Notas de aula

um tanto bagunçada ainda, sobre

Introdução às variáveis aleatórias

por J Donadelli

Sumário

§1 Variáveis Aleatórias

Se uma moeda é lançada 3 vezes com os resultados independentes. Qual é o número de caras ocorridas? Qual é a probabilidade de termos 2 caras?

resultado 1	resultado 2	resultado 3	Nº de caras
Ca	Ca	Ca	3
Ca	Ca	Co	2
Ca	Co	Ca	2
Ca	Co	Co	1
Co	Ca	Ca	2
Co	Ca	Co	1
Co	Co	Ca	1
Co	Co	Co	0

A probabilidade de ocorrerem exatamente 2 caras é 3/8. Muitas vezes estamos mais interessados numa característica numérica de um evento do que no evento propriamente dito, por exemplo

1. Qual o número de chamadas telefônicas que chegam a uma central em um intervalo de tempo?
2. Qual a distância da origem de um ponto escolhido no círculo unitário?
3. Qual a altura de um cidadão escolhido?
4. Qual o tempo de duração de uma lâmpada escolhida da linha de produção?

Essas são grandezas que dependem do resultado de um experimento, são chamadas de variáveis aleatórias.

Variável aleatória real: uma variável aleatória (v.a.) real de um modelo probabilístico $(\Omega, \mathcal{A}, \mathbb{P})$ é uma função que associa a cada elemento de um espaço amostral Ω um número real

$$X : \Omega \to \mathbb{R}$$

de modo que, para todo $t \in \mathbb{R}$,

$$[X \leqslant t] := \{\omega \in \Omega : X(\omega) \leqslant t\} \tag{1.0.1}$$

é um evento aleatório do modelo cuja probabilidade é

$$\mathbb{P}(X \leqslant t) := \mathbb{P}(\{\omega \in \Omega : X(\omega) \leqslant t\}).$$

Se uma função definida num espaço amostral a valores reais é ou não uma variável aleatória depende do espaço de eventos $\mathcal{A}$.

Comumente, usamos as letras maiúsculas finais do alfabeto X, Y, Z, W para variáveis aleatórias.

Exemplo 1. Se X é o número de caras em 3 lançamentos de uma moeda então

$$\begin{array}{lll} X((\mathrm{Ca},\mathrm{Ca},\mathrm{Ca})) = 3 & X((\mathrm{Ca},\mathrm{Co},\mathrm{Co})) = 1 & X((\mathrm{Co},\mathrm{Co},\mathrm{Ca})) = 1 \\ X((\mathrm{Ca},\mathrm{Ca},\mathrm{Co})) = 2 & X((\mathrm{Co},\mathrm{Ca},\mathrm{Ca})) = 2 & X((\mathrm{Co},\mathrm{Co},\mathrm{Co})) = 0 \\ X((\mathrm{Ca},\mathrm{Co},\mathrm{Ca})) = 2 & X((\mathrm{Co},\mathrm{Ca},\mathrm{Co})) = 1 & \end{array}$$

e

$$[X \leqslant t] = \begin{cases} \emptyset, & \text{se } t < 0 \\ \big\{(\mathrm{Co},\mathrm{Co},\mathrm{Co})\big\}, & \text{se } 0 \leqslant t < 1 \\ \big\{(\mathrm{Co},\mathrm{Co},\mathrm{Co}), (\mathrm{Co},\mathrm{Co},\mathrm{Ca}), (\mathrm{Co},\mathrm{Ca},\mathrm{Co}), (\mathrm{Ca},\mathrm{Co},\mathrm{Co})\big\}, & \text{se } 1 \leqslant t < 2 \\ \big\{(\mathrm{Co},\mathrm{Co},\mathrm{Co}), (\mathrm{Co},\mathrm{Co},\mathrm{Ca}), (\mathrm{Co},\mathrm{Ca},\mathrm{Co}), (\mathrm{Ca},\mathrm{Co},\mathrm{Co}), & \\ \qquad (\mathrm{Ca},\mathrm{Ca},\mathrm{Co}), (\mathrm{Ca},\mathrm{Co},\mathrm{Ca}), (\mathrm{Co},\mathrm{Ca},\mathrm{Ca})\big\}, & \text{se } 2 \leqslant t < 3 \\ \Omega, & \text{se } t \geqslant 3. \end{cases}$$

Aqui, nesse exemplo, deve ficar claro que X satisfaz equação (1.0.1) pois tem 2^{Ω} como espaço de eventos. ◊

Exemplo 2. Consideremos o modelo clássico para o lançamento de um dado equilibrado. Seja $X : \Omega \to \mathbb{R}$ o resultado do lançamento, isto é, $X(\omega) = \omega$ para todo $\omega \in \Omega$. Para $t < 1$, temos $[X \leqslant t] = \emptyset$. Para $1 \leqslant t < 6$, temos $[X \leqslant t] = \{1, \dots, \lfloor t \rfloor\}$. Para $t \geqslant 6$, temos $[X \leqslant t] = \Omega$. ◊

1.0.1 Proposição. *A condição equação* (1.0.1) *é trivialmente satisfeita sempre que o espaço amostral é discreto.* □

Exemplo 3 (**variável aleatória constante**). Se para todo $\omega \in \Omega$ há $c \in \mathbb{R}$ tal que $X(\omega) = c$, então X é uma v.a. (verifique) tal que $[X \leqslant t] = \Omega$ caso $t \geqslant c$ e $[X \leqslant t] = \emptyset$ caso contrário. ◊

Exemplo 4. Consideremos o modelo geométrico clássico para o lançamento de um dardo num alvo de raio 3. Seja $Y : \Omega \to \mathbb{R}$ a distância euclidiana do ponto atingido ao centro do alvo, de modo que para qualquer $0 \leqslant t \leqslant 3$ o conjunto é $[X \leqslant t]$ é o circulo de raio t, cuja área está bem definida, portanto é um elemento de $\mathcal{A}$. ◊

Se $X, Y : \Omega \to \mathbb{R}$ são variáveis aleatórias então temos, por exemplo, os conjuntos

1. $[X > 3] = \overline{[X \leqslant 3]} = \overline{\{\omega \in \Omega : X(\omega) \leqslant 3\}} = \{\omega \in \Omega : X(\omega) > 3\}$;
2. $[X = 3] = [X \leqslant 3] \cap [X \geqslant 3]$,

3. $[2 \leqslant X < 3] = [X < 3] \cap [X \geqslant 2] = \{\omega \in \Omega : 2 \leqslant X(\omega) < 3\}$;

que, de fato, são eventos aleatórios. De modo análogo ao que foi descrito nos parágrafos acima, definimos os conjuntos $[X = t]$, $[X < t]$, $[X \leqslant t]$ e $[X \geqslant t]$ para uma variável aleatória $X : \Omega \to \mathbb{R}$. Esses conjuntos são eventos aleatórios do modelo pois para quaisquer reais a, b pois podem ser escrito como resultado de operações elementares de conjuntos sobre eventos aleatórios

- $[X > a] = \overline{[X \leqslant a]}$;
- $[a < X \leqslant b] = [X > a] \cap [X \leqslant b]$;
- $[X = a] = \bigcap_{n \geqslant 1} [a - \frac{1}{n} < X \leqslant a]$;
- $[X \geqslant a] = [X > a] \cup [X = a]$;
- $[X < a] = \overline{[x \geqslant a]}$.

O que nos interessa sobre probabilidade com relação à variáveis aleatórias são derivados das probabilidades desses eventos. Por exemplo,

1. os eventos $[Y < 1]$ e $[Y \geqslant 1]$ particionam Ω e, por exemplo, assumindo que ambos têm probabilidade positiva $\mathbb{P}(X = 3) = \mathbb{P}(X = 3 \mid Y < 1)\mathbb{P}(Y < 1) + \mathbb{P}(X = 3 \mid Y \geqslant 1)\mathbb{P}(Y \geqslant 1)$ pelo teorema da probabilidade total.

2. Se X é o número de caras em 3 lançamentos de uma moeda então
$$\begin{aligned}[X = 2] &= \{(\mathrm{Ca}, \mathrm{Ca}, \mathrm{Co}), (\mathrm{Ca}, \mathrm{Co}, \mathrm{Ca}), (\mathrm{Co}, \mathrm{Ca}, \mathrm{Ca})\}\\ [X \geqslant 2] &= \{(\mathrm{Ca}, \mathrm{Ca}, \mathrm{Co}), (\mathrm{Ca}, \mathrm{Co}, \mathrm{Ca}), (\mathrm{Co}, \mathrm{Ca}, \mathrm{Ca}), (\mathrm{Ca}, \mathrm{Ca}, \mathrm{Ca})\}\end{aligned}$$
e $\mathbb{P}(X = 2) = 3/8$ e $\mathbb{P}(X \geqslant 2) = 1/2$. O evento complementar a $[X \geqslant 2]$ representa o evento $\overline{[X \geqslant 2]} = [X < 2] = \{(\mathrm{Ca}, \mathrm{Co}, \mathrm{Co}), (\mathrm{Co}, \mathrm{Ca}, \mathrm{Co}), (\mathrm{Co}, \mathrm{Co}, \mathrm{Ca}), (\mathrm{Co}, \mathrm{Co}, \mathrm{Co})\}$ que ocorre com probabilidade $\mathbb{P}(\overline{X \geqslant 2}) = 1 - \mathbb{P}(X \geqslant 2) = \frac{1}{2}$. Se A representa o evento definido por "o primeiro lançamento foi Ca" então
$$\mathbb{P}(X = 2 \mid A) = \frac{\mathbb{P}([X = 2] \cap A)}{\mathbb{P}(A)} = \frac{\mathbb{P}(\{(\mathrm{Ca}, \mathrm{Ca}, \mathrm{Co}), (\mathrm{Ca}, \mathrm{Co}, \mathrm{Ca})\})}{\mathbb{P}(\{(\mathrm{Ca}, \mathrm{Ca}, \mathrm{Ca}), (\mathrm{Ca}, \mathrm{Ca}, \mathrm{Co}), (\mathrm{Ca}, \mathrm{Co}, \mathrm{Ca}), (\mathrm{Ca}, \mathrm{Co}, \mathrm{Co})\})} = \frac{1}{2}.$$
$$\mathbb{P}(X = 2 \mid X \geqslant 2) = \frac{\mathbb{P}([X = 2] \cap [X \geqslant 2])}{\mathbb{P}(X \geqslant 2)} = \frac{\mathbb{P}(\{(\mathrm{Ca}, \mathrm{Ca}, \mathrm{Co}), (\mathrm{Ca}, \mathrm{Co}, \mathrm{Ca}), (\mathrm{Co}, \mathrm{Ca}, \mathrm{Ca})\})}{\mathbb{P}(\{(\mathrm{Ca}, \mathrm{Ca}, \mathrm{Ca}), (\mathrm{Ca}, \mathrm{Ca}, \mathrm{Co}), (\mathrm{Ca}, \mathrm{Co}, \mathrm{Ca}), (\mathrm{Co}, \mathrm{Ca}, \mathrm{Ca})\})} = \frac{3}{4}.$$

1.1 Função de uma variável aleatória.

se $X : \Omega \to \mathbb{R}$ é uma v.a. e $f : \mathbb{R} \to \mathbb{R}$ é uma função real então a função composta pode não ser uma variável aleatória, entretanto será em vários casos úteis.

Exemplo 5. Se X é uma v.a. de $(\Omega, \mathcal{A}, \mathbb{P})$ então podemos definir $Z : \Omega \to \mathbb{R}$ por $Z(\omega) = X(\omega)^2$ para todo $\omega \in \Omega$. A função Z também é uma v.a. pois para todo t não negativo temos $[Z \leqslant t] = [-\sqrt{t} \leqslant X \leqslant \sqrt{t}] \in \mathcal{A}$ para todo $t \in \mathbb{R}$.

Podemos definir $W : \Omega \to \mathbb{R}$ por $W(\omega) = X(\omega)^3$ para todo $\omega \in \Omega$. A função W também é uma v.a. pois $[W \leqslant t] = [X \leqslant \sqrt[3]{t}] \in \mathcal{A}$ para todo $t \in \mathbb{R}$.

Se X é uma v.a. de $(\Omega, \mathcal{A}, \mathbb{P})$ e $a, b \in \mathbb{R}$ $(a \neq 0)$ então $Y : \Omega \to \mathbb{R}$ dada por $Y(\omega) = a \cdot X(\omega) + b$ é uma variável aleatória. De fato, para todo real t temos $[Y \leqslant t] = [X \leqslant (t-b)/a] \in \mathcal{A}$. ◊

Em geral, quando f é "bem comportada", a composta é uma v.a.:

1.1.1 Teorema. *Se X é uma variável aleatória de $(\Omega, \mathcal{A}, \mathbb{P})$ e $f : \mathbb{R} \to \mathbb{R}$ é contínua (ou contínua por partes) então $f(X)$ é uma variável aleatória.*

Esse resultado será provado mais adiante. Uma consequência imediata é:

1.1.2 Corolário. *X^k é uma variável aleatória para todo $k \in \mathbb{N}$.* □

Operações aritméticas com variáveis aleatórias são bem comportadas.

1.1.3 Teorema. *Se X e Y são variáveis aleatórias de um espaço de probabilidades $(\Omega, \mathcal{A}, \mathbb{P})$ então $X + Y$ e $X \cdot Y$ também são variáveis aleatória.*

Demonstração. Sejam X e Y variáveis aleatórias definidas em $(\Omega, \mathcal{A}, \mathbb{P})$.

Para a soma $X + Y(\omega) := X(\omega) + Y(\omega)$ temos que, para qualquer $t \in \mathbb{R}$

$$[X + Y < t] = \bigcup_{r \in \mathbb{Q}} [X < r] \cap [Y < t - r]$$

e o lado direito está em $\mathcal{A}$. Para verificar a igualdade de conjuntos tomemos $\omega \in [X + Y < t]$. Então $X(\omega) + Y(\omega) < t$. Tomemos um racional $r \in \mathbb{Q}$ tal que $X(\omega) < r < t - Y(\omega)$. De $X(\omega) < r$ temos $\omega \in [X < r]$ e de $r < t - Y(\omega)$ temos que $\omega \in [Y < t - r]$. Por outro lado, se $\omega \in [X < r] \cap [Y < t - r]$ então, claramente, $\omega \in [X + Y < t]$.

Para o produto $X \cdot Y(\omega) := X(\omega) \cdot Y(\omega)$ basta notar que

$$X \cdot Y = \frac{1}{2}\Big((X + Y)^2 - X^2 - Y^2\Big)$$

e o lado direito é uma variável aleatória pois, pelo exemplo 5 e pelo parágrafo acima $(X+Y)^2 - X^2 - Y^2$ é variável aleatória e, para concluir, usamos o resultado apresentado no exemplo 5. □

1.1.4 Teorema. *Se $X_1, X_2, \ldots$ são variáveis aleatórias de um espaço de probabilidades $(\Omega, \mathcal{A}, \mathbb{P})$ tais que existe o limite $\lim_{n\to\infty} X_n(\omega)$ para todo $\omega \in \Omega$, então*

$$X = \lim_{n\to\infty} X_n$$

é uma variável aleatória de $(\Omega, \mathcal{A}, \mathbb{P})$.

Uma definição não usual para X^{-1}: notemos que a função inversa de X pode não estar definida pois X não é, necessariamente, injetora (ela pode ser considerada sobrejetora pois podemos restringir o contradomínio à imagem). Entretanto, definimos X^{-1} para todo $A \subset \mathbb{R}$ por

$$X^{-1}(A) := \{\omega \in \Omega : X(\omega) \in A\}.$$

Uma propriedade importante dessa definição é que ela preserva e comuta com as operações sobre conjuntos (verifique):

$$X^{-1}(\overline{A}) = \overline{X^{-1}(A)}, \quad X^{-1}\left(\bigcup_n A_n\right) = \bigcup_n X^{-1}(A_n) \quad e \quad X^{-1}\left(\bigcap_n A_n\right) = \bigcap_n X^{-1}(A_n).$$

Também usamos a notação

$$[X \in A] := X^{-1}(A)$$

o qual é um evento aleatório sempre que A é boreliano. Daí temos, na definição de variável aleatória real, que $[X \leqslant t] = X^{-1}((-\infty, t]) \in \mathcal{A}$ é equivalente a dizer que $X^{-1}(A) \in \mathcal{A}$ para todo boreliano A, portanto, nesse caso faz sentido falar da probabilidade $\mathbb{P}(X \in A)$.

Variável aleatória indicadora. denotamos por $\mathbf{1}_A$ a função indicadora da ocorrência do evento aleatório A, isto é, para todo $\omega \in \Omega$

$$\mathbf{1}_A(\omega) = \begin{cases} 1 & \text{se } \omega \in A; \\ 0 & \text{se } \omega \notin A. \end{cases}$$

Assim,

$$[\mathbf{1}_A \leqslant t] = \begin{cases} \Omega, & \text{se } t \geqslant 1 \\ \overline{A}, & \text{se } 0 \leqslant t < 1 \\ \emptyset, & \text{caso contrário} \end{cases}$$

de modo que essa função é uma variável aleatória. Notemos que o próprio evento pode ser dado por uma variável aleatória. Por exemplo, se X é o resultado do lançamento de um dado, então $\mathbf{1}_{[X>3]}$ vale 1 se o resultado do lançamento é maior que 3 e 0 se o resultado do lançamento é menor ou igual a 3.

Exercício 1. Considere os eventos A e B de um espaço de probabilidade. Prove que

1. $\mathbf{1}_\Omega = 1$
2. $\mathbf{1}_{A\cup B} = \mathbf{1}_A + \mathbf{1}_B - \mathbf{1}_{A\cap B}$.
3. $\mathbf{1}_{\overline{A}} = 1 - \mathbf{1}_A$.
4. $A \subset B \Rightarrow \mathbf{1}_A \leqslant \mathbf{1}_B$.
5. $\mathbf{1}_{A\cap B} = \mathbf{1}_A \cdot \mathbf{1}_B$.

1.2 Independência de variáveis aleatórias reais. dizemos que X e Y são variáveis aleatórias independentes se os eventos $[X \leqslant a]$ e $[X \leqslant b]$ são independentes para quaisquer $a, b \in \mathbb{R}$.

1.3 Função de distribuição acumulada. Seja $(\Omega, \mathcal{E}, \mathbb{P})$ um espaço de probabilidade e X uma variável aleatória nesse espaço. Uma maneira simples de descrever as propriedades probabilísticas de X é dada pela sua função de distribuição acumulada.

A função de distribuição acumulada (f.d.a.) de X é a função $F_X : \mathbb{R} \to \mathbb{R}$ dada por

$$F_X(t) := \mathbb{P}(X \leqslant t).$$

É imediato da definição de f.d.a. que

$$0 \leqslant F_X(t) \leqslant 1.$$

Agora, observemos que se $x < y$ então $[X \leqslant x] \subset [X \leqslant y]$ logo $\mathbb{P}(X \leqslant x) \leqslant \mathbb{P}(X \leqslant y)$, ou seja, $F_X(x) \leqslant F_X(y)$, resumindo

$$x < y \Longrightarrow F_X(x) \leqslant F_X(y)$$

o que significa que F_X é não-decrescente. Além dessas duas propriedades, também valem

$$\lim_{t\to+\infty} F_X(t) = 1, \ \lim_{t\to-\infty} F_X(t) = 0 \ \text{ e } \ \lim_{t\to a^+} F_X(t) = F_X(a)$$

com as quais temos quatro propriedades que *caracterizam* funções de distribuição, i.e., qualquer função real F que satisfaz

(F1) $0 \leqslant F(t) \leqslant 1$ para todo $t \in \mathbb{R}$;

(F2) $x < y \Longrightarrow F(x) \leqslant F(y)$ para todos $x, y \in \mathbb{R}$;

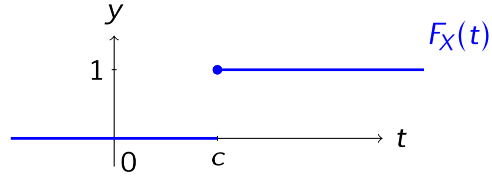
(F3) $\lim_{t\to+\infty} F(t) = 1$ e $\lim_{t\to-\infty} F(t) = 0$;

(F4) F é contínua à direta: $\lim_{t\to a^+} F(t) = F(a)$, para todo $a \in \mathbb{R}$.

é a f.d.a. de alguma variável aleatória.

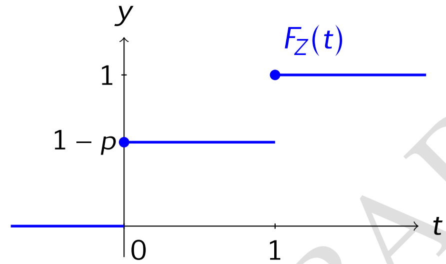
Por exemplo, no caso de variável aleatória constante, $X(\omega) = c$, algum $c \in \mathbb{R}$ e todo $\omega \in \Omega$

$$F_X(t) = \begin{cases} 0, & \text{se } t < c \\ 1, & \text{se } t \geqslant c \end{cases}$$



Exemplo 6. Consideremos uma moeda com probabilidade p de resultar Ca num lançamento. Seja $Z : \Omega \to \mathbb{R}$ a variável indicadora do evento {Ca}, ou seja, Z é dada por $Z(\text{Ca}) = 1$ e $Z(\text{Co}) = 0$.

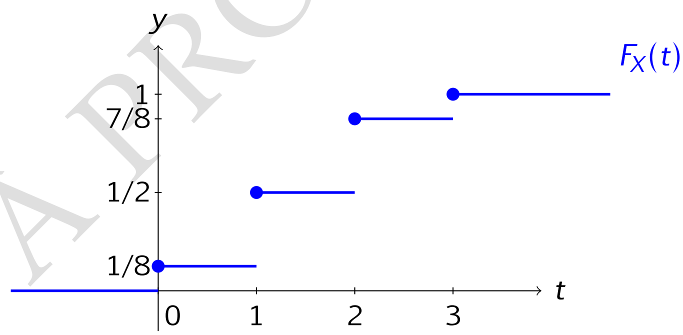
$$F_Z(t) = \begin{cases} 0, & \text{se } t < 0 \\ 1-p & \text{se } 0 \leqslant t < 1 \\ 1, & \text{se } t \geqslant 1 \end{cases}$$



◊

Exemplo 7. No caso dos 3 lançamentos de uma moeda, se X é o número de caras

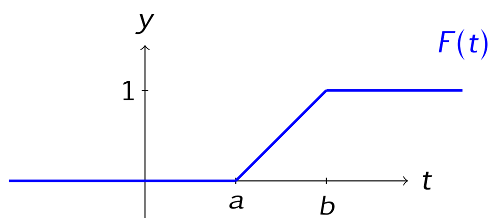
$$F_X(t) = \begin{cases} 0, & \text{se } t < 0 \\ 1/8, & \text{se } 0 \leqslant t < 1 \\ 1/2, & \text{se } 1 \leqslant t < 2 \\ 7/8, & \text{se } 2 \leqslant t < 3 \\ 1, & \text{se } t \geqslant 3 \end{cases}$$



◊

Exemplo 8. Seja F dada por

$$F(t) = \begin{cases} 0, & \text{se } t < a \\ \frac{t-a}{b-a}, & \text{se } a \leqslant t < b \\ 1, & \text{se } t \geqslant b \end{cases}$$



F é contínua em toda a reta, portanto é continua a direita; $0 \leqslant F(t) \leqslant 1$ para todo real t; os limites quando $t \to +\infty$ e quando $t \to -\infty$ são, respectivamente, 1 e 0. Assim, essa função e uma f.d.a. de uma variável aleatória U. ◊

A variável aleatória U do exemplo acima pode ser vista como a coordenada de um ponto escolhido no intervalo $[a,b]$ no modelo geométrico clássico $([a,b], \mathcal{B}([a,b]), \mathbb{P})$. Para $[c,d] \subset [a,b]$ a probabilidade de sortear um ponto no intervalo $[c,d]$ é calculada usando a f.d.a. do seguinte modo: a probabilidade de $[X < c]$ é $F(c)$, a probabilidade de $[X > d]$ é $1 - F(d)$, um desses eventos ocorre com probabilidade

$$\mathbb{P}([X < c] \uplus [X > d]) = F(c) + 1 - F(d) = 1 - \frac{d-c}{b-a}$$

portanto, $\mathbb{P}(X \in [c, d]) = (d-c)/(b-a)$, que é o que esperávamos. No caso particular de $a = 0$ e $b = 1$ temos que $U(\omega) = \omega$ e no intervalo $[0, 1]$ a f.d.a. vale $F_U(t) = t$.

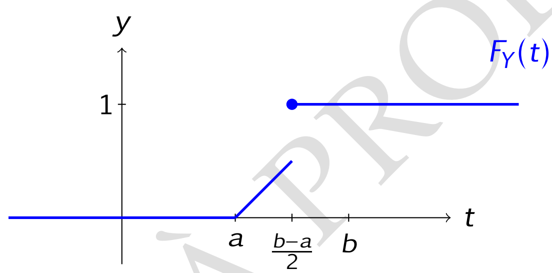
Exemplo 9. Definamos a variável aleatória Y por

$$Y(\omega) = \min\left\{U(\omega), \frac{b-a}{2}\right\}$$

em que U é a variável aleatória do exemplo anterior. Usando a definição de Y é imediato que

$$\mathbb{P}\left(Y \leqslant \frac{b-a}{2}\right) = 1.$$

Para $t < (b-a)/2$, $Y(\omega) \leqslant t$ se e só se $U(\omega) \leqslant t$, logo $F_Y(t) = F_U(t)$ e o gráfico de F_Y é



◊

Pode ser provado que essas propriedades caracterizam uma f.d.a.

