Probabilidades e Estatística E[1]

[1] Última actualização a 11 de Abril de 2016

1. Introdução à Teoria das Probabilidades
2. Variáveis aleatórias e suas distribuições de probabilidade
   (a) Definição de variável aleatória;
   (b) Momentos de variáveis aleatórias;
   (c) Algumas distribuições importantes
   (d) Vectores aleatórios
   (e) Geração de números pseudo-aleatórios
3. Teorema Limite Central
4. Estimação pontual
5. Estimação por intervalo de confiança
6. Testes de hipóteses
7. Regressão linear simples

Página da disciplina:

https://clip.unl.pt/

- Programa
- Regras de avaliação
- Horário de atendimento
- Horário das aulas
- Documentação de apoio

Número de maneiras diferentes de escolher k elementos, de um conjunto de n elementos:

		Há **repetição**?	
		Sim	**Não**
Interessa a **ordem**?	**Sim**	Arranjos com rep.: $^{n}A'_{k} = n^{k}$	Arranjos: $^{n}A_{k} = \frac{n!}{(n-k)!}$
	Não	Comb. com rep. $^{n}C'_{k} = \frac{(n+k-1)!}{(n-1)!k!}$	Combinações $^{n}C_{k} = \binom{n}{k} = \frac{n!}{(n-k)!k!}$

Sejam $A = \{a_1, \ldots, a_n\}$ e $B = \{b_1, \ldots, b_m\}$ conjuntos com n e m elementos distintos, respectivamente.

Produto cartesiano:

Designa-se por **produto cartesiano** o conjunto de pares (a_i, b_j) em que o primeiro elemento provém de A e o segundo de B e representa-se por $A \times B$.

O número de elementos de $A \times B$ é

$$\boxed{\#(A \times B) = n \times m.}$$

1. Os escritórios de uma empresa estão equipados com telefones funcionando internamente como extensões identificadas por uma sequência de 3 algarismos, dos quais o primeiro não é zero. Quantos telefones podem ser identificados?

4. Um centro comercial tem 8 portas. De quantas maneiras distintas se pode
   - (a) entrar e sair do centro comercial?
   - (b) entrar por uma porta e sair por outra?

6. De quantos modos diferentes é possível dispor numa fila para fotografia, 3 homens e 2 mulheres se:
   - (a) os homens e as mulheres puderem ocupar indistintamente qualquer lugar?
   - (b) se um dos homens, o mais alto por exemplo, ficar no meio e todos os restantes indistintamente em qualquer lugar?
   - (c) se ficarem alternadamente homens e mulheres, nunca dois homens seguidos ou duas mulheres seguidas?

7. Quatro livros de Matemática, seis de Física e dois de Química, todos diferentes, devem ser arrumados numa prateleira. Quantas arrumações diferentes são possíveis, se:
   - (a) os livros de cada matéria ficarem todos juntos?
   - (b) apenas os livros de Matemática ficarem juntos?

8 Mostre que $\boxed{n! + (n+1)! = (n+2)n!}$ e que $\boxed{(n+1)! - n! = n \times n!}$.

9 Vinte e cinco membros de uma sociedade devem eleger um presidente, um secretário e um tesoureiro. Supondo que qualquer um dos vinte e cinco membros é elegível para qualquer dos cargos, quantas são as hipóteses distintas de eleição?

15 De quantas maneiras distintas se poderá formar uma comissão, com três elementos escolhidos de entre os vinte e cinco membros de uma sociedade?

Definição (Experiência aleatória)

Uma **experiência aleatória** *é uma experiência na qual:*

- *todos os possíveis resultados da experiência são conhecidos à partida;*
- *para qualquer realização da experiência não se sabe, antes desta ocorrer, qual dos seus possíveis resultados vai acontecer;*
- *a experiência pode sempre ser repetida sob idênticas condições.*

Exemplos:

- resultado do lançamento de uma moeda ao ar (assumindo que não "aterra" de lado!);
- nº de *pintas* após o lançamento de um dado;
- o tempo de vida de uma lâmpada;

Definição (Espaço de resultados ou universo)

Chamamos **espaço de resultados** *ou* **universo**, *e representamos por* Ω, *ao conjunto de todos os possíveis resultados de uma experiência aleatória.*

Exemplos:

- resultado do lançamento de uma moeda ao ar (assumindo que não "aterra" de lado!). $\Omega = \{\text{Cara, Coroa}\}$;
- nº de *pintas* após o lançamento de um dado. $\Omega = \{1, 2, 3, 4, 5, 6\}$;
- o tempo de vida de uma lâmpada. $\Omega = \{t \in \mathbb{R} : t \geq 0\} = \mathbb{R}_0^+$;

Definição (Acontecimento)

Um **acontecimento** *é um subconjunto do espaço de resultados, Ω.*

Notas:

- Cada acontecimento formado por apenas um ponto amostral é designado por **acontecimento simples** ou **elementar**.
- Ao conjunto $\varnothing$ chamamos **acontecimento impossível**
- Ao conjunto Ω chamamos **acontecimento certo**.

Definição (Sub-acontecimento)

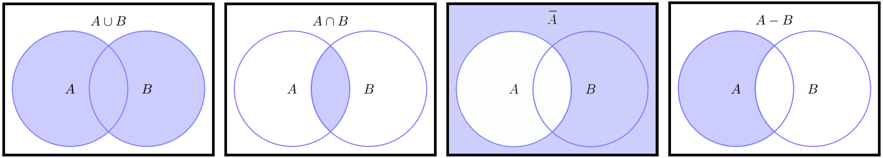
A *é* **sub-acontecimento** *de* B, *e escreve-se* $A \subseteq B$, *se e só se a realização de* A *implica a realização de* B.

Principais Operações:

Podemos aplicar as operações usuais sobre conjuntos de modo a obter outros acontecimentos de interesse. As operações mais usuais são:

- A **união** de dois acontecimentos A e B, $A \cup B$;
- A **intersecção** de dois acontecimentos A e B, $A \cap B$;
- O **complementar** do acontecimento A, $\overline{A}$;
- A **diferença** dos acontecimentos A e B, $A - B$ $(= A \cap \overline{B})$;

Diagramas de Venn:



Propriedades importantes:

- $A \cup B = B \cup A$ (Prop. comutativa)
  $A \cap B = B \cap A$ (Prop. comutativa)
- $(A \cup B) \cup C = A \cup (B \cup C)$ (Prop. associativa)
  $(A \cap B) \cap C = A \cap (B \cap C)$ (Prop. associativa)
- $(A \cup B) \cap C = (A \cap C) \cup (B \cap C)$ (Prop. distributiva)
  $(A \cap B) \cup C = (A \cup C) \cap (B \cup C)$ (Prop. distributiva)
- $\overline{\overline{A}} = A$ (dupla negação)
- $(A \cup A) = (A \cap A) = A$ (idempotência)

Leis de De Morgan:

- $\overline{A \cup B} = \overline{A} \cap \overline{B}$;
- $\overline{A \cap B} = \overline{A} \cup \overline{B}$;

Definição (Acontecimentos disjuntos ou mutuamente exclusivos)

Se $A \cap B = \varnothing$, dizemos que os acontecimentos A e B são **disjuntos** *ou* **mutuamente exclusivos**.

Definição (Definição clássica, ou de Laplace, de Probabilidade)

Se uma experiência aleatória tem a si associado um número finito N *de resultados, mutuamente exclusivos e igualmente prováveis, então a probabilidade de qualquer acontecimento* A, $P(A)$, *é dada por:*

$$\boxed{P\left(A\right) = \frac{N_A}{N} = \frac{n^o \ de \ resultados \ favoráveis \ a \ A}{n^o \ de \ resultados \ possíveis}.}$$

Exemplo: Considere a experiência aleatória que consiste no lançamento de um dado e registo do número da face virada para cima. Definam-se os acontecimentos: A - O número é menor que 3; B - O número é par.

Determine a probabilidade de: Ocorrer pelo menos um dos acontecimentos;

Ocorrer A e B;

Ocorrer A mas não B;

Não ocorrer A.

Definição (Definição frequencista de Probabilidade)

A probabilidade de ocorrer A é avaliada a partir de informação existente sobre A, sendo dada pelo limite da frequência relativa com que se observou A, isto é,

$$\boxed{P\left(A\right) = \lim_{n\to\infty} \frac{n_A}{n},}$$

onde n_A representa o número de observações de A e n o número de realizações da experiência aleatória.

Para valores elevados de n, $P(A) \approx \dfrac{n_A}{n}$.

Definição (Definição Axiomática de Probabilidade)

Seja $\mathcal{A}$ uma família de acontecimentos, fechada para as operações usuais. A Probabilidade é uma função cujo domínio é $\mathcal{A}$ e que verifica as seguintes condições ou axiomas:

1. $P(A) \geq 0$, *qualquer que seja o acontecimento* $A \in \mathcal{A}$;
2. $P(\Omega) = 1$;
3. *Se* A *e* B *são acontecimentos disjuntos, isto é, se* $A \cap B = \varnothing$ *então* $P(A \cup B) = P(A) + P(B)$.

Esta axiomática não contempla situações em que temos de considerar uma infinidade numerável de acontecimentos. É usual substituir $\boxed{\mathbf{3}}$ por

3. Se A_1, A_2,. . . são acontecimentos disjuntos dois a dois, então

$$P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i).$$

Sejam A e B dois acontecimentos. Então são consequências imediatas dos axiomas os seguintes resultados:

1. $P(\varnothing) = 0$;
2. Se $A \subseteq B$ então $P(A) \leq P(B)$;
3. $P(\bar{A}) = 1 - P(A)$;
4. $P(A) \in [0, 1]$;
5. $P(A - B) = P(A \cap \overline{B}) = P(A) - P(A \cap B)$;
6. $P(A \cup B) = P(A) + P(B) - P(A \cap B)$;

7. Dados os acontecimentos A_i, $i = 1, \ldots, n$,

$$
\begin{aligned}
P\left(\cup_{i=1}^{n} A_i\right) = & P(A_1 \cup A_2 \cup \ldots \cup A_n) = \\
& = \sum_{i=1}^{n} P\left(A_i\right) - \sum_{i \neq j} P\left(A_i \cap A_j\right) \\
& \quad + \sum_{i \neq j \neq k} P\left(A_i \cap A_j \cap A_k\right) - \ldots + (-1)^{n-1} P\left(\cap_{i=1}^{n} A_i\right)
\end{aligned}
$$

Definição (Acontecimentos incompatíveis)

Dois **acontecimentos** A e B *dizem-se* **incompatíveis** *se*

$$P\left(A \cap B\right) = 0.$$

Realizou-se um ensaio clínico para testar um novo medicamento. Escolheram-se 200 doentes com a mesma doença. 100 desses doentes foram tratados com o novo medicamento e os restantes 100 com o medicamento convencional. Os resultados foram os seguintes:

	Melhorou (M)	Não melhorou ($\overline{M}$)	
Medicamento experimental (E)	69	31	100
Medicamento convencional ($\overline{E}$)	58	42	100
	127	73	200

1. Qual a probabilidade de um doente, escolhido ao acaso,
   (a) melhorar?
   (b) tomar o medicamento experimental e melhorar?
2. Qual a probabilidade de um doente, que tomou o medicamento experimental, melhorar?

