beta_0,\beta_1) = 2\sum_{i=1}^{n}(y_i - \beta_0 - \beta_1\, x_i)(-1) \\ \frac{\partial}{\partial\beta_1} SQ(\beta_0,\beta_1) = 2\sum_{i=1}^{n}(y_i - \beta_0 - \beta_1\, x_i)(-x_i) \end{array}$$

$$\begin{array}{lcl} \frac{\partial}{\partial\beta_0} SQ(\beta_0,\beta_1) = 0 & \Rightarrow & \sum_{i=1}^{n} y_i = n\,\beta_0 + \beta_1 \sum_{i=1}^{n} x_i \\ \frac{\partial}{\partial\beta_1} SQ(\beta_0,\beta_1) = 0 & \Rightarrow & \sum_{i=1}^{n} x_i y_i = \beta_0 \sum_{i=1}^{n} x_i + \beta_1 \sum_{i=1}^{n} x_i^2 \end{array}$$

A solução desse sistema de equações pode ser expressa por $\beta_0 = \bar y - \hat\beta_1 \bar x$ e $\beta_1 = \frac{\sum_{i=1}^{n} x_i y_i - n\,\bar x\,\bar y}{\sum_{i=1}^{n} x_i^2 - n\,\bar x^2}$, podendo-se provar que este é ponto de mínimo, visto que a matriz hessiana avaliada neste ponto é definida positiva.

Portanto, os estimadores de mínimos quadrados de β_0 e β_1 são

$$\hat\beta_0 = \bar Y - \hat\beta_1 \bar x \qquad \text{e} \qquad \hat\beta_1 = \frac{\sum_{i=1}^{n} x_i Y_i - n\,\bar x\,\bar Y}{\sum_{i=1}^{n} x_i^2 - n\,\bar x^2}.$$

Consequentemente, a equação ou recta de regressão é estimada por

$$\widehat Y \equiv \widehat E(Y|x) = \hat\beta_0 + \hat\beta_1\, x,$$

i.e., dado um valor x, o valor esperado de Y é estimado por $\hat\beta_0 + \hat\beta_1\, x$. A estimação pontual de $E(Y|x)$ deve restringir-se ao domínio dos valores observados na amostra da variável explicativa x.

Estimadores de máxima verosimilhança

Supondo que os erros aleatórios são tais que $\epsilon_i \sim N(0,\sigma^2)$, $i{=}1,\ldots,n$, tem-se que $Y_i \sim N(\beta_0 + \beta_1\, x_i, \sigma^2)$ e portanto a função de verosimilhança associada ao modelo de regressão linear simples é

$$L(\beta_0,\beta_1,\sigma^2|\{y_i,x_i\}) = \prod_{i=1}^{n}\left[\frac{1}{\sqrt{2\,\pi}\,\sigma}\exp\!\left(-\frac{1}{2\,\sigma^2}(y_i - \beta_0 - \beta_1 x_i)^2\right)\right]$$

A maximização da função acima em relação aos parâmetros β_0 e β_1 equivale a maximizar $-\sum_{i=1}^n (y_i - \beta_0 - \beta_1 x_i)^2 = -SQ(\beta_0, \beta_1)$, ou seja, minimizar a soma de quadrados dos desvios médios. Por conseguinte, os estimadores de máxima verosimilhança de β_0 e β_1 são os estimadores de mínimos quadrados dos parâmetros, $\hat{\beta}_0$ e $\hat{\beta}_1$. Além disso, pode-se provar que o EMV de σ^2 é $\tilde{\sigma}^2 = \frac{1}{n}\sum_{i=1}^n (Y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)^2$.

Propriedades dos estimadores dos mínimos quadrados

- Estimador $\hat{\beta}_1$:
$$\hat{\beta}_1 = \frac{\sum_{i=1}^n x_i Y_i - n\,\bar{x}\,\bar{Y}}{\sum_{i=1}^n x_i^2 - n\,\bar{x}^2} = \frac{\sum_{i=1}^n (x_i - \bar{x}) Y_i}{\sum_{i=1}^n (x_i - \bar{x})^2} = \sum_{i=1}^n k_i Y_i,$$
onde $k_i = \frac{x_i - \bar{x}}{\sum_{i=1}^n (x_i - \bar{x})^2}$, com $\sum_{i=1}^n k_i = 0$, $\sum_{i=1}^n k_i x_i = 1$ e $\sum_{i=1}^n k_i^2 = \frac{1}{\sum_{i=1}^n (x_i - \bar{x})^2} = \frac{1}{\sum_{i=1}^n x_i^2 - n\,\bar{x}^2}$.
Logo,
  - $E(\hat{\beta}_1) = \sum_{i=1}^n k_i\, E(Y_i) = \beta_0 \sum_{i=1}^n k_i + \beta_1 \sum_{i=1}^n k_i x_i = \beta_1$.
  - $Var(\hat{\beta}_1) = \sum_{i=1}^n k_i^2\, Var(Y_i) = \sigma^2 (\sum_{i=1}^n x_i^2 - n\,\bar{x}^2)^{-1}$.