1.3.1 Mais propriedades de uma f.d.a.: A função de distribuição acumulada tem várias propriedades que ajudam no cálculo de probabilidades, algumas são dadas na proposição abaixo. Usamos a notação

$$F(a-) := \lim_{t \to a^-} F(t).$$

1.3.2 Proposição. *A função de distribuição acumulada F de uma variável aleatória X satisfaz*

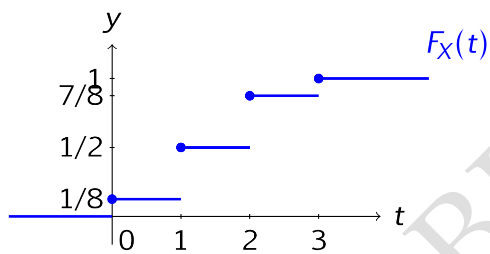
1. $1 - F(t) = \mathbb{P}(X > t)$, *para todo* $t \in \mathbb{R}$;
2. $F(b) - F(a) = \mathbb{P}(a < X \leqslant b)$, *para quaisquer* $a < b$;
3. $F(a) - F(a-) = \mathbb{P}(X = a)$, *para todo real* a.

1.3.3 Corolário. *F é contínua em $a \in \mathbb{R}$ se, e só se, $\mathbb{P}(X = a) = 0$.*

As descontinuidades das funções de distribuição acumulada são do tipo salto. Se F não é contínua em a então o salto em a é de grandeza $\mathbb{P}(X=a) = F(a) - F(a-)$. No exemplo abaixo, o salto em $t=2$ é $7/8 - 1/2 = 3/8 = \mathbb{P}(X=2)$. Notemos que a soma dos saltos de tamanho $\geqslant 1/n$ não deve ser maior que 1, portanto, há no máximo n desses saltos; desse fato podemos concluir que há no máximo um número enumerável de pontos de descontinuidade em qualquer função de distribuição acumulada.

Exemplo 10. No caso dos 3 lançamentos de uma moeda, X é o número de caras

$$F_X(t) = \begin{cases} 0, & \text{se } t < 0 \\ 1/8, & \text{se } 0 \leqslant t < 1 \\ 1/2, & \text{se } 1 \leqslant t < 2 \\ 7/8, & \text{se } 2 \leqslant t < 3 \\ 1, & \text{se } t \geqslant 3 \end{cases}$$

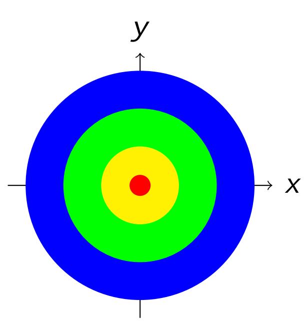


Usando as propriedades de uma função de distribuição temos, por exemplo

- $\mathbb{P}(X > 2) = 1 - F(2) = 1 - 7/8 = 3/8$;
- $\mathbb{P}(X > 3) = 1 - F(3) = 0$;
- $\mathbb{P}(0{,}5 < X \leqslant 2{,}5) = F(2{,}5) - F(0{,}5) = 7/8 - 1/8 = 3/4$;
- $\mathbb{P}(X = 1) = F(1) - F(1-) = 1/2 - 1/8 = 3/8$;
- $\mathbb{P}(X = 1{,}8) = F(1{,}8) - F(1{,}8-) = 1/2 - 1/2 = 0$;
- $\mathbb{P}(X = -1) = F(-1) - F(-1-) = 0 - 0 = 0$;
- $\mathbb{P}(X = 7) = F(7) - F(7-) = 1 - 1 = 0$.

◊

Exemplo 11. Consideremos o exemplo em que um dardo acerta aleatoriamente um alvo composto de círculos concêntricos de raios 1/4, 1, 2 e 3 que supomos centrados na origem de um sistema cartesiano.

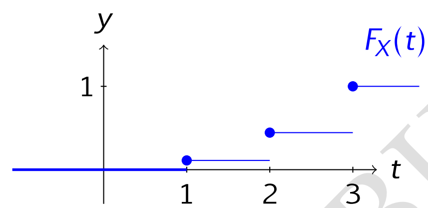


Para cada $k = 1, 2, 3$, considere as regiões A_k da figura acima do seguinte modo:

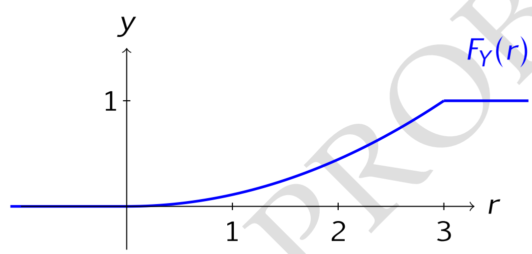
$$
\begin{aligned}
A_1 &= \left\{(x,y)\colon x^2 + y^2 \leqslant 1\right\} \text{ (Vermelho + Amarelo)} \\
A_2 &= \left\{(x,y)\colon 1 < x^2 + y^2 \leqslant 4\right\} \quad \text{(Verde)} \\
A_3 &= \left\{(x,y)\colon 4 < x^2 + y^2 \leqslant 9\right\} \quad \text{(Azul)}
\end{aligned}
$$

Se $X(\omega) = k$ quando $\omega \in A_k$ então $\mathbb{P}(X = k) = \mathbb{P}(A_k) = (2k-1)/9$ e X tem função de distribuição acumulada

$$
F_X(t) = \begin{cases} 0, & \text{se } t < 1 \\ \frac{\lfloor t \rfloor^2}{9}, & \text{se } 1 \leqslant t \leqslant 3 \\ 1, & \text{se } t > 3 \end{cases}
$$



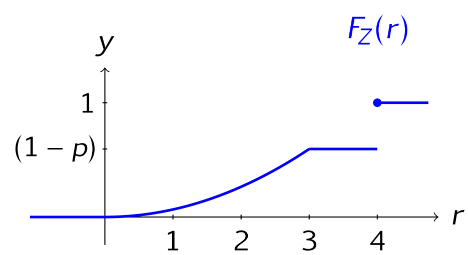
Se $\omega = (x,y) \in \Omega$ e $Y(\omega) = \sqrt{x^2 + y^2}$ é a distância do ponto atingido ao centro, a função de distribuição acumulada de Y é $F_Y(r) = r^2/9$ se $0 \leqslant r \leqslant 3$



Agora, suponha que o jogador erre o alvo com probabilidade p, para algum $p > 0$ fixo. A pontuação Z é, caso acerte, 1 se acertou o verde, 2 o azul e 3 o vermelho ou, caso erre o alvo, 4. Então

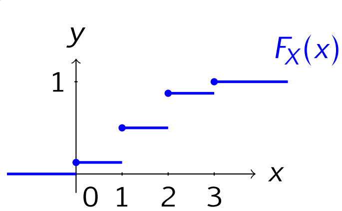
$$
\mathbb{P}(Z \leqslant r) = \mathbb{P}(Z \leqslant r \mid \text{acertou})\mathbb{P}(\text{acertou}) + \mathbb{P}(Z \leqslant r \mid \text{errou})\mathbb{P}(\text{errou})
$$

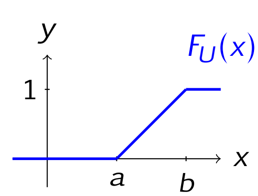
$$
F_Z(r) = \begin{cases} 0 & \text{se } r < 0, \\ (1-p)F_Y(r) & \text{se } 0 \leqslant r < 4, \\ 1 & \text{se } r \geqslant 4. \end{cases}
$$

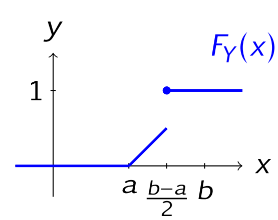


◊

1.4 Variáveis aleatórias discretas e contínuas. Lembrando os gráficos das f.d.a.'s dos exemplos 7, 8 e 9







o primeiro corresponde a uma variável aleatória discreta, o segundo a uma variável aleatória contínua e o terceiro corresponde a uma variável aleatória que não é discreta nem contínua.

Variável aleatória discreta: é uma variável aleatória X que assume valores num conjunto enumerável (finito ou infinito) $I = \{x_0, x_1, \dots\} \subset \mathbb{R}$. A função dada por $p_X(x_i) := \mathbb{P}(X = x_i)$ para todo $x_i \in I$ é chamada *função de massa de probabilidade* ou, simplesmente, *função de probabilidade* de X.

Claramente,

$$\sum_{x_i \in I} \mathbb{P}(X = x_i) = 1$$

e qualquer evento que envolve X tem sua probabilidade determinada pelos valores $p_X(x_i)$. A função de distribuição acumulada é dada por

$$F_X(t) = \sum_{i : x_i \leqslant t} p_X(x_i).$$

Ademais, F_X é uma função escada com saltos que ocorrem nos pontos $x_i \in I$.

Variável aleatória (absolutamente) contínua: é uma variável aleatória X para a qual existe uma função $f_X : \mathbb{R} \to [0, \infty)$ tal que

$$F_X(t) = \int_{-\infty}^{t} f_x(x)\,\mathrm{d}x$$

para todo $t \in \mathbb{R}$. A função dada por f_X é chamada *função de densidade de probabilidade* ou, simplesmente, *densidade* de X.

Claramente, a integral de f_X em $\mathbb{R}$ é igual a 1. Reciprocamente, uma função f não negativa cuja integral em $\mathbb{R}$ é igual a 1 é densidade de alguma v.a. pois se definirmos $F : \mathbb{R} \to \mathbb{R}$ por $F(t) = \int_{-\infty}^{t} f(x)\,\mathrm{d}x$ valem as propriedades **F1**, **F2**, **F3** e **F4** que caracterizam uma f.d.a.

Sendo F_X a integral indefinida de uma função ela é uma função contínua, de fato F_X é *absolutamente contínua*[1] e, portanto, $F_x' = f_X$ em quase todo ponto[2]. Na prática, podemos verificar que X admite uma densidade se F_X é contínua e se tem derivada no interior de uma quantidade enumerável de intervalos fechados cuja união é $\mathbb{R}$.

[1] g é absolutamente contínua em $[a, b]$ se para todo $\varepsilon > 0$, existe $\delta > 0$, para toda coleção finita de intervalos disjuntos $(x_1, y_1), \dots, (x_n, y_n)$

$$\sum_{i=1}^{n} |x_i - y_i| \Rightarrow \sum_{i=1}^{n} |g(x_i) - g(y_i)| < \delta.$$

(1) g é absolutamente contínua se e só se for dada por uma integral indefinida [Teo. 13 do cap. 5 de Royden, *Real Analysys*].

(2) Se g é absolutamente contínua em $[a, b]$ então tem derivada em quase todo ponto de $[a, b]$ [Cor. 11 do cap. 5 de Royden, *Real Analysys*].

[2] Ou seja, o conjuntos dos pontos t tais que $F_x'(t) \neq f_X(t)$ tem medida nula, o que significa que para qualquer $\varepsilon > 0$ (por menor que seja) tal conjunto está contido numa reunião de intervalos de comprimento total menor que ε.

Nem discreta, nem contínua: uma mistura de ambas, é o caso da variável aleatória Y dada por $Y(\omega) = \min\{Z(\omega), (b-a)/2\}$ apresentada no exemplo 9 temos

$$\mathbb{P}\left(Y = \frac{b-a}{2}\right) = F_Y\left(\frac{b-a}{2}\right) - F_Y\left(\frac{b-a^-}{2}\right) = \frac{a+b}{2(b-a)}.$$

Se Y fosse uma variável aleatória contínua esse valor deveria ser 0 (por quê?) Essa variável aleatória não é discreta porque sua imagem não é enumerável. Essa variável é dita do tipo *mista*. Ainda, é possível ocorrer de uma v.a. não ser de nenhum dos três tipos; possivelmente, o caso mais conhecido é da *função de Cantor*[3] Uma variável aleatória cuja distribuição é a função de Cantor é dita *singular*: X é singular se F_X é contínua (não absolutamente) e $F_X(t) = 0$ em quase todo ponto. Informalmente falando, *toda variável aleatória é uma mistura de discreta, absolutamente contínua e singular.*

Na próxima seção apresentamos uma ferramenta que permite um tratamento unificado de variáveis discretas e contínuas.

1.5 Variáveis aleatórias discretas.

Uma variável aleatória X é discreta se assume valores num conjunto $\mathrm{Im}(X) \subset \mathbb{R}$ discreto (enumerável finito ou infinito). A função de distribuição acumulada é dada por

$$F_X(t) = \sum_{x \in \mathrm{Im}(X): x \leqslant t} \mathbb{P}(X = x)$$

em que a soma é sobre todo $x \in \mathrm{Im}(X)$ tal que $x \leqslant t$.

Função de massa de probabilidade: (f.m.p.) ou, simplesmente, função de probabilidade da variável aleatória X é a função $f_X : \mathbb{R} \to [0,1]$ dada por

$$f_X(a) := \mathbb{P}(X = a)$$

de modo que

$$\sum_{a \in \mathrm{Im}(X)} f_X(a) = 1.$$

e para qualquer $S \subset \mathbb{R}$

$$\mathbb{P}(X \in S) = \sum_{a \in S} f_X(a).$$

Por exemplo, no caso do lançamento de 3 moedas, o número de caras tem função de probabilidade

$$f_X(a) = \begin{cases} 0 & \text{se } a \neq 0, 1, 2, 3 \\ 3/8 & \text{se } a = 1 \text{ ou } x = 2 \\ 1/8 & \text{se } a = 0 \text{ ou } x = 3. \end{cases}$$

[3] seção 2.2 de Barry James, *Probabilidade: um curso de nível intermediário*.

Exemplo 12. Dado um inteiro positivo n, lançamos uma moeda com probabilidade até que resulte cara ou complete n lançamentos.

A probabilidade de cara é $p \in (0, 1)$ e os resultados dos lançamentos são independentes. O número de lançamentos é uma variável aleatória que denotamos por X. Qual a probabilidade de $[X = k]$?

Cada resultado possível com $k < n$ lançamentos tem probabilidade $\mathbb{P}(X = k) = p(1 - p)^{k-1}$ e são dois resultado possíveis com n lançamentos, um que termina com cara e outro que termina com coroa, portanto $\mathbb{P}(X = n) = (1 - p)^{n-1}$. ◊

Exercício 2. Lançamos uma moeda equilibrada n vezes, independentemente. O número de caras é uma variável aleatória que denotamos por X. Determine $\mathbb{P}(X = k)$.

1.5.1 Principais modelos discretos:

Distribuição uniforme discreta: dado S de cardinalidade k dizemos que X tem distribuição uniforme discreta em S, fato denotado por $X \sim \mathrm{U}(k)$, se

$$f_X(x) = \begin{cases} \frac{1}{k}, & \text{se } x \in S \\ 0, & \text{caso contrário.} \end{cases}$$

Distribuição de Bernoulli:: na prática, ocorrem muitas situações com experimentos para os quais nos interessa apenas dois resultados, por exemplo

1. uma peça é classificada como *boa* ou *defeituosa*;
2. o resultado de um exame médico é *positivo* ou *negativo*;
3. um paciente submetido a um tratamento é *curado* ou *não* da doença;
4. um entrevistado *concorda* ou *não concorda* com a afirmação feita;
5. no lançamento de um dado *ocorre* ou *não ocorre* a face "5".

Nessas situações podemos representar, genericamente, os resultados do experimento com o espaço amostral $\Omega = \{\text{sucesso}, \text{fracasso}\}$ e o modelo probabilístico fica determinado dado $p = \mathbb{P}(\text{sucesso}) \in [0, 1]$. Esses experimentos recebem o nome de *Ensaios de Bernoulli* e a variável aleatória indicadora do evento "sucesso" é uma *variável aleatória de Bernoulli* com parâmetro p.

A notação

$$X \sim \mathrm{Bernoulli}(p)$$

indica que X é uma variável aleatória de Bernoulli com parâmetro p; ela assume dois valores: 1 se ocorre sucesso e 0 se ocorre fracasso; sua f.m.p. é dada por

$$\mathrm{be}_p(x) = \begin{cases} p^x(1 - p)^{1-x} & \text{se } x \in \{0, 1\} \\ 0 & \text{nos outros casos.} \end{cases}$$

Distribuição binomial:: consideremos n repetições *independentes* de um Ensaio de Bernoulli. Seja X o número de sucessos nas repetições.

Por exemplo, um dado equilibrado é lançado 3 vezes. Assumindo independência dos resultados, qual é a probabilidade de se obter a face 5 duas vezes? Se S denota *sucesso*, i.e., "ocorre face 5" e F denota *fracasso*, "não ocorrer face 5" então podemos associar a cada resposta do experimento um elemento de $\Omega = \{SSS, SSF, SFS, FSS, SFF, FSF, FFS, FFF\}$ onde atribuímos $p = \mathbb{P}(S) = 1/6$ e $1 - p = \mathbb{P}(F) = 5/6$. A função de massa de probabilidade para o número X de sucessos é, usando independência das respostas,

ω	x	$f_X(x)$
FFF	0	$(1-p)^3$
SFF, FSF, FFS	1	$3p(1-p)^2$
SSF, SFS, FS	2	$3p^2(1-p)$
SSS	3	p^3

e podemos escrever essa função como $f_X(x) = \binom{3}{x}p^x(1-p)^{3-x}$ para todo $x \in \{0,1,2,3\}$ e $f_X(x) = 0$ nos outros casos. Assim, $\mathbb{P}(X = 2) = \binom{3}{2}p^2(1-p)^{3-2} = 0{,}0694$.

Uma *variável aleatória binomial* de parâmetros $n \in \mathbb{N}$ e $p \in (0,1)$ é uma variável aleatória com f.m.p.

$$\mathrm{bi}_{n,p}(x) = \begin{cases} \binom{n}{x}p^x(1-p)^{n-x}, & \text{se } x \in \{0,1,2,\dots,n\} \\ 0 & \text{nos outros casos} \end{cases}$$

que pode ser vista como o número de sucessos em n ensaios independentes de Bernoulli e com mesma probabilidade p de sucesso. A notação

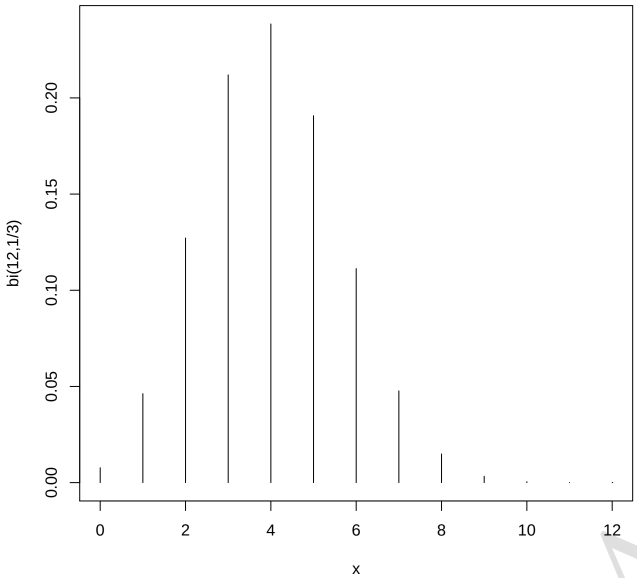
$$X \sim \mathrm{Binomial}(n, p)$$

indica que X é uma variável aleatória com distribuição binomial com parâmetros n e p.

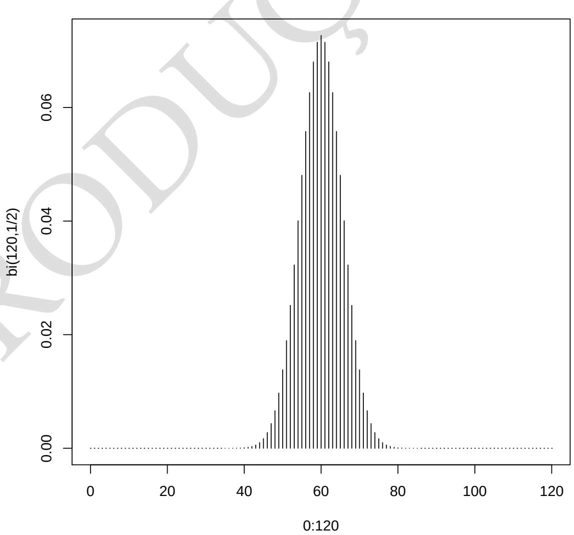
Exemplo 13. Numa prova com 12 questões de múltipla escolha, com 3 alternativas, se todas as respostas forem escolhidas aleatoriamente, então o número de acertos é $X \sim \mathrm{Binomial}(12, \frac{1}{3})$ e a função de massa de probabilidade é

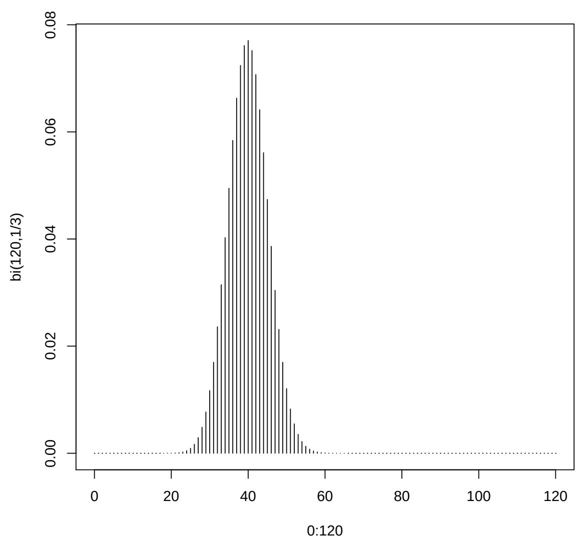
$$\mathrm{bi}_{12,\frac{1}{3}}(x) = \binom{12}{x}\left(\frac{1}{3}\right)^x\left(\frac{2}{3}\right)^{n-x}$$

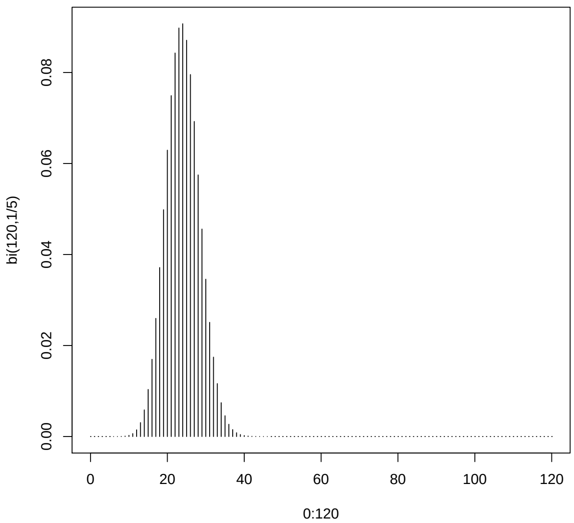
tem o seguinte *gráfico de barras*

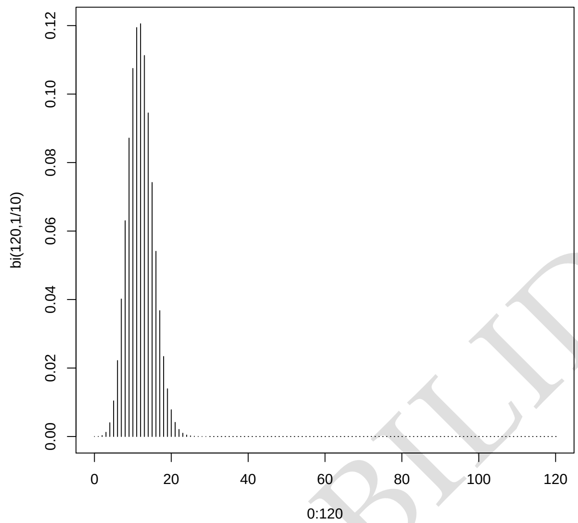


A seguir, respectivamente, os gráficos mostram a distribuição nos casos de uma prova com 120 questões e 2 alternativas, uma prova com 120 questões e 3 alternativas e uma prova com 120 questões e 5 alternativas e, finalmente, 10 alternativas









◊

Exemplo 14. Três bolas são aleatoriamente retiradas de um saco com 20 bolas numeradas de 1 a 20. Supondo que todas as $\binom{20}{3}$ seleções são equiprováveis, com que probabilidade pelo menos uma tem número maior ou igual a 18?

