Parte 2A

Distribuições discretas

Distribuições discretas

Variáveis aleatórias e leis de probabilidade.
Medidas de localização, dispersão e forma.
Distribuições univariadas mais comuns.
Pares aleatórios e vectores aleatórios.
Distribuição multinomial.

O que é uma variável aleatória? Qual o propósito?

Experiência aleatória: dois lançamentos de um dado (com faces de 1 a 6)

Espaço de resultados: $\Omega = \{ (i\,,j) : i = 1, 2, \ldots, 6,\ j = 1, 2, \ldots, 6\}$; $\#\Omega = 36$

Espaço de acontecimentos: $(\Omega, \mathcal{A})$, com $\mathcal{A} = \mathcal{P}(\Omega)$.

Se o dado é equilibrado, temos a regra de Laplace,

$$\boxed{P(A) = \frac{\# A}{36},\ \forall A \subseteq \Omega}$$

e $(\Omega, \mathcal{A}, P)$ é um espaço de probabilidade.

Podemos estar interessados no estudo da soma das componentes i e j do resultado (i, j), ou da maior delas, ou da diferença absoluta das mesmas, etc.

No caso da soma, associamos a cada resultado $\omega = (i, j)$ o valor $i + j$ e basta saber quais as probabilidades correspondentes a cada valor $i + j$ possível. À função que a cada resultado $\omega = (i, j) \in \Omega$ faz corresponder o valor real $i + j$ chamamos variável aleatória (v.a.). A que propósito? Para transportar a medida P, definida em $(\Omega, \mathcal{A})$, para $(\mathbb{R}, \mathcal{B})$. Assim descartamos Ω e passamos para $\mathbb{R}$. ☺

Variável aleatória

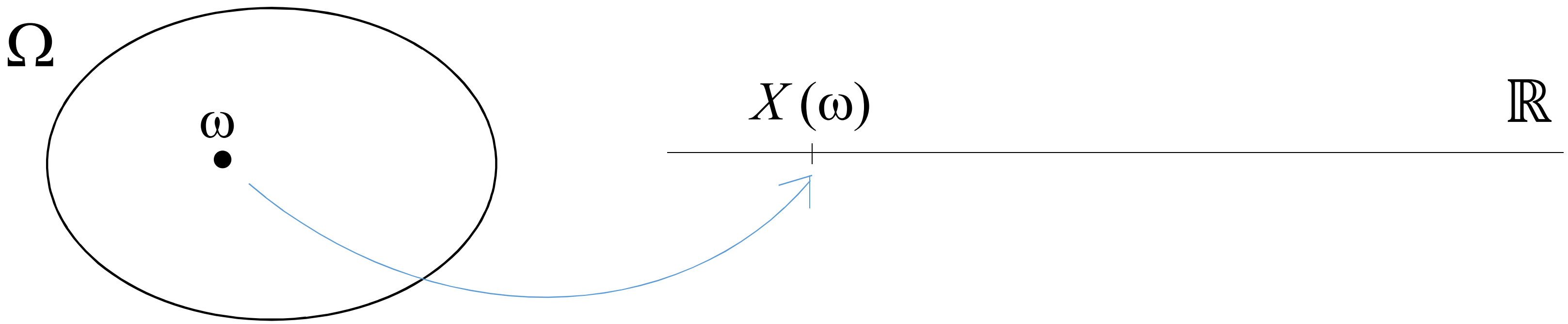
Seja $(\Omega, \mathcal{A}, P)$ um espaço de probabilidades.

Variável aleatória é uma função $X : \Omega \to \mathbb{R}$

que a cada $\omega \in \Omega$ faz corresponder um valor real, $X(\omega)$,

tal que $X^{-1}(B) \in \mathcal{A}$, para qualquer boreliano B de $\mathbb{R}$.

Émile Borel
1871-1956



Cada boreliano B de $\mathbb{R}$ vai ser probabilizável, e a sua probabilidade será a da sua imagem inversa por X, i.e., será $P(\, X^{-1}(B) \,)$

Exemplo 4: X – "soma das pintas no lançamento de 2 dados" é a função que a cada resultado elementar $\omega = (i, j)$ faz corresponder o valor $i + j$, com $i, j = 1, 2, \ldots, 6$. Temos $\Omega = \{ (i\ , j) : i = 1, 2, ..., 6,\ j = 1, 2, ..., 6\}$ e

$$\begin{aligned} X:\ \ \Omega &\to \mathbb{R} \\ (i, j) &\mapsto i + j \end{aligned}$$

Neste caso, o contradomínio de X (também chamado suporte de X, e representado por $\mathrm{supp}(X)$) é o conjunto $\{2, 3, \ldots, 12\}$, que tem 11 elementos. Neste caso o suporte tem um nº finito de elementos e diz-se então que a v.a. X é "discreta".

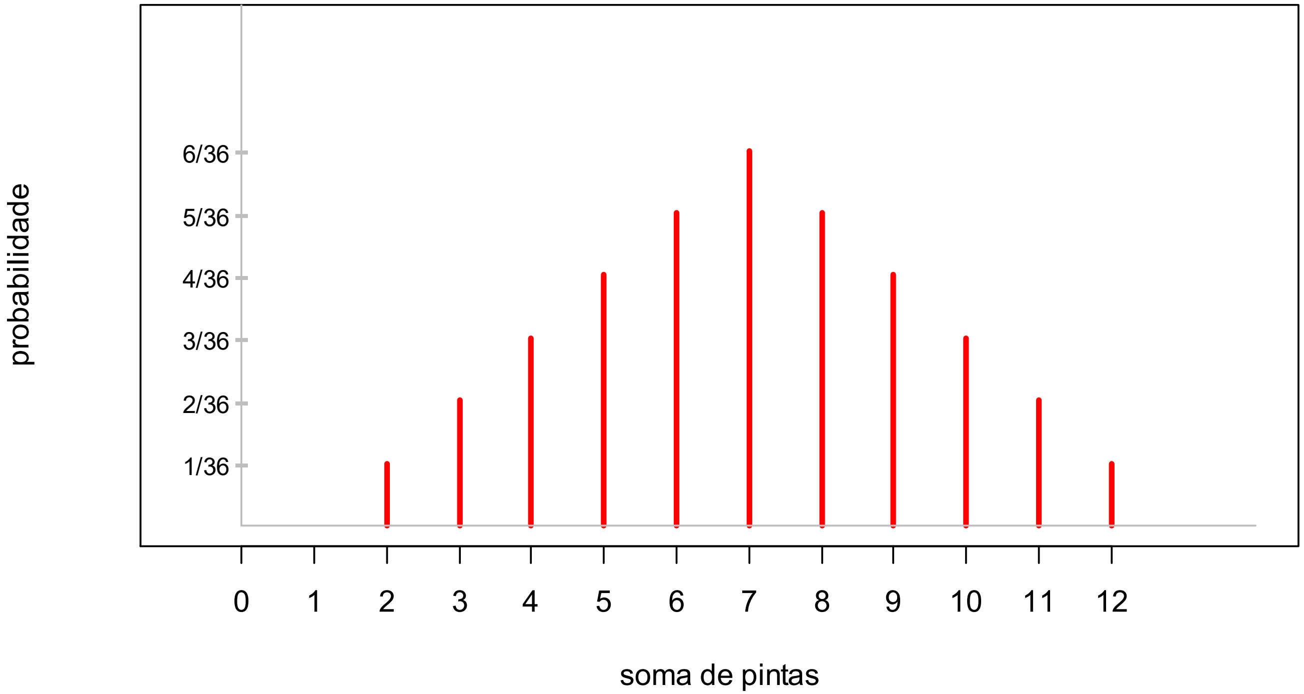
Se o contradomínio (ou suporte) de X for um conjunto **contável** (i.e., finito ou numerável), diz-se que X é uma v.a. discreta. Nesse caso, a cada x (do suporte) atribuímos a probabilidade do conjunto $\{\omega: X(\omega) = x\}$, i.e., do acontecimento $X^{-1}(x)$

Exemplo 4 (cont.): X – "soma das pintas lançando 2 dados equilibrados".
X é v.a. discreta pois o suporte de X é $\{2, 3, 4, \ldots, 12\}$, que tem um nº finito de elementos (são 11). E então

$$
\begin{aligned}
P(X=2) &= P(\ \{(1,1)\}\ ) = 1/36\\
P(X=3) &= P(\ \{(1,2), (2,1)\}) = 2/36\\
P(X=4) &= P(\ \{(1,3), (3,1), (2,2)\}\ ) = 3/36\\
&\vdots
\end{aligned}
$$

Aos valores 2, 3, 4, …, 12 do suporte de X, associamos as probabilidades $1/36$, $2/36$, $3/36$, $4/36$, $5/36$, $6/36$, $5/36$, $4/36$, $3/36$, $2/36$, $1/36$, respetivamente (cuja soma é 1, claro, pois $P(\Omega) = 1$).

Exemplo 4 (cont.): X – "soma das pintas lançando 2 dados equilibrados".
A representação gráfica da função que aos valores 2, 3, 4, …, 12 (do suporte de X) associa as probabilidades $1/36$, $2/36$, $3/36$, $4/36$, $5/36$, $6/36$, $5/36$, $4/36$, $3/36$, $2/36$, $1/36$, é a seguinte:



Exemplo 5: A experiência consiste num tiro ao alvo ($\mathbb{R}^2$), com resultado dado pelas coordenadas a e b do ponto de impacto da bala ($\Omega = \mathbb{R}^2$). A *mouche* é o ponto $(0,0)$. A pontuação do tiro será 10, 9, …, 0, consoante a distância à *mouche* seja ≤10*cm*, de 10*cm* a 20*cm*, …, >100*cm*. Classifique as v.a.

X – distância do ponto de impacto à *mouche* – não discreta

Y – pontuação obtida com o tiro – discreta

$$X : \Omega \to \mathbb{R}, \quad (a,b) \mapsto \sqrt{a^2 + b^2}$$

$$Y : \Omega \to \mathbb{R}, \quad (a,b) \mapsto \begin{cases} 10 & , \ se \quad \sqrt{a^2+b^2} \le 10 \\ 9 & , \ se \quad 10 < \sqrt{a^2+b^2} \le 20 \\ \vdots & \\ 0 & , \ se \quad \sqrt{a^2+b^2} > 100 \end{cases}$$

supp(X) = [0 , + ∞ [

supp(Y) = { 0 , 1, …, 10 }

Exemplo 6: $X =$ "nº caras em 3 lançamentos de moeda equilbrada"

$\Omega =$ {CCC, CCE, CEC, CEE, ECC, ECE, EEC, EEE}

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

CCC $\to$ 3
CCE $\to$ 2
CEC $\to$ 2
CEE $\to$ 1
ECC $\to$ 2
ECE $\to$ 1
EEC $\to$ 1
EEE $\to$ 0

contradomínio
(ou suporte) de X:
$\{ 0, 1, 2, 3 \} \subset \mathbb{R}$

A distribuição ou lei de probabilidade de X fica definida por

$$X: \begin{cases} 0 & 1 & 2 & 3 \\ 1/8 & 3/8 & 3/8 & 1/8 \end{cases}$$

Binomial $(n\ ,\ p)$, com $n = 3$, $p = 0.5$

$P(X = 2) = P$ ({CCE, CEC,ECC}) $= 3/8$

$P(X \geq 2) = P\ (X = 2$ ou $3) =$

$= P$ ({CCE, CEC,ECC,CCC}) $=$

$= 4/8$

$= P\ (X = 2) + P\ (X = 3)$

Variáveis aleatórias discretas

- são as que têm suporte finito ou numerável, $\{x_1, \ldots, x_n\}$ ou $\{x_1, x_2, \ldots\}$

- ficam identificadas pela função massa de probabilidade (fmp), que atribui a cada valor x_i a correspondente probabilidade p_i, ou seja, por

$$\boxed{X : \begin{cases} x_1 & x_2 & x_3 & \cdots \\ p_1 & p_2 & p_3 & \cdots \end{cases}}$$

ou

$$\boxed{p_i = P(X = x_i)\,, \quad i \in I}$$

I é um conjunto de índices em nº finito ou numerável

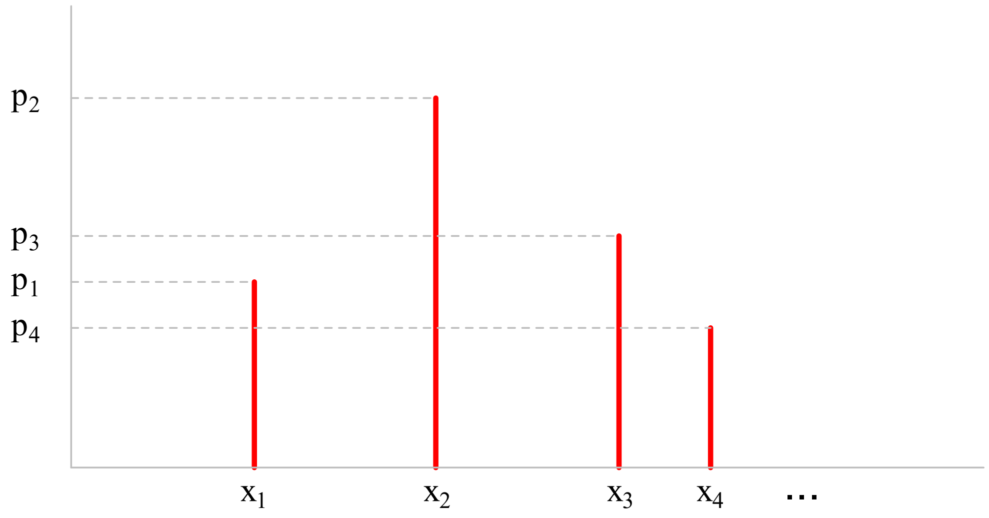
Exemplo 6 (cont.):

$$X : \begin{cases} 0 & 1 & 2 & 3 \\ 1/8 & 3/8 & 3/8 & 1/8 \end{cases}$$

ou

$$p_i = P(X = i) = \binom{3}{i}\frac{1}{8}, \quad i = 0,1,2,3$$

com representação gráfica
(se tal for possível…)



A fmp satisfaz às seguintes propriedades, que decorrem da axiomática da probabilidade:

$$(i) \qquad p_i \geq 0$$

p_i não negativos (I)

$$(ii) \qquad \sum_i p_i = 1$$

p_i têm soma unitária (II)

$$(iii) \qquad P(X \in B) = \sum_{i:x_i \in B} P(X = x_i)$$

aditividade (III)

Exemplo 7: em 2 lançamentos de uma moeda equilibrada, a fmp da v.a. X = "nº de caras" pode representar-se por

$$X:\begin{cases} 0 & 1 & 2 \\ 1/4 & 1/2 & 1/4 \end{cases} \quad \text{ou} \quad P(X=i)=\binom{2}{i}\frac{1}{4}\ ,\ i=0,1,2$$

E então, por (iii), temos por exemplo,

$$P(X\in[1,2])=P(1\le X\le 2)=P(X=1)+P(X=2)=3/4$$

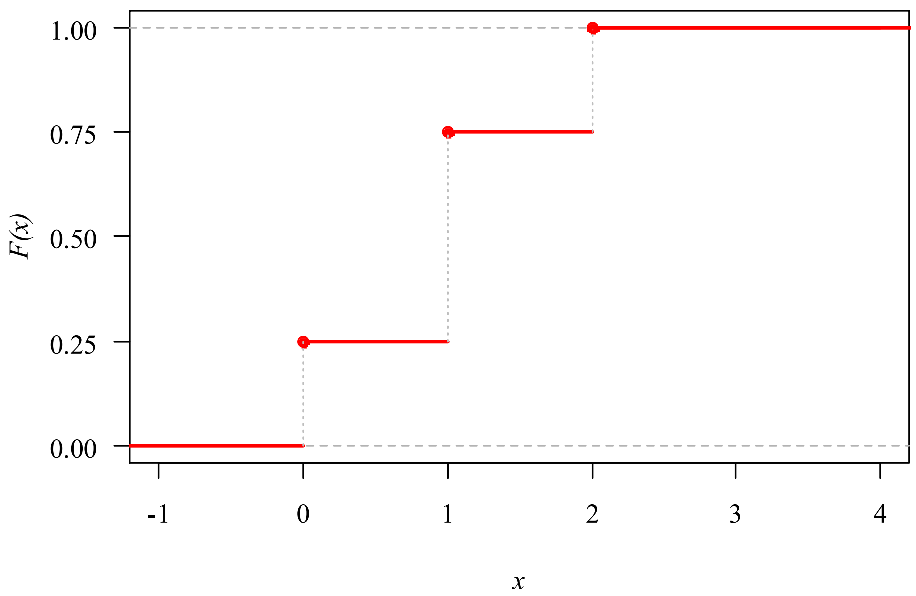
Nota: Se Ω for finito ou infinito numerável, então qualquer v.a. definida em Ω será discreta; se Ω for infinito não numerável, as v.a. podem ser discretas ou não (vd. *exemplo* 5).

