<u>Capítulo 5</u> - Distribuições conjuntas de probabilidades e complementos

5.1 Duas variáveis aleatórias discretas. Distribuições conjuntas, marginais e condicionais. Independência

Em relação a uma mesma experiência podem ser consideradas várias variáveis aleatórias.

Exemplos:

a) Transmissão/recepção de sinais em que cada um é classificado como: alta, média ou baixa qualidade. Considerem-se as v.a.

X - nº de sinais de alta qualidade
Y - nº de sinais de baixa qualidade

Trata-se de duas variáveis discretas.

b) Fabrico de peças cilíndricas

X - comprimento da peça
Y - diâmetro da peça

Trata-se de duas variáveis contínuas.

Nos Capítulos 3 e 4 estudou-se a distribuição de probabilidade de v.a. (discretas ou contínuas), consideradas isoladamente.

Para duas ou mais v.a. o seu comportamento simultâneo é estudado usando as chamadas <u>distribuições de probabilidade conjuntas</u>.

Definição: Dadas duas variáveis aleatórias discretas, chama-se função de probabilidade conjunta a $f_{XY}(x,y) = P(X = x, Y = y)$, tal que

(1) $f_{XY}(x,y) \geq 0$, $\forall_{(x,y)}$ (2) $\sum_x \sum_y f_{XY}(x,y) = 1$

Exemplo: Lançamento de dois dados perfeitos. Seja

X - nº de vezes que sai a face 6 x=0,1,2
Y - nº de vezes que sai a face 5 y=0,1,2

$$P(X=0,Y=0) = \frac{4}{6}\times\frac{4}{6}$$

$$P(X=1,Y=0) = \frac{4}{6}\times\frac{1}{6}\times 2 = P(X=0,Y=1)$$

$$P(X=1,Y=1) = \frac{1}{6}\times\frac{1}{6}\times 2$$

$$P(X=2,Y=0) = \frac{1}{6}\times\frac{1}{6} = P(X=0,Y=2)$$

$$P(X=1,Y=2) = 0 = P(X=2,Y=1)$$

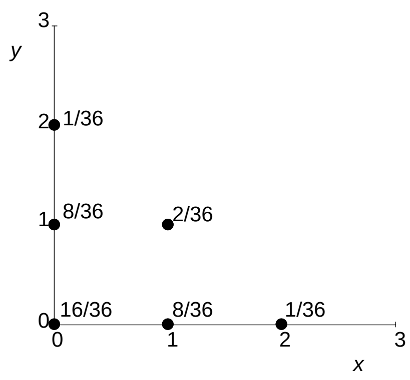
$$P(X=2,Y=2) = 0$$

$$\sum_{x=0}^{2}\sum_{y=0}^{2} P(X=x,Y=y) = 1$$

Organização da função de probabilidade conjunta em tabela:

$Y \setminus X$	0	1	2
0	16/36	8/36	1/36
1	8/36	2/36	0
2	1/36	0	0

Gráfico:



A partir da tabela podem calcular-se também as probabilidades (chamadas marginais) para *X* ou *Y*. Por exemplo:

$$P(X=0)=f_{XY}(0,0)+f_{XY}(0,1)+f_{XY}(0,2)=$$
$$=\sum_{y=0}^{2} f_{XY}(0,y)$$

Estas probabilidades podem ser acrescentadas à tabela da função de probabilidade conjunta:

$Y \setminus X$	0	1	2	$P(Y=y)$
0	16/36	8/36	1/36	25/36
1	8/36	2/36	0	10/36
2	1/36	0	0	1/36
$P(X=x)$	25/36	10/36	1/36	1

<u>Nota</u>: verificar que $X \sim Bin\left(2,\frac{1}{6}\right)$ e $Y \sim Bin\left(2,\frac{1}{6}\right)$ (diz-se que *X* e *Y* são identicamente distribuídas)

Definição: Se *X* e *Y* forem v.a. discretas com função de probabilidade conjunta $f_{XY}(x,y)$, as funções de probabilidade marginais de *X* e *Y* são

$$f_X(x)=P(X=x)=\sum_y f_{XY}(x,y)$$
$$f_Y(y)=P(Y=y)=\sum_x f_{XY}(x,y)$$

Nota: dada uma função de probabilidade conjunta, $E(X)$, $V(X)$, $E(Y)$ e $V(Y)$ podem ser calculados usando a f.p. conjunta ou as f.p. marginais:

$$E(X)=\mu_X=\sum_x x f_X(x)=\sum_x x\left(\sum_y f_{XY}(x,y)\right)=$$
$$=\sum_x\sum_y x f_{XY}(x,y)$$

Definição: Dadas as v.a. discretas *X* e *Y* com f.p. conjunta $f_{XY}(x,y)$, a função de probabilidade condicionada de *Y* dado *X*=*x* (tal que $f_X(x)>0$) é

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

e verifica

(1) $f_{Y|x}(y)\geq 0,\ \forall_y$ (2) $\sum_y f_{Y|x}(y)=1$

(tem as propriedades de uma f.p.)