Se a maior bola selecionada é maior ou igual a 18, então menos uma tem número maior ou igual a 18 e vice-versa. Seja X a maior bola selecionada; X é uma variável aleatória. Queremos determinar com que probabilidade ocorre $[X = 18]$ ou $[X = 19]$ ou $[X = 20]$, ou seja

$$\mathbb{P}(X \geqslant 18) = \mathbb{P}([X = 18] \cup [X = 19] \cup [X = 20]) = \mathbb{P}(X = 18) + \mathbb{P}(X = 19) + \mathbb{P}(X = 20)$$

a igualdade vem do fato dos eventos serem independentes. A quantidade de seleções de três bolas de modo que a maior seja igual a i é $\binom{i-1}{2}$ pois há $i-1$ bolas menores, das quais escolhemos 2. Portanto

$$\mathbb{P}(X = i) = \frac{\binom{i-1}{2}}{\binom{20}{3}}$$

portanto

$$\mathbb{P}(X \geqslant 18) = \frac{\binom{17}{2}}{\binom{20}{3}} + \frac{\binom{18}{2}}{\binom{20}{3}} + \frac{\binom{19}{2}}{\binom{20}{3}} \approx 0{,}403.$$

◊

Exemplo 15. Um equipamento resiste a um teste de choque com probabilidade 3/4. Qual é probabilidade de que em 4 equipamentos testados 2 equipamentos sobrevivam ao choque? Se $X \sim \mathrm{Binomial}(4, \frac{3}{4})$

$$\mathrm{bi}_{4,\frac{3}{4}}(2) = \binom{4}{2}\left(\frac{3}{4}\right)^2\left(\frac{1}{4}\right)^2 = \frac{27}{128} \approx 0{,}21.$$

◊

Exemplo 16. Uma febre atinge 25% dos rebanhos bovinos. Para tratamento foram desenvolvidas três vacinas, que denominaremos $V1$, $V2$ e $V3$ e que foram inoculadas em 10, 17 e 23 animais, respectivamente. Após um período de observação, o número desses animais que ficaram doentes foram, respectivamente, 0, 1 e 2. Há, nessas informações alguma evidência de eficácia das vacinas?

Consideremos três grupos dos mesmos tamanhos compostos por animais não vacinados e vamos calcular a probabilidade dessas populações se saírem tão bem quanto as vacinadas. A probabilidade de um indivíduo ficar doente é 1/4 e se X é a quantidade de animais que ficam doente em um rebanho de tamanho n então $X \sim \text{Binomial}(n, \frac{1}{4})$ e a probabilidade com que x animais fiquem doentes é

$$\text{bi}_{n,\frac{1}{4}}(x) = \binom{n}{x}(1/4)^x(3/4)^{n-x}$$

e para as populações consideradas

	população	proporção de doentes
1.	$n = 10,\ x = 0$	$\mathbb{P}(X = 0) = 0{,}05631351$
2.	$n = 17,\ x = 1$	$\mathbb{P}(X \leqslant 1) = 0{,}05011298$
4.	$n = 23,\ x = 2$	$\mathbb{P}(X \leqslant 2) = 0{,}04920334$

Portanto, a chance de um rebanho não vacinado ser sair melhor que um vacinado é aproximadamente 5%. A vacina $V3$ se sai ligeiramente melhor que as outras. ◊

Exemplo 17. Um fabricante garante que seu produto tem uma taxa de itens defeituosos de 3%. Numa seleção de 20 itens a serem inspecionados, qual é a probabilidade de ao menos um ser defeituoso? Se X é a quantidade de itens defeituosos

$$\mathbb{P}(X \geqslant 1) = 1 - \mathbb{P}(X = 0) = 1 - \binom{20}{0}(0{,}03)^0(0{,}97)^{20} = 0{,}4562.$$

Se 10 carregamentos por mês deixam a fabrica e de cada carregamento 20 itens são inspecionados, com que probabilidade 3 carregamentos tem pelo menos um item defeituoso?

$$\binom{10}{3}(0{,}4562)^3(1 - 0{,}4562)^7 = 0{,}1602.$$

◊

Seja $X \sim \text{Binomial}(n, p)$. Para n e p fixos, quando x varia de 0 a n o valor de $\text{bi}_{n,p}(x)$ cresce monotonicamente e depois decresce monotonicamente. De fato,

$$\frac{\text{bi}_{n,p}(x)}{\text{bi}_{n,p}(x-1)} = \frac{(n - x + 1)p}{(1 - p)x} \tag{1.5.1}$$

que é crescente se $(n - x + 1)p > (1 - p)x$ ou, equivalentemente, $x < (n + 1)p$.

1.5.2 Proposição. *Seja $X \sim \text{Binomial}(n, p)$. Para n e p fixos, quando x varia de 0 a n o valor de $\text{bi}_{n,p}(x)$ cresce monotonicamente e depois decresce monotonicamente atingindo o máximo quando $x = \lfloor (n + 1)p \rfloor$.* □

Distribuição de Poisson:: a f.m.p. de uma variável aleatória de Poisson expressa a probabilidade de ocorrência de um determinado número de eventos num intervalo de tempo fixo (ou região do espaço), se estes eventos ocorrem com uma taxa média conhecida e independentemente do tempo desde a última ocorrência. Por exemplo:

1. o número de erros de impressão numa página de livro;

2. o número de chamadas que chega a um *call center*;

3. o número de partículas α descarregadas por um material radioativo em um período fixo de tempo.

Há vários exemplos curiosos de fenômenos aleatórios com essa distribuição na literatura. Começaremos com o seguinte exemplo do livro do Feller: Na segunda guerra mundial a cidade de Londres foi intensamente bombardeada pelos alemães. Para determinar se as bombas tinham um alvo ou foram lançadas aleatoriamente os ingleses dividiram o sul da cidade em 576 pequenas regiões, de tamanho $0{,}25\,\mathrm{km}^2$. O total de bombas que atingiram a região foi 537, o que dá uma taxa de 0,9323 bombas por região; se n_k é o número de regiões que receberam k bombas, a contagem foi

k	0	1	2	3	4	5 ou mais
n_k	229	211	93	35	7	1

ao qual o modelo de Poisson se ajusta impressionantemente bem, o que levou-os a acreditar que o bombardeio foi aleatório.

William Sealy Gosset, um químico e matemático formado em Oxford, foi contratado, em 1899, pela famosa cervejaria *Arthur Guinness and Son* em Dublin; sua tarefa era para aperfeiçoar o processo de produção de cerveja. Gosset (que publicou artigos sob o pseudônimo de *Student*, porque o seu empregador proibiu publicações por funcionários depois que um colega de trabalho havia divulgado segredos comerciais) trabalhou com o modelo de Poisson para a contagem de células de levedura. No artigo sobre tal trabalho, Gosset discutiu "como a dispersão nas contagens de colonias de levedura foi semelhante ao limite exponencial da distribuição binomial".

Outra aplicação curiosa desta distribuição foi feita por Ladislau Bortkiewicz em 1898, quando lhe foi dada a tarefa de investigar o número de soldados no exército russo morto acidentalmente por coice de cavalo.

A notação

$$X \sim \text{Poisson}(\lambda)$$

indica que X é uma *variável aleatória de Poisson* com parâmetro $\lambda > 0$, ela conta o número de ocorrências de um determinado evento que ocorre a uma taxa λ e cuja f.m.p. é dada por

$$\mathrm{po}_\lambda(x) = \begin{cases} \frac{e^{-\lambda}\lambda^x}{x!}, & \text{se } x \in \mathbb{N} \\ 0, & \text{nos outros casos.} \end{cases}$$

A distribuição de Poisson pode ser derivada como um caso limite para a distribuição binomial quando o número de ensaios tende ao infinito e a taxa média de ocorrências permanece fixa. Por isso, pode ser usado como uma aproximação da distribuição binomial se n for suficientemente grande e p suficientemente pequena. A taxa de ocorrência de um evento está relacionada com a probabilidade de um evento ocorrer em pequenos subintervalos, por pequeno entendemos o suficiente para que a probabilidade de um evento ocorrer duas vezes nesse intervalo é insignificante. Dividimos o intervalo inteiro em n subintervalos $I_1, \dots, I_n$ de igual tamanho com $n > \lambda$. Isto significa que a probabilidade de ocorrência do evento em um intervalo I_k, para cada k é igual a λ/n. Agora, assume-se que as ocorrências do evento em todo o intervalo pode ser visto como n ensaios de Bernoulli com parâmetro λ/n. Suponha que em n ensaios independentes de Bernoulli com probabilidade de sucesso $p = p(n) = \lambda/n$, para uma constante $\lambda > 0$. A probabilidade de x sucessos é $\mathrm{bi}_{n,\frac{\lambda}{n}}(x)$. Para $x = 0$

$$\mathrm{bi}_{n,\frac{\lambda}{n}}(0) = \left(1 - \frac{\lambda}{n}\right)^n$$

portanto, para n suficientemente grande $\mathrm{bi}_{n,p}(0) \approx e^{-\lambda}$ (pela definição de e como limite de uma sequência). Agora, usando equação (1.5.1) podemos concluir que

$$\begin{aligned} \mathrm{bi}_{n,\frac{\lambda}{n}}(1) &\approx \mathrm{bi}_{n,\frac{\lambda}{n}}(0)\lambda &&\approx \lambda e^{-\lambda} \\ \mathrm{bi}_{n,\frac{\lambda}{n}}(2) &\approx \mathrm{bi}_{n,\frac{\lambda}{n}}(1)\lambda/2 &&\approx (\lambda^2/2)e^{-\lambda} \\ \mathrm{bi}_{n,\frac{\lambda}{n}}(3) &\approx \mathrm{bi}_{n,\frac{\lambda}{n}}(2)\lambda/6 &&\approx (\lambda^3/3!)e^{-\lambda} \end{aligned}$$

que podemos estender usando indução para

$$\mathrm{bi}_{n,\frac{\lambda}{n}}(x) \approx \frac{\lambda^x e^{-\lambda}}{x!} = \mathrm{po}_\lambda(x)$$

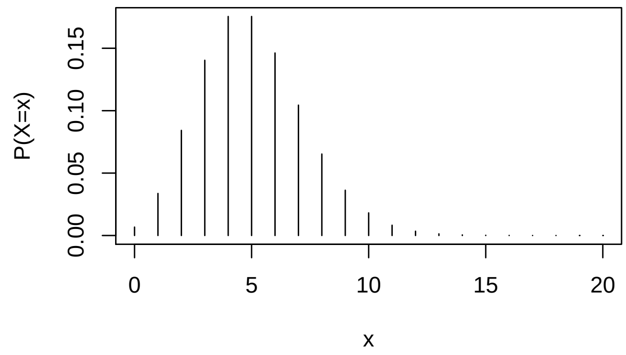
que é conhecido como a *aproximação de Poisson para a distribuição binomial*; em resumo, fixado λ e fixado x, se $Y_n \sim \mathrm{Binomial}(n, \lambda/n)$ e $X \sim \mathrm{Poisson}(\lambda)$, então

$$\mathbb{P}(X = x) = \lim_{n\to\infty} \mathbb{P}(Y_n = x)$$

e uma prova pode ser vista neste link.

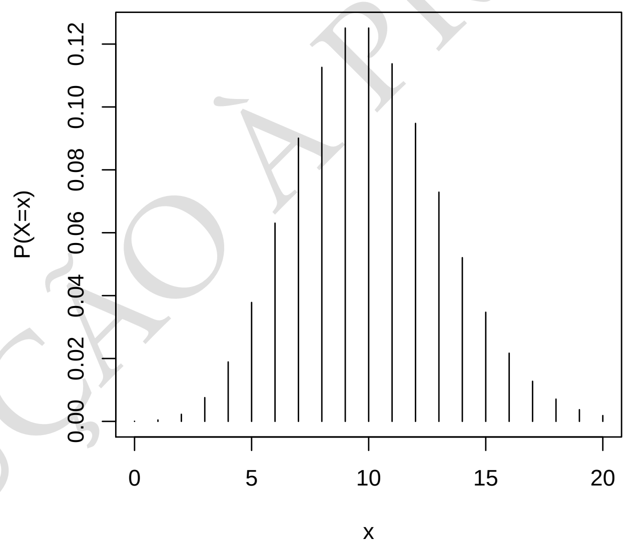
Exemplo 18. Um telefone recebe em média 5 chamadas por minuto. Supondo que a distribuição de Poisson seja um modelo adequado para essa situação, qual a probabilidade com que o telefone não receba chamadas durante um intervalo de 1 minuto?

$$\mathrm{po}_5(0) = \frac{5^0 e^{-5}}{0!} = 0{,}0067$$



◊

Exemplo 19. O número de partículas que contaminam a superfície de um CD no processo de fabricação tem distribuição de Poisson. O número médio de partículas é 0,1 particulas/cm^2 e a área de um CD é 100cm^2. Seja X o número de partículas num CD; $X \sim \text{Poisson}(10)$.



1. a probabilidade de ter 12 partículas é

$$\mathbb{P}(X = 12) = \frac{e^{-10}10^{12}}{12!} = 0{,}095$$

2. a probabilidade de ter 0 partículas é

$$\mathbb{P}(X = 0) = \frac{e^{-10}10^{0}}{0!} = 4{,}54 \times 10^{-5}$$

3. a probabilidade de ter $\leqslant 12$ partículas é

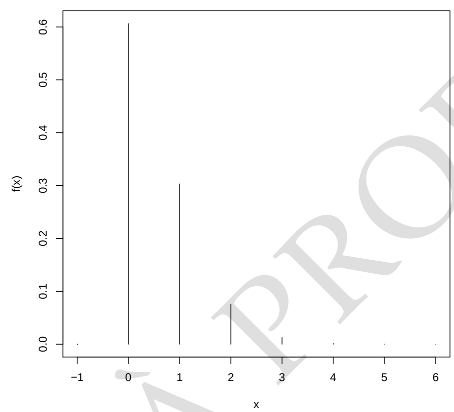
$$\mathbb{P}(X \leqslant 12) = \sum_{x \leqslant 12} \frac{e^{-10} 10^x}{x!} = 0{,}792$$

$\lozenge$

Exemplo 20. Suponha que essas notas tenham erros tipográficos por página que segue uma distribuição de Poisson com $\lambda = 1/2$. Qual é a probabilidade de haver pelo menos um erro nessa página? Se X é o número de erros por página

$$\mathbb{P}(X \geqslant 1) = 1 - \mathbb{P}(X = 0) = 1 - e^{-1/2} \approx 0{,}393.$$

$\lozenge$

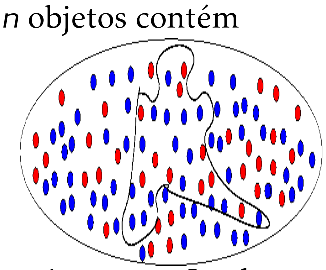


Exercício 3. Prove que a f.m.p. de $X \sim \text{Poisson}(\lambda)$ satisfaz

$$\text{po}_\lambda(x+1) = \frac{\lambda}{x+1} \text{po}_\lambda(x).$$

Distribuição hipergeométrica:: uma coleção de n objetos contém

1. a objetos azuis,
2. $n - a$ objetos vermelhos.



Uma amostra com s elementos é selecionada aleatoriamente. Qual a probabilidade da amostra conter x ($x \leqslant a$) bolas azuis? O número de bolas azuis é uma variável aleatória hipergeométrica. Uma *variável aleatória hipergeométrica* com parâmetros n, a, s tem f.m.p.

$$\text{hg}_{n,a,s}(x) = \begin{cases} \dfrac{\binom{a}{x}\binom{n-a}{s-x}}{\binom{n}{s}}, & \text{se } x \in \left\{ \max\{0, s-(n-a)\}, \ldots, \min\{a, n\} \right\} \\ 0, & \text{nos outros casos.} \end{cases}$$

Exemplo 21. Qual a probabilidade de acertar 4 dos 6 números sorteados na mega-sena se todos os resultados são igualmente prováveis? Os parâmetros são $a = s = 6$ e $n = 60$, logo

$$\mathrm{hg}_{60,6,6}(4) = \frac{\binom{6}{4}\binom{54}{2}}{\binom{60}{6}} = 0{,}0004287524.$$

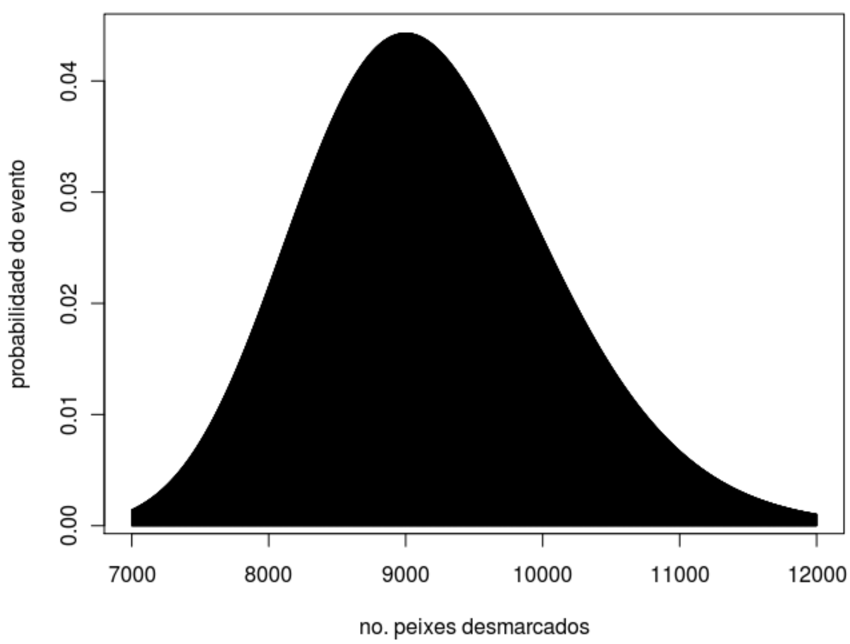
◊

Exemplo 22. Um comprador de componentes elétricos compra os componentes em lote composto de 10 componentes. A política de controle de qualidade é inspecionar 3 componentes escolhidos aleatoriamente e comprar o lote somente se os 3 não apresentarem defeitos. Se 30% dos lotes têm 4 componentes com defeito e 70% apenas 1, qual é a proporção de lotes aceitos? Consideremos os eventos: A definido por "aceita um lote" e B definido por "lote com 4 peças com defeito".

$$\mathbb{P}(A) = \mathbb{P}(A \mid B)\mathbb{P}(B) + \mathbb{P}\left(A \,\middle|\, \overline{B}\right)\mathbb{P}(\overline{B}) = \frac{\binom{4}{0}\binom{6}{3}}{\binom{10}{3}}\frac{3}{10} + \frac{\binom{1}{0}\binom{9}{3}}{\binom{10}{3}}\frac{7}{10} = \frac{54}{100}.$$

◊

Exemplo 23 (**Estimativa de máxima verossimilhança**). Num lago 1000 peixes foram capturados, marcados e devolvidos. Uma nova captura de 1000 peixes é feita e é constatado que 100 deles estão marcados. O que pode ser dito a respeito do tamanho da população de peixes no lago? A probabilidade do evento em função do número (desconhecido) de peixes desmarcados tem *gráfico de barras*



sugere uma população de aproximadamente 9.000 + 1.000 peixes (máxima verossimilhança — estimativa que maximiza a probabilidade do evento ocorrido) e, de fato, essa estimativa por ser feita de modo análogo a equação (1.5.1)). ◊

Distribuição geométrica:: uma variável aleatória tem *distribuição geométrica* com parâmetro p se tem f.m.p. dada por

$$\mathrm{ge}_p(x) = \begin{cases} (1-p)^{x-1}p, & \text{se } x \in \{1,2,\dots\} \\ 0, & \text{nos outros casos.} \end{cases}$$

que correspondente ao número de ensaios de Bernoulli independentes com parâmetro p até ocorrer um sucesso. A notação

$$X \sim \text{Geometrica}(p)$$

indica que X é uma variável aleatória geométrica com parâmetro p.

Exemplo 24. Uma urna contém N bolas brancas e M bolas pretas. As bolas são selecionadas aleatoriamente, uma de cada vez, até que saia uma bola preta. Se supormos que cada bola retirada é substituída antes da próxima retirada, qual é a probabilidade com que sejam necessárias exatamente i retiradas?

Se X é o número de retiradas até sair bola preta

$$\mathbb{P}(X=i) = \left(\frac{N}{N+M}\right)^{i-1}\left(\frac{M}{N+M}\right) = \frac{MN^{i-1}}{(N+M)^i}.$$

Com que probabilidade são necessárias pelo menos k retiradas?

$$\begin{aligned} \mathbb{P}(X \geqslant k) &= \sum_{i=k}^{\infty}\left(\frac{N}{N+M}\right)^{i-1}\left(\frac{M}{N+M}\right) \\ &= \left(\frac{M}{N+M}\right)\sum_{i=k}^{\infty}\left(\frac{N}{N+M}\right)^{i-1} \\ &= \left(\frac{M}{N+M}\right)\frac{\left(\frac{N}{N+M}\right)^{i-1}}{1-\left(\frac{N}{N+M}\right)} \\ &= \left(\frac{N}{N+M}\right)^{k-1} \end{aligned}$$

Exemplo 25 (**colecionador de cupons**). Considere a seguinte situação: há N tipos de cupons. Um colecionador compra a cada unidade de tempo um cupom aleatório. A probabilidade de obter em cada vez um cupom específico é $1/N$, independentemente das aquisições anteriores.

Uma variável aleatória de interesse é o número de unidades de tempo T até o colecionador reunir pelo menos um cupom de cada tipo. Essa variável aleatória variável toma valores no em $\{N, N+1, \ldots\} \cup \{\infty\}$.

Assumindo que o colecionador já esteja de posse de k tipos diferentes de cupons, $0 \leqslant k < N$, a quantidade de tempo até a aquisição de um tipo novo de cupom é uma variável aleatória geométrica com parâmetro $(N-k)/N$.

Distribuição binomial negativa:: Uma *variável aleatória binomial negativa* com parâmetros p e r tem f.m.p. dada por

$$\mathrm{nbi}_{p,r}(x) = \begin{cases} \binom{x-1}{r-1} p^r (1-p)^{x-r}, & \text{se } x \in \{r+1, r+2, \dots\} \\ 0, & \text{nos outros casos.} \end{cases}$$

que correspondente ao número de ensaios de Bernoulli independentes com parâmetro p até ocorrer um total de r sucessos.

1.6 Variáveis aleatórias contínuas.

Uma variável aleatória X é (absolutamente) *contínua* se

$$F_X(t) = \int_{-\infty}^{t} f_X(x)\,\mathrm{d}x \qquad (\forall t \in \mathbb{R})$$

para alguma função integrável $f_X : \mathbb{R} \to [0, +\infty)$ tal que $\int_{-\infty}^{\infty} f_X(x)\,\mathrm{d}x = 1$.

Função de densidade:: é qualquer função integrável $f : \mathbb{R} \to [0, +\infty)$ satisfazendo $\int_{-\infty}^{+\infty} f(x)\,\mathrm{d}x = 1$. A função f_X é chamada de *função de densidade de probabilidade* (f.d.p) da variável X.

Para nós, sempre valerá que: *uma variável aleatória X admite uma função de densidade se a distribuição F_X é contínua e derivável em todo ponto da reta exceto por um número enumerável deles.* Ademais, $f_X(x)$ é a derivada de $F'_X(x)$ nos pontos x em que a derivada existe, ou seja, exceto para no máximo uma quantidade enumerável de pontos $x \in \mathbb{R}$ temos

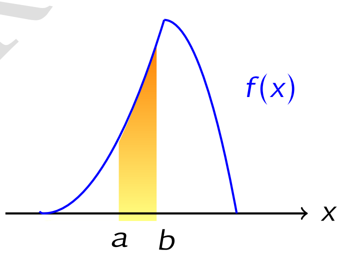
$$f_X(x) = \frac{\mathrm{d}}{\mathrm{d}x} F_X(x).$$

Assim, basta conhecer a f.d.p. de X para conhecer sua distribuição e vice-versa.

Exercício 4. Mostre que para uma função de densidade f vale que $F(t) := \int_{-\infty}^{t} f(x)\,\mathrm{d}x$ satisfaz as quatro propriedades descritas na pagina 7 e que caracterizam uma função de distribuição.

Para uma variável aleatória contínua X a f.d.a. F_X é contínua em toda reta, portanto, pelo item 3 da proposição 1.3.2 temos que $\mathbb{P}(X = a) = 0$ para todo $a \in \mathbb{R}$, logo $P(a \leqslant X \leqslant b) = P(a \leqslant X < b) = P(a < X < b) = P(a < X \leqslant b)$ e

$$P(a < X \leqslant b) = F_X(b) - F_X(a) = \int_{-\infty}^{b} f_X(x)\,\mathrm{d}x - \int_{-\infty}^{a} f_X(x)\,\mathrm{d}x = \int_{a}^{b} f_X(x)\,\mathrm{d}x.$$



$\mathbb{P}(a \leqslant X \leqslant b)$ = área da região delimitada pelo gráfico, por $x = a$ e $x = b$ e pelo eixo x no intervalo $[a, b]$.

Daí temos que para qualquer $\varepsilon > 0$ pequeno

$$\mathbb{P}\left(a - \frac{\varepsilon}{2} < X < a + \frac{\varepsilon}{2}\right) = \int_{a-\frac{\varepsilon}{2}}^{a+\frac{\varepsilon}{2}} f_X(x)\,\mathrm{d}x \approx \varepsilon f_X(a)$$

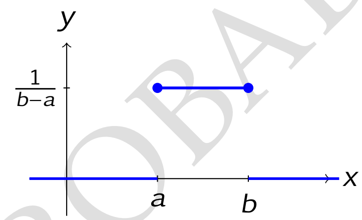
ou seja, X assume valor numa vizinhança de a de diâmetro ϵ muito pequeno com probabilidade aproximadamente $\varepsilon f_X(a)$.