Podemos também identificar qualquer v.a. X pela função de distribuição (fd), definida por

$$F(x) = P(X \leq x), \; x \in \mathbb{R}$$

Exemplo 7 (cont.): a fd da v.a. $X : \begin{cases} 0 & 1 & 2 \\ 1/4 & 1/2 & 1/4 \end{cases}$ com suporte {0,1,2} é

$$F(x) = \begin{cases} 0 & , \quad x < 0 \\ 1/4 & , \quad 0 \leq x < 1 \\ 3/4 & , \quad 1 \leq x < 2 \\ 1 & , \quad x \geq 2 \end{cases}$$



Nota: Esta fd é "em escada"; o conjunto dos pontos de salto é o suporte de X; as amplitudes dos saltos são as probabilidades p_i; a partir de uma das funções (fmp ou fd) pode deduzir-se a outra.

As fd são não decrescentes, contínuas à direita e tais que

$$\lim_{x\to-\infty} F(x) = 0, \qquad \lim_{x\to+\infty} F(x) = 1$$

No caso discreto, a fd (em escada) é dada por

$$F(x) = \sum_{i:x_i \leq x} p_i$$

É válida a seguinte fórmula (para qualquer v.a. X) para $a < b$:

$$P(a < X \leq b) = F(b) - F(a)$$

Demonstração:

Seja $B =]-\infty, b]$, $A =]-\infty, a]$ e $a < b$. Então, por ser $A \subseteq B$ e pela propriedade VII: $P(B \backslash A) = P(B) - P(A \cap B)$,
temos $P(a < X \leq b) = P(X \in]a,b]) = P(X \in B \setminus A) = P(X \in B) - P(X \in A) = F(b) - F(a)$

Distribuições discretas

- uniforme (em n pontos)
- hipergeométrica (n, N, p)
  amostragem sem reposição em populações dicotómicas
- binomial (n, p)
  nº de sucessos em n provas de Bernoulli independentes (sendo p a probabilidade de sucesso em cada prova)
  amostragem com reposição em populações dicotómicas
- geométrica (p)
  nº de provas de Bernoulli necessárias até que ocorra o 1º sucesso
- Poisson (λ)
  nº de fenómenos que ocorrem por unidade de tempo ou espaço, de forma aleatória

Distribuição uniforme (discreta)

Diz-se que a v.a. X tem distribuição uniforme em n pontos, a_1 , a_2 , ..., a_n , e escreve-se $X \frown U\{ a_1 , a_2 , ..., a_n \}$, se tiver fmp

$$\boxed{P(X = a_j) = \frac{1}{n}, \quad j = 1, 2, \ldots, n}$$

No caso particular $n = 1$, i.e., $X \frown U\{a\}$, a v.a. X diz-se "degenerada no ponto a". Por exemplo, se a v.a. X é a soma do nº de caras com o nº de coroas em 10 lançamentos de uma moeda, temos $X \frown U\{10\}$

Exemplo 8: A v.a. X que representa o resultado (nº de pintas) do lançamento de um dado equilibrado tem distribuição $U\{1, 2, 3, 4, 5, 6\}$, com fmp

$$P(X = j) = \frac{1}{6}, \quad j = 1, 2, \ldots, 6$$

Distribuição hipergeométrica

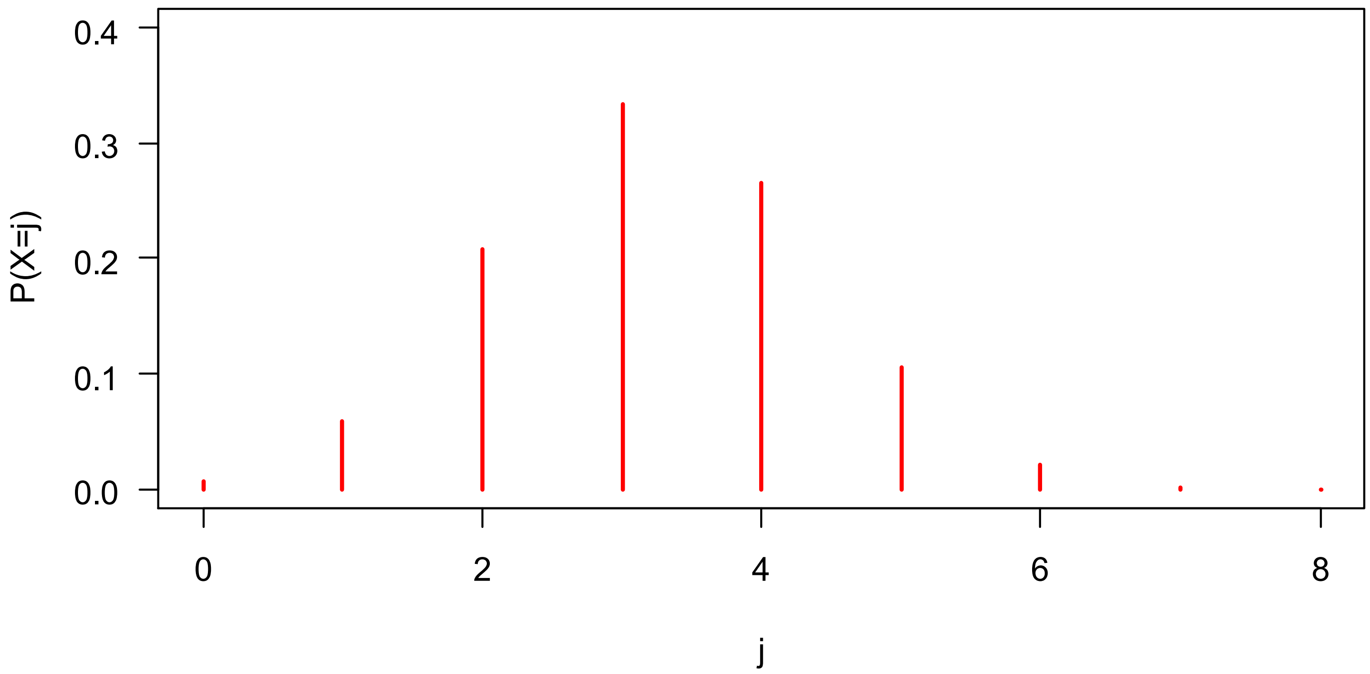
`dhyper, phyper, ...`

Modelo para a v.a. X que representa o nº de bolas brancas obtidas em n extrações sem reposição de uma urna com um total de N bolas, das quais uma proporção p são brancas, (i.e., Np são brancas) e as restantes são pretas. Escreve-se $X \frown$ HG(n, N, p) e a correspondente fmp é

$$\boxed{P(X = j) = \frac{\binom{Np}{j}\binom{N-Np}{n-j}}{\binom{N}{n}}, \quad \begin{array}{l} j = 0, 1, 2, \ldots, n \\ j \leq Np \\ n - j \leq N - Np \end{array}}$$

Numa população com uma proporção p de elementos com dada característica, X representa o nº de elementos com essa característica obtidos em n extrações sem reposição.

Exemplo 9: no caso de 25 bolas, das quais 8 são brancas, se retirarmos 10 bolas sem reposição, temos a distribuição HG(10, 25, 0.32) para o nº de bolas brancas extraídas, com a representação gráfica (o suporte desta v.a. é {0,1,2,…,8}) seguinte



$N = 25$

$Np = 8$

$p = 8/25 = 0.32$

$n = 10$

```
plot(0:8,dhyper(0:8,8,17,10),col=2,type="h",xlab="j",ylab="P(X=j)",ylim=c(0,0.4))
```

Distribuição binomial

`dbinom, pbinom, ...`

Modelo para a v.a. X que representa o "nº de caras em n lançamentos de uma moeda-p", ou seja, o "nº de *sucessos* em n repetições (independentes) de uma experiência aleatória" (chamadas *provas de Bernoulli*), na qual p é a probabilidade de *sucesso*. Escreve-se $X \frown \mathrm{bi}(n, p)$ e a fmp é

$$\boxed{P(X=j)=\binom{n}{j}p^j(1-p)^{n-j}, \quad j=0,1,2,\ldots,n}$$

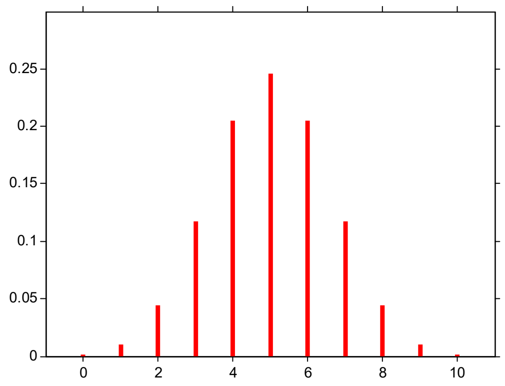
A distribuição $\mathrm{bi}(1,p)$ chama-se Bernoulli(p)

J. Bernoulli

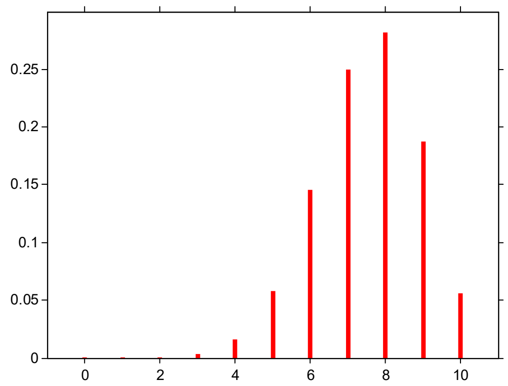
Numa população com uma proporção p de elementos com dada característica, X representa o nº de elementos com essa característica obtidos em n extrações com reposição.

Distribuição bi(n,p) – fmp

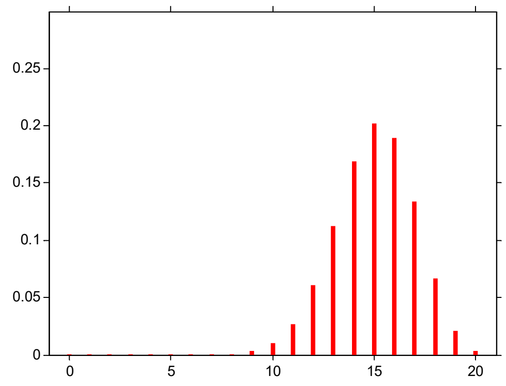
$$\mu = np$$
$$\sigma^2 = np(1-p)$$



$n = 10,\ \ p = 0.5$



$n = 10,\ \ p = 0.75$



$n = 20,\ \ p = 0.75$

Exercício: Considere distribuições bi$(10, p)$ com $p = 0.1, 0.25, 0.5, 0.75, 0.9$

(i) Determine as fmp correspondentes e represente-as graficamente

(ii) Observe em que casos há simetria (em torno de que valor?)