- Estimador $\hat{\beta}_0$:
$$\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{x} = \frac{1}{n}\sum_{i=1}^n Y_i - \sum_{i=1}^n k_i Y_i \bar{x} = \sum_{i=1}^n w_i Y_i,$$
onde $w_i = (1/n - k_i\,\bar{x})$, com $\sum_{i=1}^n w_i = 1$, $\sum_{i=1}^n w_i x_i = 0$ e $\sum_{i=1}^n w_i^2 = (\frac{1}{n} + \frac{\bar{x}^2}{\sum_{i=1}^n x_i^2 - n\,\bar{x}^2})$.
Logo,
  - $E(\hat{\beta}_0) = \sum_{i=1}^n w_i\, E(Y_i) = \beta_0 \sum_{i=1}^n w_i + \beta_1 \sum_{i=1}^n w_i x_i = \beta_0$.
  - $Var(\hat{\beta}_0) = \sum_{i=1}^n w_i^2\, Var(Y_i) = \sigma^2(\frac{1}{n} + \frac{\bar{x}^2}{\sum_{i=1}^n x_i^2 - n\,\bar{x}^2})$.

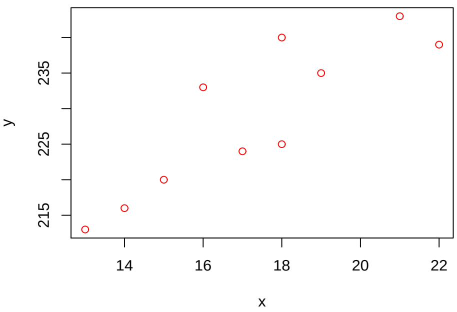
Note-se que $\hat{\beta}_0$ e $\hat{\beta}_1$ são combinações lineares dos Y_i e estimadores centrados de β_0 e β_1, respectivamente.

- Estimador $\hat{\sigma}^2$:
Seja SQE a soma de quadrados dos resíduos $Y_i - \hat{Y}_i$, $i = 1, \ldots, n$, onde $\hat{Y}_i \equiv \hat{\beta}_0 + \hat{\beta}_1 x_i$, isto é,
$$\begin{aligned} SQE &= \textstyle\sum_{i=1}^n (Y_i - \hat{Y}_i)^2 = \sum_{i=1}^n (Y_i - \bar{Y} + \hat{\beta}_1\,\bar{x} - \hat{\beta}_1 x_i)^2 \\ &= \textstyle\sum_{i=1}^n (Y_i - \bar{Y})^2 - \hat{\beta}_1^2 \sum_{i=1}^n (x_i - \bar{x})^2 = SQT - SQR, \end{aligned}$$
onde $SQT = \sum_{i=1}^n (Y_i - \bar{Y})^2$ e $SQR = \hat{\beta}_1^2 \sum_{i=1}^n (x_i - \bar{x})^2$ são conhecidas por somas de quadrados total e da regressão, respectivamente.
Pode-se provar que $E(SQT) = (n-1)\sigma^2 + \beta_1^2 \sum_{i=1}^n (x_i - \bar{x})^2$ e $E(SQR) = \sigma^2 + \beta_1^2 \sum_{i=1}^n (x_i - \bar{x})^2$ e portanto um estimador centrado de σ^2 é
$$\hat{\sigma}^2 = \frac{SQE}{n-2} = \frac{1}{n-2}\left[\left(\sum_{i=1}^n Y_i^2 - n\,\bar{Y}^2\right) - \hat{\beta}_1^2\left(\sum_{i=1}^n x_i^2 - n\,\bar{x}^2\right)\right].$$