Definição (Probabilidade condicional)

A **probabilidade condicional** *de* A *dado* B *é*

$$P(A|B) = \frac{P(A \cap B)}{P(B)}, \qquad \text{se } P(B) > 0. \tag{1}$$

Observação: Resulta de (1) o seguinte teorema,

Teorema (Teorema da Probabilidade Composta)

Sejam A *e* B *dois acontecimentos tais que* $P(B) > 0$. *Então,*

$$P\left(A \cap B\right) = P\left(A \,|B\,\right) P\left(B\right).$$

Nalguns casos verifica-se que $\boxed{P(A|B) = P(A)}$, ou seja, o conhecimento da ocorrência de B não afecta a probabilidade de A ocorrer.

Definição (Acontecimentos Independentes)

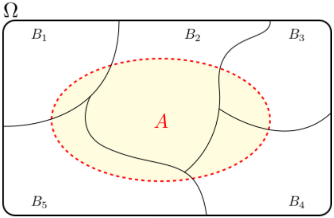
Dois acontecimentos A e B dizem-se **independentes** *se e só se*

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

Teorema

Se A e B são acontecimentos independentes, também são independentes A e $\bar{B}$, $\bar{A}$ e B e ainda $\bar{A}$ e $\bar{B}$.

Nota: Dizemos que $\{B_1, B_2, \ldots, B_n\}$ é uma **partição do espaço de resultados**, Ω, quando $B_i \cap B_j = \varnothing \ (i \neq j)$ e $\cup_{i=1}^{n} B_i = \Omega$.



Teorema (Teorema da probabilidade total)

Seja $\{B_1, B_2, \ldots, B_n\}$ *uma* **partição do espaço de resultados** Ω, *com* $P(B_i) > 0, \ \forall i$. *Dado um qualquer acontecimento* A, *tem-se*

$$P(A) = P(A|B_1)\,P(B_1) + \ldots + P(A|B_n)\,P(B_n)$$

Teorema (Teorema de Bayes)

Seja $\{B_1, B_2, \ldots, B_n\}$ *uma partição do espaço de resultados* Ω*, com* $P(B_i) > 0,\ \forall i$*. Dado um qualquer acontecimento* A*, com* $P(A) > 0$*,*

$$P(B_j \mid A) = \frac{P(A \mid B_j)\, P(B_j)}{P(A)} = \frac{P(A \mid B_j)\, P(B_j)}{\sum_{i=1}^{n} P(A \mid B_i)\, P(B_i)}.$$

Três máquinas A, B e C **produzem** botões, respectivamente, 15%, 25% e 60% da produção total. As percentagens de botões **defeituosos** fabricados por estas máquinas são respectivamente 5%, 7% e 4%.

Diga, justificando, se a seguinte afirmação é verdadeira ou falsa: Se ao acaso, da produção total de botões, for encontrado um defeituoso, a probabilidade de ele ter sido produzido pela máquina B é de cerca de 36%.

Resolução:
Sejam A, B, C e D os seguintes acontecimentos:
A - Botão produzido pela máquina A; B - Botão produzido pela máquina B;
C - Botão produzido pela máquina C; D - Botão com defeito;

De acordo com o enunciado, $P(A) = 0.15$, $P(B) = 0.25$, $P(C) = 0.6$, $P(D|A) = 0.05$, $P(D|B) = 0.07$ e $P(D|C) = 0.04$.

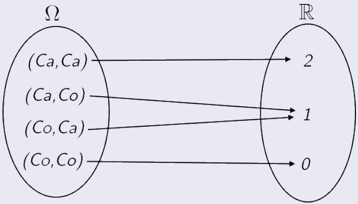
Usando o **Teorema de Bayes**, $P(B|D) = 175/490 \simeq 36\%$.

Exemplo

Considere a experiência que consiste no lançamento de 2 moedas equilibradas, e registo da face voltada para cima. O espaço de resultados é

$$\Omega = \{(Ca, Ca), (Ca, Co), (Co, Ca), (Co, Co)\}.$$

Podemos, **por exemplo**, *atribuir*



Seja X a aplicação $X : \Omega \rightarrow \mathbb{R}$ que associa a cada elemento $\omega \in \Omega$ um valor real $X(\omega)$. **Observação:** Neste exemplo, X = número de caras.

Definição

- *A imagem do* **acontecimento** A *por* X *é o conjunto de valores reais:*
$$X(A) = \{X(\omega) : \omega \in A\}.$$
- *A imagem inversa de* $D \subset \mathbb{R}$ *por* X *é o subconjunto de* Ω*:*
$$X^{-1}(D) = \{\omega \in \Omega : X(\omega) \in D\}.$$

Exemplo: Se $A = \{$(Ca,Ca), (Ca,Co)$\}$, a imagem de A por X é $X(A) = \{1, 2\}$.

Exemplo: Sejam $D_1 = \{0, 2\}$, $D_2 =]0, 2]$, $D_3 = [3, +\infty[$ e $D_4 =]-\infty, 0]$.
Então, $X^{-1}(D_1) = \{(Ca, Ca), (Co, Co)\}$, $X^{-1}(D_2) = \{(Ca, Ca), (Co, Ca), (Ca, Co)\}$, $X^{-1}(D_3) = \varnothing$, e $X^{-1}(D_4) = \{(Co, Co)\}$.

Definição (Variável aleatória)

Uma **variável aleatória** X *(v.a.) é uma aplicação* $X : \Omega \to \mathbb{R}$ *tal que,*

$$A_x = \{\omega \in \Omega : X(\omega) \leq x\} = X^{-1}(]-\infty, x]), \quad \forall x \in \mathbb{R},$$

é um acontecimento.

Repare que:

$$A_x = X^{-1}(]-\infty; x]) = \begin{cases} \varnothing, & x < 0 \\ \{(Co, Co)\} & 0 \leq x < 1 \\ \{(Co, Co), (Ca, Co), (Co, Ca)\} & 1 \leq x < 2 \\ \Omega & x \geq 2 \end{cases}$$

Como calcular a probabilidade de se observarem valores de X?

Exemplo:

$$P(X=1)=P(X^{-1}(\{1\}))=P(\{(Ca,Co),(Co,Ca)\})=\frac{1}{2}$$

Proposição

Se $X_1, X_2, \ldots, X_m$ são m variáveis aleatórias e $h:\mathbb{R}^m \to \mathbb{R}$ é uma função contínua, então

$$Y=h\left(X_1, X_2, \ldots, X_m\right)$$

é uma v.a..

Definição (Função de distribuição)

Define-se **função de distribuição** *da variável aleatória* X *como:*

$$F(x) = P(X \leq x) = P(X^{-1}(]-\infty, x])) = P(\{\omega : X(\omega) \leq x\}), \quad \forall x \in \mathbb{R}.$$

Propriedades da função de distribuição:

1. $\lim_{x \to -\infty} F(x) = 0$ e $\lim_{x \to +\infty} F(x) = 1$;
2. F é contínua à direita, isto é, $\lim_{x \to a^+} F(x) = F(a)$;
3. F é não decrescente, isto é, se $x < y$, $F(x) \leq F(y)$.

Proposição

Seja X *uma v.a. com função de distribuição* F*. Tem-se:*

$$P(X = x) = P(X \leq x) - P(X < x) = F(x) - F(x^-), \quad \forall x \in \mathbb{R}.$$

Considere o conjunto,

$$D = \{a \in \mathbb{R} : \; P(X = a) > 0\}.$$

Nota: D é o conjunto de pontos de descontinuidade da função de distribuição F.

Definição (Variável aleatória discreta)

Uma v.a. X *diz-se do* **tipo discreto** *ou simplesmente* **discreta** *se o conjunto* D *é quanto muito numerável, e se* $P(X \in D) = 1$.
O conjunto D *é designado o* **suporte** *de* X.

Definição (Função de probabilidade)

Seja X uma v.a. discreta. Chama-se função de probabilidade (f.p.) de X à função definida pelo conjunto dos valores de D e pelas respectivas probabilidades, isto é, por (x_i, p_i) onde $x_i \in D$ e $p_i = P(X = x_i)$.

Propriedades da função de probabilidade:

1. $P(X = x) > 0, \ \forall x \in D$;
2. $\sum_{x \in D} P(X = x) = 1$.

Cálculo de probabilidades:

$$\boxed{P(X \in I) = P(X \in D \cap I) = \sum_{x \in (D \cap I)} P(X = x), \quad \text{para qualquer } I \subseteq \mathbb{R}.}$$

Definição (Valor médio)

O **valor médio** *ou* **valor esperado** *ou* **média** *de uma v.a.* X *é:*

$$\mu = \mathrm{E}(X) = \begin{cases} \sum_{x \in D} x P(X = x) & \text{se } X \text{ é uma v.a. } \textbf{discreta} \\ & \text{com suporte } D; \\ \int_{-\infty}^{+\infty} x f(x) dx & \text{se } X \text{ é uma v.a. } \textbf{contínua}^2; \end{cases}$$

Observações:

- O **valor médio** não é necessariamente um valor observável de X e expressa o centro de massa (ponto de equilíbrio) da função de probabilidade/função densidade.
- É uma medida (ou parâmetro) de **localização** de X.

[2]Nota: O caso em que X é uma v.a. contínua será estudado posteriormente.