Definição: O <u>valor esperado condicionado</u> de *Y* dado *X*=*x* é

$$E(Y|x)=\mu_{Y|x}=\sum_y y f_{Y|x}(y)$$

e a respectiva <u>variância condicionada</u> é

$$V(Y|x)=\sum_y \left(y-\mu_{Y|x}\right)^2 f_{Y|x}(y)$$

Nota: Definições semelhantes para *X* dado *Y*=*y*.

Exemplo (cont.): *Y*|*X*=1

$$P(Y=0|X=1)=\frac{8/36}{10/36}=\frac{8}{10}$$
$$P(Y=1|X=1)=\frac{2/36}{10/36}=\frac{2}{10}$$
$$P(Y=2|X=1)=\frac{0}{10/36}=0$$

$$E(Y|X=1)=0\times\frac{8}{10}+1\times\frac{2}{10}=\frac{2}{10}$$

Y|*X*=0

$$P(Y=0|X=0)=\frac{16/36}{25/36}=\frac{16}{25}$$
$$P(Y=1|X=0)=\frac{8/36}{25/36}=\frac{8}{25}$$
$$P(Y=2|X=0)=\frac{1/36}{25/36}=\frac{1}{25}$$

$$E(Y|X=0)=0\times\frac{16}{25}+1\times\frac{8}{25}+2\times\frac{1}{25}=\frac{10}{25}$$

O conceito de independência de acontecimentos pode ser estendido a variáveis aleatórias.

Definição: Dadas duas v.a. discretas *X* e *Y*, se uma das seguintes condições se verificar então as outras também se verificam e as v.a. são independentes

(1) $f_{XY}(x,y) = f_X(x) f_Y(y), \quad \forall_{x,y}$

(2) $f_{Y|x}(y) = f_Y(y), \quad \forall_{x,y}$ com $f_X(x) > 0$

(3) $f_{X|y}(x) = f_X(x), \quad \forall_{x,y}$ com $f_Y(y) > 0$

Exemplo (cont.): Serão *X* e *Y* independentes?
Não, porque, por exemplo,

$$P(X = 2, Y = 2) = 0 \neq P(X = 2)P(Y = 2)$$

Consequência da definição de independência: Se *X* e *Y* são independentes, então

$$P(X \in A \text{ e } Y \in B) = P(X \in A) \times P(Y \in B)$$

5.2 Duas variáveis aleatórias contínuas. Distribuições conjuntas, marginais e condicionais. Independência

Do mesmo modo que para caracterizar uma v.a. contínua foi necessário introduzir o conceito de função de densidade de probabilidade, define-se para duas variáveis aleatórias contínuas a função de densidade de probabilidade conjunta.

Definição: A função de densidade de probabilidade conjunta para duas v.a. contínuas, *X* e *Y*, representada por $f_{XY}(x,y)$, satisfaz:

(1) $f_{XY}(x,y) \geq 0 \quad \forall_{(x,y)}$

(2) $\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} f_{XY}(x,y)dxdy = 1$

(3) $P((X,Y) \in R) = \iint_R f_{XY}(x,y)dxdy \quad \forall$ região *R*

Definição: Dadas duas v.a. contínuas, *X* e *Y*, com f.d.p. conjunta $f_{XY}(x,y)$, as funções de densidade de probabilidade marginais de *X* e *Y* são dadas por

$$f_X(x) = \int_{-\infty}^{+\infty} f_{XY}(x,y)dy$$

$$f_Y(y) = \int_{-\infty}^{+\infty} f_{XY}(x,y)dx$$

Definição: Dadas duas v.a. contínuas, *X* e *Y*, com f.d.p. conjunta $f_{XY}(x,y)$, a função de densidade de probabilidade condicionada de *Y* dado *X*=*x* (tal que $f_X(x) > 0$) é

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

e verifica:

(1) $f_{Y|x}(y) \geq 0, \ \forall_y$ (2) $\int_{-\infty}^{+\infty} f_{Y|x}(y)dy = 1$

(3) $P(Y \in B | X = x) = \int_B f_{Y|x}(y)dy$

Definição: O valor esperado condicionado de *Y* dado *X*=*x* é

$$E(Y|x) = \mu_{Y|x} = \int_{-\infty}^{+\infty} y f_{Y|x}(y)dy$$

e a respectiva variância condicionada é

$$V(Y|x) = \int_{-\infty}^{+\infty} (y - \mu_{Y|x})^2 f_{Y|x}(y)dy$$

Nota: Definições semelhantes para *X* dado *Y*=*y*.