(iii) Calcule $P(X \leq 7)$, $P(X < 7)$, $P(3 < X \leq 7)$, $P(3 < X < 7)$, $P(X \geq 7)$

(iv) Calcule P (X ser ímpar)

No `R`, há 4 funções para cada lei de probabilidade implementada, com o nome da distribuição antecedido de um prefixo (`d`, `p`, `q`, `r`), por exemplo: `dbinom, pbinom, ..., dhyper, phyper, ...`

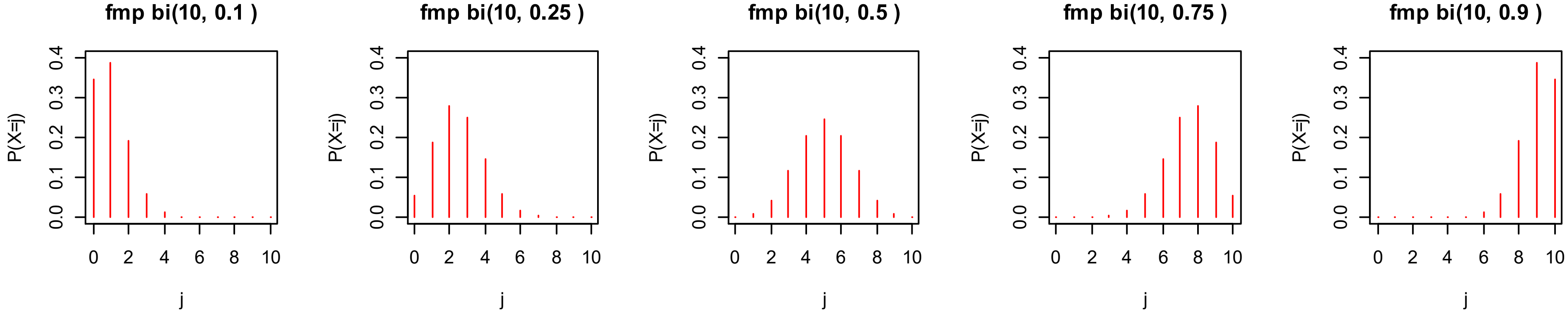
- `d` – **d**ensidade / massa de probabilidade (fdp / fmp)
- `p` – **p**robabilidade acumulada, i.e., distribuição (fd) e variante
- `q` – **q**uantil, i.e., inversa (generalizada) da fd
- `r` – simulação de amostras / `random`

Exemplo 10:

```
dbinom(4:7,10,0.4)            # calcula P(X = 4), ..., P(X = 7) no caso X ⌢ bi(10, 0.4)
pbinom(7,10,0.4)              # calcula P(X ≤ 7) no caso X ⌢ bi(10, 0.4)
sum(dbinom(0:7,10,0.4))       # o mesmo que o anterior
pbinom(7,10,0.4,lower=F)      # calcula P(X > 7) no caso X ⌢ bi(10, 0.4)
1-pbinom(7,10,0.4)            # o mesmo que o anterior
qbinom(0.75,2,0.5)            # fd inversa; F(1) = 0.75, logo min{F-1(0.75)} = 1
rbinom(12,4,0.5)              # simula 12 vezes o "nº caras em 4 lanç. de uma moeda-½"
```

Resolução do exercício:

```
# (i) fmp da bi(10,0.5):
fmp <- dbinom(0:10,10,0.5); names(fmp) <- 0:10; fmp; round(fmp,4)
0      1      2      3      4      5      6      7      8      9      10
0.0010 0.0098 0.0439 0.1172 0.2051 0.2461 0.2051 0.1172 0.0439 0.0098 0.0010
# os 5 gráficos (plot) de uma vez, em linha (matriz 1×5), com mfrow=c(1,5)
par(mfrow=c(1,5))
for (i in c(0.1,0.25,0.5,0.75,0.9))
plot(0:10,dbinom(0:10,10,i),type="h",col=2,ylim=c(0,0.4),xlab="j",
       ylab="P(X=j)", main=paste("fmp bi(10,",i,")"))
# (ii) notar que a simetria ocorre no caso p=0.5, em torno do ponto 5 (=10/2)
# para n=11, p=0.5, ver que há simetria em torno do ponto 5.5=11/2, fazendo:
par(mfrow=c(1,1)); plot(0:11,dbinom(0:11,11,0.5),type="h",col=2)
```



Resolução (cont.): Caso bi$(10,p)$ com $p = 0.25$ (para os outros casos é idêntico)

(iii) Calcule $P(X \leq 7)$, $P(X < 7)$, $P(3 < X \leq 7)$, $P(3 < X < 7)$, $P(X \geq 7)$

(iv) Calcule P (X ser ímpar)

`pbinom(x, n, p)` devolve $P(X \leq x)$	no caso $X \frown$ bi(n, p)
`pbinom(x, n, p, lower.tail=F)` devolve $P(X > x)$	no caso $X \frown$ bi(n, p)

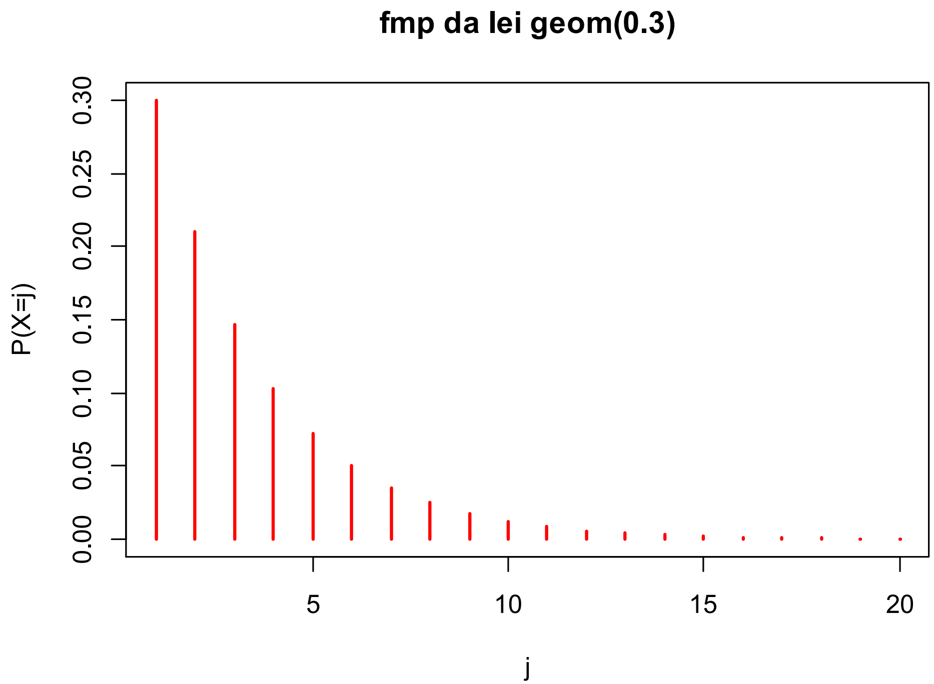
```
## (iii) recordar a fórmula P(a < X ≤ b) = F(b)-F(a)
## calcular P(X ≤ 7), P(X < 7), P(3 < X ≤ 7), P(3 < X < 7), P(X ≥ 7)
pbinom(7,10,0.25)                      # P(X≤7)
pbinom(7,10,c(0.25,0.5,0.75))          # P(X≤7), para p = 0.25,0.5,0.75
pbinom(6,10,0.25)                      # P(X<7) porque P(X<7)=P(X≤6)
pbinom(7,10,0.25)-pbinom(3,10,0.25)    # P(3<X≤7)
pbinom(6,10,0.25)-pbinom(3,10,0.25)    # P(3<X<7) = P(3<X≤6)
pbinom(6,10,0.25, lower=F)             # P(X≥7) porque P(X≥7)=P(X>6)

## (iv) calcular P(X ser ímpar)
sum(dbinom(0:4*2+1,10,0.25))           # 0:4*2+1 ou seq(1,9,by=1)
[1] 0.4995117
```

Distribuição geométrica

`dgeom, pgeom, ...`

Modelo para a v.a. X que representa o "nº lançamentos necessários (de uma moeda-p) até se observar uma cara", ou seja, o "nº de provas de Bernoulli até se observar um *sucesso*". Escreve-se $X \frown \text{geom}(p)$ e a fmp é

$$\boxed{P(X=j)=(1-p)^{j-1}p, \quad j=1,2,\ldots}$$



Variante: Y ="nº de insucessos antes de obter o 1º sucesso" (implementada no R), i.e., $Y=X-1$.

Generalização (distribuição de Pascal, ou binomial negativa): X_k = "nº de provas até obter o k-ésimo sucesso" ou Y_k = "nº de insucessos antes de obter o k-ésimo sucesso"; $Y_k=X_k-k$.

Distribuição de Poisson

`dpois, ppois, ...`

Modelo para a v.a. X que representa o nº de fenómenos que ocorrem no tempo (ou no espaço) de forma aleatória.

Siméon Denis Poisson
1781-1840

Exemplos:

nº de parasitas num hospedeiro por unidade de superfície,
nº de colónias de bactérias por unidade de volume numa solução,
nº de partículas radioativas emitidas por unidade de tempo,
nº de defeitos por unidade de superfície num tecido,
nº de partículas de poeira num dado volume de ar,
nº de gralhas por página de um livro, etc.

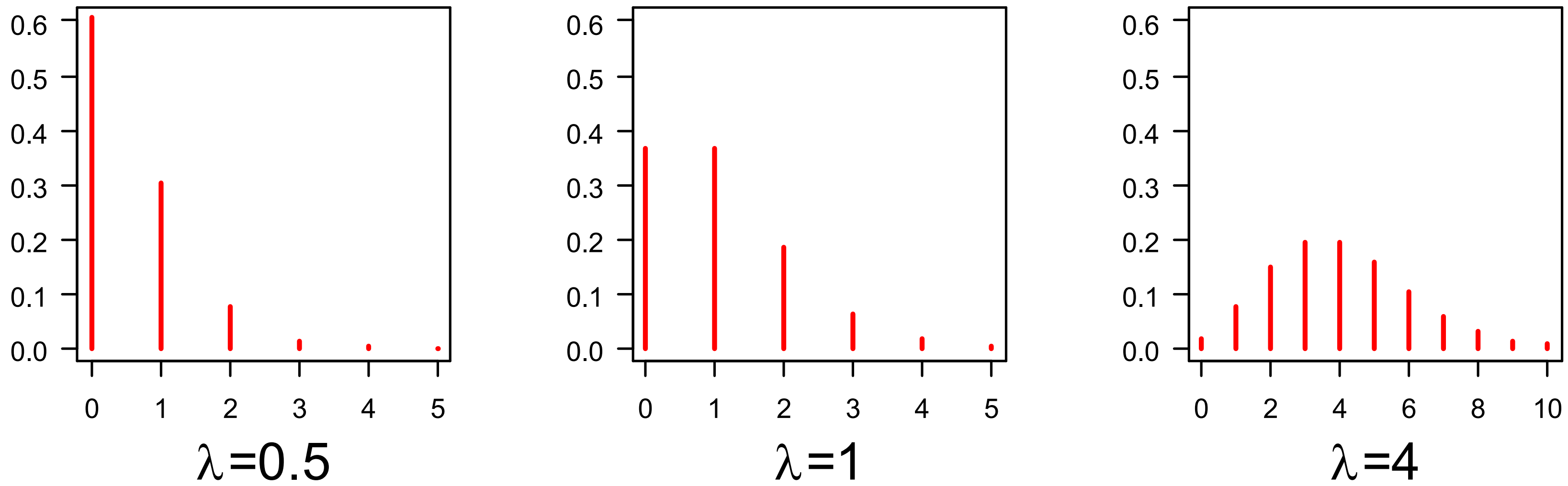
Escreve-se $X \frown \text{Poisson}(\lambda)$ e a fmp é

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

$\lambda > 0$ é um parâmetro que representa o nº médio de fenómenos por unidade de tempo (ou espaço)

Distribuição Poisson(λ) – fmp

$\mu = \lambda$
$\sigma^2 = \lambda$



Nota: Num "processo de Poisson" o nº de fenómenos que ocorrem em t unidades de tempo (ou de espaço) tem distribuição Poisson(λt). Por exemplo, se o nº de partículas radioativas emitidas por segundo tiver distribuição Poisson(0.5), então o nº de partículas radioativas emitidas num minuto (60 segundos) terá distribuição Poisson(30).

Exercício: O nº de organismos aquáticos por *cm*³ num dado lago tem distribuição Poisson(2). Qual a probabilidade de numa amostra de 10 *cm*³ de água do lago haver

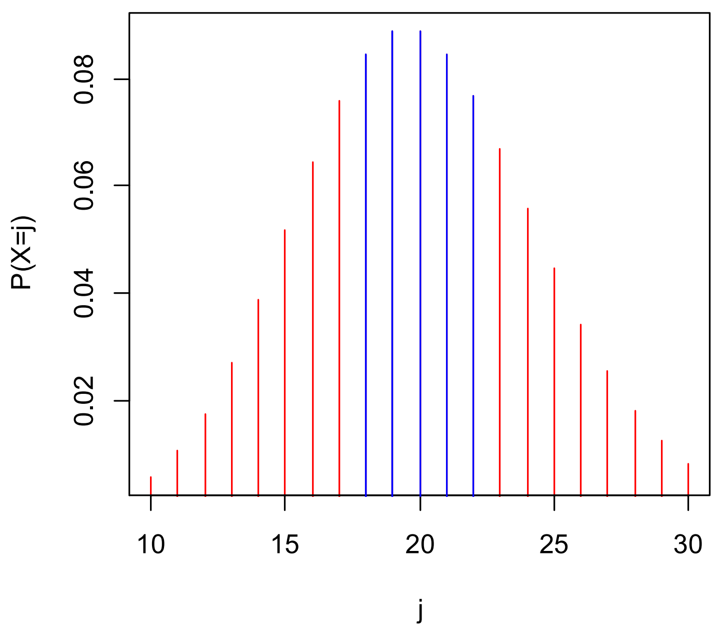
(i) pelo menos 18 organismos?

(ii) entre 18 e 22 organismos (*inclusive*)?

Resolução:

O nº de organismos em 10 *cm*³ (v.a. X) tem distribuição Poisson(2×10). Calcula-se $P(X \geq 18)$ e $P(18 \leq X \leq 22)$, notando que $P(X \geq 18) = P(X > 17)$ e $P(18 \leq X \leq 22) = P(17 < X \leq 22)$:

```
ppois(17,20,lower.tail=F)
[1] 0.7029716
ppois(22,20)-ppois(17,20)
[1] 0.4235829
sum(dpois(18:22,20))
[1] 0.4235829
```



Aproximação da binomial à Poisson:

O modelo Poisson(λ) aparece como limite do modelo $\mathrm{bi}(n,p)$, com $n \to \infty$ e $p \to 0$, sendo $\lambda = np$ constante, i.e.,

$$\Updownarrow$$

$$p = \frac{\lambda}{n}$$

$$\boxed{\binom{n}{j} p^j (1-p)^{n-j} \xrightarrow[p=\lambda/n]{n\to\infty} e^{-\lambda}\frac{\lambda^j}{j!}, \quad j = 0,1,2,\ldots}$$

e por isso (como $p \approx 0$) a Poisson é também conhecida por lei dos acontecimentos raros

Exemplo 11: Num decaimento radioactivo, a v.a. $X =$ "nº de partículas libertadas por unidade de tempo" ajusta-se bem ao modelo Poisson(λ) ...

... divida-se a unidade de tempo (e.g., 1 *min*) em n intervalos de amplitude h (com n grande), de modo a que haja 0 ou 1 partículas libertadas em cada intervalo. Se os números de partículas (0 ou 1) libertadas em cada intervalo forem mutuamente independentes, então o número de partículas libertadas por unidade de tempo tem distribuição $\mathrm{bi}(n,p)$, que por sua vez é aproximada pela Poisson(λ), com $\lambda = np$ (na prática, λ é dado pelo número médio de partículas libertadas por unidade de tempo).