1. Considere a v.a. X com função de probabilidade

$$X \left\{ \begin{array}{ccc} 0 & 1 & 2 \\ 0.25 & 0.5 & 0.25 \end{array} \right.$$

Então, o **valor médio** de X é

$$\mathrm{E}(X) = 0 \times P(X = 0) + 1 \times P(X = 1) + 2 \times P(X = 2) = 1$$

2. Considere a v.a. X com função de probabilidade

$$X \left\{ \begin{array}{ccc} 0 & 2 & 4 \\ 0.3 & 0.5 & 0.2 \end{array} \right.$$

o **valor médio** de X é

$$\begin{aligned} \mathrm{E}(X) &= 0 \times P(X = 0) + 2 \times P(X = 2) + 4 \times P(X = 4) = \\ &= 0 + 1 + 0.8 = 1.8 \end{aligned}$$

Definição

Seja X uma v.a. e g uma função real. Então o **valor médio** *de $g(X)$ é:*

$$\mathrm{E}(g(X)) = \begin{cases} \sum\limits_{x \in D} g(x)P(X = x) & \text{se } X \text{ é uma v.a. } \textbf{discreta}; \\ \\ \int_{-\infty}^{+\infty} g(x)f(x)dx & \text{se } X \text{ é uma v.a. } \textbf{contínua}; \end{cases}$$

Observação: O valor médio só está definido se o somatório/integral for absolutamente convergente.

Considere a v.a. X com função de probabilidade

$$X \begin{cases} 0 & 2 & 4 \\ 0.3 & 0.5 & 0.2 \end{cases}$$

- o **valor médio** de $g(X) = X^2$ é

$$\begin{aligned} E(X^2) &= 0^2 \times P(X=0) + 2^2 \times P(X=2) + 4^2 \times P(X=4) = \\ &= 0 \times 0.3 + 4 \times 0.5 + 16 \times 0.2 = 5.2 \end{aligned}$$

- o **valor médio** de $g(X) = \ln(1+X)$ é

$$\begin{aligned} E(\ln(1+X)) = \ln(1+0)P(X=0) + \ln(1+2)P(X=2) + \ln(1+4)P(X=4) = \\ = 0.871 \end{aligned}$$

- o **valor médio** de $g(X) = \frac{X-4}{2}$ é

$$\begin{aligned} E\left(\frac{X-4}{2}\right) &= \frac{0-4}{2} \times P(X=0) + \frac{2-4}{2} \times P(X=2) + \frac{4-4}{2} \times P(X=4) = \\ &= -2 \times 0.3 - 1 \times 0.5 + 0 \times 0.2 = -0.6 - 0.5 = -1.1 \end{aligned}$$

Proposição

Sejam X e Y variáveis aleatórias, a e b, constantes reais. Então:

- $\mathrm{E}(a) = a$;
- $\mathrm{E}(aX + b) = a\mathrm{E}(X) + b$;
- $\mathrm{E}(X \pm Y) = \mathrm{E}(X) \pm \mathrm{E}(Y)$;
- $\mathrm{E}\left(g(X) \pm h(Y)\right) = \mathrm{E}\left(g(X)\right) \pm \mathrm{E}\left(h(Y)\right)$

Exemplo: o **valor médio** de $g(X) = \frac{X-4}{2}$ é

$$\mathrm{E}\left(\frac{X-4}{2}\right) = \mathrm{E}\left(\frac{1}{2}X - 2\right) = \frac{1}{2}\mathrm{E}(X) - 2$$

Definição (Variância e desvio padrão)

Seja X uma v.a. com valor médio μ. Define-se a **variância** *de X por*

$$\sigma^2 = \mathrm{V}(X) = \mathrm{E}((X-\mu)^2).$$

A $\sigma = \sqrt{V(X)}$ chamamos **desvio padrão** *da v.a. X.*

Exemplo: Considere a v.a. X com função de probabilidade

$$X \begin{cases} 0 & 2 & 4 \\ 0.3 & 0.5 & 0.2 \end{cases}$$

A variância de X é

$$\begin{aligned} \mathrm{V}(X) &= (0-1.8)^2 \times P(X=0) + (2-1.8)^2 \times P(X=2) + (4-1.8)^2 \times P(X=4) = \\ &= (0-1.8)^2 \times 0.3 + (2-1.8)^2 \times 0.5 + (4-1.8)^2 \times 0.2 = 1.96 \end{aligned}$$

O desvio padrão de X é $\sigma = \sqrt{1.96} = 1.4$

Proposição

Seja X é uma v.a., para a qual $V(X) < \infty$. Então,

$$V(X) = \mathrm{E}\left(X^2\right) - \mathrm{E}^2(X).$$

Exemplo: Considere a v.a. X com função de probabilidade

$$X \begin{cases} 0 & 2 & 4 \\ 0.3 & 0.5 & 0.2 \end{cases}$$

A variância de X é $\mathrm{V}(X) = \mathrm{E}\left(X^2\right) - \mathrm{E}^2(X) = 5.2 - 1.8^2 = 1.96$

Proposição

Seja X uma variável aleatória, a e b, constantes reais. Então:

- $\mathrm{V}(a) = 0$;
- $\mathrm{V}(aX + b) = a^2\mathrm{V}(X)$;

Definição (Coeficiente de variação)

O Coeficiente de variação de X, com suporte positivo é, $\quad \mathrm{CV} = \frac{\sigma}{\mu}$.

Definição (Moda)

A Moda, representada por mo, é o valor que maximiza a função de probabilidade ou a função densidade de probabilidade, desde que seja único.

Exemplo: Considere a v.a. X com função de probabilidade

$$X \begin{cases} 0 & 2 & 4 \\ 0.3 & 0.5 & 0.2 \end{cases}$$

O Coeficiente de variação de X é $\mathrm{CV}(X) = \sqrt{1.96}/1.8 = 7/9 = 0.778$ e a moda é 2.

Definição (Coeficiente de assimetria)

$$\gamma_1 = \frac{\mathrm{E}((X-\mu)^3)}{\sigma^3}.$$

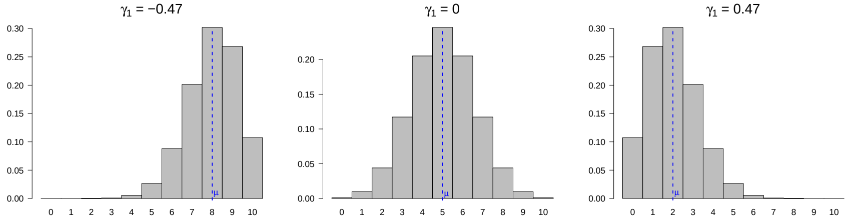


Figura: Função de probabilidade de 3 variáveis aleatórias discretas. A linha azul vertical indica o valor de μ.

Definição (Coeficiente de achatamento ou kurtosis:)

$$\gamma_2 = \frac{\mathrm{E}((X-\mu)^4)}{\sigma^4}.$$

Nota: *A* $\gamma_2 - 3$ *chamamos* **excesso de kurtosis**.

Exemplo: Considere a v.a. X com função de probabilidade

$$X \begin{cases} 0 & 2 & 4 \\ 0.3 & 0.5 & 0.2 \end{cases}$$

Temos

$$\mathrm{E}((X-1.8)^3) = 0.384, \qquad \mathrm{E}((X-1.8)^4) = 7.8352,$$
$$\gamma_1 = 0.1399 \qquad \gamma_2 = 2.0396$$

Nota: Podemos também calcular o numerador da fórmula de γ_1 e γ_2 usando os resultados $E((X-\mu)^3) = E(X^3) - 3\mu E(X^2) + 2\mu^3$, $E((X-\mu)^4) = E(X^4) - 4\mu E(X^3) + 2\mu^3 + 6\mu^2 E(X^2) - 3\mu^4$

Definição (Distribuição uniforme discreta)

Dizemos que a variável aleatória X segue uma distribuição **Uniforme Discreta** *de parâmetro n e escrevemos $X \sim Unif(n)$ ou $X \sim U(n)$, se a função de probabilidade de X é dada por:*

$$X \left\{ \begin{array}{cccc} 1 & 2 & \ldots & n \\ \frac{1}{n} & \frac{1}{n} & \ldots & \frac{1}{n} \end{array} \right. \qquad ou \qquad P(X = x) = \frac{1}{n}\,, \quad x = 1, \ldots, n.$$

Proposição

Seja a v.a. $X \sim U(n)$. Então, $E(X) = \frac{n+1}{2}$, $V(X) = \frac{n^2-1}{12}$ e

$$F(x) = \begin{cases} 0, & x < 1 \\ \frac{k}{n}, & k \leq x < k+1, \quad k = 1, \ldots, n-1 \\ 1, & x \geq n \end{cases} .$$

$\boxed{\textbf{Exemplo:}}$ Seja X a variável aleatória que representa o número de pontos que saem no lançamento de um dado equilibrado, com as faces numeradas de 1 a 6.

Temos então,

- $X \sim U(6)$;
- $E(X) = 7/2 = 3.5$;
- $V(X) = (6^2 - 1)/12 = 35/12$;
- A probabilidade de sair um número maior ou igual a 5 é $P(X \geq 5) = 2/6 = 1/3$.

Exemplo:

- Num aquário existem 9 peixes, dos quais 5 estão **saudáveis** (S) e os restantes 4 estão **doentes** (D).
- Considere a experiência aleatória: **extracção ao acaso e sem reposição de 3 peixes e registo do seu estado de saúde**.
- Associada a esta experiência aleatória, considere a variável aleatória

X - número de peixes saudáveis na amostra extraída de 3 peixes.

Determine a função de probabilidade de X.