Retomando o exemplo 8, a variável aleatória Z com distribuição dada por

$$F_Z(t) = \begin{cases} 0, & \text{se } t < a \\ \frac{t-a}{b-a}, & \text{se } a \leqslant t < b \\ 1, & \text{se } t \geqslant b \end{cases}$$

é contínua e tem derivada em todo ponto, exceto a e b e

$$f_Z(x) = \begin{cases} 0, & \text{se } x < a \text{ ou } x > b \\ \frac{1}{b-a}, & \text{se } a \leqslant x \leqslant b \end{cases}$$



1.6.1 Observação. *Os valores $f(a)$ e $f(b)$, no caso acima, podem ser arbitrários pois, para quaisquer que sejam esses valores, a integral $\int_{-\infty}^{t} f(x)\,\mathrm{d}x$ ainda vale $F_Z(t)$.*

1.6.2 Principais modelos contínuos: Retomemos alguns fatos, se X é v.a. contínua valem

1. $f_X(x) \geqslant 0$ para todo $x \in \mathbb{R}$;
2. $\int_{-\infty}^{+\infty} f_X(x)\mathrm{d}x = 1$;
3. $\mathbb{P}(a \leqslant X \leqslant b) = \int_a^b f_X(x)\mathrm{d}x = F_X(b) - F_X(a)$ para todo $a \leqslant b$;
4. $\mathbb{P}(X = a) = 0$, para qualquer $a \in \mathbb{R}$.

Distribuição uniforme contínua:: uma variável aleatória contínua X é *uniforme* no intervalo $[a, b]$, para $a < b$, se sua f.d.p. é

$$f(x) = \begin{cases} \frac{1}{b-a}, & \text{se } x \in [a, b] \\ 0 & \text{caso contrário} \end{cases}$$

e denotamos esse fato por $X \sim \text{Uniforme}(a, b)$.

Nesse caso, a probabilidade de X estar num subintervalo de $[a, b]$ é proporcional ao comprimento de tal subintervalo; de fato, para $y < z$ reais

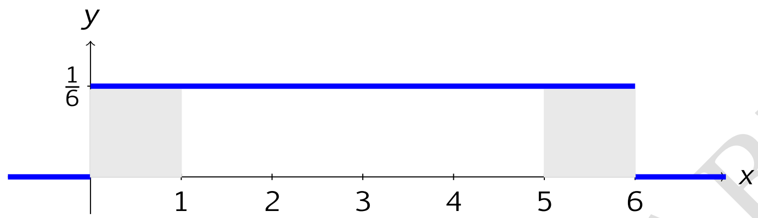
$$\mathbb{P}(y \leqslant X \leqslant z) = \int_y^z \frac{1}{b-a}\,\mathrm{d}x = \frac{z-y}{b-a}.$$

Exemplo 26. Num teste, tubos de PVC de 6 m são submetidos a grande pressão d'água até que o primeiro vazamento ocorra. A distância do início do tubo até o vazamento é uniformemente distribuída. Qual a probabilidade de que o vazamento esteja a no máximo 1 m das extremidades?

Seja $X \sim$ Uniforme$(0,6)$ a distância do início do tubo até o vazamento. A probabilidade procurada é

$$\mathbb{P}([0 \leqslant X \leqslant 1] \cup [5 \leqslant X \leqslant 6]) = \mathbb{P}(0 \leqslant X \leqslant 1) + \mathbb{P}(5 \leqslant X \leqslant 6) = \int_0^1 \frac{1}{6} \mathrm{d}x + \int_5^6 \frac{1}{6} \mathrm{d}x = \frac{1}{3}$$

que corresponde à área da região destacada em cinza



O valor médio de $X \sim$ Uniforme(a,b) é

$$\mathbb{E}[X] = \int_a^b x \frac{1}{b-a} \mathrm{d}x = \frac{a+b}{2}$$

a variância

$$\mathrm{Var}[X] = \int_a^b x^2 \frac{1}{b-a} \mathrm{d}x - \left(\frac{a+b}{2}\right)^2 = \frac{(b-a)^2}{12}$$

de modo que o desvio padrão é $\approx 0{,}29(b-a)$.

Distribuição exponencial:: Uma variável aleatória contínua X é *exponencial* com parâmetro $\lambda > 0$, e denotamos esse fato por

$$X \sim \text{Exponencial}(\lambda)$$

se sua função de densidade é

$$f_X(x) = \begin{cases} \lambda \mathrm{e}^{-\lambda x}, & \text{se } x \geqslant 0 \\ 0 & \text{caso contrário.} \end{cases}$$

A função de distribuição é

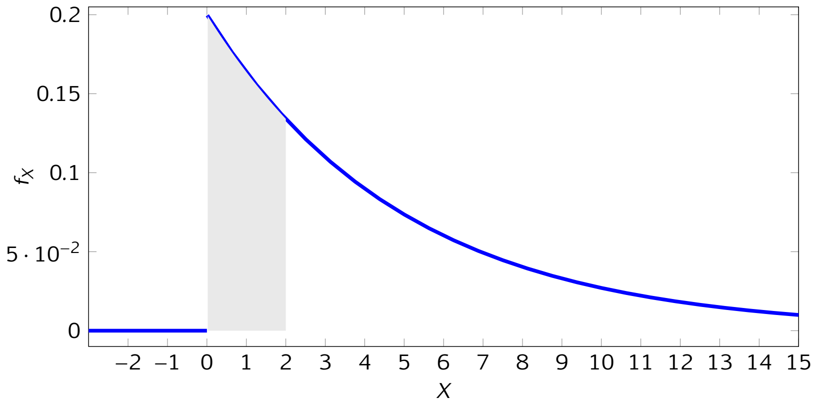
$$F_X(t) = \int_0^t \lambda \mathrm{e}^{-\lambda x} \, \mathrm{d}x = 1 - \mathrm{e}^{-\lambda x}$$

portanto $\mathbb{P}(X > a) = \mathrm{e}^{-\lambda a}$.

Variáveis aleatórias exponenciais são muitas vezes utilizados para modelar a distribuição da quantidade de tempo decorrido até que algum evento particular ocorra. Por exemplo, seja $T \sim$ Exponencial$(0{,}2)$ o intervalo de tempo (em minutos) entre emissões consecutivas de uma fonte radioativa. A probabilidade de haver uma emissão em até 2 min é $\mathbb{P}(T \leqslant 2) = 1 - \mathbb{P}(T > 2) = 1 - \mathrm{e}^{0{,}2 \cdot 2} \approx 0{,}33$ ou, de outro modo

$$\mathbb{P}(T < 2) = \int_0^2 0{,}2\mathrm{e}^{-0{,}2x} \, \mathrm{d}x = 1 - \mathrm{e}^{0{,}2 \cdot 2}$$

que corresponde a área da região em cinza no gráfico



A probabilidade do intervalo ser maior que 7 dado que foi maior que 5

$$\begin{aligned}\mathbb{P}(T>7 \mid T>5) &= \frac{\mathbb{P}([T>7]\cap[T>5])}{\mathbb{P}(T>5)} = \frac{\mathbb{P}(T>7)}{\mathbb{P}(T>5)} \\ &= \frac{\mathrm{e}^{-0,2\cdot 7}}{\mathrm{e}^{-0,2\cdot 5}} = \mathrm{e}^{-0,2\cdot 2} = \mathbb{P}(T>2)\end{aligned}$$

Dizemos que uma variável aleatória T é *sem memória* se

$$\mathbb{P}(T>s+t \mid T>t) = \mathbb{P}(T>s) \tag{1.6.1}$$

para quaisquer $s,t \geqslant 0$. Se X tem distribuição exponencial então é sem memória. Por outro lado, uma variável aleatória contínua sem memória tem distribuição exponencial.

Para ver como isso funciona, imagine que no instante $t_0 = 0$ ligamos um despertador que irá tocar depois de um tempo T que é distribuído exponencialmente com parâmetro λ. Suponha que tivemos que sair e ao voltar, no instante t, descobrimos que o despertador ainda não tocou. Seja S o tempo que decorre partir de então (i.e, observado $[T>t]$) até o despertador tocar.

$$\mathbb{P}(S>s \mid T>t) = \mathbb{P}(T>s+t \mid T>t)$$

que por equação (1.6.1) é $\mathbb{P}(T>t)$. A coisa importante de se notar é que a distribuição do tempo até ocorrer o evento não depende do instante inicial t. A distribuição exponencial é única com essa propriedade.

Exercício 5. Prove que $T \sim \text{Exponencial}(\lambda)$ é sem memória.

Exemplo 27. Suponha que um sistema contenha componentes cujo tempo até falhar é $T \sim \text{Exponencial}(1/5)$ em anos. Se 5 desses componentes são instalados em cada sistema, qual é a probabilidade de que pelo menos 2 componentes ainda estejam funcionando após 8 anos?

Se X é a quantidade de componentes funcionando após 8 anos então $X \sim \text{Binomial}(5,p)$ com $p = \mathbb{P}(T>8) = \mathrm{e}^{-8/5} \approx 0{,}2$, assim

$$\mathbb{P}(X \geqslant 2) = \sum_{x=2}^{5} \mathrm{bi}_{5,p}(x) = 1 - \sum_{x=0}^{1} \mathrm{bi}_{5,p}(x) \approx 0{,}26.$$

A esperança de $T \sim \text{Exponencial}(\lambda)$ é

$$\mathbb{E}[T] = \int_0^\infty \lambda x e^{-\lambda x} dx = \lambda \left(-\frac{e^{-\lambda x}(\lambda x + 1)}{\lambda^2} \right) \Big|_0^\infty = \frac{1}{\lambda},$$

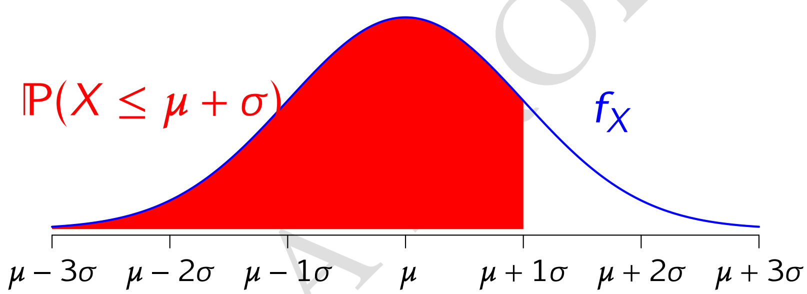
e a variância

$$\sigma^2 = \text{Var}[T] = \int_0^\infty \lambda x^2 e^{-\lambda x} dx - \left(\frac{1}{\lambda}\right)^2 = \frac{1}{\lambda^2}.$$

A distribuição normal: A variável aleatória X tem distribuição *normal* com parâmetros μ e σ^2, abreviado por $X \sim \mathcal{N}(\mu; \sigma^2)$, se sua função densidade de probabilidade é dada por

$$f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

para todo $x \in \mathbb{R}$.



O problema com o qual nos deparamos agora é que

$$\mathbb{P}(a < X < b) = \int_a^b f_X(x)\, dx$$

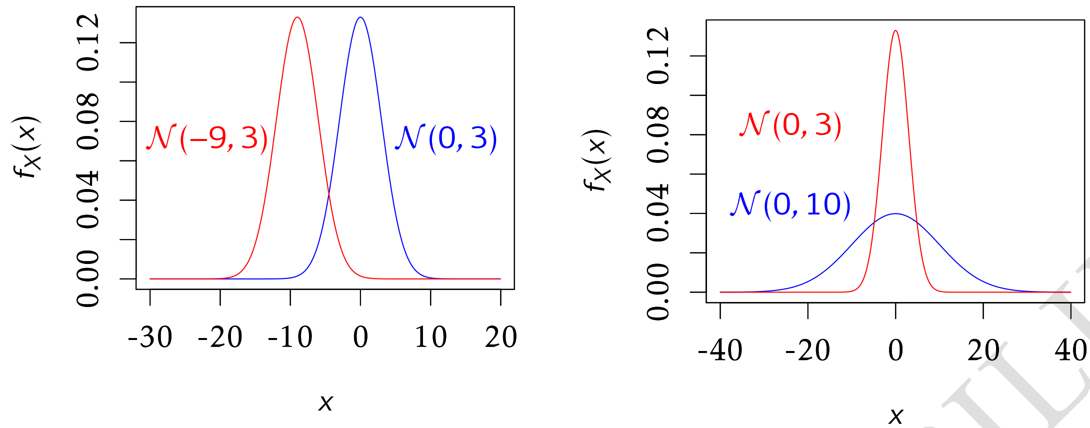
não tem solução analítica.

A distribuição normal tem as seguintes propriedades

1. $\mathbb{E}[X] = \mu$ e $\text{Var}[X] = \sigma^2$;
2. $f_X(x) \to 0$ quando $x \to \pm\infty$;
3. μ é ponto de máximo de $f_X(x)$ e $\mu - \sigma$ e $\mu + \sigma$ são pontos de inflexão de $f_X(x)$;
4. o gráfico $(x, f_X(x))$ é simétrico com relação a $x = \mu$;

Exercício 6. Prove as propriedades 2, 3 e 4.

Os gráficos abaixo mostram a função de densidade de normais com parâmetros diferentes



Uma propriedade interessante, e muito útil, é a seguinte

1.6.3 Proposição. *Se* $X \sim \mathcal{N}(\mu;\sigma^2)$ *então* $aX+b \sim \mathcal{N}(a\mu+b;a^2\sigma^2)$, *para quaisquer* $a,b \in \mathbb{R}$.

Demonstração. Escrevemos $Y = aX+b$ e temos

$$F_Y(x) = \mathbb{P}(Y \leqslant x) = \mathbb{P}(aX+b \leqslant x) = \mathbb{P}\left(X \leqslant \frac{x-b}{a}\right) = F_X\left(\frac{x-b}{a}\right)$$

se $a > 0$, logo a densidade de Y é

$$F_y'(x) = \frac{1}{a}F_X'\left(\frac{x-b}{a}\right) = \frac{1}{a\sigma\sqrt{2\pi}}e^{-\frac{((x-b)/a-\mu)^2}{2\sigma^2}} = \frac{1}{a\sigma\sqrt{2\pi}}e^{-\frac{(x-(b+a\mu))^2}{2a^2\sigma^2}}$$

portanto $Y \sim \mathcal{N}(a\mu+b;a^2\sigma^2)$, caso $a < 0$ a mesma conclusão vale e deixamos a verificação como exercício. □

Agora, sabemos pela proposgição 1.6.3 acima que para quaisquer $a,b \in \mathbb{R}$

$$\text{se } X \sim \mathcal{N}(\mu;\sigma^2) \text{ então } aX+b \sim \mathcal{N}(a\mu+b;a^2\sigma^2)$$

de modo que se tomarmos $a = 1/\sigma$ e $b = -\mu/\sigma$ temos

$$\text{se } X \sim \mathcal{N}(\mu;\sigma^2) \text{ então } \tfrac{X-\mu}{\sigma} \sim \mathcal{N}\left(\tfrac{1}{\sigma}\mu - \tfrac{\mu}{\sigma};\left(\tfrac{1}{\sigma}\right)^2\sigma^2\right), \text{ portanto, } \tfrac{X-\mu}{\sigma} \sim \mathcal{N}(0;1)$$

Com isso, temos a seguinte consequência da proposição 1.6.3

1.6.4 Corolário. *Se* $X \sim \mathcal{N}(\mu;\sigma^2)$ *então* $\frac{X-\mu}{\sigma} \sim \mathcal{N}(0;1)$. □

Se $a < x < b$ então $(a-\mu)/\sigma < (x-\mu)/\sigma < (b-\mu)/\sigma$, portanto

$$\mathbb{P}(a < X < b) = \mathbb{P}\left(\frac{a-\mu}{\sigma} < \frac{X-\mu}{\sigma} < \frac{b-\mu}{\sigma}\right).$$

Variável aleatória padronizada:: a variável aleatória padronizada de $X \sim \mathcal{N}(\mu;\sigma^2)$ é

$$Z_X := \frac{X-\mu}{\sigma}$$

e temos $Z_X \sim \mathcal{N}(0;1)$. Agora,

$$\boxed{\mathbb{P}(a < X < b) = \mathbb{P}\left(\frac{a-\mu}{\sigma} < Z_X < \frac{b-\mu}{\sigma}\right)}$$

de modo que para calcular probabilidades que envolvem uma variável aleatória normal $\mathcal{N}(\mu;\sigma^2)$ basta conhecermos a função de distribuição de $Z \sim \mathcal{N}(0;1)$

$$\Phi(x) := \mathbb{P}(Z \leqslant x) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{x} \mathrm{e}^{-t^2/2}\,\mathrm{d}t$$

(é costume usar $\Phi(x)$ para denotar a f.d.a. de uma variável aleatória com distribuição normal).

Exercício 7. Prove que $X \sim \mathcal{N}(0;1)$ tem média 0 e variância 1. Deduze desse fato a média e a variância de $Y \sim \mathcal{N}(\mu;\sigma^2)$.

Tabela da distribuição normal padrão

Para $Z \sim \mathcal{N}(0;1)$, a tabela que usamos é da forma

x	0,00	0,01	0,02	0,03	0,04	$\cdots$
$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$
1,3	$\cdots$	$\cdots$	0,9066	$\cdots$	0,9099	$\cdots$
$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$

e para calcularmos $\mathbb{P}(Z \leqslant 1,32)$ decompomos 1,32 = 1,3+0,02 (parte inteira e primeira casa decimal + segunda casa decimal), em seguida $\mathbb{P}(Z \leqslant 1,32)$ é lido na linha indexada por 1,3 com a coluna indexada por 0,02, no caso $\mathbb{P}(Z \leqslant 1,32) = 0,9066$. Analogamente, $\mathbb{P}(Z \leqslant 1,34) = 0,9099$.

Vejamos mais alguns exemplos de consulta à tabela da distribuição normal

1. quanto é $\mathbb{P}(0 < Z \leqslant 1{,}71)$?

   $\mathbb{P}(0 < Z \leqslant 1{,}71) = \mathbb{P}(Z \leqslant 1{,}71) - \mathbb{P}(Z < 0) = 0{,}9564 - 0{,}5 = 0{,}4564$

2. quanto é $\mathbb{P}(0{,}32 \leqslant Z \leqslant 1{,}71)$?

   $\mathbb{P}(0{,}32 \leqslant Z \leqslant 1{,}71) = \mathbb{P}(Z \leqslant 1{,}71) - \mathbb{P}(Z < 0{,}32) = 0{,}9564 - 0{,}6255 = 0{,}3309$

3. quanto é $\mathbb{P}(Z \leqslant -1{,}71)$?

   $\mathbb{P}(Z \leqslant -1{,}71) = \mathbb{P}(Z \geqslant 1{,}71) = 1 - \mathbb{P}(Z < 1{,}71) = 1 - 0{,}9564 = 0{,}0436$

4. quanto é $\mathbb{P}(-1{,}71 \leqslant Z \leqslant 1{,}71)$?

$$\mathbb{P}(-1{,}71 \leqslant Z \leqslant 1{,}71) = \mathbb{P}(Z \leqslant 1{,}71) - \mathbb{P}(Z < -1{,}71) = 0{,}9564 - 0{,}0436 = 0{,}9128$$

5. ou seja, genericamente, se $Z \sim \mathcal{N}(0;1)$ então para $y \geqslant x \geqslant 0$ reais temos

- $\mathbb{P}(Z \leqslant x) = \Phi(x)$
- $\mathbb{P}(y \leqslant Z \leqslant x) = \Phi(x) - \Phi(y)$
- $\mathbb{P}(Z \leqslant -x) = \Phi(-x) = 1 - \Phi(x)$
- $\mathbb{P}(-x \leqslant Z \leqslant x) = \Phi(x) - \Phi(-x) = 2\Phi(x) - 1$

6. Como encontrar o valor z da distribuição $\mathcal{N}(0;1)$ tal que $\mathbb{P}(Z \leqslant z) = 0,975$?

Consultando a parte relevante da tabela

x	0,00	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
1,8	0,9641	0,9649	0,9656	0,9664	0,9671	0,9678	0,9686	0,9693	0,9699	0,9706
1,9	0,9713	0,9719	0,9726	0,9732	0,9738	0,9744	0,9750	0,9756	0,9761	0,9767
2,0	0,9772	0,9778	0,9783	0,9788	0,9793	0,9798	0,9803	0,9808	0,9812	0,9817

obtemos $z = 1,96$.

7. Qual é o z tal que $\mathbb{P}(0 < Z \leqslant z) = 0,4664$? Como da tabela determinamos probabilidade de forma $\mathbb{P}(Z \leqslant z)$ basta lembrar que, por simetria $\mathbb{P}(Z \leqslant 0) = 0{,}5$ logo, se somarmos $0,5 + 0,4664$ temos $\mathbb{P}(Z \leqslant z) = 0,5 + 0,4664$, portanto $z = 1{,}83$.

O z tal que $\mathbb{P}(Z \geqslant z) = 0,0228$ e determinado por

$$1 - 0,0228 = 0,9772 \Longrightarrow z = 2$$

Distribuição normal padrão $\mathcal{N}(0;1)$ - Tabela f.d.a.

$\mathbb{P}(Z \leq x) = \Phi(x)$

x	0,00	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
0,0	0,5000	0,5040	0,5080	0,5120	0,5160	0,5199	0,5239	0,5279	0,5319	0,5359
0,1	0,5398	0,5438	0,5478	0,5517	0,5557	0,5596	0,5636	0,5675	0,5714	0,5753
0,2	0,5793	0,5832	0,5871	0,5910	0,5948	0,5987	0,6026	0,6064	0,6103	0,6141
0,3	0,6179	0,6217	0,6255	0,6293	0,6331	0,6368	0,6406	0,6443	0,6480	0,6517
0,4	0,6554	0,6591	0,6628	0,6664	0,6700	0,6736	0,6772	0,6808	0,6844	0,6879
0,5	0,6915	0,6950	0,6985	0,7019	0,7054	0,7088	0,7123	0,7157	0,7190	0,7224
0,6	0,7257	0,7291	0,7324	0,7357	0,7389	0,7422	0,7454	0,7486	0,7517	0,7549
0,7	0,7580	0,7611	0,7642	0,7673	0,7703	0,7734	0,7764	0,7794	0,7823	0,7852
0,8	0,7881	0,7910	0,7939	0,7967	0,7995	0,8023	0,8051	0,8078	0,8106	0,8133
0,9	0,8159	0,8186	0,8212	0,8238	0,8264	0,8289	0,8315	0,8340	0,8365	0,8389
1,0	0,8413	0,8438	0,8461	0,8485	0,8508	0,8531	0,8554	0,8577	0,8599	0,8621
1,1	0,8643	0,8665	0,8686	0,8708	0,8729	0,8749	0,8770	0,8790	0,8810	0,8830
1,2	0,8849	0,8869	0,8888	0,8907	0,8925	0,8944	0,8962	0,8980	0,8997	0,9015
1,3	0,9032	0,9049	0,9066	0,9082	0,9099	0,9115	0,9131	0,9147	0,9162	0,9177
1,4	0,9192	0,9207	0,9222	0,9236	0,9251	0,9265	0,9279	0,9292	0,9306	0,9319
1,5	0,9332	0,9345	0,9357	0,9370	0,9382	0,9394	0,9406	0,9418	0,9429	0,9441
1,6	0,9452	0,9463	0,9474	0,9484	0,9495	0,9505	0,9515	0,9525	0,9535	0,9545
1,7	0,9554	0,9564	0,9573	0,9582	0,9591	0,9599	0,9608	0,9616	0,9625	0,9633
1,8	0,9641	0,9649	0,9656	0,9664	0,9671	0,9678	0,9686	0,9693	0,9699	0,9706
1,9	0,9713	0,9719	0,9726	0,9732	0,9738	0,9744	0,9750	0,9756	0,9761	0,9767
2,0	0,9772	0,9778	0,9783	0,9788	0,9793	0,9798	0,9803	0,9808	0,9812	0,9817
2,1	0,9821	0,9826	0,9830	0,9834	0,9838	0,9842	0,9846	0,9850	0,9854	0,9857
2,2	0,9861	0,9864	0,9868	0,9871	0,9875	0,9878	0,9881	0,9884	0,9887	0,9890
2,3	0,9893	0,9896	0,9898	0,9901	0,9904	0,9906	0,9909	0,9911	0,9913	0,9916
2,4	0,9918	0,9920	0,9922	0,9925	0,9927	0,9929	0,9931	0,9932	0,9934	0,9936

Exemplo 28. O tempo gasto no exame vestibular de uma universidade tem distribuição normal, com média 120 min e desvio padrão 15 min.

1. Qual é a probabilidade com que um candidato termine o exame antes de 100 minutos?

   Se X é o tempo gasto no exame vestibular, então $X \sim \mathcal{N}(120;15^2)$ logo

   $$\mathbb{P}(X < 100) = \mathbb{P}\left(Z_X \leqslant \frac{100-120}{15}\right) = \mathbb{P}(Z_X \leqslant -1{,}33) = 1 - \mathbb{P}(Z_X < 1{,}33)$$

   usando a tabela $1 - \mathbb{P}(Z < 1,33) = 1 - 0,9082 = 0{,}0918$.

2. Qual deve ser o tempo de prova de modo a permitir que 95% dos vestibulandos terminem no prazo estipulado?

Para determinar o tempo de prova de modo a permitir que 95% dos vestibulandos terminem no prazo estipulado, devemos encontrar x tal que $\mathbb{P}(X < x) = 0{,}95$, ou seja, tal que

$$\mathbb{P}\left(Z_X \leqslant \frac{x-120}{15}\right) = 0{,}95.$$

Pela tabela $\mathbb{P}(Z \leqslant 1{,}64) = 0{,}95$ portanto

$$\frac{x-120}{15} = 1{,}64$$

ou seja $x = 144{,}6\,\text{min}$.