Demonstração: Como $\lambda = np$ é constante, temos $p = \lambda/n$, donde

$$\lim_{n\to\infty}\binom{n}{j}p^j(1-p)^{n-j} = \lim_{n\to\infty}\binom{n}{j}\left(\tfrac{\lambda}{n}\right)^j\left(1-\tfrac{\lambda}{n}\right)^{n-j} = \lim_{n\to\infty}\frac{n(n-1)\ldots(n-j+1)}{j!}\,\frac{\lambda^j}{n^j}\left(1-\tfrac{\lambda}{n}\right)^{n-j} =$$

$$= \frac{\lambda^j}{j!}\lim_{n\to\infty}\left(1-\tfrac{1}{n}\right)\left(1-\tfrac{2}{n}\right)\ldots\left(1-\tfrac{j-1}{n}\right)\left(1-\tfrac{\lambda}{n}\right)^n\left(1-\tfrac{\lambda}{n}\right)^{-j} =$$

$$= \frac{\lambda^j}{j!}\lim_{n\to\infty}\left(1-\tfrac{\lambda}{n}\right)^n = e^{-\lambda}\frac{\lambda^j}{j!}, \quad j = 0, 1, 2, \ldots$$

$$\boxed{\lim_{n\to+\infty}\left(1+\frac{x}{n}\right)^n = e^x}$$

Por exemplo, para $n = 100$, $p = 0.02$ temos $\lambda = np = 2$; comparamos as fmp bi(100, 0.02) e Poisson(2):

```
round(dbinom(0:10,100,0.02),5)
 [1] 0.13262  0.27065  0.27341  0.18228  0.09021  0.03535  0.01142  0.00313  0.00074  0.00015  0.00003
round(dpois(0:10,2),5)
 [1] 0.13534  0.27067  0.27067  0.18045  0.09022  0.03609  0.01203  0.00344  0.00086  0.00019  0.00004
```

Aproximação da hipergeométrica à binomial

O modelo $\mathrm{bi}(n,p)$ aparece como limite do $\mathrm{HG}(n, N, p)$, quando $N \to \infty$ (n e j fixos):

$$\boxed{\underbrace{\frac{\binom{Np}{j}\binom{N-Np}{n-j}}{\binom{N}{n}}}_{\text{probabilidade de } j \text{ sucessos em tiragens sem reposição}} \xrightarrow[N\to\infty]{} \underbrace{\binom{n}{j} p^j (1-p)^{n-j}}_{\text{probabilidade de } j \text{ sucessos em tiragens com reposição}}, \; j = 0, 1, 2, \ldots, n}$$

Na prática, como n (nº de extrações) está fixo e $N \to \infty$, este resultado permite concluir que, no caso de N (nº de elementos da população) ser grande em comparação com n, é indiferente fazer as extrações com ou sem reposição.

Demonstração:

$$\frac{\binom{Np}{j}\binom{N-Np}{n-j}}{\binom{N}{n}} = \binom{n}{j}\frac{Np(Np-1)\ldots(Np-j+1)}{N(N-1)\ldots(N-j+1)} \times \frac{(N-Np)(N-Np-1)\ldots(N-Np-(n-j)+1)}{(N-j)(N-j-1)\ldots(N-n+1)} =$$

$$= \binom{n}{j} p \frac{p-\frac{1}{N}}{1-\frac{1}{N}} \frac{p-\frac{2}{N}}{1-\frac{2}{N}} \ldots \frac{p-\frac{j-1}{N}}{1-\frac{j-1}{N}} \times \frac{1-p}{1-\frac{j}{N}} \frac{1-p-\frac{1}{N}}{1-\frac{j+1}{N}} \ldots \frac{1-p-\frac{n-j-1}{N}}{1-\frac{n-1}{N}} \xrightarrow[N\to\infty]{} \binom{n}{j} p^j (1-p)^{n-j}, j = 0,1,2,\ldots,n$$

Por exemplo, numa população com 1000 homens e 2000 mulheres (i.e., na proporção de 1 para 2), retira-se uma amostra (sem reposição) de 10 pessoas e Y representa o nº de homens na amostra. Compara-se a distribuição HG(10, 3000, 1/3) com a bi(10, 1/3) através das fmp (N =3000, n =10, p =1/3):

```
round(dhyper(0:10,1000,2000,10), 4)
 [1] 0.0172 0.0864 0.1951 0.2605 0.2279 0.1366 0.0567 0.0161 0.0030 0.0003 0.0000
round(dbinom(0:10,10,1/3), 4)
 [1] 0.0173 0.0867 0.1951 0.2601 0.2276 0.1366 0.0569 0.0163 0.0030 0.0003 0.0000
```

Valor médio (ou valor esperado) de uma v.a. X (discreta)

é uma "média pesada" dos valores $x_i \in Supp(X)$, sendo os pesos as correspondentes probabilidades p_i ; representa-se por $E(X)$, μ_X ou μ, i.e.,

$$\boxed{\mu = E(X) = \sum_i x_i p_i}$$

, desde que $\sum_i |x_i| p_i < +\infty$

(i.e., desde que a série seja absolutamente convergente)

O valor médio de X só existe se a série $\sum_i x_i p_i$ for absolutamente convergente.

(se a série for convergente mas não absolutamente convergente, a soma depende da ordenação dos termos)

Exemplo 12: O valor médio do "número de pintas" (v.a. X) no lançamento de um dado equilibrado é 3.5, pois temos a fmp dada por

$$X:\begin{cases} 1 & 2 & 3 & 4 & 5 & 6 \\ 1/6 & 1/6 & 1/6 & 1/6 & 1/6 & 1/6 \end{cases}$$

ou

$$p_i = P(X=i) = \frac{1}{6}, \quad i = 1, 2, \ldots, 6$$

donde

$$\mu = E(X) = \sum_i x_i p_i = \sum_{i=1}^{6} i\,\frac{1}{6} = (1+2+\ldots+6)\,\frac{1}{6} = \frac{7\times 6}{2}\,\frac{1}{6} = \frac{7}{2}$$

Note que o valor médio não tem que pertencer ao suporte da v.a.

Valor médio de $h(X)$, sendo X discreta

Dada uma função h, definimos analogamente o valor médio da v.a. $Y = h(X)$ por

$$\boxed{E(h(X)) = \sum_i h(x_i)\, p_i} \quad , \text{ se } \quad \sum_i |\, h(x_i) \,|\, p_i < +\infty$$

Exemplo 12 (cont.): Cálculo do valor médio de X^2

$$E(X^2) = \sum_i x_i^{\,2}\, p_i = (1^2 + 2^2 + \ldots + 6^2)\,\frac{1}{6} = \frac{6 \times 7 \times 13}{6}\,\frac{1}{6} = \frac{91}{6}$$

$$\boxed{\sum_{i=1}^{n} i^2 = \frac{n(n+1)(2n+1)}{6}}$$

Interpretação (frequencista) do valor médio

```
# simulação (r = 6×10⁵ lanç do dado equilibrado)
pintas <- sample(1:6, 6*10^5, rep=T)
table(pintas)
pintas
     1      2      3      4      5      6
100242 100309 100244 100146  99345  99714

mean(pintas)
[1] 3.495308

mean(pintas^2)
[1] 15.13223
91/6
[1] 15.16667
```

A **média do nº de pintas** nas r simulações é aproximadamente igual ao valor médio teórico, $E(X) = 3.5$

A **média dos quadrados do nº pintas** nas r simulações é aproximadamente igual ao valor médio $E(X^2) = 91/6$

Valor médio (populacional/teórico)

$$E(X) = \sum_i x_i p_i = \sum_{i=1}^{6} i \tfrac{1}{6}$$

$$\bar{x} = \sum_i x_i f_i = \sum_{i=1}^{6} i\, f_i$$

média (amostral/empírica)

$$f_1 = \frac{100242}{6\times 10^5}$$

$$f_2 = \frac{100309}{6\times 10^5}$$

$$\vdots$$

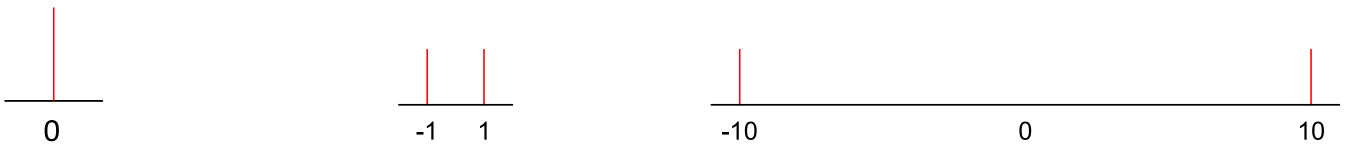
LGN: A média amostral converge – em certo sentido – para o valor médio teórico

O valor médio não identifica a v.a.

Duas v.a. podem ter o mesmo valor médio μ e serem muito diferentes. Podem diferir quanto à dispersão (i.e., quanto à maneira como os seus valores se distribuem em torno de algum valor central, como μ). Por exemplo, as v.a.

$$\boxed{W:\begin{cases} 0 \\ 1 \end{cases}} \qquad \boxed{X:\begin{cases} -1 & 1 \\ 0.5 & 0.5 \end{cases}} \qquad \boxed{Y:\begin{cases} -10 & 10 \\ 0.5 & 0.5 \end{cases}} \qquad Y = 10\,X$$

têm todas valor médio nulo, mas diferem quanto à dispersão.
Y tem dispersão maior do que X e esta tem dispersão maior do que W.



$W \frown U\{0\}$ $\qquad$ $X \frown U\{-1, 1\}$ $\qquad$ $Y \frown U\{-10, 10\}$

Variância e desvio padrão de uma v.a. X (discreta)

A dispersão de uma v.a. X pode medir-se pela variância, representada por Var(X) , σ_X^2 ou σ^2, dada por

$$\boxed{\sigma^2 = \mathrm{Var}(X) = E((X-\mu)^2)}$$

ou pelo desvio padrão, σ (a raiz quadrada da variância):

$$\boxed{\sigma = \sqrt{\mathrm{Var}(X)}}$$

No exemplo anterior,

$$\boxed{W : \begin{cases} 0 \\ 1 \end{cases}} \qquad \boxed{X : \begin{cases} -1 & 1 \\ 0.5 & 0.5 \end{cases}} \qquad \boxed{Y : \begin{cases} -10 & 10 \\ 0.5 & 0.5 \end{cases}}$$

$$\sigma_W^2 = 0 \qquad \sigma_X^2 = (-1)^2 \tfrac{1}{2} + 1^2 \tfrac{1}{2} = 1 \qquad \sigma_Y^2 = (-10)^2 \tfrac{1}{2} + 10^2 \tfrac{1}{2} = 100$$

$$\sigma_W = 0 \qquad \sigma_X = 1 \qquad \sigma_Y = 10$$

Características teóricas (v.a. discretas)

nome	símbolo	fórmula	
valor médio	μ	$E(X) = \sum_i x_i p_i$	se a série convergir absolutamente
variância	σ^2	$\mathrm{Var}(X) = E((X-\mu)^2) = \sum_i (x_i - \mu)^2 p_i$	
desvio padrão	σ	$\sqrt{\mathrm{Var}(X)}$	
moda(s)		valor(es) x_i correspondente(s) ao maior p_i	
$\vdots$			

Nota: Nem sempre o valor médio existe. No caso de X ter fmp

$$p_i = P(X = 2^i) = \tfrac{1}{2^i}, \quad i = 1, 2, \ldots$$

temos $E(X)=1+1+1+1+\ldots=+\infty$. E a variância também não existe.

Exemplo 13: (i) cálculo de valores médios

$\mathrm{U}\{1, 2, \ldots, n\}$: $\mu = E(X) = (1+2+\ldots+n)/n = (n+1)/2$

$\mathrm{bi}(1, p)$: $\mu = E(X) = 0 \times (1-p) + 1 \times p = p$

$\mathrm{Poisson}(\lambda)$: $\mu = E(X) = \sum_{j \geq 0} j\, e^{-\lambda} \frac{\lambda^j}{j!} = \lambda \sum_{j \geq 1} e^{-\lambda} \frac{\lambda^{j-1}}{(j-1)!} = \lambda$

Propriedades (valor médio, variância e desvio padrão):

$$\mathrm{var}(X) \geq 0 \qquad E(a+bX) = a + bE(X)$$ → Linearidade do valor médio

$$\mathrm{var}(X) = E(X^2) - \mu^2 \qquad \mathrm{var}(a+bX) = b^2\,\mathrm{var}(X) \qquad \sigma_{a+bX} = |b|\,\sigma_X$$

Fáceis de provar, por exemplo:

$$E(a+bX) = \sum_i (a + b\,x_i)p_i = \sum_i (a\,p_i + b\,x_i p_i) = a\sum_i p_i + b\sum_i x_i p_i = a + bE(X)$$

$$\mathrm{Var}(a+bX) = E((a+bX-a-b\mu)^2) = E((bX-b\mu)^2) = E(b^2(X-\mu)^2) = b^2E((X-\mu)^2) = b^2\mathrm{Var}(X)$$

Exemplo 12 (cont.): Calculando a variância pela fórmula acima, no caso $X \sim U\{1, 2, \ldots, 6\}$ temos

$$\mathrm{Var}(X) = E(X^2) - E^2(X) = \tfrac{91}{6} - \tfrac{7^2}{2^2} = \tfrac{35}{12}$$

Exemplo 13: (ii) cálculo de variâncias

U$\{1, 2, \ldots, n\}$:

$$\begin{cases} E(X^2) = \frac{1}{n}(1^2 + 2^2 + \ldots + n^2) = \frac{1}{6}(n+1)(2n+1) \\ \therefore\ \sigma^2 = \mathrm{Var}(X) = E\left(X^2\right) - E^2(X) = \frac{1}{12}(n^2-1) \end{cases}$$

bi$(1, p)$:

$$E(X^2) = 0^2(1-p) + 1^2 p = p \ \therefore\ \sigma^2 = p - p^2 = p(1-p)$$

Poisson(λ):

$$\begin{cases} E(X(X-1)) = \sum_{j\geq 0} j(j-1)\, e^{-\lambda} \dfrac{\lambda^j}{j!} = \sum_{j\geq 2} \lambda^2\, e^{-\lambda} \dfrac{\lambda^{j-2}}{(j-2)!} = \lambda^2 \\ \therefore \sigma^2 = E(X(X-1) + X) - E^2(X) = \lambda^2 + \lambda - \lambda^2 = \lambda \end{cases}$$

- O valor médio (tal como a moda) é uma medida de localização; acompanha as mudanças de localização de X:

$$E(a + X) = a + E(X)$$

- A variância (tal como o desvio padrão) é uma medida de dispersão; mede a dispersão em torno do valor médio μ; é invariante para mudanças de localização:

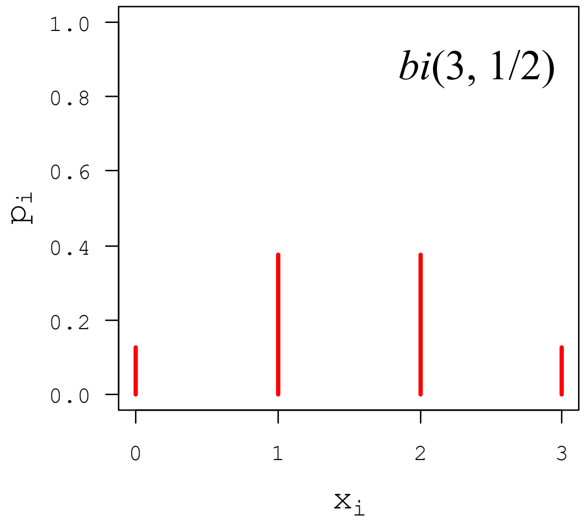
$$\mathrm{Var}(a + X) = \mathrm{Var}(X)$$

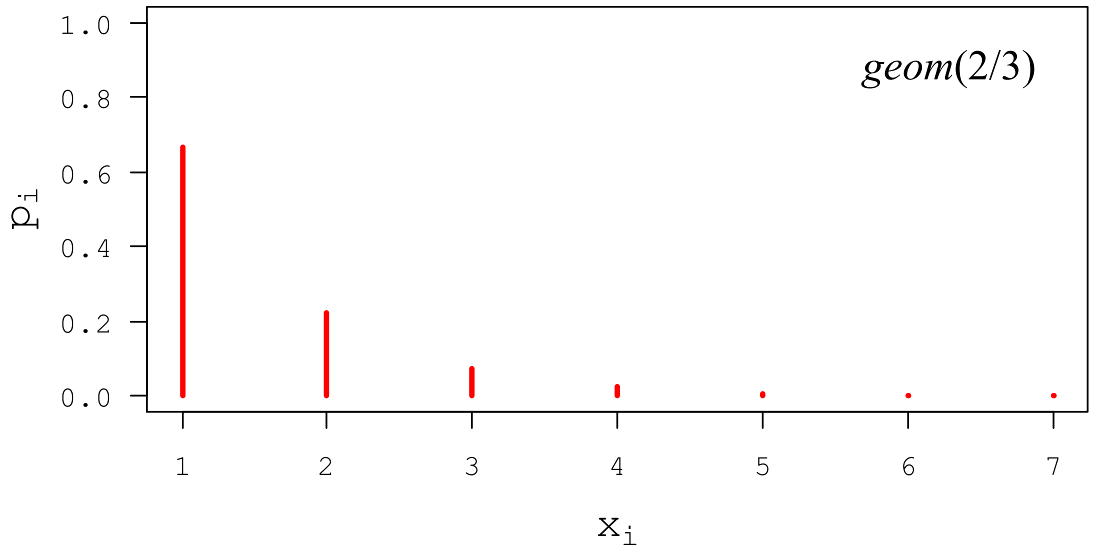
Nota: Há outras medidas de localização e de dispersão; por exemplo o desvio absoluto médio, $E(\,|X - E(X)|\,)$

O valor médio e desvio padrão não identificam a v.a.