Definição (Distribuição Hipergeométrica)

Considere uma população de N *elementos, dos quais* M *possuem determinada característica e os restantes* $(N-M)$ *não a possuem* (**dicotomia**)*. Considere a experiência aleatória que consiste em seleccionar ao acaso e* **sem reposição** n *elementos (amostra). Associada a esta experiência aleatória, defina a v.a.* X *- nº de elementos com a característica, na amostra seleccionada sem reposição. Esta v.a.* X *tem uma função de probabilidade,*

$$P(X=k)=\frac{\binom{M}{k}\binom{N-M}{n-k}}{\binom{N}{n}}, \qquad max(0, M+n-N) \leq k \leq min(M, n).$$

e diz-se ter distribuição **Hipergeométrica** *de parâmetros* (N, M, n) *(pode ser escrito abreviadamente* $X \sim H(N, M, n)$*).*

Observação: Considere a v.a. $X \sim \mathrm{H}(N, M, n)$. Então:

$$E(X)=n\frac{M}{N}; \qquad V(X)=n\frac{M}{N^2(N-1)}(N-M)(N-n);$$

Distribuição de Bernoulli

Definição (Prova de Bernoulli)

Experiência aleatória com dois possíveis resultados ("Sucesso" ou "Insucesso").

Definição (Distribuição de Bernoulli)

Seja X a v.a. que toma o valor 1 se o resultado da experiência é "Sucesso" e 0 se é "Insucesso", traduzindo a **dicotomia** *dos resultados. Denotando $p = P(\text{“}Sucesso\text{”}) > 0$, a função de probabilidade de X é:*

$$X \begin{cases} 0 & 1 \\ 1-p & p \end{cases}, \qquad 0 < p < 1$$

Dizemos que X segue uma distribuição de **Bernoulli** *de parâmetro p e escrevemos $X \sim Ber(p)$.*

Proposição

Seja a v.a. $X \sim Ber(p)$. Então: $E(X) = p$ e $V(X) = p(1-p)$.

Exercício: Num determinado percurso de avião, a probabilidade de uma pessoa qualquer que aí viaje pedir uma refeição vegetariana é de 0.2. Supondo que em determinado dia viajam 10 pessoas no avião, calcule a probabilidade de:

(a) Ninguém pedir refeição vegetariana.

(b) Todos pedirem refeição vegetariana.

(c) Pelo menos uma pedir refeição vegetariana.

(d) Duas pessoas pedirem a refeição vegetariana.

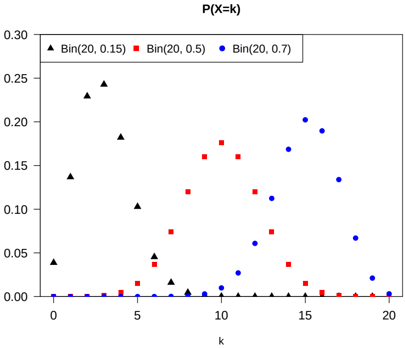
Definição (Distribuição Binomial)

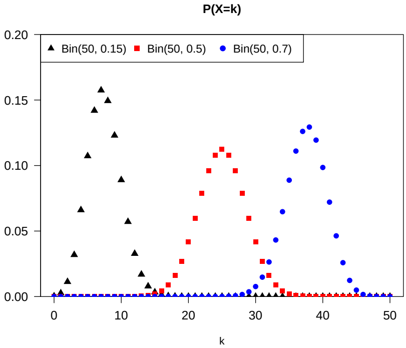
Considere uma sucessão de provas de Bernoulli independentes, onde em cada prova a probabilidade de "sucesso", p, é constante. A v.a.

$$X = \text{ "número de sucessos em } n \text{ provas de Bernoulli"}$$

segue uma distribuição **Binomial** *de parâmetros n e p, e escrevemos $X \sim B(n,p)$. A função de probabilidade é:*

$$P(X = k) = \binom{n}{k} p^k (1-p)^{n-k}, \qquad k = 0, 1, \ldots, n \qquad 0 < p < 1$$





Proposição

Considere a v.a. $X \sim B(n, p)$. *Então,*

$$E(X) = np \qquad \text{e} \qquad V(X) = np(1-p);$$

Teorema (Propriedade aditiva da distribuição Binomial)

Sejam $X_1, X_2, \ldots, X_m$, m *v.a.'s independentes tais que* $X_i \sim B(n_i, p)$, $i = 1, \ldots, m$. *Então, a* **soma**

$$S_m = X_1 + X_2 + \ldots + X_m = \sum_{i=1}^{m} X_i$$

tem também distribuição Binomial, isto é,

$$S_m = \sum_{i=1}^{m} X_i \ \sim \ B(n_1 + \ldots + n_m, p).$$

Diferença entre as distribuições Hipergeométrica e Binomial

- Hipergeométrica: Sucessivas extracções **não** são independentes;
- Binomial: Sucessivas extracções são independentes;

Aproximação da Hipergeométrica pela Binomial

Exemplo: Num lago com $N = 1000$ peixes, $M = 50$ estão saudáveis e os restantes $N - M = 950$ não estão saudáveis.

Seja X o número de peixes saudáveis numa amostra de $n = 30$ peixes, extraídos ao acaso **sem reposição**.

Então $X \sim H(1000, 50, 30)$

$$P(X=4) = \frac{\binom{50}{4}\binom{950}{26}}{\binom{1000}{30}} = 0.04386$$

Seja Y o número de peixes saudáveis numa amostra de $n = 30$ peixes, extraídos ao acaso **com reposição**.

Então $Y \sim B(30, 0.05)$

$$P(Y=4) = \binom{30}{4} 0.05^4 0.95^{26} = 0.04514$$

Aproximação da distribuição Hipergeométrica pela Binomial: Seja $X \sim H(N, M, n)$. Se $\frac{n}{N} \leq 0.1$ (o tamanho da amostra for muito pequeno, em relação ao tamanho da população), podemos aproximar a função de probabilidade (f.p.) de X pela f.p. da distribuição $B(n, p = M/N)$.

42 A probabilidade de um atirador acertar no alvo é 0.6. Calcule a probabilidade de:

(a) em cinco tiros, acertar três.
(b) acertar pela terceira vez ao quinto tiro.
(c) serem necessários exactamente 10 tiros para acertar um.
(d) necessitar de, pelo menos, 4 tiros para acertar 2.

Solução: 0.3456, 0.20736, 0.000157, 0.352

Definição (Distribuição Geométrica)

Considere uma sucessão de **provas de Bernoulli** *(experiências aleatórias cujo resultado é sucesso ou insucesso)* **independentes**, *onde em cada prova a probabilidade de "sucesso", p, é constante. A v.a.*

$$X = \text{"número de provas necessárias até ocorrer o primeiro sucesso"}$$

tem **distribuição Geométrica** *de parâmetro p, e escrevemos $X \sim G(p)$.*

Definição (Função de probabilidade)

A função de probabilidade é:

$$P(X = x) = p(1-p)^{x-1}, \qquad x = 1, 2, \ldots, \qquad 0 < p < 1$$

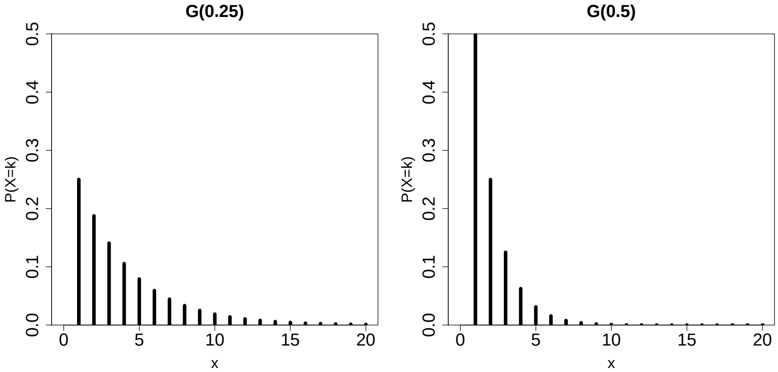


Figura: Função de probabilidade de uma v.a. $G(p)$, com $p = 0.25$ (esquerda) e $p = 0.5$ (direita)

Distribuição Geométrica

A função de distribuição é:

$$F(x) = P(X \leq x) = \sum_{j=1}^{[x]} P(X = j) = 1 - (1-p)^{[x]},\ x \geq 1,\ 0 < p < 1.$$

Proposição

$E(X) = \frac{1}{p}, \quad V(X) = \frac{1-p}{p^2}$

Como as provas de Bernoulli são independentes, a contagem do número de provas necessárias até ao próximo sucesso pode ser recomeçada em qualquer prova, sem que isso altere a distribuição da variável aleatória.

Proposição (Propriedade da falta de memória)

Para qualquer $k > v, \quad P(X > k | X > v) = P(X > k - v).$

Distribuição de Poisson

Definição: Processo de Poisson

Seja X a variável aleatória que conta o número de ocorrências de um acontecimento num dado intervalo de tempo[3] de duração t . Temos um processo de Poisson de parâmetro $\lambda > 0$, quando:

1. A probabilidade p de ocorrer exactamente um acontecimento num intervalo de amplitude arbitrariamente pequena d é proporcional à sua duração, isto é, $p = \lambda d$;
2. A probabilidade de ocorrer mais do que um acontecimento num intervalo de amplitude arbitrariamente pequena é aproximadamente igual a zero;
3. O número de acontecimentos que ocorrem em dois intervalos disjuntos são independentes.
4. O número de ocorrências em dois intervalos com a mesma duração, têm a mesma distribuição.

[3] Note que podemos também considerar uma área, um volume, etc.

Distribuição de Poisson

Função de probabilidade do processo de Poisson

Seja $X(t)$ o número de ocorrências de um acontecimento num dado intervalo de tempo de duração t. Então:

$$P(X(t)=x)=e^{-\lambda t}\frac{(\lambda t)^x}{x!}, \qquad x=0,1,2,\ldots, \quad \lambda>0.$$

Nota: Num processo de Poisson, os acontecimentos acorrem a uma taxa média λ, por unidade de tempo.

Exemplo (p. 47): Num processo de fabricação de placas de vidro produzem-se pequenas bolhas que se distribuem aleatoriamente pelas placas, com uma densidade média de 0.4 bolhas$/m^2$. Admitamos que $X(t)$ regista o número de bolhas observadas em placas com $t\ m^2$ e que $X(t)$ é um processo de Poisson de parâmetro $\lambda=0.4$ bolhas$/m^2$. Determine a probabilidade de, numa placa com $4.5m^2$, haver pelo menos uma bolha.