3. Qual é o intervalo central de tempo, tal que 80% dos estudantes gastam para completar o exame?

   O intervalo central de tempo, tal que 80% dos estudantes gastam para completar o exame é $I = (\mu - \epsilon, \mu + \epsilon)$ tal que $\mathbb{P}(I) = 0{,}8$

$$\mathbb{P}(a \leqslant X \leqslant b) = 0{,}8 \Leftrightarrow \mathbb{P}\left(\frac{a-120}{15} \leqslant Z_X \leqslant \frac{b-120}{15}\right) = 0{,}8$$

   Pela tabela $\mathbb{P}(-1{,}28 \leqslant Z \leqslant 1{,}28) = 0{,}80$, portanto $a = 100{,}8$ e $b = 139{,}2$ minutos.

Exemplo 29. Um sistema considera que um sinal digital será transmitido quando a tensão exceder $0,9\,\text{V}$. Na detecção do sinal o ruído tem distribuição $\mathcal{N}(0;0{,}45)$. Qual a probabilidade de detectar um sinal quando nada tiver sido enviado?

Se $R \sim \mathcal{N}(0;0{,}45)$ é a tensão do ruído, então

$$\mathbb{P}(R > 0{,}9) = \mathbb{P}\left(\frac{R}{0{,}45} > \frac{0{,}9}{0{,}45}\right) = \mathbb{P}(Z_X > 2) = 1 - 0{,}97725 = 0{,}02275.$$

O intervalo central que inclui 99% de todas as leituras de ruído é dado por x tal que

$$\mathbb{P}(-x < R < x) = \mathbb{P}\left(\frac{-x}{0{,}45} < \frac{R}{0{,}45} < \frac{x}{0{,}45}\right)\mathbb{P}\left(\frac{-x}{0{,}45} < Z_X < \frac{x}{0{,}45}\right) = 0{,}99.$$

De acordo com a tabela, $x/0{,}45 = 2{,}58$, ou seja, $x = 1{,}16$.

Suponha que quando um sinal é transmitido a média da variável aleatória R mude para $1{,}8\,\text{V}$. Qual a probabilidade do sinal não ser detectado? Seja S a tensão quando um sinal é transmitido.

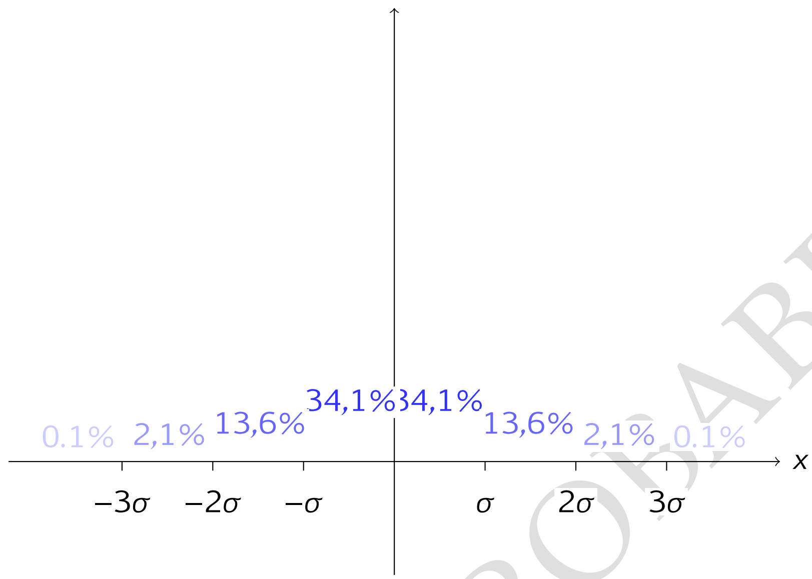
$$\mathbb{P}(S < 0{,}9) = \mathbb{P}\left(\frac{S-1{,}8}{0{,}45} < \frac{0{,}9-1{,}8}{0{,}45}\right) = \mathbb{P}(Z < -2) = 0{,}02275.$$

Essa é a probabilidade com que um sinal é perdido.

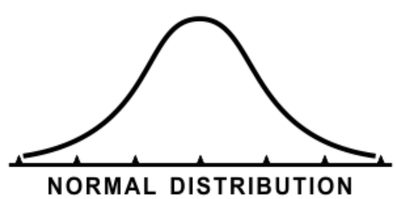
Concentração em torno de μ

Sejam $k \in \mathbb{N}$, $X \sim \mathcal{N}(\mu; \sigma^2)$ e $Z = (X - \mu)/\sigma$ então

$$\mathbb{P}(\mu - k\sigma \leqslant X \leqslant \mu + k\sigma) = \mathbb{P}(-k \leqslant Z \leqslant k) = \mathbb{P}(Z < k) - \mathbb{P}(Z < -k).$$



- Para $k = 1$, $\mathbb{P}(\mu - \sigma \leqslant X \leqslant \mu + \sigma) = \mathbb{P}(-1 \leqslant Z \leqslant 1) = 0{,}682$.
- Para $k = 2$, $\mathbb{P}(\mu - 2\sigma \leqslant X \leqslant \mu + 2\sigma) = \mathbb{P}(-2 \leqslant Z \leqslant 2) = 0{,}954$.
- Para $k = 3$, $\mathbb{P}(\mu - 3\sigma \leqslant X \leqslant \mu + 3\sigma) = \mathbb{P}(-3 \leqslant Z \leqslant 3) = 0{,}997$.



1.6.5 Esperança e variância de uma variável aleatória discreta: Se X é uma variável aleatória discreta com função de massa de probabilidade f_X então a esperança (também chamada de valor médio ou valor esperado) da variável aleatória X é, simplesmente, a média ponderada dos valores da função de probabilidade

$$\mathbb{E}[X] := \sum_x x \cdot f_X(x) \tag{1.6.2}$$

onde a soma é sobre os valores de x tais que $f_X(x) > 0$, ou seja, é a média do valores x na imagem de X ponderada pela probabilidade com que X assume esse valor, a saber $f_X(x) = \mathbb{P}(X = x)$.

O suporte de uma função é o conjunto de pontos do domínio em que a função é diferente de 0. Quando o suporte de f_X é finito equação (1.6.2) fica bem definida. Por exemplo, no caso de uma variável aleatória constante, digamos $X(\omega) = c$ para algum $c \in \mathbb{R}$, temos

$$\mathbb{E}[X] = \sum_x x \cdot f_X(x) = cf_X(c) = c.$$

Se $\mathbf{1}_A$ é a *variável aleatória indicadora* da ocorrência do evento A (definição na pág. 6) então $\mathbb{E}[\mathbf{1}_A] = 1\mathbb{P}(A) + 0\mathbb{P}(\overline{A}) = \mathbb{P}(A)$. Notemos que $\mathbf{1}_A \sim$ Bernoulli(p) para $p = \mathbb{P}(A)$ e, de fato, vale

1.6.6 Proposição. *Se* $X \sim$ Bernoulli(p) *então* $\mathbb{E}[X] = p$.

Exemplo 30. Seja X o resultado de um lançamento de um dado,

$$\mathbb{E}[X] = 1\frac{1}{6} + 2\frac{1}{6} + 3\frac{1}{6} + 4\frac{1}{6} + 5\frac{1}{6} + 6\frac{1}{6} = \frac{7}{2}.$$

Qual a probabilidade $\mathbb{P}(X = 7/2)$?

Esse exemplo chama a atenção para o fato de que o *valor esperado* para o resultado do lançamento de dum dado é um valor que não ocorre como imagem de X. Além disso, como os resultados têm a mesma probabilidade o valor esperado é a média no sentido usual, a soma de seis valores dividida por seis.

Exemplo 31. Num jogo de azar em cada aposta você ganha \$1.000.000 com probabilidade p e \$10 com probabilidade $1 - p$. Se Y é o ganho numa aposta, então a esperança de ganho numa aposta é

$$\mathbb{E}[Y] = 10^6 p + 10(1 - p).$$

No caso de $p = 1/2$, temos $\mathbb{E}[Y] = 500.005$, qual é a probabilidade de você ganhar \$500.005 numa aposta?

No exemplo anterior, se $p = 1/100$ então a probabilidade de ganhar um valor alto é muito pequeno, comparado com a probabilidade ganhar 10 reais. Apesar de que, com grande probabilidade, o ganho numa aposta seja de \$10 ainda assim o valor esperado de ganho numa única aposta é grande, $\mathbb{E}[Y] = 10.009{,}90$.

Exemplo 32. Num jogo com 3 moedas, você ganha \$5 se ocorrerem três caras ou três coroas, você perde \$3 se ocorrer uma ou duas caras, se Y é o ganho numa rodada então a esperança do ganho é

$$\mathbb{E}[Y] = 5\frac{1}{4} - 3\frac{3}{4} = -1.$$

Com que probabilidade um jogador ganha \$7 em três rodadas consecutivas e independentes?

Exemplo 33. Qual é o valor médio da soma dos pontos no lançamento de dois dados? O espaço amostral é composto por 36 eventos elementares igualmente prováveis, se X é o resultado da soma dos lançamentos, então sua função de massa de probabilidade é

x	2	3	4	5	6	7	8	9	10	11	12
$p(x)$	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{3}{36}$	$\frac{4}{36}$	$\frac{5}{36}$	$\frac{6}{36}$	$\frac{5}{36}$	$\frac{4}{36}$	$\frac{3}{36}$	$\frac{2}{36}$	$\frac{1}{36}$

O valor esperado da soma é

$$\mathbb{E}[X] = 2\frac{1}{36} + 3\frac{2}{36} + \cdots + 11\frac{2}{36} + 12\frac{1}{36} = \frac{252}{36} = 7.$$

Exemplo 34 ($\mathbb{E}[X] = \infty$). Numa urna estão 1 bola branca e 1 bola preta; uma bola é escolhida ao acaso, se for preta ela é devolvida e mais uma bola preta é colocada na urna e o sorteio é repetido, se sair bola branca o experimento termina. Se X é o número de rodadas até terminar então

$$\mathbb{P}(X = k) = \frac{1}{k(k+1)}$$

e a média é

$$\sum_{k=1}^{\infty} k\frac{1}{k(k+1)} = \sum_{k=2}^{\infty} \frac{1}{k} = \infty$$

(essa é a série harmônica).

1.6.7 Observação. *No caso do último exemplo dizemos que a variável aleatória não tem esperança finita.*

1.6.8 Teorema. *Se X é uma variável aleatória com esperança finita e g uma função real a valores reais, então $g(X)$ é uma variável aleatória e*

$$\boxed{\mathbb{E}[g(X)] = \sum_{x} g(x) f_X(x)}$$

com a soma sobre todo x tal que $f_X(x) > 0$.

Demonstração. Definamos $Y := g(X)$ e temos

$$\mathbb{E}[Y] = \sum_{y} y\,\mathbb{P}(Y = y) = \sum_{y} y\,\mathbb{P}(g(X) = y).$$

Se $\omega \in [g(X) = y]$ então existe um $x \in \mathbb{R}$ tal que $X(\omega) = x$ e $g(x) = y$ e, se $\omega \in [X = x]$ para algum x tal que $g(x) = y$ então $\omega \in [g(X) = y]$, portanto

$$[g(X) = y] = \bigcup_{\substack{x \\ g(x)=y}} [X = x]$$

em que a união é sobre todo x tal que $g(x) = y$ e é uma união de eventos mutuamente exclusivos, de modo que

$$\mathbb{E}[Y] = \sum_{y} y\,\mathbb{P}(g(X) = y) = \sum_{y} y \left(\sum_{\substack{x \\ g(x)=y}} \mathbb{P}(X = x) \right) = \sum_{y} \sum_{\substack{x \\ g(x)=y}} g(x)\mathbb{P}(X = x) = \sum_{x} g(x)\mathbb{P}(X = x)$$

que é o resultado que queríamos obter. □

1.6.9 Corolário (linearidade da esperança). *Para quaisquer* $a, b \in \mathbb{R}$, *vale* $\mathbb{E}[aX + b] = a\mathbb{E}[X] + b$.

Demonstração. Basta tomar $g(x) = ax + b$ no teorema. □

1.6.10 Proposição. *Se* $X \sim \text{Binomial}(n, p)$ *e* k *é um inteiro positivo então*

$$\mathbb{E}\left[X^k\right] = np\mathbb{E}\left[(Y+1)^{k-1}\right]$$

em que $Y \sim \text{Binomial}(n-1, p)$.

Demonstração. Pelo teorema 1.6.8 acima

$$\mathbb{E}\left[X^k\right] = \sum_{i=0}^{n} i^k \binom{n}{i} p^i (1-p)^{n-i}$$

e usando

$$i\binom{n}{i} = n\binom{n-1}{i-1}$$

de modo que

$$\mathbb{E}\left[X^k\right] = \sum_{i=1}^{n} i^{k-1} n\binom{n-1}{i-1} p^i (1-p)^{n-i} = np\sum_{i=1}^{n} i^{k-1}\binom{n-1}{i-1} p^{i-1}(1-p)^{n-i}$$

fazendo a mudança de variável $j = i - 1$

$$\mathbb{E}\left[X^k\right] = np\sum_{j=0}^{n-1} (j+1)^{k-1}\binom{n-1}{j} p^j (1-p)^{n-1-j}$$

e resta observar que a soma acima $\sum_j (j+1)^{k-1}\binom{n-1}{j} p^j (1-p)^{n-1-j}$ é exatamente $\mathbb{E}\left[(Y+1)^{k-1}\right]$ para $Y \sim \text{Binomial}(n-1, p)$, também pelo teorema 1.6.8 acima. □

1.6.11 Corolário. *Se* $X \sim \text{Binomial}(n, p)$ *então* $\mathbb{E}[X] = np$.

Demonstração. Basta tomar $k = 1$ na proposição 1.6.10. □

Exemplo 35. Seja X a variável aleatória que descreve o número de carros lavados num lava-rápido em 1 hora, cuja função de probabilidade é

x	4	5	6	7	8	9
$f_X(x)$	1/12	1/12	1/4	1/4	1/6	1/6

Se para x carros lavados o atendente recebe $2x - 1$ reais do gerente, então o ganho médio, por hora, é

$$\mathbb{E}[2X - 1] = \sum_{x=4}^{9} (2x-1) f_X(x) = 12{,}67.$$

Variância: A variância da variável aleatória X é uma medida de quão dispersos estão os valores que a variável assume com relação ao valor médio. Por exemplo, se $\mathbb{P}(X = 100) = \mathbb{P}(X = -100) = 1/2$ e $\mathbb{P}(Y = 1) = \mathbb{P}(Y = -1) = 1/2$ então X e Y têm valor esperado 0 mas Y assume valores mais próximos da média que X. Essa característica da distribuição dos valores em torno da média não é capturada pela esperança. Uma opção é computar o valor médio da distância entre um valor de X e o valor médio $\mathbb{E}[X]$, isto é, a esperança de $|X - \mathbb{E}[X]|$, entretanto, uma solução mais conveniente do ponto de vista matemático é calcular o valor médio do quadrado desses desvios $|X - \mathbb{E}[X]|^2$.

a variância de uma variável aleatória X de esperança finita é o valor esperado da variável aleatória $g(X) = (X - \mathbb{E}[X])^2$.

$$\mathrm{Var}[X] := \mathbb{E}\left[(X - \mathbb{E}[X])^2\right]$$

Da definição temos

$$\mathrm{Var}[X] = \sum_x (x - \mathbb{E}[X])^2 f_X(x)$$

e o seguinte exercício fornece um modo, em geral, mais fácil para computar a variância.

Exercício 8. Prove que

$$\mathrm{Var}[X] = \mathbb{E}\left[X^2\right] - (\mathbb{E}[X])^2.$$

Desvio padrão:: é definido como a raiz quadrada positiva da variância

$$\sigma_X = \sqrt{\mathrm{Var}(X)}.$$

Suponhamos que X é, por exemplo, a quantidade de refrigerante engarrafada por uma máquina de uma fábrica em ml (mililitros). Então $\mathrm{Var}[X]$ é a dispersão dos valores de X com respeito a média em ml^2, o desvio padrão é uma grandeza em ml.

Exemplo 36. Se X é o resultado do lançamento de um dado

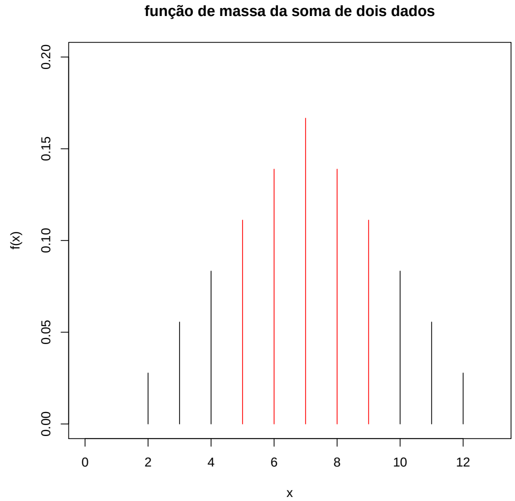
$$\mathbb{E}\left[X^2\right] = \frac{1+4+9+16+25+36}{6} = \frac{91}{6}$$

e a variância no valor resultante do lançamento de dado é $91/6 - (7/2)^2 = 35/12 \approx 2,91$ (veja o exemplo 30). O desvio padrão é $\approx 1,7$.

Exemplo 37. Retomando o exemplo 33, se X é o resultado da soma do lançamento de dois dados, então o valor esperado da soma é 7 e o da variância é

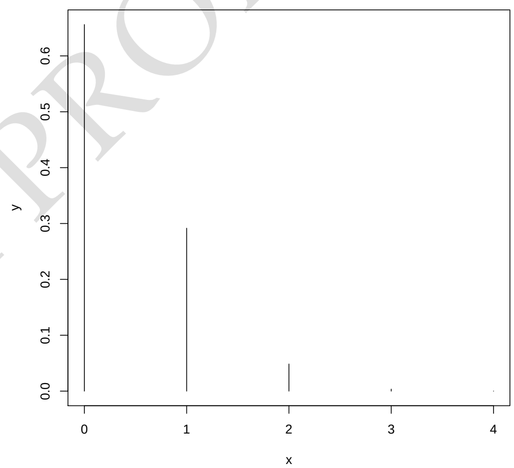
$$\mathrm{Var}(X) = (2-7)^2(1/36) + (3-7)^2(2/36) + \cdots + (11-7)^2(2/36) + (12-7)^2(1/36) = \frac{210}{36} = 5,83.$$

O desvio padrão vale $\sigma_X \approx 2,41$. Notemos que no intervalo $(\mathbb{E}[X] - \sigma_X, \mathbb{E}[X] + \sigma_X) = \{5,6,7,8,9\}$ está concentrado 2/3 da massa de probabilidade.



Exemplo 38. Um canal digital transmite informação em pacotes de 4 bits. Os bit podem ser recebidos com erro e X denota o número de bits errados num pacote, com função de massa de probabilidade

x	$f(x)$
0	$0,6561$
1	$0,2916$
2	$0,0486$
3	$0,0036$
4	$0,0001$



e o valor médio do número de bits errados é

$$\begin{aligned}\mathbb{E}[X] &= 0 \cdot f(0) + 1 \cdot f(1) + 2 \cdot f(2) + 3 \cdot f(3) + 4 \cdot f(4) \\ &= 0 \cdot 0,6561 + 1 \cdot 0,2916 + 2 \cdot 0,0486 + 3 \cdot 0,0036 + 4 \cdot 0,0001 \\ &= 0,4\end{aligned}$$

e a variância

$$\begin{aligned}\mathrm{Var}[X] &= \sum_{x=0}^{4} (x - \mathbb{E}[X])^2 f(x) \\ &= 0,16 \cdot 0,6561 + 0,36 \cdot 0,2916 + 2,56 \cdot 0,0486 + 6,76 \cdot 0,0036 + 12,96 \cdot 0,001296 \\ &= 0,36\end{aligned}$$

portanto o desvio padrão é $0,6$.

Exemplo 39 (**variância de uma variável Bernoulli**). Se $X \sim \text{Bernoulli}(p)$ então

$$\text{Var}[X] = \mathbb{E}\left[X^2\right] - (\mathbb{E}\,[X])^2 = p - p^2$$

pois $\mathbb{E}\left[X^2\right] = 1^2 p + 0^2(1-p) = p$.

Exemplo 40 (**variância de uma variável binomial**). Se $X \sim \text{Binomial}(n,p)$ então, pela proposição 1.6.10 e seu corolário

$$\text{Var}[X] = \mathbb{E}\left[X^2\right] - (\mathbb{E}\,[X])^2 = np\mathbb{E}\,[Y+1] - (np)^2$$

com $Y \sim \text{Binomial}(n,p)$. Pela linearidade da esperança $\mathbb{E}\,[Y+1] = \mathbb{E}\,[Y] + 1$, logo $\mathbb{E}\,[Y+1] = (n-1)p+1$. Assim,

$$\text{Var}[X] = np(n-1)p + np - (np)^2 = np - np^2 = np(1-p).$$

Exercício 9 (**esperança e variância de uma variável Poisson**). A esperança e a variância de $X \sim \text{Poisson}(\lambda)$ é

$$\mathbb{E}\,[X] = \text{Var}[X] = \lambda.$$

Use que $\sum_{j \geqslant 0} \frac{\lambda^j}{j!} = \mathrm{e}^{\lambda}$ para provar que $\mathbb{E}\,[X] = \lambda$. Prove que $\mathbb{E}\left[X^2\right] = \lambda(\lambda+1)$. Conclua que $\text{Var}[X] = \lambda$.

Exercício 10 (**esperança e variância de uma variável hipergeométrica**). A média e a variância de uma variável aleatória hipergeométrica são dadas por

$$\mathbb{E}\,[X] = sp \qquad \text{e} \qquad \text{Var}[X] = sp(1-p)\frac{n-s}{n-1}$$

em que $p = \frac{a}{n}$. A prova dessas fórmulas é similar à prova para variável aleatória binomial, segue da identidade $\mathbb{E}\left[X^k\right] = \frac{sa}{n}\mathbb{E}\left[(Y+1)^k\right]$ em que Y é uma variável aleatória hipergeométrica com parâmetros $n-1, a-1, s-1$. Use as identidades

$$i\binom{a}{i} = a\binom{a-1}{i-1} \qquad \text{e} \qquad s\binom{n}{s} = n\binom{n-1}{a-1}$$

e obtenha delas $\mathbb{E}\left[X^k\right] = \frac{as}{n}\mathbb{E}\left[(Y+1)^k\right]$. Em seguida, deduza a esperança e variância.

1.6.12 Proposição. *Se X é uma variável aleatória discreta com esperança finita e $a, b \in R$, então*

$$\text{Var}[aX+b] = a^2\text{Var}[X].$$

Demonstração. Sejam X, a e b como no enunciado. Abaixo usamos várias vezes a linearidade da esperança, corolário 1.6.9.