Exemplo 9.1: A resistência de uma certa fibra sintética (Y) é suposta estar relacionada com a percentagem de algodão (x). Para avaliar essa conjectura tomou-se uma amostra aleatória de 10 peças da fibra produzidas sob as mesmas condições, obtendo-se os seguintes dados:

y	213	220	216	225	235	218	239	243	233	240
x	13	15	14	18	19	17	22	21	16	18

$\sum_{i=1}^{10} x_i = 173$
$\sum_{i=1}^{10} y_i = 2288$
$\sum_{i=1}^{10} x_i^2 = 3069$
$\sum_{i=1}^{10} y_i^2 = 524510$
$\sum_{i=1}^{10} x_i y_i = 39825$



As estimativas de mínimos quadrados de β_0 e β_1 são

$$\hat{\beta}_1 = \frac{39825 - 10 \times 17.3 \times 228.8}{3069 - 10 \times 17.3^2} = 3.188$$
$$\hat{\beta}_0 = 228.8 - 3.188 \times 17.3 = 173.65$$

Consequentemente, a equação ou recta de regressão estimada é

$$\widehat{Y} \equiv \widehat{E}(Y|x) = 173.65 + 3.188\, x,$$

sendo 3.188 a variação na resistência média da fibra sintética por cada incremento de 1% na percentagem de algodão.

A estimativa da variância dos erros aleatórios é

$$\hat{\sigma}^2 = \frac{1}{8}\left[(524510 - 10 \times 228.8^2) - 3.188^2(3069 - 10 \times 17.3^2)\right] = 30.27.$$

Inferências adicionais no modelo de regressão linear simples

Parâmetro β_1.

Como $\hat{\beta}_1 = \sum_{i=1}^n k_i Y_i$ é uma combinação linear de normais independentes, $E(\hat{\beta}_1) = \beta_1$ e $Var(\hat{\beta}_1) = \sigma^2(\sum_{i=1}^n x_i^2 - n\,\bar{x}^2)^{-1}$, então

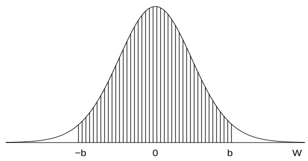
$$\hat{\beta}_1 \sim N\left(\beta_1, \frac{\sigma^2}{\sum_{i=1}^n x_i^2 - n\,\bar{x}^2}\right),$$

e, por conseguinte,

$$W = \frac{\hat{\beta}_1 - \beta_1}{\sqrt{\frac{\hat{\sigma}^2}{\sum_{i=1}^n x_i^2 - n\,\bar{x}^2}}} \sim t_{(n-2)}.$$

Considerando W acima como variável fulcral na construção de um intervalo de confiança a $100(1-\alpha)\%$ para β_1, tem-se que

$$P(a < W < b) = \gamma = 1-\alpha,$$



onde $b = -a = F^{-1}_{t_{(n-2)}}(1-\frac{\alpha}{2})$, e

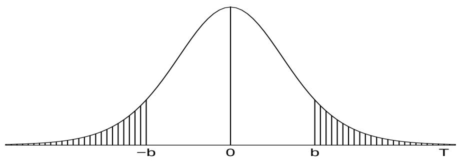
$$P\left(\hat{\beta}_1 - b\sqrt{\frac{\hat{\sigma}^2}{\sum_{i=1}^n x_i^2 - n\,\bar{x}^2}} < \beta_1 < \hat{\beta}_1 + b\sqrt{\frac{\hat{\sigma}^2}{\sum_{i=1}^n x_i^2 - n\,\bar{x}^2}}\right) = 1-\alpha$$

Logo, um intervalo (aleatório) de confiança a $100(1-\alpha)\%$ para β_1 é

$$IAC(\beta_1, 1-\alpha) = \left(\hat{\beta}_1 \pm b\sqrt{\frac{\hat{\sigma}^2}{\sum_{i=1}^n x_i^2 - n\,\bar{x}^2}}\right).$$

Teste de hipóteses:

1. Hipóteses: $H_0 : \beta_1 = \beta_1^0$ versus $H_1 : \beta_1 \neq \beta_1^0$.
2. Estatística do teste: $T = \frac{\hat{\beta}_1 - \beta_1^0}{\sqrt{\frac{\hat{\sigma}^2}{\sum_{i=1}^n x_i^2 - n\,\bar{x}^2}}} \overset{H_0}{\sim} t_{(n-2)}$, cujo valor observado é denotado por t_0.
3. Região crítica bilateral: Fixado α, $RC_\alpha = (-\infty, -b) \cup (b, \infty)$, onde $b = F^{-1}_{t_{(n-2)}}(1 - \frac{\alpha}{2})$.