A **definição de independência** coincide com a que foi dada para as variáveis discretas mas em que $f_{XY}(x,y)$, $f_X(x)$, $f_Y(y)$, $f_{Y|x}(y)$ e $f_{X|y}(x)$ são funções de densidade de probabilidade.

5.3 Covariância e correlação

Pretende-se uma medida que meça algum aspecto da variação conjunta de duas variáveis aleatórias.

Definição: a covariância entre duas v.a., *X* e *Y*, é dada por

$$\text{cov}(X,Y) = \sigma_{XY} = E\left[(X-\mu_X)(Y-\mu_Y)\right]$$

Se ambas as v.a. forem discretas tem-se

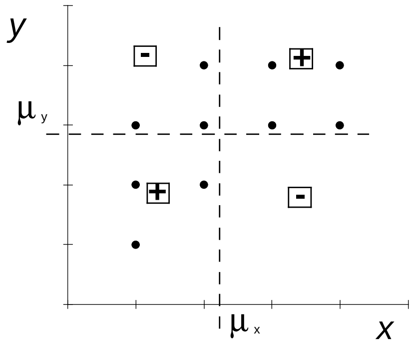
$$\sigma_{XY} = \sum_x \sum_y (x-\mu_X)(y-\mu_Y) f_{XY}(x,y)$$

Se ambas as v.a. forem contínuas tem-se

$$\sigma_{XY} = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} (x-\mu_X)(y-\mu_Y) f_{XY}(x,y)dxdy$$

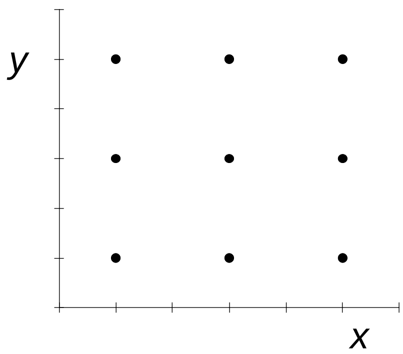
Interpretação do sinal da covariância:

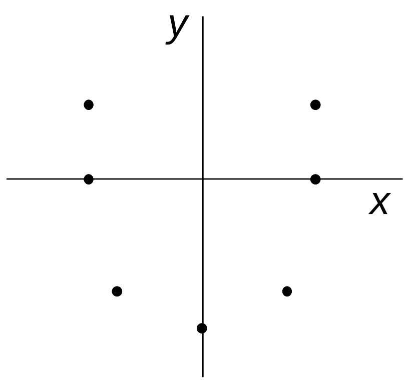
$\text{cov}(X,Y) > 0$:



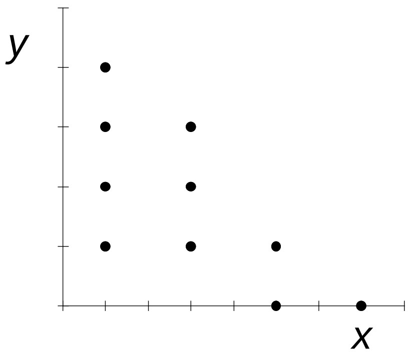
Nota: pontos com iguais probabilidades.

$\text{cov}(X,Y) = 0$:





$\text{cov}(X,Y) < 0$:



Propriedades da covariância:

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

2) $\text{cov}(X,Y) = \text{cov}(Y,X)$ **3)** $\text{cov}(X,X) = V(X)$

4) $\text{cov}(aX,Y) = a\,\text{cov}(X,Y)$

5) $\text{cov}(X+Z,Y) = \text{cov}(X,Y) + \text{cov}(Z,Y)$

6) $\text{cov}\left(\sum_{i=1}^n X_i, \sum_{j=1}^n Y_j\right) = \sum_{i=1}^n \sum_{j=1}^n \text{cov}\left(X_i, Y_j\right)$

Exemplo (cont.): Calcular a covariância entre *X* e *Y*

$Y \setminus X$	0	1	2
0	16/36	8/36	1/36
1	8/36	2/36	0
2	1/36	0	0

$$E(XY) = 0\times 0\times \frac{16}{36} + \ldots + 1\times 1\times \frac{2}{36} + \ldots + 2\times 2\times 0 = $$
$$= \frac{2}{36}$$

$$X \text{ e } Y \sim Bin\left(2, \frac{1}{6}\right) \Rightarrow E(X) = E(Y) = 2\times \frac{1}{6} = \frac{1}{3}$$

logo $\qquad \operatorname{cov}(X,Y) = \frac{2}{36} - \frac{1}{9} = -\frac{1}{18}$

Significado: em média há uma tendência para *y* decrescer quando *x* cresce e vice-versa.

O valor absoluto da covariância não é interpretável porque esta pode ser arbitrariamente alterada por uma mudança de escala (propriedade 4).