Da definição $\text{Var}[aX+b] = \mathbb{E}\left[(aX+b)^2\right] - (\mathbb{E}[aX+b])^2$. Do primeiro termo deduzimos

$$\mathbb{E}\left[(aX+b)^2\right] = \mathbb{E}\left[a^2X^2 + 2abX + b^2\right] = \mathbb{E}\left[a^2X^2\right] + \mathbb{E}[2abX] + \mathbb{E}\left[b^2\right] = a^2\mathbb{E}\left[X^2\right] + 2ab\mathbb{E}[X] + b^2$$

e do segunda termo $(\mathbb{E}[aX+b])^2 = (a\mathbb{E}[X]+b)^2 = a^2(\mathbb{E}[X])^2 + 2ab\mathbb{E}[X] + b^2$ portanto

$$\text{Var}(aX+b) = a^2\mathbb{E}\left[X^2\right] - a^2(\mathbb{E}[X])^2 = a^2\left(\mathbb{E}\left[X^2\right] - (\mathbb{E}[X])^2\right) = a^2\text{Var}[X].$$

□

1.6.13 Esperança e variância de variáveis aleatórias contínuas: Se X é uma variável aleatória com função de densidade de probabilidade f_X então o valor médio (ou valor esperado, ou esperança) da variável aleatória X é

$$\mathbb{E}[X] := \int_{-\infty}^{+\infty} x f_X(x)\,\mathrm{d}x \tag{1.6.3}$$

1.6.14 Observação (uma justificativa informal para equação (1.6.3)). *A definição de valor médio no caso discreto é intuitiva. No caso contínuo podemos justificar, ingenuamente, da seguinte maneira: Sejam* $I_n = (y_n, y_{n+1}]$, *para* $n \in \mathbb{Z}$, *uma coleção de intervalos centrados em* x_n *que* particiona *a reta e que, por simplicidade, supomos todos do mesmo comprimento* ε. *Definimos a variável aleatória discreta* Y *sobre o mesmo espaço amostral dada por*

$$Y = \sum_n x_n \mathbf{1}_{[X \in I_n]}$$

que assume os valores x_n $(n \in \mathbb{Z})$. *Assim* $[Y = x_n] = [X \in I_n]$ *e a esperança de* Y *é*

$$\mathbb{E}[Y] = \sum_n x_n \mathbb{P}(Y = x_n) = \sum_n x_n \mathbb{P}(X \in I_n).$$

Notemos que se $\omega \in [X \in I_n]$, *então* $X(\omega) \in I_n$ *e* $Y(\omega) = x_n$, *logo*

$$|Y(\omega) - X(\omega)| \leqslant \frac{|y_{n+1} - y_n|}{2} = \frac{\varepsilon}{2} \quad (\forall \omega \in \Omega).$$

Disso, a definição de esperança para a variável X *deve satisfazer* $|\mathbb{E}[X] - \mathbb{E}[Y]| \leqslant \frac{\varepsilon}{2}$, *logo* $EX = \lim_{\varepsilon \to 0} \mathbb{E}[Y]$

$$\mathbb{E}[X] = \lim_{\varepsilon \to 0} \sum_n x_n \mathbb{P}(y_n < X \leqslant y_{n+1}) = \lim_{\varepsilon \to 0} \sum_n x_n (F_X(y_{n+1}) - F_X(y_n)) = \int_{-\infty}^{+\infty} x\,\mathrm{d}F_X = \int_{-\infty}^{+\infty} x f_X(x)\,\mathrm{d}x$$

(lembremos que $F_X' = f_X$*).*

Por exemplo, seja T o tempo de vida útil de um equipamento eletrônico em horas. T tem f.d.p.

$$f(t) = \begin{cases} \frac{20.000}{t^3} & \text{se } t > 100 \\ 0 & \text{caso contrário.} \end{cases}$$

O tempo médio de vida é

$$\mathbb{E}[T] = \int_{-\infty}^{+\infty} x f(x)\,\mathrm{d}x = \int_{100}^{+\infty} \frac{20.000}{x^2}\,\mathrm{d}x = 200 \text{ horas.}$$

Exemplo 41 ($\mathbb{E}[X] = \infty$). Seja X uma variável com f.d.p.

$$f(x) = \begin{cases} \frac{10}{x^2} & \text{se } x > 10 \\ 0 & \text{caso contrário.} \end{cases}$$

X tem esperança

$$\mathbb{E}[X] = \int_{10}^{\infty} x\frac{10}{x^2}\,\mathrm{d}x = \int_{10}^{\infty} \frac{10}{x}\,\mathrm{d}x = 10\ln(x)\Big|_{10}^{\infty} = \infty$$

1.6.15 Observação. *No caso acima dizemos que a variável aleatória não tem esperança finita.*

1.6.16 Teorema. *Se X é uma variável aleatória com esperança finita e função de densidade de probabilidade f_X e g uma função real a valores reais contínua, então $g(X)$ é uma variável aleatória contínua e*

$$\boxed{\mathbb{E}[g(X)] = \int_{-\infty}^{+\infty} g(x)f(x)\,\mathrm{d}x}$$

Demonstração. Começamos deduzindo que, para X variável aleatória contínua

$$\begin{aligned} \mathbb{P}(X > a) &= 1 - F_X(a) = \int_{-\infty}^{+\infty} f_X(u)\,\mathrm{d}u - \int_{-\infty}^{a} f_X(u)\,\mathrm{d}u \\ &= \int_{-\infty}^{a} f_X(u)\,\mathrm{d}u + \int_{a}^{+\infty} f_X(u)\,\mathrm{d}u - \int_{-\infty}^{a} f_X(u)\,\mathrm{d}u \\ &= \int_{a}^{+\infty} f_X(u)\,\mathrm{d}u \end{aligned}$$

agora, se $X \geqslant 0$ então

$$\begin{aligned} \int_{0}^{+\infty} \mathbb{P}(X > x)\,\mathrm{d}x &= \int_{0}^{+\infty}\int_{x}^{+\infty} f_X(u)\,\mathrm{d}u\,\mathrm{d}x \\ &= \int_{0}^{+\infty}\int_{0}^{u} f_X(u)\,\mathrm{d}x\,\mathrm{d}u \\ &= \int_{0}^{+\infty} u f_X(u)\,\mathrm{d}u = \mathbb{E}[X]. \end{aligned}$$

Para provar o teorema, primeiro assumiremos que $g(x) \geqslant 0$ para todo x. Então,

$$\mathbb{E}[g(X)] = \int_{0}^{+\infty} \mathbb{P}(g(X) > x)\,\mathrm{d}x = \int_{0}^{+\infty}\int_{B} f(u)\,\mathrm{d}u\,\mathrm{d}x$$

em que $B = \{u \in \mathbb{R} : g(u) > x\}$

$$\mathbb{E}\left[g(X)\right] = \int_0^{+\infty} \int_B f(u)\,\mathrm{d}u\,\mathrm{d}x = \int_0^{+\infty} \int_0^{g(u)} f(u)\,\mathrm{d}x\,\mathrm{d}(u) = \int_0^{+\infty} g(u)f(u)\,\mathrm{d}u$$

que prova o enunciado pelo teorema para g não negativa. Pra finalizar, se g assume valores reais então definimos as variáveis aleatórias não negativas

$$g^+(x) = \max\{g(x), 0\} \quad \text{e} \quad g^-(x) = \max\{-g(x), 0\}$$

e temos que $g(x) = g^+(x) - g^-(x)$, portanto,

$$\begin{aligned}
\mathbb{E}\left[g(X)\right] &= \mathbb{E}\left[g^+(X)\right] - \mathbb{E}\left[g^-(X)\right] \\
&= \int_0^{+\infty} g^+(u)f(u)\,\mathrm{d}u - \int_0^{+\infty} g^-(u)f(u)\,\mathrm{d}u \\
&= \int_0^{+\infty} g(u)f(u)\,\mathrm{d}u.
\end{aligned}$$

□

1.6.17 Corolário (linearidade da esperança). *Para quaisquer* $a, b \in \mathbb{R}$, *vale* $\mathbb{E}\left[aX + b\right] = a\mathbb{E}\left[X\right] + b$.

Demonstração. Basta tomar $g(x) = ax + b$ no teorema. □

Por exemplo, seja Z uma variável aleatória com f.d.p

$$f(x) = \begin{cases} \frac{x^2}{3}, & \text{se } -1 < x < 2 \\ 0 & \text{caso contrário.} \end{cases}$$

e $g : \mathbb{R} \to \mathbb{R}$ é dada por $g(x) = 4x + 3$ então

$$\mathbb{E}\left[g(Z)\right] = \mathbb{E}\left[4Z + 3\right] = \int_{-1}^{2} (4x + 3)\frac{x^2}{3}\,\mathrm{d}x = 8.$$

Variância

A **variância** da variável aleatória contínua X é dada pelo valor esperado de $g(X) = (X - \mathbb{E}\left[X\right])^2$, sempre que $\mathbb{E}\left[X\right] < \infty$

$$\boxed{\mathrm{Var}[X] := \mathbb{E}\left[X - \mathbb{E}\left[X\right]\right]^2 = \mathbb{E}\left[X^2\right] - \mathbb{E}\left[X\right]^2}$$

donde temos

$$\mathrm{Var}[X] = \int_{-\infty}^{+\infty} (x - \mathbb{E}\left[X\right])^2 f(x)\,\mathrm{d}x = \int_{-\infty}^{+\infty} x^2 f(x)\,\mathrm{d}x - \left(\int_{-\infty}^{+\infty} x f(x)\,\mathrm{d}x\right)^2$$

O **desvio padrão** é definido como a raiz quadrada positiva da variância, caso seja finita,

$$\sigma_X = \sqrt{\mathrm{Var}[X]}.$$

A variância tem a mesma propriedade do caso discreto.

1.6.18 Proposição. *Se X, Y são variáveis aleatórias contínuas com esperança finita e $a, b \in R$, então*

$$\mathrm{Var}[aX + b] = a^2 \mathrm{Var}[X].$$

Exercício 11. Num jogo de apostas, se o ganho é x e a perda é y em cada rodada, então o ganho médio é

$$x \cdot \mathbb{P}(\text{ocorrencias favoraveis}) + y \cdot \mathbb{P}(\text{ocorrencias desfavoraveis}).$$

Uma variável aleatória U não-negativa em função de distribuição acumulada F e densidade $f = F'$. Um jogo lhe é oferecido da seguinte forma: você pode escolher um número não negativo c, se $U > c$ então você ganha a quantidade c, caso contrário, você não ganha nada. Como exemplo, suponha que U é a altura (medida em cm) da próxima pessoa entrando em uma estação ferroviária pública específica. Se você escolher $c = 100$, então você quase certamente ganha essa quantia. Um valor de $c = 200$ dobraria a quantia se você ganhar, mas reduz drasticamente a sua probabilidade de ganhar. Encontre uma equação para caracterizar o valor de c que maximiza o ganho médio.

Soma de variáveis aleatórias: Sejam X e Y variáveis aleatórias de um modelo probabilístico. A soma delas é a variável aleatória $X + Y$ dada por

$$(X + Y)(\omega) := X(\omega) + Y(\omega).$$

Se $h(x, y) = ax + by$, para reais a, b, então

$$\begin{aligned}
\mathbb{E}[h(X, Y)] &= \sum_x \sum_y (ax + by) f_{X,Y}(x, y) \\
&= \sum_x ax \sum_y f_{X,Y}(x, y) + \sum_y by \sum_y f_{X,Y}(x, y) \\
&= \sum_x ax f_X(x) + \sum_y by f_Y(y) \\
&= a\mathbb{E}[X] + b\mathbb{E}[Y].
\end{aligned}$$

assim como

$$\begin{aligned}
\mathbb{E}[h(X, Y)] &= \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} (ax + by) f_{X,Y}(x, y)\,\mathrm{d}x\,\mathrm{d}y \\
&= \int_{-\infty}^{+\infty} ax \left(\int_{-\infty}^{+\infty} f_{X,Y}(x, y)\,\mathrm{d} \right) \mathrm{d}x + \int_{-\infty}^{+\infty} by \left(\int_{-\infty}^{+\infty} f_{X,Y}(x, y)\,\mathrm{d}x \right) \mathrm{d}y \\
&= \int_{-\infty}^{+\infty} ax f_X(x)\,\mathrm{d}x + \int_{-\infty}^{+\infty} by f_Y(y)\,\mathrm{d}y \\
&= a\mathbb{E}[X] + b\mathbb{E}[Y].
\end{aligned}$$

1.6.19 Teorema (linearidade da esperança). *Se X e Y são variáveis aleatórias de um modelo probabilístico e a e b são reais então*

$$\boxed{\mathbb{E}[aX + bY] = a\mathbb{E}[X] + b\mathbb{E}[Y]}\,.$$

Retomando o exemplo 33, se dois dados honestos são lançados e X é o resultado do primeiro dado e Y o resultado do segundo dado, então o valor esperado para a soma é $E(X+Y) = \mathbb{E}[X] + \mathbb{E}[Y] = \frac{7}{2} + \frac{7}{2} = 7$.

Exercício 12. Se $X_1, \dots, X_n$ são variáveis aleatórias discretas com esperança finita do mesmo modelo probabilístico então $\mathbb{E}[X_1 + \cdots + X_n] = \mathbb{E}[X_1] + \cdots + \mathbb{E}[X_n]$.

Por exemplo, se n dados honestos são lançados, então o valor esperado para a soma dos resultados é $\mathbb{E}[X_1 + \cdots + X_n] = \mathbb{E}[X_1] + \cdots + \mathbb{E}[X_n] = n\frac{7}{2}$, em que X_i é o valor do resultado de um lançamento.

Exemplo 42 (**Esperança de uma variável aleatória** Binomial(n,p)). Sejam $Y_i, Y_2, \dots, Y_n$ variáveis aleatórias com distribuição de Bernoulli com parâmetro p. Então $X = Y_i + Y_2 + \cdots + Y_n$ conta o número de sucesso em n ensaios de Bernoulli. Para todo i temos $\mathbb{E}[Y_i] = p$, portanto $\mathbb{E}[X] = np$.

A proposição 1.5.2, página 18, sugere que os maiores valores de $\mathrm{bi}_{n,p}$ (os valores mais prováveis de uma variável aleatória binomial) estão em torno do valor médio (reveja os gráficos do exemplo 13, página 15).

Sejam X e Y variáveis aleatórias, com esperança finita e definidas sobre o mesmo espaço amostral. O produto delas é $X \cdot Y(\omega) := X(\omega) \cdot Y(\omega)$. Em geral *não vale* $\mathbb{E}[X \cdot Y] = \mathbb{E}[X]\mathbb{E}[Y]$.

Exemplo 43 ($\mathbb{E}[X \cdot X] \neq \mathbb{E}[X]\mathbb{E}[X]$). Seja X o resultado do lançamento de um dado, como no exemplo 30.

$$\mathbb{E}\left[X^2\right] = \frac{1+4+9+16+25+36}{6} = \frac{91}{6}.$$

Notemos que, com o resultado do exemplo 30 podemos concluir que $\mathbb{E}[X \cdot X] \neq \mathbb{E}[X]\mathbb{E}[X]$.

1.6.20 Proposição. *Se X, Y são variáveis aleatórias com esperança finita e $a, b \in R$, então* $\mathrm{Var}[X+Y] = \mathrm{Var}[X] + \mathrm{Var}[Y] + 2(\mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y])$.

Demonstração. Sejam X, Y, a, b como no enunciado e vamos usar a notação $\mu_X = \mathbb{E}[X]$ e $\mu_Y = \mathbb{E}[Y]$. Abaixo usamos várias vezes a linearidade da esperança.

Usando a definição $\mathrm{Var}(X+Y) = \mathbb{E}\left[(X+Y-\mathbb{E}[X+Y])^2\right]$

$$\begin{aligned}
\mathrm{Var}(X+Y) &= \mathbb{E}\left[(X+Y-\mathbb{E}[X+Y])^2\right] \\
&= \mathbb{E}\left[(X-\mathbb{E}[X]) + (Y - E(Y))\right]^2 \\
&= \mathbb{E}\left[(X-\mathbb{E}[X])^2 + (Y-\mathbb{E}[Y])^2 + 2((X-\mathbb{E}[X])(Y-\mathbb{E}[Y])\right] \\
&= \mathbb{E}\left[(X-\mathbb{E}[X])^2\right] + \mathbb{E}\left[(Y-\mathbb{E}[Y])^2\right] + \mathbb{E}\left[2((X-\mathbb{E}[X])(Y-\mathbb{E}[Y])\right] \\
&= \mathrm{Var}[X] + \mathrm{Var}(Y) + \mathbb{E}\left[2((X-\mathbb{E}[X])(Y-\mathbb{E}[Y])\right] \\
&= \mathrm{Var}[X] + \mathrm{Var}(Y) + 2(\mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y])
\end{aligned}$$

pois $\mathbb{E}[2((X - \mathbb{E}[X])(Y - \mathbb{E}[Y])] = 2\mathbb{E}[XY - X\mathbb{E}[Y] - Y\mathbb{E}[X] + \mathbb{E}[X]\mathbb{E}[Y]]$ e como $\mathbb{E}[X\mathbb{E}[Y]] = \mathbb{E}[Y\mathbb{E}[X]] = \mathbb{E}[X]\mathbb{E}[Y]$ os termos do meio se cancelam. □

O termo $\mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y]$ no item 2 do resultado acima é chamado de **covariância** de X e Y e denotado $\mathrm{cov}(X, Y)$.

Exercício 13. Verifique que se $\mathbb{E}[X \cdot Y] = \mathbb{E}[X]\mathbb{E}[Y]$ então $\mathrm{Var}(X + Y) = \mathrm{Var}[X] + \mathrm{Var}(Y)$.

Soma de variáveis aleatórias uniforme

$X, Y \sim$ Uniforme$([0, 1])$

$$f_{X+Y}(z) = \int_0^1 1 \cdot f_Y(z - x)\,\mathrm{d}x$$

$0 \leqslant z \leqslant 1$, $f_{X+Y}(z) = \int_0^z dx = z$
$1 < z \leqslant 2$, $f_{X+Y}(z) = \int_{z-1}^1 dx = 2 - z$

Soma de variáveis aleatórias Poisson

$X \sim$ Poisson(λ_X)
$Y \sim$ Poisson(λ_Y)

$$f_{X+Y}(z) = \sum_x f_X(x) f_Y(z - x) = \frac{\mathrm{e}^{-(\lambda_X + \lambda_Y)}}{z!}(\lambda_X + \lambda_Y)^z$$

Soma de variáveis aleatórias Binomial

$X \sim$ Binomial(n, p)
$Y \sim$ Binomial(m, p)
$X + Y \sim$ Binomial$(n + m, p)$

Soma de variáveis aleatórias normais

Vimos que se $X \sim \mathcal{N}(\mu; \sigma^2)$ então $aX + b \sim \mathcal{N}(a\mu + b; a^2\sigma^2)$, para quaisquer $a, b \in \mathbb{R}$. De fato, vale uma afirmação mais geral:

1.6.21 Teorema. *Sejam $X_1, X_2, \ldots, X_n$ variáveis aleatórias independentes com $X_i \sim \mathcal{N}(\mu_i; \sigma_i^2)$, então*

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

para quaisquer constantes $c_1, c_2, \ldots, c_n \in \mathbb{R}$.

1.6.22 Corolário. *Se* $M_n = \frac{\sum_i X_i}{n}$ *então*

$$M_n \sim \mathcal{N}\left(\frac{\sum_i \mu_i}{n}; \frac{\sum_i {\sigma_i}^2}{n^2}\right).$$

Caso as variáveis tenham a mesma distribuição

$$M_n \sim \mathcal{N}\left(\mu; \frac{\sigma^2}{n}\right).$$

Exemplo 44. Na fabricação de placas retangulares há pequenas perturbações de modo que o comprimento e a largura em centímetros de uma placa escolhida ao acaso são variáveis aleatórias $C \sim \mathcal{N}(2; 0{,}1^2)$ e $L \sim \mathcal{N}(5; 0{,}2^2)$, respectivamente. Qual a probabilidade do perímetro exceder 15 cm?

Se Y é a variável aleatória pra o perímetro de uma placa escolhida ao acaso, então $Y = 2C + 2L$ e pelo teorema acima $Y \sim \mathcal{N}(14; 0{,}2)$ logo

$$\mathbb{P}(Y > 15) = \mathbb{P}\left(\frac{Y - 14}{\sqrt{0,2}} > \frac{15 - 14}{\sqrt{0{,}2}}\right) = \mathbb{P}(Z > 2{,}236) = 0{,}0129$$

Exemplo 45. O engarrafamento de um refrigerante de 300ml tem variações de modo que o volume do líquido numa garrafa é uma variável aleatória com distribuição $\mathcal{N}(300; 25^2)$. Numa inspeção, 10 garrafas são selecionadas e o volume de cada garrafa, $V_1, V_2, \dots, V_{10}$ é medido, de modo que se a **média amostral**

$$M_n = \frac{V_1 + V_2 + \cdots + V_{10}}{10}$$

for menor que 290 (ml) então a engarrafadora é multada. Qual é a probabilidade de multa?

$M_n \sim \mathcal{N}(300; \frac{25^2}{10})$ de modo que

$$\mathbb{P}(M_n < 290) = \mathbb{P}(Z < -1{,}26) = 0,1038$$

Definimos $XY(\omega) := X(\omega)Y(\omega)$. Se $\mathbf{1}_{[X=a]\cap[Y=b]}(\omega)$ é a variável aleatória indicadora de $\omega \in [X = a] \cap [Y = b]$, então, variando a, b nos reais tais que $f_X(a), f_Y(b) > 0$, temos

$$XY(\omega) = \sum_a \sum_b a \cdot b \cdot \mathbf{1}_{[X=a]\cap[Y=b]}(\omega) \qquad (\forall \omega \in \Omega)$$

portanto

$$\begin{aligned}\mathbb{E}[XY] &= \sum_a \sum_b ab\mathbb{E}\left[\mathbf{1}_{[X=a]\cap[Y=b]}\right] \\ &= \sum_a \sum_b ab\mathbb{P}([X = a] \cap [Y = b])\end{aligned}$$

agora, se assumimos X e Y independentes, então

$$
\begin{aligned}
\mathbb{E}[XY] &= \sum_a \sum_b ab\mathbb{P}([X=a]\cap[Y=b]) \\
&= \sum_a \sum_b ab\mathbb{P}(X=a)\mathbb{P}(Y=b) \\
&= \sum_a aP(X=a)\sum_b b\mathbb{P}(Y=b) \\
&= \sum_a af_X(a)\sum_b bf_X(b) \\
&= \mathbb{E}[X]\,\mathbb{E}[Y].
\end{aligned}
$$

ou seja, provamos que

1.6.23 Proposição. *Se X e Y são variáveis aleatórias independentes então $\mathbb{E}[XY] = \mathbb{E}[X]\,\mathbb{E}[Y]$*

Como corolário segue da proposição 1.6.12 que

1.6.24 Corolário. *Se X e Y são variáveis aleatórias independentes então $\mathrm{Var}(X+Y) = \mathrm{Var}[X] + \mathrm{Var}(Y)$.*

1.6.25 Observação. *Esses conceitos desenvolvidos até aqui podem ser estendidos, de modo natural, ao caso de 3 ou mais variável aleatória.*

Desigualdades de Markov e Bienaymé–Chebyshev: Seja λ um real positivo e Z uma variável aleatória que assume valores não negativos. Definamos a variável aleatória Y por

$$
Y(\omega) := \begin{cases} \lambda, & \text{se } Z(\omega) \geqslant \lambda \\ 0, & \text{nos outros casos.} \end{cases}
$$

Da definição temos $Y \leqslant Z$ donde deduzimos que $\mathbb{E}[Y] \leqslant \mathbb{E}[Z]$. Também, o valor esperado de Y é

$$
\mathbb{E}[Y] = \lambda\,\mathbb{P}(Y=\lambda) + 0\,\mathbb{P}(Y\neq\lambda) = \lambda\,\mathbb{P}(Z\geqslant\lambda)
$$

logo $\mathbb{E}[Z] \leqslant \lambda\mathbb{P}(Z\geqslant\lambda)$. Disso concluímos

1.6.26 Teorema (Desigualdade de Markov). *Se λ é real positivo e Z uma variável aleatória que assume valores não negativos, então*

$$
\mathbb{P}(Z\geqslant\lambda) \leqslant \frac{\mathbb{E}[Z]}{\lambda}.
$$

Se X é uma variável aleatória, podemos tomar $Z = (X-\mathbb{E}[X])^2$ que é uma variável aleatória não negativa. Lembremos que $\mathbb{E}[Z] = \mathrm{Var}[X]$. Da desigualdade de Markov com a constante λ^2 no lugar de λ deduzimos

$$
\mathbb{P}\left((X-\mathbb{E}[X])^2 \geqslant \lambda^2\right) \leqslant \frac{\mathrm{Var}[X]}{\lambda^2}.
$$

Disso concluímos

1.6.27 Teorema (**Desigualdade de Bienaymé–Chebyshev**). *Se λ é real positivo e X uma variável aleatória então*

$$\mathbb{P}(|X - \mathbb{E}[X]| \geqslant \lambda) \leqslant \frac{\mathrm{Var}[X]}{\lambda^2}. \tag{1.6.4}$$

Se fizermos $\lambda = k\sigma$ na equação (1.6.4) (σ é o desvio padrão de X) obtemos a probabilidade de X desviar de $\mathbb{E}[X]$ por pelo menos k desvios padrão

$$\mathbb{P}(|X - \mathbb{E}[X]| \geqslant k\sigma) \leqslant \frac{1}{k^2}. \tag{1.6.5}$$

Exemplo 46. Numa doença rara, um paciente se recupera com probabilidade 0,4. Numa escolha aleatória de 15 pacientes doentes, seja X o número de sobreviventes. Então, de 3 a 9 pacientes sobrevivem com probabilidade

$$\begin{aligned}\mathbb{P}(3 \leqslant X \leqslant 9) &= \sum_{x=3}^{9} \binom{15}{x} (0,4)^x (0,6)^{15-x} \\ &= 0{,}94\end{aligned}$$

Agora, usando a desigualdade de Bienaymé–Chebyshev, equação equação (1.6.5), para determinar $X \in (\mu - 2\sigma, \mu + 2\sigma)$, em que $\mu = \mathbb{E}[X]$, temos

$$\mathbb{P}(|X - \mu| \geqslant 2\sigma) \leqslant \frac{1}{4}$$

portanto $X \in (\mu - 2\sigma, \mu + 2\sigma)$ com probabilidade $3/4 = 0{,}75$. A variável aleatória X tem média e variância dadas, respectivamente, por $\mu = 6$ e $\sigma^2 = 3{,}6$, portanto

$$\mathbb{P}(2{,}20 \leqslant X \leqslant 9{,}89) = 0{,}75$$

e como X é inteiro $[3 \leqslant X \leqslant 9] = [2{,}20 \leqslant X \leqslant 9{,}89]$.

A desigualdade de Bienaymé–Chebyshev vale para qualquer distribuição e, por isso, a estimativa não é muito precisa. Compare os dois valores obtidos no exemplo acima. No caso de $X \sim \mathcal{N}(\mu; \sigma^2)$, por exemplo, sabemos que $\mathbb{P}(\mu - 2\sigma \leqslant X \leqslant \mu + 2\sigma) \approx 0{,}95$ enquanto que equação (1.6.5) garante apenas $\mathbb{P}(\mu - 2\sigma \leqslant X \leqslant \mu + 2\sigma) = 0{,}75$.