4. Conclusão: Se $t_0 \in RC_\alpha$, rejeita-se H_0 ao nível de significância de $100\alpha\%$. Caso contrário, não se rejeita H_0 a $100\alpha\%$.

Parâmetro β_0.

Como $\hat{\beta}_0 = \sum_{i=1}^n w_i Y_i$ é uma combinação linear de normais independentes, $E(\hat{\beta}_0) = \beta_0$ e $Var(\hat{\beta}_0) = \sigma^2(\frac{1}{n} + \frac{\bar{x}^2}{\sum_{i=1}^n x_i^2 - n\,\bar{x}^2})$, então

$$W = \frac{\hat{\beta}_0 - \beta_0}{\sqrt{\hat{\sigma}^2(\frac{1}{n} + \frac{\bar{x}^2}{\sum_{i=1}^n x_i^2 - n\,\bar{x}^2})}} \sim t_{(n-2)}.$$

Considerando W acima como variável fulcral na construção de um intervalo de confiança a $100(1-\alpha)\%$ para β_0, tem-se que

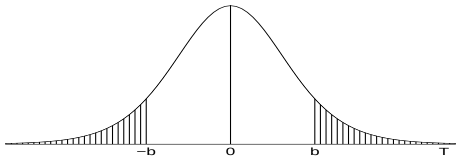
$$P(a < W < b) = \gamma = 1-\alpha,$$

onde $b = -a = F^{-1}_{t_{(n-2)}}(1-\frac{\alpha}{2})$ e, consequentemente, um intervalo (aleatório) de confiança a $100(1-\alpha)\%$ para β_0 é dado por

$$IAC(\beta_0, 1-\alpha) = \left(\hat{\beta}_0 \pm b\sqrt{\hat{\sigma}^2\left(\frac{1}{n} + \frac{\bar{x}^2}{\sum_{i=1}^n x_i^2 - n\,\bar{x}^2}\right)}\right).$$

Teste de hipóteses:

1. Hipóteses: $H_0 : \beta_0 = \beta_0^0$ versus $H_1 : \beta_0 \neq \beta_0^0$.
2. Estatística do teste: $T = \frac{\hat\beta_0 - \beta_0^0}{\sqrt{\hat\sigma^2(\frac{1}{n} + \frac{\bar x^2}{\sum_{i=1}^n x_i^2 - n\,\bar x^2})}} \overset{H_0}{\sim} t_{(n-2)}$, cujo valor observado é denotado por t_0.
3. Região crítica bilateral: Fixado α, $RC_\alpha = (-\infty, -b) \cup (b, \infty)$, onde $b = F^{-1}_{t_{(n-2)}}(1 - \frac{\alpha}{2})$.



4. Conclusão: Se $t_0 \in RC_\alpha$, rejeita-se H_0 ao nível de significância de $100\alpha\%$. Caso contrário, não se rejeita H_0 a $100\alpha\%$.

Estimação de $E(Y|x_0)$.

Dado um valor x_0 da variável explicativa, um estimador pontual do valor esperado de Y é

$$\widehat{Y}_0 \equiv \widehat{E}(Y|x_0) = \hat\beta_0 + \hat\beta_1\, x_0 = \sum_{i=1}^n \bigg(\frac{1}{n} + k_i(x_0 - \bar x)\bigg) Y_i.$$

Como $\widehat{Y}_0$ é uma combinação linear de normais e

- $E(\widehat{Y}_0) = E(\hat\beta_0) + E(\hat\beta_1)x_0 = \beta_0 + \beta_1\, x_0$,
- $Var(\widehat{Y}_0) = \cdots = \sigma^2\bigg(\frac{1}{n} + \frac{(\bar x - x_0)^2}{\sum_{i=1}^n x_i^2 - n\,\bar x^2}\bigg)$,

$$\Rightarrow \quad W = \frac{\hat\beta_0 + \hat\beta_1\, x_0 - (\beta_0 + \beta_1\, x_0)}{\sqrt{\hat\sigma^2(\frac{1}{n} + \frac{(\bar x - x_0)^2}{\sum_{i=1}^n x_i^2 - n\,\bar x^2})}} \ \sim \ t_{(n-2)}.$$