Distribuição de Poisson

Distribuição de Poisson

A variável aleatória X segue uma distribuição de Poisson de parâmetro λ, e escrevemos $X \sim P(\lambda)$, se a função de probabilidade de X é:

$$P(X = x) = e^{-\lambda} \frac{\lambda^x}{x!}, \qquad x = 0, 1, 2, \ldots, \quad \lambda > 0.$$

Teorema (Propriedade aditiva da Poisson)

Sejam $X_1, X_2, \ldots, X_n$ variáveis aleatórias **independentes** *com distribuição Poisson, $X_i \sim P(\lambda_i)$, $i = 1, \ldots, n$. Então,*

$$S_n = \sum_{i=1}^{n} X_i \sim P(\lambda_1 + \ldots + \lambda_n)$$

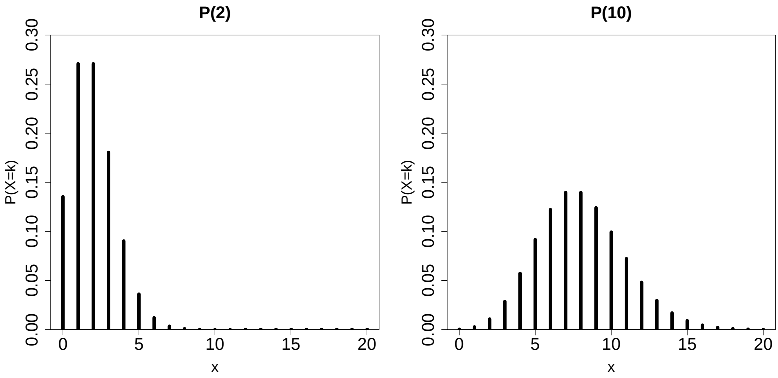


Figura: Função de probabilidade de uma v.a. $P(\lambda)$, com $\lambda = 2$ (esquerda) e $\lambda = 10$ (direita)

Seja X uma v.a. tal que $X \sim B(n,p)$. É possível de verificar que

$$\lim_{\substack{n\to\infty \\ np\to\lambda}} \binom{n}{x} p^x (1-p)^{(n-x)} = e^{-\lambda} \frac{\lambda^x}{x!}, \quad x = 0, 1, 2, \ldots$$

Então, caso $n \geq 50$ e $np \leq 5$, podemos aproximar a distribuição **Binomial** pela distribuição de **Poisson** com $\lambda = np$, isto é, podemos usar a aproximação da função de probabilidade

$$\binom{n}{x} p^x (1-p)^{(n-x)} \approx e^{-np} \frac{(np)^x}{x!}.$$

Exemplo: Numa feira popular a probabilidade de uma pessoa contrair uma intoxicação alimentar é de 0.0005. Determine a probabilidade de, em 300 pessoas, 2 ficarem intoxicadas.

- Considere a v.a. $X =$ número de pessoas que contraem uma intoxicação alimentar na feira, entre as 300 pessoas.
- $X \sim B(300, 0.0005)$;
- $P(X=2) = \binom{300}{2} 0.0005^2 (1-0.0005)^{300-2} = 0.009659984$ (valor exacto);
- $P(X=2) \approx e^{-0.15} \frac{0.15^2}{2!} = 0.009682965$ (valor aproximado);

Nota: Só devemos usar a aproximação quando não conseguimos calcular o valor exacto da probabilidade.

Sejam $X_1, X_2, \ldots, X_p$ p variáveis aleatórias. Então $\mathbf{X} = (X_1, X_2, \ldots, X_p)$ é um vector aleatório de dimensão p. Vamos restringir-nos apenas aos vectores aleatórios com dois elementos ($p = 2$), ditos **pares aleatórios**, representados por (X, Y).

Definição (Par aleatório discreto)

(X, Y) *é um* **par aleatório discreto** *se e só se* X *e* Y *são variáveis aleatórias discretas.*

Então, o conjunto (suporte)

$$D = \{(x, y) \in \mathbb{R}^2 : P(X = x, Y = y) > 0\}$$

é, quando muito, numerável e $P((X, Y) \in D) = 1$.

Par aleatório

Definição (Função de probabilidade conjunta)

Seja (X,Y) *um* **par aleatório discreto** *tomando valores no conjunto*

$$D = \{\, (x,y) \in \mathbb{R}^2 : \; P(X=x, Y=y) > 0 \,\}.$$

Chamamos **função de probabilidade conjunta** *(f.p.c.) de* (X,Y) *à função:*

$$P(X=x, Y=y), \quad i = 1,2,\ldots, \; j = 1,2,\ldots$$

Propriedades da função de probabilidade conjunta:

- $0 \leq P(X=x, Y=y) \leq 1, \quad \forall (x,y)$
- $\sum\sum_{(x,y)\in D} P(X=x, Y=y) = 1$

Exercício: verifique se $P(X = x, Y = y) = (x + 2y)/24$, $x = -1, 1$, $y = 1, 2, 3$ é uma função de probabilidade.

Observação: Quando o conjunto D é finito e pequeno é costume representarmos a função de probabilidade conjunta (f.p.c.) numa tabela.

Exemplo: Seja (X, Y) um par aleatório discreto com a seguinte f.p.c.:

$X \setminus Y$	0	1	2
0	$1/4 = P(X=0, Y=0)$	$1/8 = P(X=0, Y=1)$	$0 = P(X=0, Y=2)$
1	$1/8 = P(X=1, Y=0)$	$1/8 = P(X=1, Y=1)$	$1/8 = P(X=1, Y=2)$
2	$0 = P(X=2, Y=0)$	$0 = P(X=2, Y=1)$	$1/4 = P(X=2, Y=2)$

Cálculo de probabilidades: Para qualquer $I \subseteq \mathbb{R}^2$,

$$\boxed{P\left((X,Y) \in I\right) = P\left((X,Y) \in (D \cap I)\right) = \sum_{(x,y)\in(D\cap I)}\sum P(X = x, Y = y).}$$

Definição (função de probabilidade marginal)

Dado um par aleatório discreto (X,Y), definimos a **função de probabilidade marginal de** X *como*

$$P\left(X = x\right) = \sum_{y} P\left(X = x, Y = y\right),$$

e a **função de probabilidade marginal de** Y *como*

$$P\left(Y = y\right) = \sum_{x} P\left(X = x, Y = y\right).$$

Nota: *Nos somatórios anteriores, basta considerar valores x e y tais que $(x,y) \in D$.*

Par aleatório

Definição (Função de probabilidade condicional de X dado $Y = y$ (fixo))

$$P(X = x|Y = y) = \frac{P(X = x, Y = y)}{P(Y = y)}, \quad \textit{se} \quad P(Y = y) > 0.$$

De modo análogo, podemos definir a **função de probabilidade condicional** de Y dado $X = x$.

Definição (Independência entre v.a.'s discretas)

As v.a.'s X e Y dizem-se **independentes** *se e só se,*

$$P(X = x, Y = y) = P(X = x)P(Y = y), \qquad \forall (x, y) \in \mathbb{R}^2.$$

Nota: É suficiente verificar a definição anterior para qualquer $(x, y) \in D$.

Exemplo: Seja (X, Y) um par aleatório discreto com a seguinte f.p.c.:

$X \setminus Y$	0	1	2	
0	1/4	1/8	0	3/8
1	1/8	1/8	1/8	3/8
2	0	0	1/4	1/4
	3/8	1/4	3/8	

1. Qual a probabilidade de X ser maior que Y?
2. Calcule $P(X \leq 1,\, Y > 0)$.
3. Determine a função de probabilidade de $X|Y = 2$ e calcule $E(X|Y = 2)$.
4. X e Y são v.a.'s independentes? Justifique a sua resposta.

Definição

Seja (X, Y) um par aleatório e $g: \mathbb{R}^2 \to \mathbb{R}$ uma função. Define-se o **valor médio** *ou* **valor esperado** *ou* **média** *de $g(X, Y)$ como:*

$$E(g(X,Y)) = \begin{cases} \displaystyle\sum\sum_{(x,y)\in D} g(x,y)P(X=x, Y=y) & \text{se } X \text{ e } Y \text{ são} \\ & \text{v.a.'s discretas;} \\ \displaystyle\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} g(x,y) f_{X,Y}(x,y)\,dx\,dy & \text{se } X \text{ e } Y \text{ são} \\ \text{não é estudado} & \text{v.a.'s contínuas;} \end{cases}$$

Definição (Covariância)

Seja (X, Y) *um par aleatório. Define-se* **covariância** *entre* X *e* Y *por:*

$$\mathrm{Cov}\,(X, Y) = \mathrm{E}\left[(X - \mathrm{E}(X))\,(Y - \mathrm{E}(Y))\right],$$

Proposição

Caso exista a $\mathrm{Cov}(X, Y)$*, esta pode ser calculada através da fórmula:*

$$\mathrm{Cov}(X, Y) = \mathrm{E}(XY) - \mathrm{E}(X)\mathrm{E}(Y).$$

Proposição

Se X *e* Y *são* **independentes**, $\mathrm{E}(XY) = \mathrm{E}(X)\mathrm{E}(Y)$ *e* $\mathrm{Cov}(X, Y) = 0$.

Interpretação da Covariância

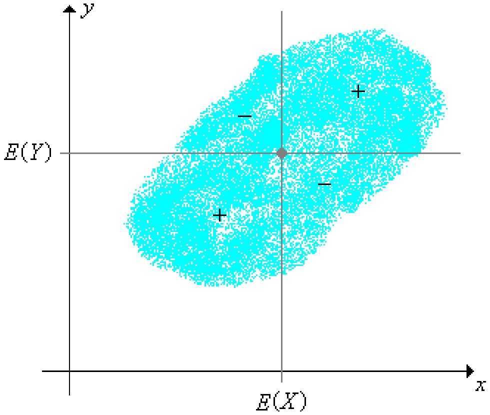


Figura: Fonte: http://www.math.uah.edu/stat/expect/Covariance.png

Proposição

Sejam X, Y e Z *v.a.'s,* a e b, *constantes reais. Então:*

- $\text{Cov}(X,Y) = \text{Cov}(Y,X)$;
- $\text{Cov}(X,X) = V(X)$;
- $\text{Cov}\,(X+Y,Z) = \text{Cov}\,(X,Z) + \text{Cov}\,(Y,Z)$;
- $\text{Cov}\,(aX+b,Y) = a\text{Cov}\,(X,Y)$.