Duas v.a. podem ter o mesmo valor médio e variância mas diferirem quanto à forma, por exemplo, uma pode ser mais assimétrica do que a outra. Há várias medidas de assimetria

As leis $bi(3, 0.5)$ e $geom(2/3)$ têm ambas valor médio 3/2 e variância 3/4, mas diferem na assimetria



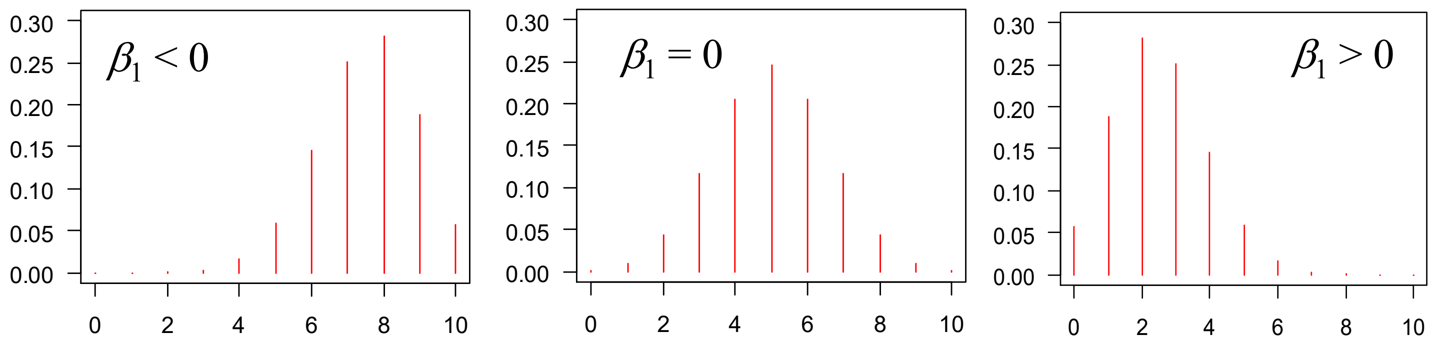


Coeficiente de assimetria β_1

- O coeficiente β_1 é uma medida de assimetria, definido por

$$\boxed{\beta_1 = E((X-\mu)^3/\sigma^3) = \frac{1}{\sigma^3}\sum_i (x_i-\mu)^3 p_i}$$

Este coeficiente é invariante para mudanças de localização e escala, i.e., X e $a+bX$ $(b>0)$ têm o mesmo β_1



Nota: X simétrica $\Rightarrow$ $\beta_1 = 0$, mas a recíproca não é verdadeira (vd. exercício nº 37)

Distribuições discretas – formulário

modelo	parâmetros	v. médio μ	variância σ^2	assimetria β_1
$bi(n,p)$	$0<p<1$ $n \in \mathbb{N}$	np	$np(1-p)$	$\dfrac{1-2p}{\sqrt{np(1-p)}}$
$geom(p)$	$0<p<1$	$\dfrac{1}{p}$	$\dfrac{1-p}{p^2}$	$\dfrac{2-p}{\sqrt{1-p}}$
$HG(n, N, p)$	$0<p<1$ $n \leq N$ $N \in \mathbb{N}$	np	$np(1-p)\dfrac{N-n}{N-1}$	$\cdots$
$Poisson(\lambda)$	$\lambda > 0$	λ	λ	$\dfrac{1}{\sqrt{\lambda}}$
$U\{1, \ldots, n\}$	$n \in \mathbb{N}$ $n > 1$	$\dfrac{n+1}{2}$	$\dfrac{n^2-1}{12}$	0

Estudo de duas v.a. em simultâneo (par aleatório)

Exemplo 14: Em 3 lançamentos de uma moeda equilibrada, calcular a lei de probabilidade de

- X – número de "caras"
- Y – nº de mudanças de face
- X e Y em simultâneo, i.e., do par aleatório (X,Y)

Ω = {CCC, CCE, CEC, CEE, …}

$$(X,Y) : \Omega \rightarrow \mathbb{R}^2$$

CCC $\rightarrow$ (3,0)

CCE $\rightarrow$ (2,1)

$\vdots$

suporte de (X,Y) : {………} $\subset \mathbb{R}^2$

Se o suporte do par aleatório for um conjunto finito ou numerável, diz-se que o par é discreto

Exemplo 14 (cont.):

	X	Y
CCC	3	0
CCE	2	1
CEC	2	2
CEE	1	1
ECC	2	1
ECE	1	2
EEC	1	1
EEE	0	0

X \ Y	0	1	2
0			
1			
2	0	2/8	
3	1/8	0	0

A estudar:

- Distribuições marginais de X e de Y
- Distribuições condicionais
- Dependência entre X e Y

Exemplo 14 (cont.):

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

$$\boxed{X : \begin{cases} 0 & 1 & 2 & 3 \\ \frac{1}{8} & \frac{3}{8} & \frac{3}{8} & \frac{1}{8} \end{cases}}$$

$$\boxed{Y : \begin{cases} 0 & 1 & 2 \\ \frac{1}{4} & \frac{2}{4} & \frac{1}{4} \end{cases}}$$

Distribuição marginal de X

$$P(X=2) = \sum_{j=0}^{2} P(X=2, Y=j) = 0 + \tfrac{2}{8} + \tfrac{1}{8} = \tfrac{3}{8}$$

$$P(X=i) = \sum_{j=0}^{2} P(X=i, Y=j),\ i = 0,1,2,3$$

$$X \frown bi(3, 1/2)$$

Distribuição marginal de Y

$$Y \frown bi(2, 1/2)$$

Exemplo 14 (cont.):

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

$X \backslash Y$	y_1	y_2	y_3	
x_1	p_{11}	p_{12}	p_{13}	$p_{1\bullet}$
x_2	p_{21}	p_{22}	p_{23}	$p_{2\bullet}$
x_3	p_{31}	p_{32}	p_{33}	$p_{3\bullet}$
x_4	p_{41}	p_{42}	p_{43}	$p_{4\bullet}$
	$p_{\bullet 1}$	$p_{\bullet 2}$	$p_{\bullet 3}$	1

Distribuição condicional de X dado $Y = y$ (para cada y fixo):

$$P(X=i \mid Y=2) = \frac{P(X=i, Y=2)}{P(Y=2)} = \begin{cases} 0 & i = 0 \ ou \ 3 \\ 1/2 & i = 1 \\ 1/2 & i = 2 \end{cases} \qquad \therefore \ X \mid \{Y=2\} \frown \mathrm{U}\{1, 2\}$$

Fórmula geral: $$\boxed{P(X = x_i \mid Y = y_j) = \frac{P(X = x_i, Y = y_j)}{P(Y = y_j)} = \frac{p_{ij}}{p_{\bullet j}}, \ i = 1, 2, 3, 4}$$

Distribuição condicional de Y dado $X = x$ (para cada x fixo): definição análoga

Distribuição conjunta do par (X,Y)

A fmp conjunta de um par aleatório discreto $(X,\ Y)$ pode, geralmente, ser representada por uma tabela com as probabilidades conjuntas, dadas por $p_{ij} = P(X = x_i\ , Y = y_j)$, na seguinte forma

$X \backslash Y$	y_1	y_2	y_3	$\cdots$	
x_1	p_{11}	p_{12}	p_{13}	$\cdots$	$p_{1\bullet}$
x_2	p_{21}	p_{22}	p_{23}	$\cdots$	$p_{2\bullet}$
x_3	p_{31}	p_{32}	p_{33}	$\cdots$	$p_{3\bullet}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$		$\vdots$
	$p_{\bullet 1}$	$p_{\bullet 2}$	$p_{\bullet 3}$	$\cdots$	1

$$p_{ij} \geq 0$$

$$\sum_{i,j} p_{ij} = 1$$

Distribuições marginais de X e Y

$X \backslash Y$	y_1	y_2	y_3	$\cdots$	
x_1	p_{11}	p_{12}	p_{13}	$\cdots$	$p_{1\bullet}$
x_2	p_{21}	p_{22}	p_{23}	$\cdots$	$p_{2\bullet}$
x_3	p_{31}	p_{32}	p_{33}	$\cdots$	$p_{3\bullet}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$		$\vdots$
	$p_{\bullet 1}$	$p_{\bullet 2}$	$p_{\bullet 3}$	$\cdots$	1

Distribuição marginal de X:

fmp $$\boxed{p_{i\bullet} = P(X = x_i) = \sum_j P(X = x_i, Y = y_j) = \sum_j p_{ij}, \; i = 1, 2, \ldots}$$

Distribuição marginal de Y:

fmp $$\boxed{p_{\bullet j} = P(Y = y_j) = \sum_i P(X = x_i, Y = y_j) = \sum_i p_{ij}, \; j = 1, 2, \ldots}$$

Distribuições condicionais

$X \backslash Y$	y_1	y_2	y_3	$\cdots$	
x_1	p_{11}	p_{12}	p_{13}	$\cdots$	$p_{1\bullet}$
x_2	p_{21}	p_{22}	p_{23}	$\cdots$	$p_{2\bullet}$
x_3	p_{31}	p_{32}	p_{33}	$\cdots$	$p_{3\bullet}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$		$\vdots$
	$p_{\bullet 1}$	$p_{\bullet 2}$	$p_{\bullet 3}$	$\cdots$	1

Distribuição condicional de X dado $Y = y_j$ (para um y_j fixo):

fmp $$\boxed{P(X = x_i \mid Y = y_j) = \frac{P(X = x_i, Y = y_j)}{P(Y = y_j)} = \frac{p_{ij}}{p_{\bullet j}}, \quad i = 1, 2, \ldots}$$

Distribuição condicional de Y dado $X = x_i$ (para um x_i fixo):

fmp $$\boxed{P(Y = y_j \mid X = x_i) = \frac{P(X = x_i, Y = y_j)}{P(X = x_i)} = \frac{p_{ij}}{p_{i\bullet}}, \quad j = 1, 2, \ldots}$$

Par aleatório

é uma função $(X, Y) : \Omega \longrightarrow \mathbb{R}^2$, que a cada $\omega \in \Omega$ faz corresponder o par $(X(\omega), Y(\omega))$, tal que $(X,Y)^{-1}(B) \in \mathcal{A}$, para qualquer boreliano B de $\mathbb{R}^2$.

O par (X, Y) diz-se discreto se o seu suporte for finito ou numerável.

No caso discreto, o par aleatório fica identificado pela fmp conjunta ou pela fd conjunta, dadas por

- fmp conjunta: $p_{ij} = P(X = x_i \,,\, Y = y_j)$
- fd conjunta: $F(x,y) = P(X \leq x \,,\, Y \leq y)$

Vector aleatório

é uma função $(X_1, \ldots, X_n)$: $\Omega \to \mathbb{R}^n$, que a cada $\omega \in \Omega$ faz corresponder o vector $(X_1(\omega), \ldots, X_n(\omega))$, tal que $(X_1, \ldots, X_n)^{-1}(B) \in \mathcal{A}$, para qualquer boreliano B de $\mathbb{R}^n$, i.e., $B \in \mathcal{B}$.

O vector aleatório diz-se discreto se tiver suporte finito ou numerável.

No caso discreto, o vector aleatório fica identificado pela fmp conjunta ou pela fd conjunta, com fórmulas análogas às do par.

Exemplo 15: Em 120 lançamentos de um dado equilibrado, sendo X_i o nº de vezes que sai a face com i pintas, podemos considerar o vector aleatório $(X_1, \ldots, X_5)$.

Independência

As v.a. X e Y dizem-se independentes se

$$\boxed{P(X \in A, Y \in B) = P(X \in A)\ P(Y \in B),\ \forall_{A \subset \mathbb{R}\ ,\ B \subset \mathbb{R}\ ,\ A, B\ \in\ \mathcal{B}}}$$

No caso de um par aleatório discreto, esta condição equivale a ter

$$\boxed{p_{ij} = p_{i\bullet}\ p_{\bullet j}\,,\ \ \forall i,\ j}$$

As v.a. $X_1,\ \ldots,\ X_n$ dizem-se mutuamente independentes se

$$\boxed{P(X_1 \in A_1, \ldots, X_n \in A_n) = P(X_1 \in A_1)\ \ldots\ P(X_n \in A_n),\ \ \forall_{A_i\ \in\ \mathcal{B},\ i=1,2,\ldots,n}}$$

No caso discreto, equivale a ter a fmp conjunta igual ao produto das fmp marginais

Propriedades 1

Se as v.a. $X_1, \ldots, X_n$ forem mutuamente independentes, também o são $h_1(X_1), \ldots, h_n(X_n)$.*

$X_1, \ldots, X_n$
mutuamente
independentes

Demonstração:

Para quaisquer borelianos $B_1, \ldots, B_n$ de $\mathbb{R}$ temos

$$P\big(h_1(X_1) \in B_1, \ \ldots, h_n(X_n) \in B_n\big) = P\Big(X_1 \in h_1^{-1}(B_1), \ \ldots, X_n \in h_n^{-1}(B_n)\Big) =$$

$$= P\Big(X_1 \in h_1^{-1}(B_1)\Big) \ldots P\Big(X_n \in h_n^{-1}(B_n)\Big) = P\big(h_1(X_1) \in B_1\big) \ldots P\big(h_n(X_n) \in B_n\big)$$

donde $h_1(X_1), \ldots, h_n(X_n)$ são mutuamente independentes.