Considerando W acima como variável fulcral na construção de um intervalo de confiança a $100(1-\alpha)\%$ para $E(Y|x_0)$, tem-se que

$$P(a < W < b) = \gamma = 1-\alpha,$$

onde $b = -a = F^{-1}_{t_{(n-2)}}(1 - \frac{\alpha}{2})$, e

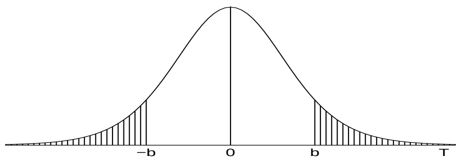
$$P\bigg(\hat\beta_0 + \hat\beta_1\, x_0 - b\sqrt{\hat\sigma^2(\frac{1}{n} + \frac{(\bar x - x_0)^2}{\sum_{i=1}^n x_i^2 - n\,\bar x^2})} < E(Y|x_0) < \hat\beta_0 + \hat\beta_1\, x_0 + b\sqrt{\hat\sigma^2(\frac{1}{n} + \frac{(\bar x - x_0)^2}{\sum_{i=1}^n x_i^2 - n\,\bar x^2})}\bigg) = 1-\alpha$$

Logo, um intervalo (aleatório) de confiança a $100(1-\alpha)\%$ para $E(Y|x_0)$ é dado por

$$IAC(E(Y|x_0), 1-\alpha) = \left(\hat{\beta}_0 + \hat{\beta}_1\, x_0 \pm b\,\sqrt{\hat{\sigma}^2\left(\frac{1}{n} + \frac{(\bar{x}-x_0)^2}{\sum_{i=1}^n x_i^2 - n\,\bar{x}^2}\right)}\right).$$

Teste de hipóteses:

1. Hipóteses: $H_0 : E(Y|x_0) = \mu^0$ versus $H_1 : E(Y|x_0) \neq \mu^0$.
2. Estatística do teste: $T = \frac{\hat{\beta}_0+\hat{\beta}_1\, x_0-\mu^0}{\sqrt{\hat{\sigma}^2\left(\frac{1}{n}+\frac{(\bar{x}-x_0)^2}{\sum_{i=1}^n x_i^2-n\,\bar{x}^2}\right)}} \overset{H_0}{\sim} t_{(n-2)}$, cujo valor observado é denotado por t_0.
3. Região crítica bilateral: Fixado α, $RC_\alpha = (-\infty, -b) \cup (b, \infty)$, onde $b = F^{-1}_{t_{(n-2)}}(1-\frac{\alpha}{2})$.



4. Conclusão: Se $t_0 \in RC_\alpha$, rejeita-se H_0 ao nível de significância de $100\alpha\%$. Caso contrário, não se rejeita H_0 a $100\alpha\%$.

Exemplo 9.1a: Teste ao nível de significância de 1% se a percentagem de algodão (x) influencia a resistência da fibra sintética (Y).

Teste de hipóteses:

1. Hipóteses: $H_0 : \beta_1 = 0$ versus $H_1 : \beta_1 \neq 0$.
2. Estatística do teste: $T = \frac{\hat{\beta}_1}{\sqrt{\frac{\hat{\sigma}^2}{\sum_{i=1}^n x_i^2-n\,\bar{x}^2}}} \overset{H_0}{\sim} t_{(n-2)}$, cujo valor observado é $t_0 = 3.188/\sqrt{30.27/76.1} = 5.054$.
3. Região crítica: Fixado $\alpha = 0.01$, $RC_\alpha = (-\infty, -3.355) \cup (3.355, \infty)$, onde $F^{-1}_{t_{(8)}}(0.995) = 3.355$.
4. Conclusão: Como $t_0 \in RC_{0.01}$, rejeita-se H_0 ao nível de significância de 1%. Note-se que o valor-P, $P = P(|T_0| \geq 5.054|H_0) = 0.00098$, com $F^{-1}_{t_{(8)}}(0.9995) = 5.041$, e portanto há forte evidência de que a percentagem de algodão influencia a resistência da fibra sintética.