Teorema

Sejam X e Y *variáveis aleatórias. Então,*

$$\text{V}(X \pm Y) = \text{V}(X) + \text{V}(Y) \pm 2\text{Cov}(X,Y)$$

Definição (Coeficiente de correlação)

Define-se o **coeficiente de correlação** *de* (X, Y) *por*

$$\rho\,(X, Y) = \frac{\mathrm{Cov}(X, Y)}{\sqrt{\mathrm{V}\,(X)\,\mathrm{V}\,(Y)}}.$$

Propriedades:

- $-1 \leq \rho\,(X, Y) \leq 1$;
- $|\rho\,(X, Y)| = 1$ se e só se $P\,(Y = a + bX) = 1$, sendo a e b constantes reais;
- Se X e Y são v.a.'s **independentes**, $\rho(X, Y) = 0$.

Interpretação da Correlação

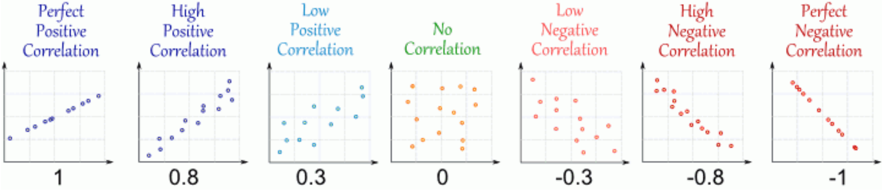


Figura: Fonte: https://www.mathsisfun.com/data/images/correlation-levels.gif

Observação: Como regra empírica, podemos considerar que existe forte relação linear positiva (negativa) entre X e Y se $\rho(X,Y) \geq 0.7$ ($\rho(X,Y) \leq -0.7$).

Definição (Variável aleatória contínua)

Uma v.a. X *diz-se do tipo (absolutamente) contínuo ou simplesmente* **contínua** *se*

$$D = \{a \in \mathbb{R} : \ P(X = a) > 0\} = \varnothing,$$

e se existe uma função não negativa, f*, tal que para* $I \subseteq \mathbb{R}$*,*

$$P(X \in I) = \int_I f(x)dx.$$

À função f *chamamos* **função densidade de probabilidade** *ou simplesmente* **função densidade***.*

Notas:

- Ao conjunto $S = \{x \in \mathbb{R} : \ f(x) > 0\}$ é chamado **suporte** de X.
- $F(x) = P(X \leq x) = P(X \in]-\infty, x])$.

Propriedades da função densidade de probabilidade:

1. $f(x) \geq 0, \quad \forall x \in \mathbb{R};$
2. $\int_{-\infty}^{+\infty} f(x)\, dx = 1.$

Observações:

- Por definição, $f(x) = F'(x)$, nos pontos onde existe derivada. Se não existir derivada, $f(x) = 0$.
- $P(a \leq X \leq b) = \int_a^b f(x)\, dx = F(b) - F(a)$
- $P(a \leq X \leq b) = P(a < X \leq b) = P(a \leq X < b) = P(a < X < b), \ a \leq b.$
- $P(X = x) = 0, \quad \forall x \in \mathbb{R}.$

Definição

Seja X uma v.a. contínua e g uma função real. Então,

- *o* **valor médio** *de X é:*

$$\mathrm{E}(X) = \int_{-\infty}^{+\infty} x f(x) dx,$$

- *o* **valor médio** *de $g(X)$ é:*

$$\mathrm{E}(g(X)) = \int_{-\infty}^{+\infty} g(x) f(x) dx,$$

Observação: O valor médio só está definido se o integral for absolutamente convergente.

Não é estudado no ano 2015/16.

Definição (Par aleatório contínuo)

Um par aleatório (X, Y) diz-se contínuo se existe uma função não negativa $f_{X,Y}$ tal que, tal que, para qualquer região $I \subset \mathbb{R}^2$,

$$P((X,Y) \in I) = \int\int_I f_{X,Y}(u,v)dudv.$$

A $f_{X,Y}$ chamamos **função densidade probabilidade conjunta** *ou apenas* **função densidade conjunta**.

Observação: A função $f_{X,Y}$ satisfaz as seguintes condições:

(i) $f_{X,Y}(x,y) \geq 0, \quad \forall (x,y) \in \mathbb{R}^2$;

(ii) $\int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f_{X,Y}(x,y)dxdy = 1$.

Definição (Função densidade de probabilidade marginal)

Define-se a **função densidade de probabilidade marginal** *de* X, *e de* Y *como:*

$$\begin{aligned} f_X(x) &= \int_{-\infty}^{+\infty} f_{(X,Y)}(x,y)\,dy, \quad \forall x \in \mathbb{R} \\ f_Y(y) &= \int_{-\infty}^{+\infty} f_{(X,Y)}(x,y)\,dx, \quad \forall y \in \mathbb{R} \end{aligned}$$

Proposição (Função densidade condicional de X)

Em todos os pontos (x, y) *onde* $f_{X,Y}$ *é contínua,* $f_Y(y) > 0$ *e é contínua, a função densidade condicional de* X*, dado* $Y = y$*, existe e calcula-se como:*

$$f_{X|Y}(x|y) = \frac{f_{X,Y}(x, y)}{f_Y(y)}.$$

Proposição (Função densidade condicional de Y)

Em todos os pontos (x, y) *onde* $f_{X,Y}$ *é contínua,* $f_X(x) > 0$ *e é contínua, a função densidade condicional de* Y*, dado* $X = x$*, existe e calcula-se como:*

$$f_{Y|X}(y|x) = \frac{f_{X,Y}(x, y)}{f_X(x)}.$$

Definição (Independência entre v.a.'s contínuas)

Seja (X, Y) *um par aleatório contínuo. As variáveis* X *e* Y *dizem-se* **independentes** *se e só se*

$$f_{X,Y}(x, y) = f_X(x)\, f_Y(y), \quad \forall (x, y) \in \mathbb{R}^2$$

Exemplo: Os tempos de vida, em centenas de horas, das duas componentes principais de um sistema de controlo são v.a.'s (X, Y) com função densidade conjunta:

$$f_{X,Y}(x, y) = \begin{cases} cx^2y & 0 < x < 3, 0 < y < 2 \\ 0 & \text{outros valores de } (x, y) \in \mathbb{R}^2 \end{cases}$$

a) Qual o valor de c?

b) Qual a probabilidade de cada uma das componentes durar mais de 100 horas?

c) Qual a probabilidade da 1ª componente durar mais de 100 horas?

d) Os tempos de vida das componentes são independentes?

Solução: a) $c = \frac{1}{18}$ b) $\frac{13}{18}$ c) $\frac{26}{27}$

Definição (Quantil)

O quantil de ordem p, q_p, da v.a. X é a solução da equação

$$F(q_p) = p \quad \Leftrightarrow \quad P(X \leq q_p) = p.$$

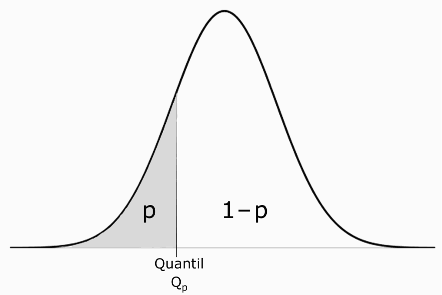


Figura: Quantil q_p de uma v.a. contínua.

Observações:

- Se X é uma v.a. **contínua** e F tem função inversa F^{-1} a equação é equivalente a

$$F(q_p) = p \quad \Leftrightarrow \quad q_p = F^{-1}(p), \quad 0 < p < 1;$$

- Se X é uma v.a. **discreta**, a equação $F(q_p) = p$ só tem solução para alguns valores de p;

- Caso a solução equação $F(q_p) = p$ não seja dada por um dos dois pontos anteriores, consideramos $q_p = \min\{x : F(x) \geq p\}$.

Definição (Mediana)

Trata-se do quantil de ordem $p = 1/2$. Costuma-se representar a mediana, da v.a. X, por $med(X)$.

Definição (Distribuição Uniforme Contínua)

Dizemos que a variável aleatória X segue uma **distribuição Uniforme** *no intervalo $[a, b]$, $a < b$, e escrevemos $X \sim U(a, b)$, se a função densidade de probabilidade de X é dada por:*

$$f(x) = \begin{cases} \frac{1}{b-a}\ , & a \leq x \leq b \\ 0\ , & \textit{caso contrário} \end{cases}$$

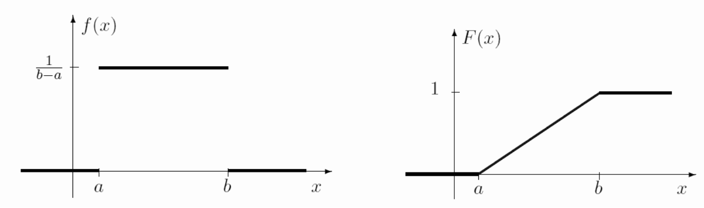


Figura: Função densidade de probabilidade de probabilidade (esquerda) e função de distribuição (direita) de uma v.a. $X \sim U(a, b)$.

Distribuição Uniforme Contínua

A distribuição Uniforme é utilizada para caracterizar uma variável que varia aleatoriamente num intervalo limitado e quaisquer dois sub-intervalos com a mesma amplitude têm a mesma probabilidade de conter uma observação.

Proposição

Considere a v.a. $X \sim U(a,b)$. *Então:*

- $E(X) = \frac{a+b}{2}$;
- $V(X) = \frac{(b-a)^2}{12}$;
- $F(x) = \begin{cases} 0, & x < a \\ \frac{x-a}{b-a}, & a \leq x < b \\ 1, & x \geq b \end{cases}$.

Simulação

É uma técnica que permite estudar o comportamento de modelos probabilísticos. É usada frequentemente em situações onde é muito difícil, ou até mesmo impossível, obter uma solução exacta do modelo. Iremos apenas referir a simulação estocástica pelo método de Monte Carlo (MMC).

O MMC é usado em diversas áreas, sendo por exemplo aplicado em:

- Resolução de integrais e equações diferenciais;
- Controlo de tráfego;
- Previsão de Índices Financeiros;
- Modelação de materiais;
- Aerodinâmica;

Geração de números pseudo-aleatórios

Um método de gerar os valores aleatórios do intervalo $]0,1[$, consiste em colocar numa urna 10 bolas numeradas de 0 a 9 e proceder à extracção, com reposição, de m bolas. Os algarismos extraídos constituem as m casas decimais do número aleatório. Mas é um método muito demorado!