Exemplo 47. Um produto é distribuído em lotes de 40 unidades. Um lote é inaceitável se três ou mais itens apresentam defeito. O departamento de controle de qualidade de um comprador adotou o plano de selecionar aleatoriamente 5 unidades de cada lote comprado e rejeitar o lote se um item inspecionado for defeituoso. Num lote com 3 itens defeituoso, a probabilidade de haver um defeituoso numa amostra de 5 é

$$\frac{\binom{3}{1}\binom{37}{4}}{\binom{40}{5}} = 0{,}3017$$

ou seja, a estratégia detecta um lote ruim em apenas 30% dos casos. Como já vimos, se X é o número de itens defeituosos na amostra, então $X \in (\mu - 2\sigma, \mu + 2\sigma)$ com probabilidade pelo menos 3/4, pela desigualdade de Bienaymé–Chebyshev. Nesse caso $\mu = 0{,}375$ e $\sigma = 0{,}558$ e temos então que 1 item em 5 é defeituoso com probabilidade $\geqslant 3/4$, ou seja, em 75% dos casos.

Exemplo 48. Consideremos n lançamentos de uma moeda com os resultados independentes. Seja X_i a variável indicadora (definida na pág. 6) do evento "ocorre cara no i-ésimo lançamento". A variável aleatória

$$S_n = X_1 + \cdots + X_n$$

é a quantidade de ocorrência de cara e $\frac{S_n}{n}$ é a fração de caras. A esperança de S_n é, usando o exercício 12

$$\mathbb{E}\left[S_n\right] = \mathbb{E}\left[\sum_{i=1}^{n} X_i\right] = \left(\sum_{i=1}^{n} \mathbb{E}\left[X_i\right]\right) = \frac{n}{2}.$$

e a variância é $\mathrm{Var}(S_n) = \mathbb{E}\left[S_n^2\right] - \mathbb{E}\left[S_n\right]^2$, assim precisamos calcular $\mathbb{E}\left[S_n^2\right]$.

$$\mathbb{E}\left[S_n^2\right] = \mathbb{E}\left[\sum_{i=1}^{n}\sum_{j=1}^{n} X_i X_j\right] = \sum_{i=1}^{n}\sum_{j=1}^{n} \mathbb{E}\left[X_i X_j\right],$$

se $i \neq j$ então

$$\mathbb{E}\left[X_i \cdot X_j\right] = \mathbb{P}(X_i \cdot X_j = 1) = \mathbb{P}([X_i = 1] \cap [X_j = 1]) = \frac{1}{2} \cdot \frac{1}{2}$$

e se $i = j$ então

$$\mathbb{E}\left[X_i \cdot X_j\right] = E(X_i^2) = P(X_i = 1) = \frac{1}{2}$$

portanto

$$\mathbb{E}\left[S_n^2\right] = \sum_{i=1}^{n}\sum_{j=1}^{n} \mathbb{E}\left[X_i X_j\right] = \sum_{i=1}^{n} \mathbb{E}\left[X_i^2\right] + \sum_{i=1}^{n}\sum_{\substack{j=1 \\ j\neq i}}^{n} \mathbb{E}\left[X_i X_j\right] = \frac{n}{2} + \frac{n(n-1)}{4}.$$

De volta à variância de S_n temos

$$\mathrm{Var}\left(S_n\right) = \mathbb{E}\left[S_n^2\right] - \mathbb{E}\left[S_n\right]^2 = \frac{n}{2} + \frac{n(n-1)}{4} - \left(\frac{n}{2}\right)^2 = \frac{n}{4}.$$

A esperança de S_n/n é, usando linearidade da esperança (corolário 1.6.9)

$$\mathbb{E}\left[\frac{S_n}{n}\right] = \frac{1}{n}\mathbb{E}\left[S_n\right] = \frac{1}{2}.$$

e a variância é (proposição 1.6.12),

$$\mathrm{Var}\left[\frac{S_n}{n}\right] = \frac{1}{n^2}\mathrm{Var}[S_n] = \frac{1}{n^2}\frac{n}{4} = \frac{1}{4n}.$$

Usando a desigualdade de Bienaymé–Chebyshev, eq. equação (1.6.4), para qualquer $\lambda > 0$

$$\mathbb{P}\left(\left|\frac{S_n}{n} - \frac{1}{2}\right| \geqslant \lambda\right) \leqslant \frac{1}{4n\lambda^2}$$

portanto

$$\mathbb{P}\left(\frac{1}{2}-\lambda \leqslant \frac{S_n}{n} \leqslant \frac{1}{2}+\lambda\right) > 1-\frac{1}{4n\lambda^2}.$$

Por exemplo, fazendo $\lambda = n^{-1/2}$ temos que o número médio de caras está no intervalo $(\frac{1}{2}-\frac{1}{\sqrt{n}}, \frac{1}{2}+\frac{1}{\sqrt{n}})$ com probabilidade maior que 3/4.

Mais que isso, a probabilidade tende a 1 quando $n \to \infty$. Esse resultado foi provado pela primeira vez em 1713, por Jacob Bernoulli sem usar a desigualdade de Bienaymé–Chebyshev, desconhecida na época.

Exemplo 49. Numa eleição, seja p a fração (desconhecida) da população que vota no candidato D. Para simplificar, assumimos que um voto em voto em D é ensaio de Bernoulli com parâmetro p. Suponha que são realizadas n entrevistas como o objetivo de estimar o valor de p. Se $V_i \sim$ Bernoulli(p) é a variável aleatória indicadora do i-ésimo voto ser para D, para $1 \leqslant i \leqslant n$, então

$$S_n = \sum_{i=1}^{n} V_i \sim \text{Binomial}(n,p)$$

é o total de entrevistados a favor de D.

Queremos determinar n para obtermos uma estimativa para p com erro de 4 pontos percentuais com pelo menos 95% de certeza, i.e.,

$$\mathbb{P}\left(\left|\frac{S_n}{n}-p\right| \geqslant 0{,}04\right) < 0{,}05. \tag{1.6.6}$$

Usando a desigualdade de Chebyshev equação (1.6.4),

$$\mathbb{P}\left(|S_n - np| \geqslant 0{,}04n\right) \leqslant \frac{np(1-p)}{0{,}0016n^2} \leqslant \frac{1}{4\cdot 0{,}0016\cdot n}$$

e fazendo o lado direito igual a 0,05 e resolvendo para n obtemos $n = 3125$ entrevistas para que S_n/n estime p dentro dos parâmetros de exigência.

Esse valor de n está superestimado, muito por causa da generalidade da desigualdade de Bienaymé–Chebyshev, ela não usa nenhuma particularidade da distribuição binomial que é bem concentrada em torno da média. A equação equação (1.6.6) equivale a

$$\mathbb{P}\left(S_n \leqslant (p-0{,}04)n\right) + \mathbb{P}\left(S_n \geqslant (p+0{,}04)n\right) < 0{,}05$$

e a soma a esquerda da desigualdade é $F_{S_n}((p-0{,}04)n) + F_{S_n}(((1-p)-0{,}04)n)$ que é máxima para $p = 0{,}5$ para todo n fixo, portanto

$$\begin{aligned}\mathbb{P}\left(\frac{S_n}{n} \leqslant p-0{,}04\right)+\mathbb{P}\left(\frac{S_n}{n} \geqslant p+0{,}04\right) &\leqslant F_{S_n}((0{,}5-0{,}04)n)+F_{S_n}((0{,}5-0{,}04)n)\\ &= 2F_{S_n}(0{,}46n)\end{aligned}$$

que é menor que $0,05$ para $n \geqslant 624$. Essa estimativa é bem próxima a de dados reais. Informações obtidas no sítio do IBOPE referentes ao 1º turno das eleições municipais de 2008 exibem a seguinte quantidade de entrevistados de acordo como o erro e o grau de confiança:

cidade	no. de eleitores	no. de entrevistas	proporção eleitores entrevistados	margem de erro	grau de confiança
São Paulo	8.198.282	1.204	0,01468%	3%	95%
Rio de Janeiro	4.579.282	1.204	0,02629%	3%	95%
Belo Horizonte	1.772.227	1.204	0,06793%	3%	95%
Santo André	533.428	504	0,09448%	4%	95%
Diadema	301.229	504	0,1673%	4%	95%
São Carlos	154.572	504	0,326%	4%	95%
Cubatão	91.693	504	0,5496%	4%	95%
Registro	41.001	504	1,2292%	4%	95%
Campinas	724.143	2.000	0,2761%	3%	99%
São José dos Campos	414.353	2.000	0,4826%	3%	99%
Ribeirão Preto	388.690	2.000	0,5145%	3%	99%

Adiante, quando estudarmos o Teorema Central do Limite, retomaremos esse exemplo e mostraremos uma regra bastante prática para estimar o tamanho n da amostra de eleitores.

Leis dos Grandes Números:

Convergência em probabilidade:: Quando X e uma sequência X_n, $n \geqslant 1$ são variáveis aleatórias de um mesmo modelo probabilístico, dizemos que X_n converge para X em probabilidade, denotado por

$$X_n \xrightarrow[n\to\infty]{\mathrm{P}} X$$

se para todo $\epsilon > 0$

$$\lim_{n\to\infty} \mathbb{P}(|X_n - X| < \epsilon) = 1.$$

1.6.28 Teorema (**Lei Fraca dos Grandes Números**). *Sejam $X_1, X_2, \ldots, X_n$ variável aleatória's independentes e identicamente distribuídas, cada uma com média $\mathbb{E}[X_i] = \mu$ finita e variância* $\mathrm{Var}[X] = \sigma^2$ *finita. Então*

$$\frac{X_1 + X_2 + \cdots + X_n}{n} \xrightarrow[n\to\infty]{\mathrm{P}} \mu$$

Demonstração. Sejam $X_1, X_2, \ldots$ uma sequência de variáveis aleatórias independentes, identicamente distribuídas e todas com valor esperado μ e variância σ^2 finitos. Pondo

$$M_n := \frac{X_1 + X_2 + \cdots + X_n}{n}.$$

temos

$$\mathbb{E}[M_n] = \mu \ \text{ e } \ \mathrm{Var}[M_n] = \frac{\sigma^2}{n}.$$

Usando a desigualdade de Chebyshev equação (1.6.4), concluímos

$$\lim_{n\to\infty} \mathbb{P}(|M_n - \mu| \geqslant \lambda) \leqslant \lim_{n\to\infty} \frac{\sigma^2}{n\lambda^2} = 0.$$

para todo $\lambda > 0$. □

Exemplo 50. Sejam $g : [0,1] \to [0,1]$ uma função contínua e $X_i \sim \text{Uniforme}([0,1])$ variáveis aleatórias independentes, $i \geqslant 1$. As variáveis aleatórias $g(X_i)$, para todo $i \geqslant 1$, são independentes de mesma distribuição, portanto, pela lei dos grandes números

$$\frac{g(X_1) + g(X_2) + \cdots + g(X_n)}{n} \xrightarrow[n\to\infty]{\mathrm{P}} \mu$$

em que $\mu = \mathbb{E}\,[g(X_i)]$ para todo i. Se $X \sim \text{Uniforme}([0,1])$ então

$$\mu = \mathbb{E}\,[g(X)] = \int_0^1 g(x) f_X(x)\,\mathrm{d}x = \int_0^1 g(x)\mathrm{d}x$$

e a variância é

$$\sigma^2 = \mathbb{E}\,[X-\mu]^2 = \int_0^1 (g(x)-\mu)^2\,\mathrm{d}x \leqslant 1$$

pois $|g(x)-\mu| \leqslant 1$, portanto,

$$\int_0^1 g(x)\,\mathrm{d}x \approx \frac{g(X_1) + g(X_2) + \cdots + g(X_n)}{n} = M_n$$

e a Bienaymé–Chebyshev nos diz quão boa é essa estimativa

$$\mathbb{P}\left(|M_n - \mu| \geqslant \varepsilon\right) \leqslant \frac{\sigma^2}{n\varepsilon^2} < \frac{1}{n\varepsilon^2}$$

para qualquer erro $\varepsilon > 0$. Por exemplo, para $\varepsilon = 0{,}001$, se $n = 100.000.000$ então

$$|M_n - \mu| = \left| M_n - \int_0^1 g(x)\mathrm{d}x \right| < 0{,}001$$

com probabilidade $> 0{,}99$.

Exercício 14. Uma aposta de 1 real tem ganho esperado de −0,0141. O que a Lei dos Grandes Números nos diz sobre os seus ganhos se você fizer um grande número apostas de 1 real? Será que ela lhe assegura que suas perdas serão pequenas? Será que ela lhe assegura que, se n for muito grande você vai perder?

Exercício 15. Pedro e Paula ambos querem cortar um pedaço de papel retangular. Como ambos são probabilistas eles determinam a forma exata do retângulo utilizando realizações de uma variável aleatória positiva, digamos U, como se segue. Pedro é preguiçoso e gera apenas uma única realização dessa variável aleatória; então ele corta um quadrado que tem comprimento e largura igual a esse valor. Paula gosta de diversidade e gera duas realizações independentes de U. Ela, então, corta um retângulo com largura igual a primeira realização e comprimento igual ao da segunda realização. (a) Serão as áreas cortadas por Pedro e Paula diferentes em média? (b) se forem, Pedro ou Paula deverá ter um retângulo com área maior?

Convergência quase-certa:: Quando X e uma sequência x_n, $n \geqslant 1$ são variáveis aleatórias de um mesmo modelo probabilístico, dizemos que X_n converge para X quase certamente, denotado por

$$X_n \xrightarrow[n\to\infty]{\text{q.c.}} X$$

se existe $A \subset \Omega$ tal que $\mathbb{P}(A) = 1$ e

$$\lim_{n\to\infty} X_n(\omega) = X(\omega), \text{ para todo } \omega \in A$$

ou,

$$\mathbb{P}\left(\lim_{n\to\infty} X_n = X\right) = 1$$

É possível provar que

$$\text{convergência quase certa} \Rightarrow \text{convergência em probabilidade}$$

e, também, exibir contra-exemplo para a recíproca.

Lei Forte dos Grandes Números:: estabelece que, sob as mesmas hipóteses da lei fraca,

$$\frac{X_1 + \cdots + X_n}{n} \xrightarrow[n\to\infty]{\text{q.c.}} \mu$$

A Lei dos Grandes Números mostra que o modelo probabilístico, e os axiomas de probabilidade em particular, é consistente com a interpretação frequentista de probabilidade.

1.7 Teorema Central do Limite. Consideremos uma sequência $X_1, X_2, \ldots$ de variáveis aleatórias *independentes* com a *mesma distribuição*, de esperança μ finita e variância $\sigma^2 > 0$ finita. A soma das n primeiras variáveis aleatórias

$$S_n := X_1 + \cdots + X_n$$

tem esperança $\mathbb{E}[S_n] = n\mu$ e variância $\mathrm{Var}[S_n] = n\sigma^2$ e a média

$$M_n := \frac{X_1 + \cdots + X_n}{n}$$

tem esperança $\mathbb{E}[M_n] = \mu$ e variância $\mathrm{Var}[M_n] = \sigma^2/n$.

Vimos, pela lei de grandes números, que $M_n \approx \mu$. O Teorema Central do Limite diz que, *para n grande*, a variável aleatória normalizada de S_n e M_n

$$Z_n := \frac{S_n - n\mu}{\sigma\sqrt{n}} = \frac{M_n - \mu}{\sigma/\sqrt{n}}$$

têm distribuição aproximadamente $\mathcal{N}(0;1)$. Portanto, $M_n = \mu + (\sigma/\sqrt{n})Z_n$, ou seja, o que veremos agora é que, de fato, $M_n \approx \mu + (\sigma/\sqrt{n})Z$ com $Z \sim \mathcal{N}(0;1)$.

Por ora, ignoremos o real significado de *para n grande* e vejamos alguns exemplos.

Exemplo 51. Lâmpadas produzidas numa fábrica têm vida útil em horas regida pela distribuição normal $\mathcal{N}(800; 40^2)$. Uma seleção aleatória simples de tamanho 16 tem vida útil média M_{16} com distribuição aproximada $\mathcal{N}(800; 40^2/16)$. A probabilidade da vida útil média ser menor que 775 horas é aproximadamente

$$\mathbb{P}(M_{16} < 775) = \mathbb{P}\left(\frac{M_{16} - 800}{40/4} < \frac{775 - 800}{40/4}\right) = \mathbb{P}(Z_{16} < -2,5) \approx 0{,}0062.$$

De fato, a aproximação aqui é uma igualdade pois a combinação linear de variáveis com distribuição normal tem distribuição normal.

Exemplo 52. As chamadas telefônicas numa empresa têm duração em minutos que segue a distribuição exponencial com parâmetro 1/3, portanto tem média 3 e variância 9. Uma amostra aleatória simples com 50 chamadas tem probabilidade da média amostral não ultrapassar 4 min dada por

$$\mathbb{P}(M_{50} \leqslant 4) = \mathbb{P}(Z_{50} \leqslant 2,36) \approx 0{,}991.$$

Convergência em distribuição:: Quando X e uma sequência X_n, $n \geqslant 1$, são variáveis aleatórias do mesmo espaço de probabilidade, dizemos que X_n converge para X em distribuição, denotado por

$$X_n \xrightarrow[n\to\infty]{\mathcal{D}} X$$

se $F_{X_n} \to F_X$ quando $n \to \infty$, isto é

$$\lim_{n\to\infty} \mathbb{P}(X_n \leqslant x) = \mathbb{P}(X \leqslant x)$$

de modo que podemos usar X como modelo probabilístico aproximado para X_n e quanto maior n melhor é a aproximação.

Formalmente, o teorema central do limite é enunciado como

1.7.1 Teorema (**Teorema Central do Limite (TCL)**). *Sejam $X_1, X_2, \dots, X_n$ variáveis aleatórias* mutuamente independentes *com a* mesma distribuição, *de esperança μ e variância $\sigma^2 > 0$ finitas. Então*

$$\frac{M_n - \mu}{\sigma/\sqrt{n}} \xrightarrow[n\to\infty]{\mathcal{D}} Z$$

para $Z \sim \mathcal{N}(0; 1)$.

Por exemplo, se uma moeda equilibrada é lançada n vezes e S_n é a quantidade de caras. Para $n = 100$ temos $\mathbb{E}[S_{100}] = 50$ e $\mathrm{Var}[S_{100}] = 25$. A probabilidade de termos mais que 55 caras é, aproximadamente,

$$\mathbb{P}(S_{100} > 55) = \mathbb{P}\left(\frac{S_{100} - 50}{5} > \frac{55 - 50}{5}\right) = \mathbb{P}(Z_{100}) > 1) \approx 0{,}16.$$

Agora, para $n = 400$, qual a probabilidade com que $S_{400} > 220$ (note-se que $220/400 = 55/10$)?

$$\mathbb{P}(S_{400} > 220) = \mathbb{P}\left(\frac{S_{400} - 200}{10} > \frac{220 - 200}{10}\right) = \mathbb{P}(Z_{400}) > 2) \approx 0{,}025.$$

Com que probabilidade $40 \leqslant S_{100} \leqslant 60$?

$$\mathbb{P}(40 \leqslant S_{100} \leqslant 60) = \mathbb{P}\left(\frac{40 - 50}{5} \leqslant \frac{S_{100} - 50}{5} \leqslant \frac{60 - 50}{5}\right) = \mathbb{P}(-2 \leqslant Z_{100}) \leqslant 2) \approx 0{,}954.$$

1.8 Aproximação para a Binomial. No caso que $X_1, \dots, X_n$ são variáveis aleatórias independentes com distribuição Bernoulli(p) temos $S_n = X_1 + \cdots + X_n \sim \text{Binomial}(n, p)$. Pondo $q = 1 - p$ e $Z_n = (M_n - \mu)/(\sigma/\sqrt{n})$ o teorema central do limite nos dá

$$\mathbb{P}(k \leqslant S_n \leqslant l) \approx \mathbb{P}\left(\frac{k - np}{\sqrt{npq}} \leqslant Z_n \leqslant \frac{l - np}{\sqrt{npq}}\right) \tag{1.8.1}$$

Por exemplo, se $X \sim \text{Binomial}(225; 0{,}2)$ então pela aproximação dada na equação (1.8.1)

$$\mathbb{P}(39 \leqslant X \leqslant 48) \approx \mathbb{P}\left(-1 \leqslant Z \leqslant 0{,}5\right) = 0{,}5328072 \tag{1.8.2}$$

ainda, o valor exato é

$$\mathbb{P}(39 \leqslant X \leqslant 48) = \sum_{j=39}^{48} \binom{225}{j} (0,2)^j (0{,}8)^{225-j} = 0{,}5852713.$$

Entretanto $0{,}0417 = \mathbb{P}(X = 39) \approx \mathbb{P}(-1 \leqslant Z \leqslant -1) = 0$ e aproximação seria melhor se fizéssemos

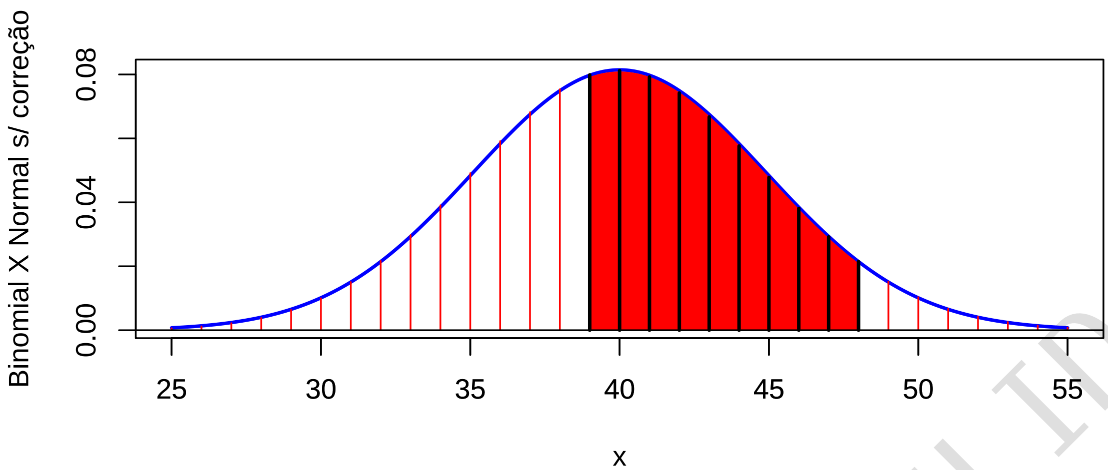
$$\mathbb{P}(X = 39) = \mathbb{P}(38,5 \leqslant X \leqslant 39,5) \approx \mathbb{P}\left(-1,083 \leqslant Z \leqslant -0,916\right) = 0{,}0403$$

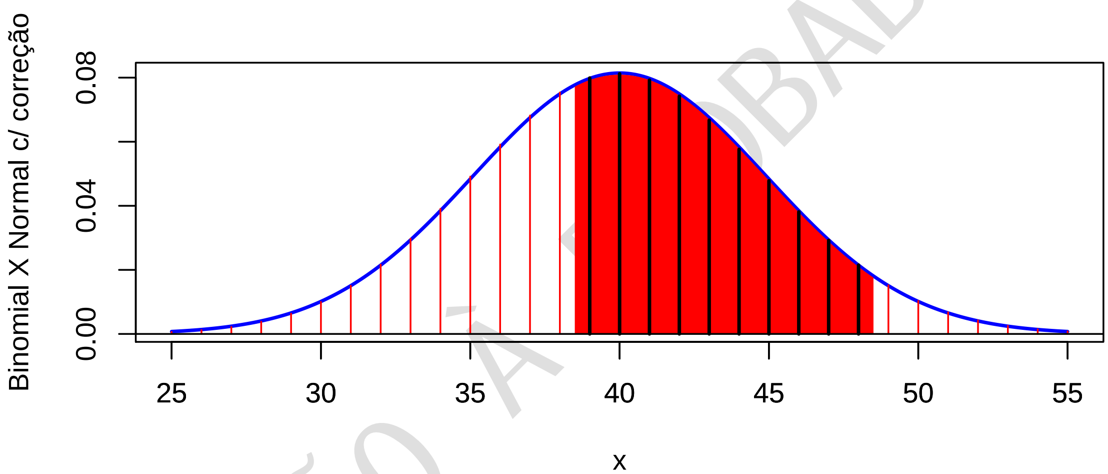
que chamamos de *correção de continuidade*, o que melhora a aproximação

$$\boxed{\mathbb{P}(k \leqslant X \leqslant l) = \mathbb{P}\left(k - \frac{1}{2} \leqslant X \leqslant l + \frac{1}{2}\right) \approx \mathbb{P}\left(\frac{k - \frac{1}{2} - np}{\sqrt{np(1-p)}} \leqslant Z \leqslant \frac{l + \frac{1}{2} - np}{\sqrt{np(1-p)}}\right)}$$

agora, com a mesma técnica, temos uma aproximação melhor equação (1.8.2)

$$\mathbb{P}(39 \leqslant X \leqslant 48) \approx \mathbb{P}\left(-1,083 \leqslant Z \leqslant 0{,}583\right) = 0{,}5806491.$$





Exemplo 53. Um teste tem 200 perguntas com 4 alternativas cada, das quais apenas uma é correta. Qual a probabilidade aproximada que o estudante acerte por chute entre 25 e 30 questões para 80 das 200 questões. Seja $X \sim \text{Binomial}(80, 1/4)$ o número de respostas certas. Usando a correção por continuidade

$$\mathbb{P}(25 \leqslant X \leqslant 30) \approx \mathbb{P}(1{,}16 \leqslant Z \leqslant 2{,}71) = 0{,}1196602.$$