Coeficiente de determinação

Definição 9.1: O *coeficiente de determinação* é uma medida relativa de ajustamento do modelo de regressão linear, representando a proporção da variação na resposta que é explicada pela variável explicativa, expresso por

$$R^2 = \frac{SQR}{SQT} = 1 - \frac{SQE}{SQT},$$

onde $SQT = \sum_{i=1}^n (Y_i - \bar{Y})^2$ e $SQR = \hat{\beta}_1^2 \sum_{i=1}^n (x_i - \bar{x})^2$ e portanto

$$R^2 = \frac{(\sum_{i=1}^n x_i Y_i - n\,\bar{x}\,\bar{Y})^2}{(\sum_{i=1}^n x_i^2 - n\,\bar{x}^2)(\sum_{i=1}^n Y_i^2 - n\,\bar{Y}^2)}.$$

O coeficiente de determinação é tal que $0 \leq R^2 \leq 1$, onde

- $R^2 \to 1$ indica bom ajustamento do modelo;
- $R^2 \to 0$ indica mau ajustamento do modelo.

1. Existem testes de hipóteses de ajustamento do modelo, *e.g.*, o teste F de falta de ajustamento (*lack-of-fit*).
2. A violação das suposições do modelo de regressão linear pode induzir a conclusões erradas sobre o modelo. Esse problema pode ser detectado através de técnicas de diagnóstico baseadas frequentemente na análise de resíduos.

Análise de resíduos na avaliação do modelo

A definição mais simples de *resíduo* é dada por

$$r_i = y_i - \hat{y}_i = y_i - \hat{\beta}_0 - \hat{\beta}_1 \, x_i,$$

onde (y_i, x_i) são os valores observados na amostra, $i = 1, \ldots, n$, enquanto os *resíduos padronizados* são por

$$r_i^s = \frac{r_i}{\sqrt{\hat{\sigma}^2}},$$

onde $\hat{\sigma}^2 = \frac{1}{n-2}[(\sum_{i=1}^n Y_i^2 - n\,\bar{Y}^2) - \hat{\beta}_1^2(\sum_{i=1}^n x_i^2 - n\,\bar{x}^2)]$.
Os gráficos de resíduos mais comuns são:

- r_i versus x_i.
- r_i versus $\hat{y}_i$.
- r_i ao longo do tempo (se fizer sentido).

A análise de gráficos de resíduos é a técnica de diagnóstico mais usada para encontrar:

- Observações discrepantes (*outliers*).
- Heterogeneidade da variância ($Var(Y_i) \neq \sigma^2$ para algum i).
- Falta de normalidade ($Y_i \nsim N(\cdot, \cdot)$).
- Dependência dos erros aleatórios ($\exists \, i \neq j, Cov(Y_i, Y_j) \neq 0$).

sem problemas perda de linearidade

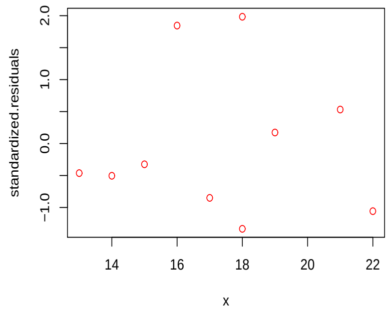
Exemplo 9.1b: Avalie o ajustamento do modelo de regressão linear simples ($Y = \beta_0 + \beta_1\, x + \epsilon$), incluindo um gráfico de resíduos.

$$r^2 = \frac{(39825 - 10 \times 17.3 \times 228.8)^2}{(3069 - 10 \times 17.3^2)(524510 - 10 \times 228.8^2)} = 0.7615.$$

Ou seja, 76.15% da variação total da resistência da fibra sintética é explicada pelo modelo de regressão linear simples com a percentagem de algodão como variável explicativa.

Gráficos de resíduos:

- $r_i^s = \frac{y_i - \hat{y}_i}{\sqrt{\hat{\sigma}^2}}$, $i = 1, \ldots, 10$.
- sem grandes problemas.



Alguns abusos no modelo de regressão:

- Selecção de variável explicativa.
  - É possível desenvolver uma relação estatisticamente significativa entre a variável resposta (Y) e a variável explicativa (x) que não faça sentido na prática.
- Extrapolação
  - A relação linear assumida para as variáveis resposta e explicativa não pode ser estendida para fora do domínio de actuação dos dados observados. Por exemplo, se os valores da variável explicativa caem em $[13, 22]$, não se deve inferir *e.g.*sobre o valor esperado da variável resposta Y quando $x_0 = 25$, a não ser que haja informação adicional sobre a validade do modelo para esse domínio estendido.

$\boxed{\text{FIM!}}$