---

* Estas $h_1, \ldots, h_n$ são funções Borel-mensuráveis, i.e., a imagem inversa de um boreliano de $\mathbb{R}$ terá que ser um acontecimento do espaço de acontecimentos $(\Omega, \mathcal{A})$

Valor médio de funções de um par aleatório

No caso de um par (X, Y) discreto, com fmp $p_{ij} = P(X{=}x_i\,,\ Y{=}y_j)$, define-se valor médio de $h(X, Y)$ por

$$\boxed{E(h(X,Y)) = \sum_i \sum_j h(x_i, y_j)\, p_{ij}}$$

(desde que a série seja absolutamente convergente)

Em particular,

$$E(X+Y) = E(X) + E(Y)$$

Demonstração:

$$E(X+Y) = \sum_i \sum_j (x_i + y_j) p_{ij} = \sum_i \sum_j x_i p_{ij} + \sum_i \sum_j y_j p_{ij} = \sum_i x_i \sum_j p_{ij} + \sum_j y_j \sum_i p_{ij} = \sum_i x_i p_{i\bullet} + \sum_j y_j p_{\bullet j} = E(X) + E(Y)$$

Propriedades 2

Dado um vector aleatório $(X_1, \ldots, X_n)$, temos (caso existam os valores médios)

2.1 $$E(X_1 + X_2 + \ldots + X_n) = E(X_1) + E(X_2) + \ldots + E(X_n)$$

2.2 $$E(a_1 X_1 + a_2 X_2 + \ldots + a_n X_n) = a_1 E(X_1) + a_2 E(X_2) + \ldots + a_n E(X_n)$$

Propriedades 3

Dado um vector aleatório $(X_1, \ldots, X_n)$, temos que se $X_1, \ldots, X_n$ forem mutuamente independentes, então

3.1 $$E(X_1 X_2 \ldots X_n) = E(X_1)E(X_2)\ldots E(X_n)$$

Demonstração ($n = 2$): $E(XY) = \sum_i \sum_j x_i y_j p_{ij} = \sum_i \sum_j x_i y_j p_{i\bullet} p_{\bullet j} = \sum_i x_i p_{i\bullet} \sum_j y_j p_{\bullet j} = E(X)E(Y)$

3.2 $$E\big(h_1(X_1)h_2(X_2)\ldots h_n(X_n)\big) = E\big(h_1(X_1)\big)E\big(h_2(X_2)\big)\ldots E\big(h_n(X_n)\big)$$

3.3 $$\mathrm{var}(X_1 + X_2 + \ldots + X_n) = \mathrm{var}(X_1) + \mathrm{var}(X_2) + \ldots + \mathrm{var}(X_n)$$

$h_1, \ldots, h_n$ são funções Borel-mensuráveis

Exemplo 16: Valor médio e variância de $X \frown \mathrm{bi}(n, p)$.

X representa o "número de caras em n lançamentos (independentes) de uma moeda equilibrada", logo é a soma dos nºs de caras em cada lançamento,

$$\boxed{X = \sum_{i=1}^{n} Y_i}$$

Estes Y_i = "nº de caras no i-ésimo lançamento" são independentes, $Y_i \frown \mathrm{bi}(1, p)$. Aplicando as propriedades sobre o valor médio da soma e sobre a variância da soma de v.a. independentes, temos então

$$E(X) = E(Y_1) + \ldots + E(Y_n) = p + \ldots + p = n\,p$$

$$\mathrm{var}(X) = \mathrm{var}(Y_1) + \ldots + \mathrm{var}(Y_n) = p\,(1-p) + \ldots + p\,(1-p) = n\,p\,(1-p)$$

e ainda $\sigma_X = \sqrt{np(1-p)}$

Covariância entre duas v.a.

A covariância entre X e Y, representada por $\operatorname{cov}(X, Y)$ define-se por

$$\boxed{\operatorname{cov}(X,Y) = E\big((X-\mu_X)(Y-\mu_Y)\big)}$$

caso este valor médio exista. Note-se que $\operatorname{cov}(X,X) = \operatorname{var}(X)$.

Propriedades 4

4.1 $$\operatorname{cov}(X,Y) = E(XY) - \mu_X \mu_Y$$

4.2 X e Y independentes $\underset{E(XY)=E(X)\,E(Y)}{\Rightarrow}$ $\operatorname{cov}(X,Y) = 0$

4.3 $$\operatorname{cov}(aX+bY,Z) = a\operatorname{cov}(X,Z) + b\operatorname{cov}(Y,Z)$$

4.5 $$\mathrm{var}(X_1 + X_2 + \ldots + X_n) = \sum_{i=1}^{n} \mathrm{var}(X_i) + 2\sum_{i<j} \mathrm{cov}(X_i, X_j)$$

Demonstração:

Considere-se $X_i^* = X_i - E(X_i)$ e recorde-se que $E\left(X_i^*\right) = 0$ e $\mathrm{var}\left(X_i^*\right) = \mathrm{var}\left(X_i\right)$. Então

$$\mathrm{var}\left(\sum_{i=1}^{n} X_i\right) = \mathrm{var}\left(\sum_{i=1}^{n} X_i^*\right) = E\left(\left(X_1^* + \ldots + X_n^*\right)^2\right) = E\left(\sum_{i=1}^{n} X_i^{*2}\right) + E\left(\sum_{i\neq j} X_i^* X_j^*\right) =$$

$$= \sum_{i=1}^{n} E\left(X_i^{*2}\right) + \sum_{i\neq j} E\left(X_i^* X_j^*\right) = \sum_{i=1}^{n} \mathrm{var}\left(X_i\right) + \sum_{i\neq j} \mathrm{cov}\left(X_i, X_j\right)$$

4.6 $$\mathrm{var}(X + Y) = \mathrm{var}(X) + \mathrm{var}(Y) + 2\,\mathrm{cov}(X, Y)$$

$$\mathrm{var}(X - Y) = \mathrm{var}(X) + \mathrm{var}(Y) - 2\,\mathrm{cov}(X, Y)$$

Desigualdade de Cauchy-Schwarz: $|E(XY)| \leq \sqrt{E(X^2)E(Y^2)}$

e a igualdade dá-se se e só se existirem constantes a e b tais que $P(Y = a + bX) = 1$

Coeficiente de correlação entre duas v.a.

A correlação entre X e Y, representada por $\rho_{X,Y}$ ou ρ define-se por

$$\boxed{\rho = \rho(X,Y) = E\left(\frac{X-\mu_X}{\sigma_X}\,\frac{Y-\mu_Y}{\sigma_Y}\right)} \quad \text{ou seja} \quad \boxed{\rho = \frac{\operatorname{cov}(X,Y)}{\sigma_X\sigma_Y}}$$

(caso exista este valor médio). Prova-se que

- $-1 \le \rho \le 1$
- X e Y independentes $\Rightarrow \rho = 0$ (a recíproca não é verdadeira)
- $\rho = \pm 1 \Leftrightarrow P(Y = a + bX) = 1$ para algum $a \in \mathbb{R}$ e $b \neq 0$
- é invariante para transformações lineares de X e de Y (a menos do sinal)

ρ mede a relação de linearidade entre X e Y

Exemplo 17:

$\boxed{\rho = 0 \;\;\not\Rightarrow\;\; X \text{ e } Y \text{ independentes}}$

$\boxed{X \text{ e } Y \text{ independentes} \;\;\Rightarrow\;\; \rho = 0}$

(i) $X \frown U\{-2,-1,1,2\}$ e $Y=X^2$ não são independentes, mas $\rho = 0$ (a relação entre X e Y é forte, mas é não linear):

X \ Y	1	4	
-2	0	1/4	1/4
-1	1/4	0	1/4
1	1/4	0	1/4
2	0	1/4	1/4
	1/2	1/2	1

$$\text{cov}(X, X^2) = E(X^3) - E(X)E(X^2) = 0$$

(note que X é uma v.a. simétrica...)

(ii) Defina-se um par (X, Y) com distribuições marginais iguais às de (i), mas com X e Y independentes:

X \ Y	1	4	
-2	1/8	1/8	1/4
-1	1/8	1/8	1/4
1	1/8	1/8	1/4
2	1/8	1/8	1/4
	1/2	1/2	1

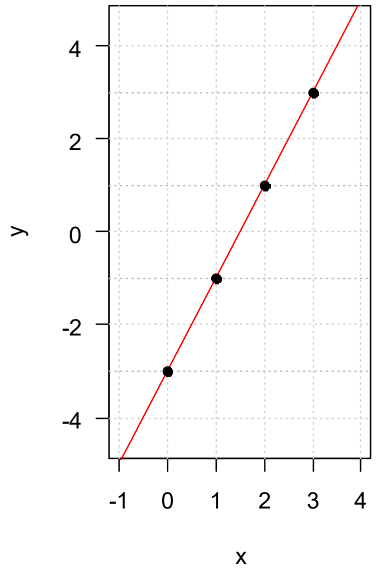
Exemplo 18: Em três lançamentos de uma moeda equilibrada, calcular a lei de probabilidade de (X, Y), sendo $X =$ nº caras $Y =$ nº caras – nº coroas

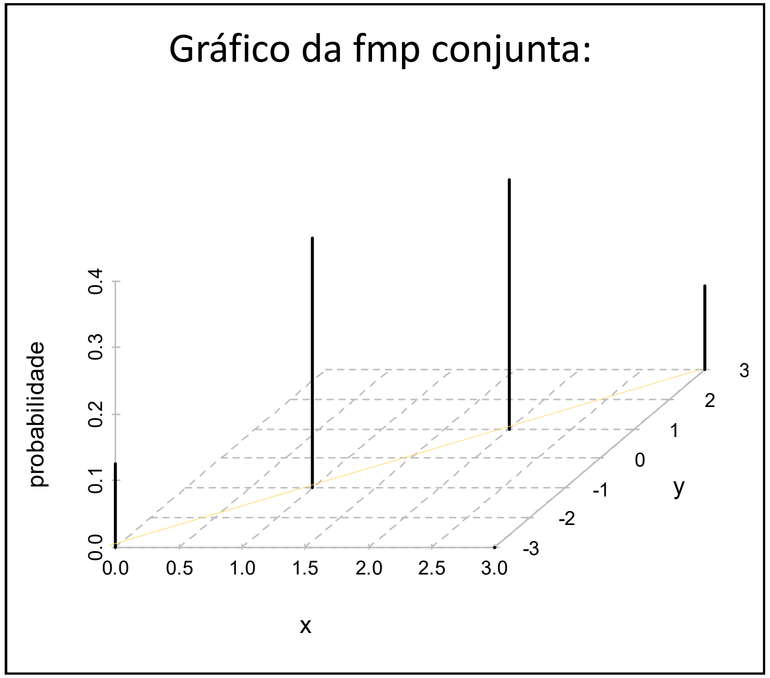
	X	Y
CCC	3	3
CCE	2	1
CEC	2	1
CEE	1	-1
ECC	2	1
ECE	1	-1
EEC	1	-1
EEE	0	-3

X \ Y	-3	-1	1	3
0	1/8	0	0	0
1	0	3/8	0	0
2	0	0	3/8	0
3	0	0	0	1/8

$$X \frown \mathrm{bi}(3,\ 1/2)$$

$$Y : \left\{ \begin{array}{cccc} -3 & -1 & 1 & 3 \\ 1/8 & 3/8 & 3/8 & 1/8 \end{array} \right.$$





Note que $Y = X - (3 - X) = 2X - 3$, donde a distribuição do par (X, Y) está concentrada na recta $y = 2x - 3$ e a correlação é igual a 1 (declive positivo; X e Y variam no mesmo sentido).

Exemplo 18 (cont.): calcule-se a correlação

$X \backslash Y$	-3	-1	1	3
0	1/8	0	0	0
1	0	3/8	0	0
2	0	0	3/8	0
3	0	0	0	1/8

$X \frown \text{bi}(3,\ 1/2) \quad \Rightarrow \quad E(X) = 3/2,\ \text{var}(X) = 3/4$

$Y: \begin{cases} -3 & -1 & 1 & 3 \\ 1/8 & 3/8 & 3/8 & 1/8 \end{cases} \quad \Rightarrow \quad E(Y) = 0$ e

$\text{var}(Y) = E(Y^2) = \frac{6}{8} + 3^2\,\frac{2}{8} = 3$

$$XY: \begin{cases} 0 & -1 & 2 & 9 \\ \frac{1}{8} & \frac{3}{8} & \frac{3}{8} & \frac{1}{8} \end{cases} \quad \Rightarrow E(XY) = -\frac{3}{8} + \frac{6}{8} + \frac{9}{8} = \frac{3}{2} \ \Rightarrow \text{cov}(X,Y) = E(XY) - 0 = \frac{3}{2}$$

donde $\rho = \dfrac{\text{cov}(X,Y)}{\sigma_X \sigma_Y} = \dfrac{3/2}{\sqrt{\frac{3}{4}}\sqrt{3}} = 1$

Exercício: Prove que a correlação é invariante pª transformações lineares das 2 v.a., a menos do sinal

Resolução: (a correlação é invariante para transformações lineares das v.a., a menos do sinal)

Recorde-se que $\mu_{a+bX} = a + b\mu_X$ e $\sigma_{a+bX} = |b|\,\sigma_X$

Então

$$\rho_{a+bX,\; c+dY} = \frac{\operatorname{cov}(a+bX, c+dY)}{\sigma_{a+bX}\sigma_{c+dY}} = \frac{E\big((a+bX-\mu_{a+bX})(c+dY-\mu_{c+dY})\big)}{|b|\,\sigma_X \;\, |d|\,\sigma_Y} =$$

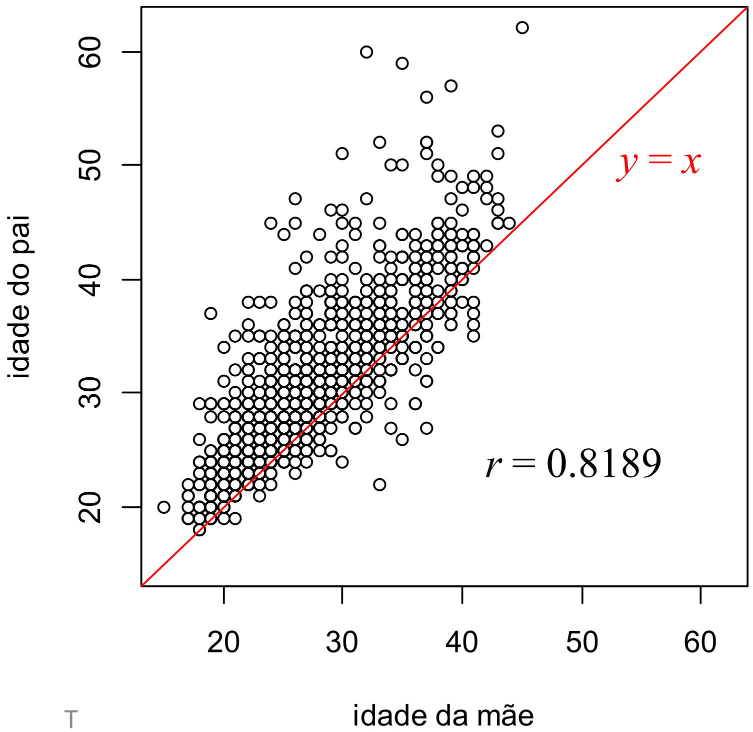
$$= \frac{E\big(bX-b\mu_X)(dY-d\mu_Y)\big)}{|bd|\,\sigma_X\sigma_Y} = \frac{bd}{|bd|}\;\frac{E\big((X-\mu_X)(Y-\mu_Y)\big)}{\sigma_X\sigma_Y} = \pm\,\rho_{X,Y}$$

Nota: Os <u>coeficientes</u>, e.g. de correlação, de assimetria, etc., são escalares (não têm unidade de medida)

Na prática, com amostras bivariadas $(x_i\,, y_i)$ ...