(o valor correto é 0,1192705)

```
Aproximacao de P(X<x) da Normal para Binomial(n,p)=( 20 , 0.5 )

    Bin(x,n,p)   Aprox.       Erro        c/ correcao  Erro        Razao erros
0   0.00000095   0.00000387 -0.00000292  0.00001076 -0.00000981  3.35976699
1   0.00002003   0.00002850 -0.00000847  0.00007196 -0.00005194  6.13204207
```

2	0.00020123	0.00017331	0.00002792	0.00039812	-0.00019689	-7.05303706
3	0.00128841	0.00087256	0.00041585	0.00182522	-0.00053680	-1.29084332
4	0.00590897	0.00364518	0.00226379	0.00695315	-0.00104418	-0.46125469
5	0.02069473	0.01267366	0.00802107	0.02208567	-0.00139094	-0.17341068
6	0.05765915	0.03681914	0.02084001	0.05876243	-0.00110328	-0.05294070
7	0.13158798	0.08985625	0.04173173	0.13177624	-0.00018826	-0.00451111
8	0.25172234	0.18554668	0.06617565	0.25116748	0.00055486	0.00838463
9	0.41190147	0.32736042	0.08454105	0.41153164	0.00036984	0.00437465
10	0.58809853	0.50000000	0.08809853	0.58846836	-0.00036984	-0.00419799
11	0.74827766	0.67263958	0.07563809	0.74883252	-0.00055486	-0.00733570
12	0.86841202	0.81445332	0.05395870	0.86822376	0.00018826	0.00348890
13	0.94234085	0.91014375	0.03219710	0.94123757	0.00110328	0.03426659
14	0.97930527	0.96318086	0.01612440	0.97791433	0.00139094	0.08626303
15	0.99409103	0.98732634	0.00676469	0.99304685	0.00104418	0.15435768
16	0.99871159	0.99635482	0.00235677	0.99817478	0.00053680	0.22777120
17	0.99979877	0.99912744	0.00067133	0.99960188	0.00019689	0.29328140
18	0.99997997	0.99982669	0.00015328	0.99992804	0.00005194	0.33883687
19	0.99999905	0.99997150	0.00002754	0.99998924	0.00000981	0.35599322
20	1.00000000	0.99999613	0.00000387	0.99999867	0.00000133	0.34301686

Voltemos ao exemplo 49: Numa eleição com dois candidatos, seja p a fração (desconhecida) da população que vota no candidato D. Para simplificar, assumimos que só há 2 respostas possíveis e um voto em D é ensaio de Bernoulli com parâmetro p. São realizadas n entrevistas: modelamos o i-ésimo entrevistado como $V_i \sim \text{Bernoulli}(p)$ que é a variável aleatória indicadora do i-ésimo voto ser para D, para todo $1 \leqslant i \leqslant n$. Temos então

$$\begin{aligned} S_n &= V_1 + \cdots + V_n \sim \text{Binomial}(n, p) \text{ é o número de votos em } D; \\ \hat{p} &= M_n = \frac{S_n}{n} \text{ é a proporção da amostra de votos em } D; \\ p &= \text{ é a proporção desconhecida da população de votos em } D \end{aligned}$$

Exemplo 54. Para uma primeira aproximação grosseira, usamos que $\hat{p} \approx p + (\sigma/\sqrt{n})Z \sim \mathcal{N}\left(p; \frac{\sigma^2}{n}\right)$, que $\sigma^2 = p(1-p) \leqslant 1/4$ e que uma variável aleatória normalmente distribuída tem probabilidade 0,95 de estar entre 2 desvios padrão da esperança, portanto

$$\mathbb{P}\left(p - 2\frac{\sigma}{\sqrt{n}} \leqslant \hat{p} \leqslant p + 2\frac{\sigma}{\sqrt{n}}\right) \geqslant \mathbb{P}\left(p - \frac{1}{\sqrt{n}} \leqslant \hat{p} \leqslant p + \frac{1}{\sqrt{n}}\right) \approx 0{,}95$$

de modo que uma estimativa com margem de erro de 4 pontos percentuais e com 95% de certeza temos $1/\sqrt{n} = 0{,}04$, logo $n = 625$. Para uma estimativa com margem de erro de 3 pontos percentuais e com 95% de certeza temos $1/\sqrt{n} = 0{,}03$ logo $n = 1112$. Para determinar o valor de n para uma estimativa com margem de erro de 3 pontos percentuais e com 99% de certeza precisamos de 3 desvios padrão, i.e.,

$$\mathbb{P}\left(p - 3\frac{\sigma}{\sqrt{n}} \leqslant \hat{p} \leqslant p + 3\frac{\sigma}{\sqrt{n}}\right) \geqslant \mathbb{P}\left(p - \frac{1{,}5}{\sqrt{n}} \leqslant \hat{p} \leqslant p + \frac{1{,}5}{\sqrt{n}}\right) \approx 0{,}99$$

portanto, temos $1{,}5/\sqrt{n} = 0{,}03$ logo $n = 1667$.

Vejamos agora esses cáculos com um pouco apurados. Sejam ε a margem de erro tolerada e γ o grau de confiança da estimativa

$$\begin{aligned}
\gamma = \mathbb{P}(-\varepsilon \leqslant p - \hat{p} \leqslant \varepsilon) &= \mathbb{P}\left(\frac{-\varepsilon}{\sqrt{p(1-p)/n}} \leqslant \frac{p-\hat{p}}{\sqrt{p(1-p)/n}} \leqslant \frac{\varepsilon}{\sqrt{p(1-p)/n}}\right) \\
&\approx \mathbb{P}\left(\frac{-\varepsilon}{\sqrt{p(1-p)/n}} \leqslant Z \leqslant \frac{\varepsilon}{\sqrt{p(1-p)/n}}\right) \\
&= \mathbb{P}\left(\frac{-\varepsilon\sqrt{n}}{\sqrt{p(1-p)}} \leqslant Z \leqslant \frac{\varepsilon\sqrt{n}}{\sqrt{p(1-p)}}\right) \\
&= 2\mathbb{P}\left(Z \leqslant \frac{\varepsilon\sqrt{n}}{\sqrt{p(1-p)}}\right) - 1
\end{aligned}$$

Queremos z_γ (obtido da tabela da distribuição da normal padrão) tal que $2\mathbb{P}\left(Z \leqslant z_\gamma\right) - 1 = \gamma$. Por exemplo, para $\gamma = 0{,}95$, de $\mathbb{P}\left(Z \leqslant z_\gamma\right) = 1 + \gamma/2 = \frac{1{,}95}{2}$ tiramos $z_\gamma = 1{,}96$. Descoberto tal z_γ precisamos escolher n de modo que

$$\frac{\varepsilon\sqrt{n}}{\sqrt{p(1-p)}} = z_\gamma$$

e se usarmos a estimativa mais conservadora $p(1-p) \leqslant 1/4$

$$\frac{\varepsilon\sqrt{n}}{\sqrt{p(1-p)}} \geqslant \frac{\varepsilon\sqrt{n}}{\sqrt{4}}$$

portanto é suficente termos n tal que

$$\frac{\varepsilon\sqrt{n}}{\sqrt{4}} = z_\gamma$$

ou seja

$$n = \frac{z_\gamma^2}{4\varepsilon^2}.$$

Exemplo 55. Para $\varepsilon = 0{,}04$ e $\gamma = 0{,}95$

$$n = \frac{1,96^2}{4 \cdot 0{,}04^2} = 600{,}25$$

Para uma estimativa com erro de 3 pontos percentuais e 95% de grau de confiança

$$n = \frac{1,96^2}{4 \cdot 0{,}03^2} = 1068.$$

Analogamente, n para uma estimativa com erro de 3 pontos percentuais com 99% de grau de confiança então $z_{0,99} = 2,57$ e

$$n = \frac{2,57^2}{4 \cdot 0{,}03^2} = 1835.$$

Informações obtidas no sítio do IBOPE referentes ao 1º turno das eleições municipais de 2008:

cidade	no. de eleitores	no. de entrevistas	proporção eleitores entrevistados	margem de erro	grau de confiança
São Paulo	8.198.282	1.204	0,01468%	3%	95%
Rio de Janeiro	4.579.282	1.204	0,02629%	3%	95%
Belo Horizonte	1.772.227	1.204	0,06793%	3%	95%
Santo André	533.428	504	0,09448%	4%	95%
Diadema	301.229	504	0,1673%	4%	95%
São Carlos	154.572	504	0,326%	4%	95%
Cubatão	91.693	504	0,5496%	4%	95%
Registro	41.001	504	1,2292%	4%	95%
Campinas	724.143	2.000	0,2761%	3%	99%
São José dos Campos	414.353	2.000	0,4826%	3%	99%
Ribeirão Preto	388.690	2.000	0,5145%	3%	99%

1.8.1 Observação. *O valor de n não depende do tamanho da população.*

No exemplo acima provamos que

$$\mathbb{P}\left(\hat{p} - 1,96\sqrt{4/n} \leqslant p \leqslant \hat{p} + 1,96\sqrt{4/n}\right) \geqslant 0{,}95$$

e dizemos que $(\hat{p} - 1,96\sqrt{4/n}\,,\ \hat{p} + 1,96\sqrt{4/n})$ é um **intervalo de confiança** para p com grau de confiança 95%. Notemos que p é um valor médio desconhecido e $\hat{p}$ é uma variável aleatória, portanto o intervalo é aleatório.

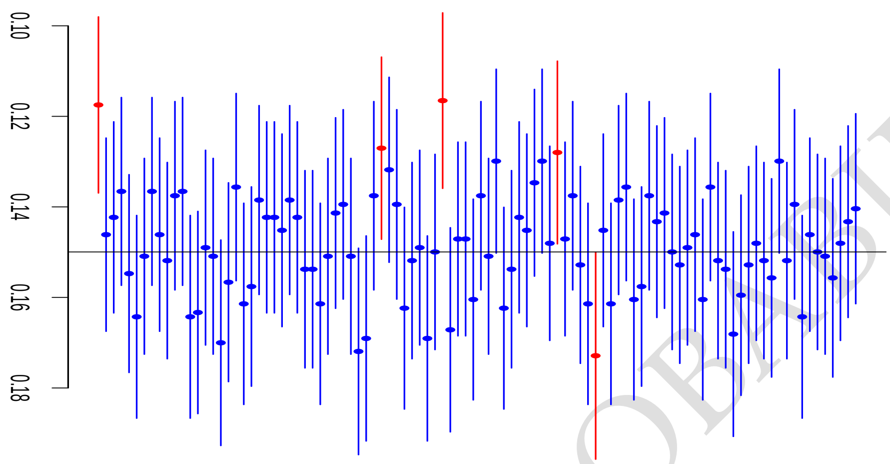
1.9 Intervalos de confiança.

Vamos usar o TCL para resolver o seguinte problema. Queremos estimar a média μ (desconhecida) de alguma característica de uma população pela média M_n de uma amostra aleatória de n indivíduos da população como, por exemplo, fizemos no exemplo 49, com erro controlado: Dados $\epsilon > 0$ e $\gamma \in (0,1)$, para cada amostra aleatória $X_1, \dots, X_n$ queremos uma estimativa intervalar $(M_n - \varepsilon\,,\ M_n + \varepsilon)$ para a média μ da população com grau de confiança γ, isto é,

$$\mathbb{P}\left(\mu \in (M_n - \varepsilon\,,\ M_n + \varepsilon)\right) \geqslant \gamma \tag{1.9.1}$$

que pode ser interpretado assim: num número grande de amostras do mesmo tamanho, se obtivermos um intervalo com grau de confiança, por exemplo 0,95, para cada uma delas, então 95% desses intervalos contém o parâmetro μ.

Por exemplo, o seguinte gráfico apresenta cem intervalos para amostras aleatórias simples de acordo com Bernoulli(0,15) de tamanho $n = 1047$, para cada amostra foi computado um intervalo de confiança de grau 0,95. Em vermelho estão os intervalos que mão contêm a média populacional 0,15:



Tamanho da amostra:: O tamanho n para uma amostra aleatória $X_1, \dots, X_n$ satisfazer equação (1.9.1) pode ser determinada como no caso binomial explicado acima.

$$\begin{aligned}
\gamma &= \mathbb{P}(M_n - \varepsilon \leqslant \mu \leqslant M_n + \varepsilon) \\
&= \mathbb{P}\left(-\frac{\varepsilon}{\sigma/\sqrt{n}} \leqslant \frac{M_n - \mu}{\sigma/\sqrt{n}} \leqslant \frac{\varepsilon}{\sigma/\sqrt{n}}\right) \\
&\approx \mathbb{P}\left(-\frac{\varepsilon}{\sigma/\sqrt{n}} \leqslant Z_n \leqslant \frac{\varepsilon}{\sigma/\sqrt{n}}\right) \\
&= \mathbb{P}\left(-\frac{\varepsilon\sqrt{n}}{\sigma^2}\sigma \leqslant Z_n \leqslant \frac{\varepsilon\sqrt{n}}{\sigma^2}\sigma\right)) \\
&= 2\mathbb{P}\left(Z_n \leqslant \frac{\varepsilon\sqrt{n}}{\sigma^2}\sigma\right) - 1
\end{aligned}$$

Tomemos o valor z_γ tal que

$$\mathbb{P}(Z \leqslant z_\gamma) = \frac{1+\gamma}{2}$$

e de $\epsilon\sqrt{n}/\sigma = z_\gamma$ deduzimos

$$n = \left(\frac{z_\gamma \sigma}{\varepsilon}\right)^2$$

e a estimativa intervalar para μ com grau de confiança γ é

$$\mathrm{IC}(\epsilon, \gamma) := \left(M_n - z_\gamma \frac{\sigma}{\sqrt{n}}\,,\ M_n + z_\gamma \frac{\sigma}{\sqrt{n}}\right)$$

Exemplo 56. A renda per-capita domiciliar numa certa região tem desvio padrão 250 reais e média desconhecida. Se desejamos estimar a renda média da população com erro 50 reais e confiabilidade $\gamma = 0{,}95$ quantos domicílios deveremos consultar? Já sabemos que $z_\gamma = 1{,}96$, então

$$n = \left(\frac{z_\gamma}{\epsilon}\right)^2 \sigma^2 = \left(\frac{1{,}96}{50}\right)^2 250^2 = 96{,}04.$$

Exemplo 57. Um provedor de internet monitora o a duração da conexão dos clientes a fim de dimensionar os seus servidores. A média e a distribuição desse tempo são desconhecidos mas o desvio padrão é $\sqrt{50}$ minutos. Numa amostra de 500 conexões o valor médio foi 25 minutos; o que podemos disser a respeito da média com grau de confiança 92%? Como o tamanho da amostra é razoavelmente grande, podemos usar o TCL e aproximar a distribuição por uma normal. Um intervalo de confiança para o tempo de conexão é

$$\left(M - z_\gamma \frac{\sigma}{\sqrt{n}}, M + z_\gamma \frac{\sigma}{\sqrt{n}}\right) = (24.45, 25.55).$$

Em virtude do uso do TCL, o intervalo acima é com grau de confiança *aproximadamente* $0{,}92$.

Na prática não conhecemos σ^2 e devemos substituí-lo por uma estimativa amostral, que pode ser

$$S^2 = \frac{1}{n-1}\sum (x_i - M_n)^2$$

Exemplo 58. O tempo de reação de um remédio pode ser considerado como tendo distribuição normal. Num teste, 20 pacientes foram sorteados e os tempo anotados:

$$\begin{array}{cccccccccc} 2,9 & 3,4 & 3,5 & 4,1 & 4,6 & 4,7 & 4,5 & 3,8 & 5,3 & 4,9 \\ 4,8 & 5,7 & 5,8 & 5,0 & 3,4 & 5,9 & 6,3 & 4,6 & 5,5 & 6,2 \end{array}$$

então, a variância amostral é $S^2 = 0{,}992079$ e o intervalo a 95% é

$$\left(M - z_{0,95}\sqrt{\frac{S^2}{n}}, M + z_{0,95}\sqrt{\frac{S^2}{n}}\right) = (4,278843\,,\ 5,211157).$$

Aproximação de Stirling: Seja $S_n = X_1 + \cdots + X_n$ com $X_i \sim$ Poisson(1) variável aleatória's independentes. Então

$$\mathbb{P}\,(n-1 < S_n \leqslant n) = \mathbb{P}\left(\frac{-1}{\sqrt{n}} < \frac{S_n - n}{\sqrt{n}} \leqslant 0\right) \approx \int_{-1/\sqrt{n}}^{0} \frac{1}{\sqrt{2\pi}} \mathrm{e}^{-x^2/2}\,\mathrm{d}x \approx \frac{1}{\sqrt{2\pi n}}$$

portanto

$$\mathbb{P}(S_n = n) = \frac{\mathrm{e}^{-n} n^n}{n!} \approx \frac{1}{\sqrt{2\pi n}}$$

donde

$$n! \approx n^{1/2+n}\mathrm{e}^{-n}\sqrt{2\pi}.$$

1.10 Distribuição condicional.

Distribuição condicional:: se X e Y são variáveis aleatórias e $x \in \mathbb{R}$ é tal que e $f_X(x) > 0$ então a função de distribuição condicionada de Y dado $[X = x]$ é a função em $y \in \mathbb{R}$ dada por

$$\boxed{F_{Y|X}(y|x) := \mathbb{P}(Y \leqslant y \mid X = x)}$$

Se X e Y são variável aleatória conjuntamente contínuas com densidade $f_{X,Y}(x, y)$

$$F_{Y|X}(y|x) := \int_{-\infty}^{y} \frac{f_{X,Y}(x, v)}{f_X(x)}\,\mathrm{d}v$$

e escrevemos $\mathbb{P}(Y \leqslant y \mid X = x)$, apesar de $\mathbb{P}(X = x) = 0$.

Função de probabilidade condicional:: a função de probabilidade de Y dado $[X = x]$ é

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

Exemplo 59. Sejam X e Y variáveis aleatórias independentes com $f_X(a) = f_Y(a) = \binom{n}{a}p^a(1-p)^{n-a}$, todo $a \in \{1, 2, \dots, n\}$, algum $p \in (0, 1)$. Então

$$\begin{aligned}
\mathbb{P}(X + Y = m) &= \sum_{i=0}^{n} \mathbb{P}(X = i,\ Y = m - i) = \sum_{i=0}^{n} \mathbb{P}(X = i)\mathbb{P}(Y = m - i) \\
&= \sum_{i=0}^{n} \binom{n}{i}\binom{n}{m-i} p^m (1-p)^{n-m+n} = \binom{2n}{m} p^m (1-p)^{2n-m}
\end{aligned}$$

pois $\binom{a+b}{k} = \sum_{i=0}^{n} \binom{a}{i}\binom{b}{k-i}$. Desse fato temos

$$\begin{aligned}
\mathbb{P}(X = k \mid X + Y = m) &= \frac{\mathbb{P}(X = k\,,\ X + Y = m)}{\mathbb{P}(X + Y = m)} \\
&= \frac{\mathbb{P}(X = k\,,\ Y = m - k)}{\mathbb{P}(X + Y = m)} \\
&= \frac{\mathbb{P}(X = k)\mathbb{P}(Y = m - k)}{\mathbb{P}(X + Y = m)} \\
&= \frac{\binom{n}{k}p^k(1-p)^{n-k}\binom{n}{m-k}p^{m-k}(1-p)^{n-m+k}}{\binom{2n}{m}p^m(1-p)^{2n-m}} \\
&= \frac{\binom{n}{k}\binom{n}{m-k}}{\binom{2n}{m}}
\end{aligned}$$

ou seja, X condicionada a $X + Y = m$ tem distribuição hipergeométrica com parâmetros $2n, n, m$.

Notemos que, caso as variável aleatória sejam independentes vale $f_{Y|X}(y|x) = f_Y(y)$ e vice-versa, i.e., a igualdade implica independência.

1.10.1 Esperança condicional: Se X e Y são variável aleatória e $x \in \mathbb{R}$ é tal que e $f_X(x) > 0$ então

$$\mathbb{E}[Y \mid X = x] := \begin{cases} \displaystyle\sum_y y f_{Y|X}(y|x) & \text{no caso discreto} \\ \displaystyle\int_{-\infty}^{+\infty} y f_{Y|X}(y|x)\,\mathrm{d}y & \text{no caso contínuo.} \end{cases}$$

Notemos que quando as variável aleatória são independentes $\mathbb{E}[Y \mid X = x] = \mathbb{E}[Y]$ (verifique).

Se $\mu : \mathbb{R} \to \mathbb{R}$ é dada por

$$\mu(x) := \mathbb{E}[Y \mid X = x]$$

então $\mu(X)$ é a **esperança condicional** de Y dado X, denotada

$$\mathbb{E}[Y \mid X]$$

que embora sugira uma média é, de fato, uma variável aleatória, a qual tem a seguinte propriedade

Exercício 16. Verifique

$$\mathbb{E}[\mathbb{E}[Y \mid X]] = \mathbb{E}[Y].$$

Logo

$$\mathbb{E}[Y] = \begin{cases} \displaystyle\sum_x \mathbb{E}[Y \mid X = x])\mathbb{P}(X = x) & \text{no caso discreto} \\ \displaystyle\int_{-\infty}^{+\infty} \mathbb{E}[Y \mid X = x]\, f_X(x)\,\mathrm{d}x & \text{no caso contínuo.} \end{cases}$$

Se $h(X, Y)$ tem esperança finita então

$$\mathbb{E}[h(X, Y) \mid Y = y] = \mathbb{E}[h(X, y) \mid Y = y].$$

Exemplo 60. Sejam $X,\ Y$ como no exemplo 59. X condicionada a $X + Y = m$ tem distribuição hipergeométrica com parâmetros $2n, n, m$, logo

$$\mathbb{E}[X \mid X + Y = m] = \frac{m}{2}.$$

Exemplo 61 (**Esperança de variável aleatória geométrica**). Seja X o número de lançamentos de uma moeda até sair cara, o que ocorre com probabilidade p, seja Y a variável aleatória indicadora de "cara no primeiro lançamento". Assim,

$$\mathbb{E}[X] = \mathbb{E}[\mathbb{E}[X \mid Y]] = \mathbb{E}[X \mid Y = 0](1 - p) + \mathbb{E}[X \mid Y = 1]\,p$$

mas $\mathbb{E}[X \mid Y = 1] = 1$ e $\mathbb{E}[X \mid Y = 0] = \mathbb{E}[()\, X + 1)$, portanto

$$\mathbb{E}[X] = \mathbb{E}[X](1 - p) + (1 - p) + p$$

que resolvendo para $\mathbb{E}[X]$ resulta em $\mathbb{E}[X] = 1/p$, como já sabíamos. Do mesmo modo, temos

$$\mathbb{E}\left[X^2\right] = \mathbb{E}\left[X^2 \mid Y = 0\right](1-p) + \mathbb{E}\left[X^2 \mid Y = 1\right]p = \mathbb{E}\left[(X+1)^2\right](1-p) + p$$

que resolvendo para $\mathbb{E}\left[X^2\right]$ resulta em

$$\mathbb{E}\left[X^2\right] = \frac{2-p^2}{p}$$

e com isso

$$\mathrm{Var}[X] = \mathbb{E}\left[X^2\right] - \mathbb{E}[X]^2 = \frac{2-p^2}{p} - \frac{1}{p}^2 = \frac{1-p^2}{p}.$$

Exemplo 62. Sejam $U_1, U_2, \dots$ variável aleatória com distribuição Uniforme((0, 1)) independentes e $m(x)$ o valor esperado para $N(x)$ que é o menor n para o qual $U_1 + \cdots + U_n > x$. Certamente,

$$\mathbb{E}[N(x)]\, U_1 = y = \begin{cases} 1, & \text{se } y > x \\ 1 + m(x-y), & \text{se } y \leqslant x. \end{cases}$$

Portanto,

$$m(x) = \int_0^1 \mathbb{E}[N(x) \mid U_1 = y]\, \mathrm{d}y = 1 + \int_0^x m(x-y)\, \mathrm{d}y = 1 + \int_0^x m(u)\, \mathrm{d}u.$$

Derivando os extremos dessa cadeia de igualdades obtemos $m'(x) = m(x)$, ou seja, $m(x) = a\mathrm{e}^x$ para alguma constante real a e como $m(0) = 1$ temos

$$m(x) = \mathrm{e}^x$$

é o número esperado de termos para que a soma de variável aleatória uniformes em $(0, 1)$ ultrapasse o valor x.

Exercício 17. Mostre que para constantes $a, b \in \mathbb{R}$

$$\mathbb{E}[aX + bY \mid Z] = a\mathbb{E}[X \mid Z] + b\mathbb{E}[Y \mid Z].$$

Exercício 18. Mostre que $\mu(X) = \mathbb{E}[Y \mid X]$ satisfaz

$$\mathbb{E}[\mu(X)g(X)] = \mathbb{E}[Yg(X)]$$

para qualquer função g pra qual as esperanças acima existam.

Exercício 19. Uma galinha bota N ovos, em que $N \sim$ Poisson(λ). Cada ovo vinga com probabilidade p e independente dos outros ovos. Calcule $\mathbb{E}[N \mid K]$, $\mathbb{E}[K]$ e $\mathbb{E}[K \mid N]$, em que K é o número de pintinhos.