Actualmente, a simulação é feita com computadores, utilizando números gerados por um processo determinístico, mas com um comportamento semelhante aos números efectivamente aleatórios. Estes números designam-se por **números pseudo-aleatórios** (NPA's).

Qualquer programa computacional que sirva para gerar NPA's deve satisfazer os seguintes requisitos:

- Deve ser **rápido**;
- Deve ter **ciclos longos**;
- Os números gerados devem poder ser **reproduzidos** (com a mesma semente) e devem apresentar um **comportamento independente** e **uniforme**;

Geração de números pseudo-aleatórios

Definição (Método Congruencial)

Permite gerar uma sequência de números inteiros no intervalo $[0, m-1]$*, através da seguinte fórmula recursiva:*

$$x_0 = \text{"semente"}$$
$$x_i = (a\, x_{i-1} + c) (\operatorname{mod} m), \qquad i = 1, 2, \ldots$$

onde a, c *e* m *são números inteiros* (a, $c < m$) *e* $x (\operatorname{mod} m)$ *representa o resto da divisão inteira de* x *por* m*. O comprimento do ciclo é menor ou igual a* m*.*

Exemplo: Seja $a = 5$, $c = 7$ e $m = 8$ e $x_0 = 0$. Os primeiros números são:

i	**0**	1	2	3	4	5	6	7	**8**	9	10	11	12	13	14	15	**16**	17	18	19
x_i	**0**	7	2	1	4	3	6	5	**0**	7	2	1	4	3	6	5	**0**	7	2	1

Neste exemplo, o comprimento do ciclo é 8.

Nota: A conversão de números inteiros em $[0, m-1]$, para valores no intervalo $]0, 1[$ é feita através da transformação: $u_i = (x_i + 1/2)/m, \quad i = 1, 2, \ldots$

Embora o método congruencial consiga gerar NPA's com qualidade aceitável, é necessário escolher convenientemente as constantes a, c e m. Por exemplo, o comprimento do ciclo é igual a m (máximo valor possível), e independente da semente x_0, se

- $a = 4c + 1$, com $c \in \mathbb{N}$ e c é ímpar;
- $m = 2^k$, com $k \in \mathbb{N}$.

Tabela: Algumas implementações do método Congruencial

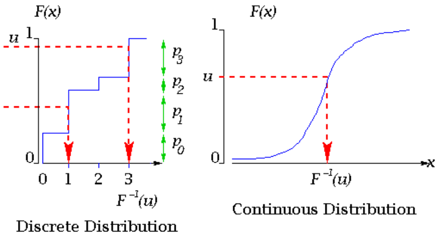
	m	a	c
Turbo Pascal 3	2^{32}	129	907633385
Borland Delphi, Virtual Pascal	2^{32}	134775813	1
Numerical Recipes in C	2^{32}	1664525	1013904223
Borland C/C++	2^{32}	22695477	1
ANSI C: Watcom, Digital Mars, CodeWarrior	2^{32}	1103515245	12345
Microsoft Visual/Quick C/C++	2^{32}	214013	2531011

Teorema (Teorema da transformação uniformizante)

Seja X uma v.a. contínua com função de distribuição F. Então

$$F(X) \sim U(0,1).$$

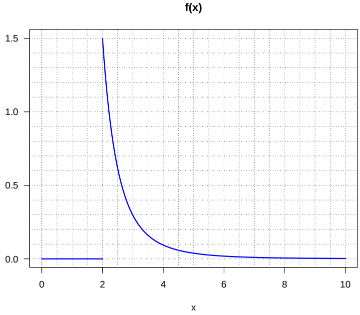
Logo, para obter uma observação de X, basta calcular $F^{-1}(u)$, $u \in]0,1[$.
Se não existir F^{-1}, consideramos $F^{-1}(u) = \inf\{x : F(x) \geq u\}$.

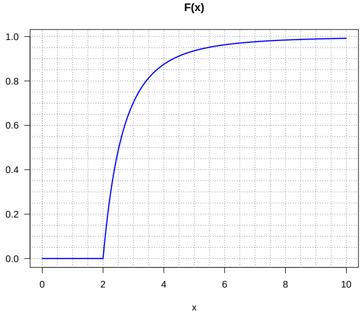


Fonte: http://www.iro.umontreal.ca/ vazquez/SimSpiders/GenerRV/Methods/inverse.gif

Exemplo: Considere a variável aleatória X com função densidade f e função de distribuição F dadas por

$$f(x) = \begin{cases} 0, & x \leq 2 \\ \frac{24}{x^4}, & x > 2 \end{cases} \qquad F(x) = \begin{cases} 0, & x \leq 2 \\ 1 - \frac{8}{x^3}, & x > 2 \end{cases}$$





Para gerar uma sequência de n NPA's é necessário calcular $x_i = 2/\sqrt[3]{1-u_i}$, $i = 1, 2, \ldots, n$ sendo u_1, u_2,$\ldots$,u_n uma sequência de NPA's uniformes em $]0, 1[$.

Seja X uma variável aleatória discreta com função de probabilidade,

$$p_j = P(X = x_j), \quad j = 1, 2, \ldots, k \qquad \text{com} \qquad \sum p_j = 1.$$

Então a sua função de distribuição é dada por:

$$F(x) = \begin{cases} 0, & x < x_1 \\ p_1, & x_1 \leq x < x_2 \\ p_1 + p_2, & x_2 \leq x < x_3 \\ \vdots & \\ 1, & x \geq x_k \end{cases}$$

Para gerar um NPA w desta distribuição, a partir de um valor u uniforme, basta escolher:

$$w = \begin{cases} x_1, & 0 < u \leq p_1 \\ x_2, & p_1 < u \leq p_1 + p_2 \\ \vdots & \\ x_k, & \sum_{i=1}^{k-1} p_i < u < 1 \end{cases}$$

Geração de números pseudo-aleatórios

1. Uma das distribuições, usada no estudo do tempo de vida de componente mecânicos, é a distribuição $Weibull(\alpha, \beta)$, com função de distribuição:

$$F(x) = \begin{cases} 0, & x \leq 0 \\ 1 - e^{-\alpha x^{\beta}}, & x > 0 \end{cases} \quad \text{com} \quad \alpha > 0, \beta > 0.$$

   (a) Indique como gerar observações da distribuição $Weibull(3, 5)$, a partir de NPA's uniformes no intervalo]0,1[.
   (b) Suponha que foram gerados os valores $u_1 = 0.2517$, $u_2 = 0.7437$ e $u_3 = 0.1911$ da $U(0, 1)$. Calcule os valores gerados da $Weibull(3, 5)$.

2. Considere a distribuição Cauchy (trata-se da distribuição t de Student com 1 grau de liberdade) com função densidade

$$f(x) = \frac{1}{\pi(1 + x^2)}, \quad x \in \mathbb{R}.$$

   (a) Verifique que a sua função de distribuição é $F(x) = \frac{1}{\pi} \arctan(x) + \frac{1}{2}$, $x \in \mathbb{R}$.
   (b) Indique como gerar observações da distribuição Cauchy, a partir de valores gerados para a distribuição $U(0, 1)$.
   (c) Imagine que gerámos os NPA's $u_1 = 0.6235$ e $u_2 = 0.4515$ da $U(0, 1)$. Quais os valores gerados para a distribuição Cauchy?

3 (Teste de P.E. 2007/08) Considere a v.a. X com função densidade, $f(x) = 2x, \quad 0 < x < 1$. Sejam $u_1 = 0.2517$ e $u_2 = 0.7437$ dois números pseudo-aleatórios da distribuição $U(0,1)$. Usando o método da Transformação Inversa, calcule dois números pseudo-aleatórios da v.a. X. Observação: $P(X \leq x) = x^2, \quad 0 < x < 1$.

Solução dos exercícios

1 (a) Seja u um NPA da distribuição $U(0,1)$. Então, $x = \sqrt[5]{-\frac{\ln(1-u)}{3}}$ é um NPA da distribuição Weibull(3,5).
(b) Os NPA gerados são: $x_1 = 0.6267$, $x_2 = 0.8538$ e $x_3 = 0.5887$.

2 (a)
(b) Seja u um NPA da distribuição $U(0,1)$. Então, $x = \tan((u - 1/2)\pi)$ é um NPA da distribuição Cauchy.
(c) Os NPA gerados são: $u_1 = 0.4087$ e $u_2 = -0.1536$.

3 $x_1 = \sqrt{0.2517} = 0.5017 \qquad x_2 = \sqrt{0.7437} = 0.8624$

Definição (Distribuição Exponencial)

Uma variável aleatória X tem **distribuição Exponencial** *de parâmetros α e δ, e escrevemos $X \sim Exp(\alpha, \delta)$, se a sua função densidade probabilidade for dada por:*

$$f(x) = \begin{cases} \frac{1}{\delta}\, e^{-(x-\alpha)/\delta}, & x > \alpha, \\ 0\ , & x \leq \alpha, \end{cases} \qquad \alpha \in \mathbb{R}, \quad \delta > 0.$$

Proposição

Seja $X \sim Exp(\alpha, \delta)$. Então,

$$F(x) = \begin{cases} 0, & x \leq \alpha, \\ 1 - \ e^{-(x-\alpha)/\delta}, & x > \alpha. \end{cases}$$

Distribuição Exponencial

Coeficientes importantes: $\boxed{E(X) = \alpha + \delta}$ e $\boxed{V(X) = \delta^2}$.

Proposição (Relação entre a distribuição Exponencial e Poisson)

Considere um acontecimento que ocorre de acordo com um Processo de Poisson de parâmetro λ, por unidade de tempo. Então, o tempo até à primeira ocorrência e o tempo entre duas ocorrências consecutivas têm distribuição $Exp(0, 1/\lambda)$.