... temos o coeficiente de correlação amostral, r, dado por fórmula empírica análoga à teórica (valores médios são substituídos por médias amostrais):

Idade da mãe e do pai numa amostra aleatória de 1214 recém-nascidos



$$r = r_{x,y} = \frac{1}{n-1}\sum_{i=1}^{n} \frac{x_i - \bar{x}}{s_x}\,\frac{y_i - \bar{y}}{s_y}$$

r é uma estimativa de ρ

As características teóricas/populacionais (valor médio μ, desvio padrão σ, correlação ρ, etc.) podem ser estimadas pelas correspondentes características empíricas/amostrais (média $\bar{x}$, desvio padrão s, correlação amostral r, etc.)

Das distribuições discretas multivariadas destaca-se a distribuição multinomial, que tem um papel relevante em Estatística. Trata-se de uma generalização da distribuição binomial.

A distribuição multinomial aplica-se em problemas que envolvem a classificação de dados em categorias ou níveis ou classes; por exemplo, classes etárias, classes sociais, económicas, de comportamento, etc.; tal como em aprendizagem automática (Machine Learning).

Bibliografia adicional: Forsyth, D. (2018). Probability and Statistics for Computer Science. Springer.

- De novo a distribuição binomial (2 categorias: sucesso e insucesso)

$$\boxed{P(X=i)=\binom{n}{i}p^i(1-p)^{n-i},\ i=0,1,\ldots,n} \quad X \frown \mathrm{bi}(n,p)$$

X = nº sucessos em n provas de Bernoulli, mutuamente independentes

p = probabilidade de sucesso em cada prova

$q=1-p$ = probabilidade de insucesso

$$\boxed{1=(p+q)^n=\sum_{i=0}^{n}\underbrace{\binom{n}{i}}_{\text{coeficientes binomiais}}p^i q^{n-i}}$$ (binómio de Newton)

Por exemplo, X = "nº de caras em n lançamentos de uma moeda-p"

E se em cada prova houver diferentes tipos de sucesso? Por exemplo, ao lançar um dado, podemos considerar 6 categorias: 5 tipos de sucesso (face 1, …, face 5) e insucesso (face 6).

- Distribuição trinomial (bivariada ≡ par aleatório); 3 categorias (mutuamente exclusivas e exaustivas)

$$\boxed{P(X=i,Y=j)=\frac{n!}{i!\,j!\,(n-i-j)!}\,p_1^{\,i}\,p_2^{\,j}\,q^{\,n-i-j}\,,\quad i+j=0,1,\ldots,n}$$

X = nº sucessos de tipo 1, Y = nº sucessos de tipo 2

p_1 = probabilidade de sucesso de tipo 1
p_2 = probabilidade de sucesso de tipo 2
$q = 1-p_1-p_2$ = probabilidade de insucesso

$$\boxed{1=(p_1+p_2+q)^n=\sum_{i,j}\underbrace{\frac{n!}{i!\,j!\,(n-i-j)!}}_{\text{coeficientes trinomiais}}\,p_1^{\,i}\,p_2^{\,j}\,q^{\,n-i-j}}$$ (trinómio)

Exemplo 19: X = "nº faces 1", Y = "nº de faces 2", em n lançamentos de um dado equilibrado ($p_1=p_2=1/6$, $q=4/6$)

- Distribuição multinomial, $(X_1 , X_2 , \ldots, X_{r\text{-}1}) \frown \mathrm{M}(n ; p_1 , p_2 , \ldots, p_{r\text{-}1})$

Experiência aleatória com espaço de resultados Ω (decomposto em r categorias, $A_1 , \ldots , A_r$ mutuamente exclusivas e exaustivas); n repetições (independentes)

A_1	sucesso de tipo 1,	$P(A_1) = p_1$
A_2	sucesso de tipo 2,	$P(A_2) = p_2$
$\vdots$	$\vdots$	$\vdots$
$A_{r\text{-}1}$	sucesso de tipo $r-1$,	$P(A_{r\text{-}1}) = p_{r\text{-}1}$
A_r	insucesso,	$P(A_r) = p_r = 1 - p_1 - \ldots - p_{r\text{-}1}$

X_k = nº sucessos de tipo k , $p_k = P(A_k)$, $k = 1, 2, \ldots, r-1$

$q = P(A_r) = p_r = 1 - p_1 - \ldots - p_{r\text{-}1}$

Qual a distribuição conjunta do vetor aleatório $(X_1 , \ldots , X_{r\text{-}1})$?

Por exemplo, X_1 = "nº faces 1", X_2 = "nº faces 2", ..., X_5 = "nº de faces 5", em n lançamentos de um dado.
Qual a distribuição de $(X_1, X_2, \ldots, X_5)$?

$(X_1, \ldots, X_{r\text{-}1}) \frown \mathrm{M}(n\,; p_1, p_2, \ldots, p_{r\text{-}1})$ (vector aleatório de dimensão $r-1$)

$$\boxed{P(X_1 = n_1, \ldots, X_{r-1} = n_{r-1}) = \frac{n!}{n_1! \ldots n_{r-1}!\, n_r!} p_1^{n_1} p_2^{n_2} \ldots p_{r-1}^{n_{r-1}} q^{n_r}, \quad n_1 + \ldots + n_{r-1} = 0, 1, \ldots, n}$$

X_k = nº sucessos de tipo k , $k = 1, 2, \ldots, r-1$

$p_k = P$(sucesso de tipo k), $q = 1 - p_1 - \ldots - p_{r\text{-}1} = P$(insucesso)

$$\boxed{1 = (p_1 + \ldots + p_{r-1} + q)^n = \sum_{n_1, \ldots, n_{r-1}} \underbrace{\frac{n!}{n_1! \ldots n_{r-1}!\, n_r!}}_{\text{coeficientes multinomiais}} p_1^{n_1} \ldots p_{r-1}^{n_{r-1}} q^{n_r}}$$

$n = n_1 + \ldots + n_r$

Exemplo 20: X_1 = "nº faces 1", ..., X_5 = "nº faces 5", em n lançamentos de um dado equilibrado.

Temos $(X_1, \ldots, X_5) \frown \mathrm{M}(n\,; 1/6, 1/6, 1/6, 1/6, 1/6)$

Propriedades (da multinomial)

$$X_k \frown \text{bi}(\, n \,,\, p_k \,)$$

$$E(X_k) = n\, p_k$$

$$\text{Var}(X_k) = n\, p_k\, (1 - p_k)$$

$$\text{cov}(X_i\,,\, X_j) = -\, n\, p_i\, p_j \;\; \color{red}{< 0}$$

$$\rho_{X_i, X_j} = \frac{\text{cov}(X_i, Y_j)}{\sigma_X \sigma_Y} = -\frac{p_i p_j}{\sqrt{(1 - p_i)(1 - p_j)}}$$

Simulação e cálculo de probabilidades para uma lei multinomial

- Para simular uma lei multinomial, usa-se `rmultinom(x, size, prob)`

Por exemplo, numa lotaria de 10^5 bilhetes, numerados de 0000 a 9999 (temos 4 dígitos de 0 a 9, extraídos com reposição), considera-se X = "nº dígitos ímpares" e Y = "nº de zeros", ou seja, temos o par trinomial $(X, Y) \frown M(4, 0.5, 0.1)$, pois $p_1 = 0.5$, $p_2 = 0.1$, $q = 0.4$. Para obter 13 simulações deste par:

```
rmultinom(13, size = 4, prob = c(5,1,4))
     [,1] [,2] [,3] [,4] [,5] [,6] [,7] [,8] [,9] [,10] [,11] [,12] [,13]
[1,]    2    2    3    2    1    2    2    4    1     2     2     2     2
[2,]    0    0    1    0    0    1    1    0    0     1     0     0     0
[3,]    2    2    0    2    3    1    1    0    3     1     2     2     2
```

- Para calcular probabilidades, usa-se `dmultinom(x, prob)`

Por exemplo, para calcular $P(X=2, Y=1)$, com $(X, Y) \frown M(4; 0.5, 0.1)$:

```
dmultinom(c(2,1,1),prob = c(5,1,4))
[1] 0.12
```

Exercícios resolvidos

Nota: (sobre o cálculo de probabilidades numa multinomial, no R)

No cálculo de probabilidades numa lei multinomial, `dmultinom(x, prob)` equivale a `dmultinom(x, size = sum(x), prob)`. Esta função não está vectorizada.

Exercício (nº 38) :

(a) (i) Calcule $P(X=2, Y=2)$, com $(X, Y) \frown \mathrm{M}(9; 0.3, 0.3)$.

(ii) Prove que X e Y são dependentes.

(b) Qual a probabilidade de em 12 lançamentos de um dado equilibrado

(i) sair 2 vezes cada face?

(ii) as faces 2 a 6 saírem em igual quantidade?

Resolução (nº 38) :

(a) (i) Cálculo de $P(X=2, Y=2)$, com $(X, Y) \frown M(9; 0.3, 0.3)$;

$n = 9; p_1 = p_2 = 0.3$; $q = 1 - 0.3 - 0.3 = 0.4$

$$\boxed{P(X=i, Y=j) = \frac{n!}{i!\,j!\,(n-i-j)!} p_1^{\,i} p_2^{\,j} q^{n-i-j}, \quad i+j = 0,1,\ldots,n}$$

$$P(X=2, Y=2) = \frac{9!}{2!\,2!\,5!} 0.3^2 0.3^2 0.4^5 \cong 0.0627$$

```
dmultinom(c(2,2,5),prob=c(0.3,0.3,0.4))
dmultinom(c(2,2,5),prob=c(3,3,4))
[1] 0.06270566
```

Por exemplo, dada uma população com 3 etnias, A, B e C, nas proporções 3:3:4, a probabilidade de numa amostra aleatória de 9 indivíduos dessa população haver dois da etnia A e dois de B (e 5 de C) é 0.0627

(ii) X e Y são dependentes pois $P(X=9, Y=9) = 0$, mas $P(X=9) > 0$ e $P(Y=9) > 0$

```
dmultinom(c(9,0,0),prob=c(3,3,4))
[1] 1.9683e-05
```

(b) (i) 12 lançamentos de um dado equilibrado; calcular P("sair 2 vezes cada face").

Temos 6 categorias; $n = 12$; $(X_1, \ldots, X_5) \frown \mathrm{M}(n\ ;\ 1/6, 1/6, 1/6, 1/6, 1/6)$

X_i = "nº de vezes que sai a face i"

$$P(X_1 = 2, \ldots, X_5 = 2) = \frac{12!}{2!\ 2!\ 2!\ 2!\ 2!\ 2!}\left(\frac{1}{6}\right)^2 \ldots \left(\frac{1}{6}\right)^2 = \frac{12!}{2^6}\left(\frac{1}{6}\right)^{12}$$

```
dmultinom(c(2,2,2,2,2,2), prob=c(1,1,1,1,1,1))
dmultinom(rep(2,6), prob=rep(1,6))
[1] 0.003438286
```

(ii) As faces 2 a 6 saem o mesmo nº de vezes sse esse nº de vezes é 0 ou 1 ou 2 (e correspondem a que a face 1 saia 12 vezes, 7 vezes e 2 vezes)

```
p <- rep(1,6)
dmultinom(c(12,rep(0,5)),prob=p)
      + dmultinom(c(7,rep(1,5)),prob=p)
      + dmultinom(rep(2,6),prob=p)
[1] 0.003481947
```

Exercício: **(nº 42)** Calcular a distribuição da soma de duas v.a., X e Y, independentes, nos casos (i) $bi(m,p)$ e $bi(n,p)$ (ii) $Poisson(\lambda)$ e $Poisson(\mu)$

Resolução:

(i) Como $X+Y$ representa o nº de sucessos em $m+n$ provas de Bernoulli(p) independentes, temos imediatamente que $X+Y \frown bi(m+n,\, p)$.

(ii) Cálculo de $P(X+Y=n)$, escrevendo o acontecimento $\{X+Y=n\}$ como união disjunta dos acontecimentos $\{X=i,\, Y=n-i\}$, com $i=0, 1, 2, \ldots, n$

$$P(X+Y=n)=\sum_{i=0}^{n} P(X=i,Y=n-i)=\sum_{i=0}^{n} P(X=i)P(Y=n-i)=$$

$$=\sum_{i=0}^{n} e^{-\lambda}\frac{\lambda^i}{i!}\, e^{-\mu}\frac{\mu^{n-i}}{(n-i)!}=e^{-\lambda-\mu}\sum_{i=0}^{n}\tfrac{1}{n!}\binom{n}{i}\lambda^i\mu^{n-i}=\tfrac{1}{n!}e^{-(\lambda+\mu)}(\lambda+\mu)^n,\quad n=0,1,2,\ldots$$

Logo $X+Y \frown Poisson(\lambda+\mu)$.

Exercício: (nº 43) Calcule $E\left(\dfrac{1}{X+1}\right)$ no caso $X \frown Poisson(\lambda)$. Particularize para $\lambda = 1$. Comente.

Resolução:

Para uma v.a. X, discreta, com fmp $p_i = P(X = x_i)$, o valor médio de $h(X)$ é dado por $E(h(X)) = \sum_i h(x_i)p_i$, desde que esta série convirja absolutamente. No caso presente, $h(X) = \frac{1}{1+X}$ com $X \frown Poisson(\lambda)$, a série é de termos positivos e é convergente:

$$E\left(\frac{1}{1+X}\right) = \sum_{i\geq 0} \frac{1}{1+i} e^{-\lambda} \frac{\lambda^i}{i!} = \frac{e^{-\lambda}}{\lambda} \sum_{i\geq 0} \frac{\lambda^{i+1}}{(i+1)!} = \frac{e^{-\lambda}}{\lambda} \sum_{j\geq 1} \frac{\lambda^j}{j!} = \frac{e^{-\lambda}}{\lambda}(e^{\lambda}-1) = \frac{1-e^{-\lambda}}{\lambda}.$$

No caso $\lambda = 1$ temos $E\left(\frac{1}{1+X}\right) = 1 - e^{-1} \approx 0.6321.$[4] E como $\frac{1}{1+E(X)} = \frac{1}{2}$, conclui-se que em geral não é válida a fórmula $E(h(X)) = h(E(X))$. Recorde-se no entanto que no caso particular $h(X) = a + bX$, a fórmula é válida.

[4]Este valor podia ter sido estimado por simulação, executando `mean(1/(1+rpois(10^7,1)))`

Exercício (Lotaria) : Numa lotaria com 10^5 bilhetes (numerados de 0000 a 9999),

(i) qual a probabilidade de sair um número com exatamente dois dígitos "ímpares" e um zero?

(ii) determine a fmp conjunta do par (X,Y), onde X é o nº de dígitos "ímpares" e Y é o nº de zeros do primeiro prémio. Calcule e identifique as distribuições marginais de X e Y.

(iii) qual a distribuição de $X+Y$?

(iv) qual a distribuição de XY ?

(v) calcule cov(X,Y) pela fórmula cov$(X,Y) = E(XY) - E(X)\,E(Y)$

Resolução (lotaria):

(i) Cálculo da probabilidade de numa lotaria com 10^5 bilhetes (nºs de 0000 a 9999) sair um número com exatamente dois dígitos "ímpares" e um zero.

Temos $n = 4$ extracções (com reposição) de uma urna com 10 bolas (dígitos de 0 a 9) e consideramos o par aleatório (X,Y) , sendo X = "nº de dígitos ímpares extraídos" e Y = "nº de zeros extraídos" ; $p_1 = ½$; $p_2 = 1/10$, ou seja,

$$(X,\ Y) \frown \mathrm{M}(\ n\ ;\ p_1\ ,\ p_2\ ) \equiv \mathrm{M}(\ 4\ ;\ 5/10\ ,\ 1/10\ )$$

$$P(X=2, Y=1) = \frac{4!}{2!\ 1!\ 1!}\left(\frac{5}{10}\right)^2\left(\frac{1}{10}\right)^1\left(\frac{4}{10}\right)^1 = \frac{4!}{2}\,\frac{100}{10000} = 0.12$$

```
dmultinom(c(2,1,1), prob=c(5,1,4))
[1] 0.12
```

(ii) Lotaria com bilhetes numerados de 0000 a 9999; cálculo da fmp conjunta do par (X,Y), X = nº de dígitos "ímpares" e Y = nº de zeros do primeiro prémio. Identificação das distribuições marginais de X e Y.

fmp conjunta do par (fórmula geral):

$$\boxed{P(X=i, Y=j) = \frac{n!}{i!\, j!\, (n-i-j)!}\, p_1^{\,i}\, p_2^{\,j}\, q^{\,n-i-j}, \quad i+j=0,1,\ldots,n}$$

Neste caso:

$$P(X=i, Y=j) = \frac{4!}{i!\;j!\,(4\text{-}i\text{-}j)!}\left(\frac{5}{10}\right)^{i}\left(\frac{1}{10}\right)^{j}\left(\frac{4}{10}\right)^{4-i-j} \qquad , \; i+j=0,\,1,\,2,\,3,\,4$$

(ii) contin.

$$P(X=i,Y=j)=\frac{4!}{i!\ j!\ (4\text{-}i\text{-}j)!}\left(\frac{5}{10}\right)^i\left(\frac{1}{10}\right)^j\left(\frac{4}{10}\right)^{4-i-j} \quad , i+j=0,\,1,\,2,\,3,\,4$$

$$\Downarrow$$

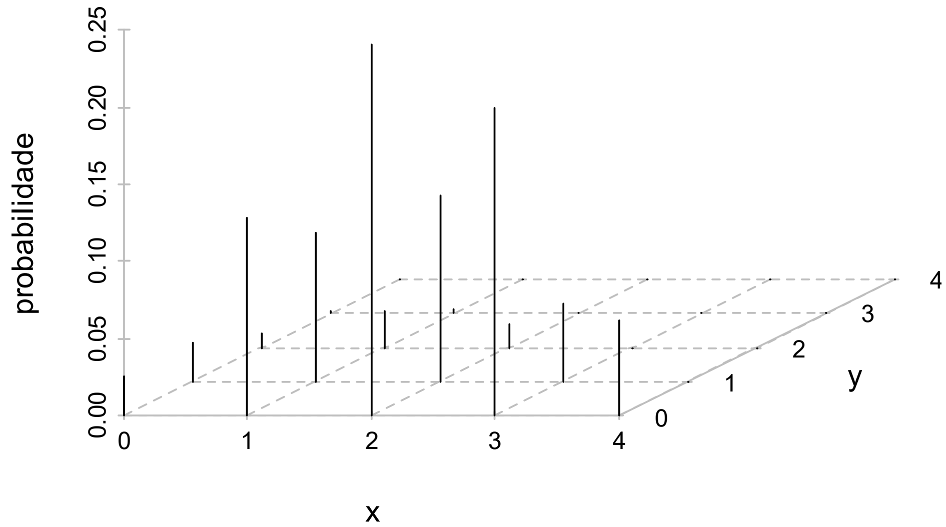
$$i+j\leq 4 \quad\Rightarrow\quad j\leq 4-\mathrm{i}$$

```
fmp <- matrix(rep(0,5*5),nr=5)
for (i in 0:4)
      { for (j in 0:(4-i))
         fmp[i+1,j+1] <- dmultinom(c(i,j,4-i-j),prob=c(5,1,4)) }
fmp
       [,1]   [,2]   [,3]   [,4]  [,5]
[1,] 0.0256 0.0256 0.0096 0.0016 1e-04
[2,] 0.1280 0.0960 0.0240 0.0020 0e+00
[3,] 0.2400 0.1200 0.0150 0.0000 0e+00
[4,] 0.2000 0.0500 0.0000 0.0000 0e+00
[5,] 0.0625 0.0000 0.0000 0.0000 0e+00
```

fmp do par (X,Y):

$X \backslash Y$	0	1	2	3	4
0	0.0256	0.0256	0.0096	0.0016	0.0001
1	0.1280	0.0960	0.0240	0.0020	0
2	0.2400	0.1200	0.0150	0	0
3	0.2000	0.0500	0	0	0
4	0.0625	0	0	0	0

```
addmargins(fmp, 1:2)
                                              Sum
    0.0256 0.0256 0.0096 0.0016 1e-04  0.0625
    0.1280 0.0960 0.0240 0.0020 0e+00  0.2500
    0.2400 0.1200 0.0150 0.0000 0e+00  0.3750
    0.2000 0.0500 0.0000 0.0000 0e+00  0.2500
    0.0625 0.0000 0.0000 0.0000 0e+00  0.0625
Sum 0.6561 0.2916 0.0486 0.0036 1e-04  1.0000

margin.table(fmp,1)
[1] 0.0625 0.2500 0.3750 0.2500 0.0625
margin.table(fmp,2)
[1] 0.6561 0.2916 0.0486 0.0036 0.0001

dbinom(0:4,4,1/2)
[1] 0.0625 0.2500 0.3750 0.2500 0.0625
dbinom(0:4,4,1/10)
[1] 0.6561 0.2916 0.0486 0.0036 0.0001
```



Marginais: $X \frown \text{bi}(4, ½)$ e $Y \frown \text{bi}(4, 1/10)$

(iii) Qual a distribuição de $X+Y$?

Como $X+Y$ representa o nº de sucessos (sucesso = sair algarismo "ímpar ou zero" em cada extracção) em 4 extracções ao acaso (com reposição) do conjunto $\{0, 1, 2, \dots, 9\}$, temos que a probabilidade de sucesso é 0.6, donde $X+Y \frown \text{bi}(\ 4\ ,\ 6/10\ )$.

(iv) Qual a distribuição de XY ?

Tabela com os produtos possíveis

X \ Y	0	1	2	3	4
0	0	0	0	0	0
1	0	1	2	3	-
2	0	2	4	-	-
3	0	3	-	-	-
4	0	-	-	-	-

XY tem suporte {0,1,2,3,4}

$P(XY=1) = P\,(X=1,\ Y=1) = 0.096$

$P(XY=2) = P\,(X=1,\ Y=2) + P(X=2,\ Y=1) = 0.0240 + 0.1200 = 0.144$

$P(XY=3) = P\,(X=1,\ Y=3) + P\,(X=3,\ Y=1) = 0.0020 + 0.0500 = 0.052$

$P(XY=4) = P\,(X=2,\ Y=2) = 0.015$

$P(XY=0) = 1\text{-}\,(0.0960 + 0.144 + 0.052 + 0.0150) = 0.693$

Logo $XY : \begin{cases} 0 & 1 & 2 & 3 & 4 \\ 0.693 & 0.096 & 0.144 & 0.052 & 0.015 \end{cases}$

(v) Calcule cov(X,Y) pela fórmula cov$(X,Y) = E(XY) - E(X)\,E(Y)$

Como $XY: \begin{cases} 0 & 1 & 2 & 3 & 4 \\ 0.693 & 0.096 & 0.144 & 0.052 & 0.015 \end{cases}$

temos então $E(XY) = 0 + 1 \times 0.096 + 2 \times 0.144 + 3 \times 0.052 + 4 \times 0.015 = 0.6$

Por outro lado, como a bi(n,p) tem valor médio np, temos

$E(X) = 4 \times 0.5 = 2$ e $E(Y) = 4 \times 0.1 = 0.4$

Logo cov$(X,Y) = E(XY) - E(X)\,E(Y) = 0.6 - 2 \times 0.4 = 0.6 - 0.8 = -0.2$

Este resultado está de acordo com a fórmula da covariância entre as v.a. num vector multinomial, $\text{cov}(X_i, X_j) = -n\,p_i\,p_j$, que no caso trinomial se reduz a

$$\text{cov}(X,Y) = -n\,p_1\,p_2 = -4 \times 0.5 \times 0.1 = -0.2$$

Exercício: (nº 30) (ii) Calcular o valor médio e variância de $X \frown Geom(p)$.

Teorema: uma série de potências com raio de convergência r é derivável termo a termo no mesmo domínio, ou seja

$$f(x)=\sum_{n} a_n x^n, |x|<r \quad \Rightarrow \quad f'(x)=\sum_{n} n a_n x^{n-1}, |x|<r$$

Resolução:

Usando este teorema temos

$$E(X)=\sum_{n\geq 1} n(1-p)^{n-1}p=\sum_{n\geq 0} n(1-p)^{n-1}p=-p\sum_{n\geq 0}\tfrac{d}{dp}(1-p)^n=-p\tfrac{d}{dp}\sum_{n\geq 0}(1-p)^n=$$

$$=-p\tfrac{d}{dp}\tfrac{1}{1-(1-p)}=-p\tfrac{d}{dp}\tfrac{1}{p}=-p(-\tfrac{1}{p^2})=\tfrac{1}{p}$$

$$E(X(X-1))=\sum_{n\geq 0} n(n-1)(1-p)^{n-1}p=p(1-p)\sum_{n\geq 0} n(n-1)(1-p)^{n-2}=p(1-p)\sum_{n\geq 0}\tfrac{d^2}{dp^2}(1-p)^n=$$

$$=p(1-p)\tfrac{d^2}{dp^2}\sum_{n\geq 0}(1-p)^n=p(1-p)\tfrac{d^2}{dp^2}\tfrac{1}{p}=p(1-p)=p(1-p)\tfrac{2}{p^3}=\tfrac{2(1-p)}{p^2}$$

donde

$$Var(X)=E(X^2)-E^2(X)=E(X(X-1))+E(X)-E^2(X)=\tfrac{2(1-p)}{p^2}+\tfrac{1}{p}-\tfrac{1}{p^2}=\tfrac{2-2p+p-1}{p^2}=\tfrac{1-p}{p^2}$$

Exercício: (nº 39) Calcular as distribuições condicionais com $(X,Y) \frown M(4;\ 1/3,\ 1/3)$.

Resolução: Temos a fmp conjunta $p_{ij} = P(X=i, Y=j) = \frac{4!}{i!\ j!\ 4\text{-}i\text{-}j!}\frac{1}{3^4}$, $\quad 0 \le i+j \le 4$, $\ i, j \in \mathbb{N}$

As leis marginais de X e Y são ambas $bi(4,1/3)$.

Calcule-se a lei de X condicional a $\{Y=j\}$, para um dado $j \in \{1, 2, 3, 4\}$:

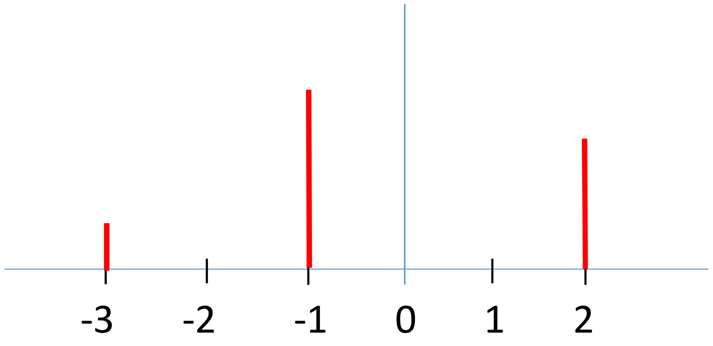
$$P(X=i \mid Y=j) = \frac{p_{ij}}{P(Y=j)} = \frac{\frac{4!}{i!\ j!\ (4-i-j)!}\frac{1}{3^4}}{\frac{4!}{j!\ (4-j)!}\frac{1}{3^j}\frac{2^{4-j}}{3^{4-j}}} = \frac{(4-j)!}{i!\ (4-j-i)!}\frac{1}{2^{4-j}} = \binom{4-j}{i}\left(\frac{1}{2}\right)^{4-j},\quad 0 \le i \le 4-j,\ i \in \mathbb{N}$$

Logo a v.a. $X \mid \{Y=j\}$ tem distribuição $bi(\ 4-j\ ,\ ½\ )$

Analogamente temos que $Y \mid \{X=i\}$ tem distribuição $bi(\ 4-i\ ,\ ½\ )$

Exercício: (nº 37) Calcule o coeficiente de assimetria da v.a. discreta X com fmp (não simétrica)

$$X:\begin{cases} -3 & -1 & 2 \\ \frac{1}{10} & \frac{5}{10} & \frac{4}{10} \end{cases}$$



Resolução:

$$\mu = -\frac{3}{10} - \frac{5}{10} + \frac{8}{10} = 0$$

$$E((X-\mu)^3) = E(X^3) = -\frac{27}{10} - \frac{5}{10} + \frac{32}{10} = 0$$

donde $\quad \beta_1 = \dfrac{1}{\sigma^3} E((X-\mu)^3) = 0$