Teorema (Propriedade da falta de memória da distribuição Exponencial)

Seja $X \sim Exp(0, \delta)$. Então,

$$P(X \geq s + t | X \geq s) = P(X \geq t).$$

Distribuição Normal

Definição (Distribuição Normal)

A variável aleatória X diz-se seguir uma **distribuição Normal** *(ou Gaussiana) de parâmetros μ e σ^2, e escrevemos $X \sim N(\mu, \sigma^2)$, se a sua função densidade probabilidade for dada por:*

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma}\, e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \qquad x \in \mathbb{R}, \quad \mu \in \mathbb{R}, \quad \sigma > 0.$$

A Função de distribuição de $X \sim N(\mu, \sigma^2)$, é definida pelo integral

$$F(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}\sigma}\, e^{-\frac{(t-\mu)^2}{2\sigma^2}} dt,$$

para o qual não se conhece **solução analítica**.

Distribuição Normal

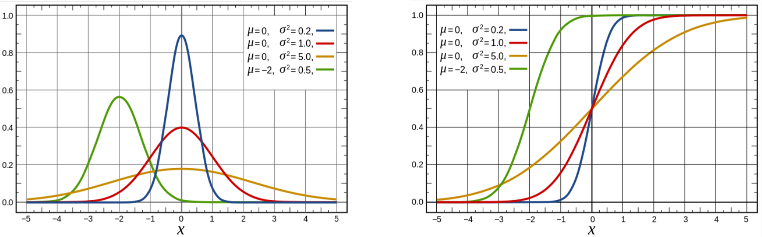


Figura: Função densidade (esquerda) e função de distribuição (direita) de $X \sim N(\mu, \sigma^2)$. Figuras retiradas de `https://en.wikipedia.org/wiki/Normal_distribution`

Proposição

Seja a v.a. $X \sim N(\mu, \sigma^2)$. *Então,*

$$E(X) = \mu \qquad e \qquad V(X) = \sigma^2.$$

Quando $\mu = 0$ e $\sigma = 1$, dizemos que a v.a. tem distribuição **Normal Reduzida** e representamos essa v.a. por Z, ou seja, $Z \sim N(0,1)$.

É usual representar as funções densidade de probabilidade e de distribuição da **Normal Reduzida** por ϕ e Φ, respectivamente.

No caso particular da **Normal Reduzida**, os valores da função de distribuição $\Phi(z) = P(Z \leq z)$ encontram-se **tabelados**.

Teorema

Se $X \sim N(\mu, \sigma^2)$ e a, b são constantes reais, com $a \neq 0$, então,

$$Y = aX + b \ \sim \ N(a\mu + b, a^2\sigma^2),$$

e

$$Z = \frac{X - \mu}{\sigma} \sim N(0,1). \quad \textit{(Normal Reduzida)}$$

Distribuição Normal

Teorema

Sejam $X_1, X_2, \ldots, X_n$ *variáveis aleatórias independentes tais que*

$$X_i \sim N\left(\mu_i, \sigma_i^2\right), \qquad i = 1, 2, \ldots, n.$$

Considerando as constantes $a_1, a_2, \ldots, a_n$*, com algum* $a_i \neq 0$*, temos:*

$$Y = a_1 X_1 + \ldots + a_n X_n \sim N\left(\mu_Y, \sigma_Y^2\right)$$

onde

$$\begin{aligned} \mu_Y &= E\left(\sum_{i=1}^{n} a_i X_i\right) = \sum_{i=1}^{n} a_i \mu_i = a_1\mu_1 + \ldots + a_n\mu_n \\ \sigma_Y^2 &= V\left(\sum_{i=1}^{n} a_i X_i\right) = \sum_{i=1}^{n} a_i^2 \sigma_i^2 = a_1^2\sigma_1^2 + \ldots + a_n^2\sigma_n^2 \end{aligned}$$

Corolário

Sejam $X_1, X_2, \ldots, X_n$ *variáveis aleatórias independentes tais que*

$$X_i \sim N\left(\mu, \sigma^2\right), \qquad i = 1, 2, \ldots, n.$$

Temos:

$$Y \;=\; X_1 + X_2 + \ldots + X_n \;\sim\; N\left(n\mu, n\sigma^2\right),$$

$$\overline{X} = \frac{X_1 + X_2 + \ldots + X_n}{n} \sim N\left(\mu, \frac{\sigma^2}{n}\right).$$

[1º teste de Estatística, 23 de Novembro de 2013]

4. Numa população, o nível de colesterol em adultos (medido em mg/dl) é uma variável aleatória, X, com distribuição Normal com parâmetros $\mu = 225$ e $\sigma = 75$.

(Extra) Calcule $P(X \leq 200)$ e $P(X \leq 350)$.

(a) Calcule a $P(200 \leq X \leq 350)$. [0.5828]

(b) Qual é o valor de colesterol, a partir do qual se encontra 2.5% da população adulta com colesterol mais elevado? [371.9973]

(c) Três adultos realizaram análises ao sangue para medir o nível de colesterol. Qual é a probabilidade da média dos níveis de colesterol dos três adultos ser inferior a 230. (Assuma que o nível de colesterol no sangue é independente de pessoa para pessoa) [0.546]

[1º Teste de Probabilidades e Estatística D 2005/06]

Um foguete espacial é constituído por 3 partes distintas, *cápsula*, *corpo* e *depósitos*. Representem as v.a.'s X, Y e W o peso da cápsula, o peso do corpo do foguete e o peso dos depósitos, respectivamente, em toneladas. Sabe-se que $X \sim N(5, 1)$, $Y \sim N(10, 2^2)$ e $W \sim N(7, 2^2)$, sendo as três variáveis independentes entre si.

1. Qual a probabilidade de o peso da cápsula estar compreendido entre 3 e 7 toneladas? [0.9544]
2. Qual o peso h que o corpo do foguete ultrapassa em 2.5% das vezes? [13.92]
3. Qual a probabilidade de o peso da cápsula mais o peso dos depósitos excederem o peso do corpo do foguete? [0.7486]

Distribuição do Qui Quadrado

Definição (Distribuição do Qui Quadrado)

Uma variável aleatória X tem distribuição do **Qui-quadrado** *com n graus de liberdade, e escrevemos $X \sim \chi^2_n$, se a sua função densidade de probabilidade for dada por:*

$$f(x) = \begin{cases} 0, & x \leq 0 \\ \frac{1}{\Gamma(n/2)2^{n/2}} e^{-x/2} x^{n/2-1}, & x > 0 \end{cases}$$

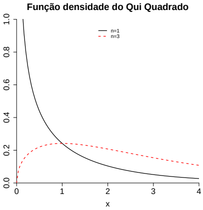


Figura: Função densidade de probabilidade de uma v.a. $X \sim \chi^2_n$.

Proposição

Considere a v.a. $X \sim \chi^2_n$. *Então,*

$$E(X) = n \quad \text{e} \quad V(X) = 2n.$$

Teorema

Sejam $X_1, X_2, \ldots, X_n$ *v.a.'s independentes com distribuição* $N(0,1)$. *Então,*

$$Y = X_1^2 + X_2^2 + \ldots + X_n^2 \sim \chi^2_n.$$

Definição (Distribuição t de Student)

Sejam $X \sim N(0,1)$ *e* $Y \sim \chi^2_n$, *com* X *e* Y *independentes. Então a seguinte variável aleatória:*

$$T = \frac{X}{\sqrt{Y/n}}$$

tem distribuição t-student com n *graus de liberdade e escrevemos* $T \sim t_n$. *A sua função densidade probabilidade é dada por:*

$$f(x) = \frac{\Gamma\left(\frac{n+1}{2}\right)}{\Gamma\left(\frac{n}{2}\right)\sqrt{n\pi}}(1 + x^2/n)^{-(n+1)/2}, \qquad x \in \mathbb{R}.$$

Distribuição t de Student

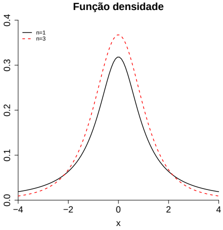


Figura: Função densidade de probabilidade de uma v.a. $X \sim t_n$.

Proposição

Seja $X \sim t_n$. *Então,*

$$E(X) = 0, \quad n > 1 \qquad e \qquad V(X) = \frac{n}{n-2}, \quad n > 2.$$

Teorema Limite Central

Teorema (Teorema Limite Central)

Seja $X_1, X_2 \ldots,$ uma sucessão de v.a.'s **i***ndependentes e* **i***denticamente* **d***istribuídas (***i.i.d.***), com valor médio μ e variância $\sigma^2 \neq 0$, finitos. Considere a v.a. $S_n = \sum_{i=1}^{n} X_i$ Então,*

$$Z_n = \frac{S_n - E(S_n)}{\sqrt{V(S_n)}} = \frac{S_n - n\mu}{\sqrt{n}\sigma} \overset{a}{\sim} N(0,1), \qquad n \to +\infty.$$

Observação: Na prática, este teorema usa-se quando temos $n \geq 30$.

Cálculo de probabilidades:

$$P(S_n \leq x) = P\left(\frac{S_n - n\mu}{\sqrt{n}\sigma} \leq \frac{x - n\mu}{\sqrt{n}\sigma}\right) \approx \Phi\left(\frac{x - n\mu}{\sqrt{n}\sigma}\right)$$

Exemplo

Num estudo sobre vendas num hipermercado, concluiu-se que a procura diária de arroz é uma v.a. com valor médio 40kg e desvio-padrão 5kg. Tendo sido encomendado 14500kg de arroz para venda no próximo ano, qual a probabilidade deste stock cobrir a procura de arroz nesse período?

Nota: considere que o hipermercado está aberto 364 dias.

Se no quociente que define Z_n atrás dividirmos tanto o numerador como o denominador por n, o T.L.C. passa a ser enunciado não em relação ao total, S_n, mas em relação à média das variáveis aleatórias X_i,

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i.$$

Teorema Limite Central

Corolário

Seja X uma v.a. com distribuição **Binomial** *de parâmetros n e p, isto é, $X \sim B(n, p)$. Se $n \geq 30$ e p tal que $np > 5$ e $n(1-p) > 5$, então:*

$$\frac{X - np}{\sqrt{np(1-p)}} \overset{a}{\sim} N(0, 1)$$

Corolário

Seja X uma v.a. com distribuição **Poisson** *de parâmetro λ, isto é, $X \sim P(\lambda)$. Se $\lambda > 5$, então:*

$$\frac{X - \lambda}{\sqrt{\lambda}} \overset{a}{\sim} N(0, 1)$$