Medida cujo sinal dá a mesma indicação mas que é adimensional e cujo valor absoluto já é interpretável:

Definição: A correlação ou coeficiente de correlação entre duas v.a. *X* e *Y* é

$$\rho_{XY} = \frac{\operatorname{cov}(X,Y)}{\sqrt{V(X)V(Y)}} = \frac{\sigma_{XY}}{\sigma_X \sigma_Y}$$

Propriedades:

1) $-1 \le \rho_{XY} \le 1, \quad \forall_{X,Y}$

2) a) $\rho_{XY} = 1$ sse $Y = aX + b$ com $a > 0$

b) $\rho_{XY} = -1$ sse $Y = aX + b$ com $a < 0$

Demonstração: 1) a) $\qquad \rho_{XY} \ge -1$

Considere-se a v.a. auxiliar $\frac{X}{\sigma_X} + \frac{Y}{\sigma_Y}$, sabe-se que

$V\left(\frac{X}{\sigma_X} + \frac{Y}{\sigma_Y}\right) \ge 0, \quad \forall_{X,Y}$, por outro lado

$$V\left(\frac{X}{\sigma_X} + \frac{Y}{\sigma_Y}\right) = E\left[\left(\frac{X}{\sigma_X} + \frac{Y}{\sigma_Y}\right)^2\right] - \left[E\left(\frac{X}{\sigma_X} + \frac{Y}{\sigma_Y}\right)\right]^2 =$$
$$= \frac{E(X^2)}{\sigma_X^2} + \frac{E(Y^2)}{\sigma_Y^2} + 2\frac{E(XY)}{\sigma_X\sigma_Y} - \frac{\mu_X^2}{\sigma_X^2} - \frac{\mu_Y^2}{\sigma_Y^2} - 2\frac{\mu_X\mu_Y}{\sigma_X\sigma_Y} =$$
$$= \frac{E(X^2) - \mu_X^2}{\sigma_X^2} + \frac{E(Y^2) - \mu_Y^2}{\sigma_Y^2} + 2\frac{E(XY) - \mu_X\mu_Y}{\sigma_X\sigma_Y} =$$
$$= 1 + 1 + 2\rho_{XY} \ge 0 \Leftrightarrow \rho_{XY} \ge -1$$

b) $\rho_{XY} \le 1$ Repetir a demonstração anterior com a v.a. auxiliar $\frac{X}{\sigma_X} - \frac{Y}{\sigma_Y}$

2) b) Pela demonstração anterior,

$$V\left(\frac{X}{\sigma_X} + \frac{Y}{\sigma_Y}\right) = 0 \Leftrightarrow \rho_{XY} = -1$$

mas $V\left(\frac{X}{\sigma_X} + \frac{Y}{\sigma_Y}\right) = 0 \Leftrightarrow \frac{X}{\sigma_X} + \frac{Y}{\sigma_Y} = \text{constante}$, o que demonstra o resultado.

Teorema: Se *X* e *Y* são independentes então

$$\sigma_{XY} = \rho_{XY} = 0$$

Demonstração: Vamos admitir que *X* e *Y* são ambas discretas (seria semelhante para ambas contínuas, substituindo os integrais por somatórios)

X e Y são independentes sse

$$f_{XY}(x,y) = f_X(x) f_Y(y), \quad \forall_{x,y}$$

logo

$$E(XY) = \sum_x \sum_y xy f_{XY}(x,y) =$$
$$= \sum_x \sum_y xy f_X(x) f_Y(y) =$$
$$= \left(\sum_x x f_X(x) \right) \left(\sum_y y f_Y(y) \right) = E(X)E(Y)$$

pelo que $\sigma_{XY} = 0$ e $\rho_{XY} = 0$

Muito importante: a proposição inversa não é verdadeira, isto é

$\rho_{XY} = \sigma_{XY} = 0$ **não implica** X e Y independentes

Exemplo:

$Y \setminus X$	-1	0	1	$P(Y=y)$
-1	0	1/6	0	1/6
0	1/12	1/2	1/12	2/3
1	0	1/6	0	1/6
$P(X=x)$	1/12	5/6	1/12	1

Tem-se $E(X) = E(Y) = 0$ e $E(XY) = 0$, pelo que $\sigma_{XY} = 0$, mas X e Y não são independentes.

5.4 Combinações lineares de variáveis aleatórias

Definição: dadas p v.a., $X_1, X_2, \ldots, X_p$ e p constantes $c_1, c_2, \ldots, c_p$,

$$Y = c_1 X_1 + c_2 X_2 + \ldots + c_p X_p$$

é uma combinação linear de $X_1, X_2, \ldots, X_p$.

Valor esperado duma combinação linear:

1) $E(c_1 X_1 + c_2 X_2) = c_1 E(X_1) + c_2 E(X_2)$

Demonstração:

$$E(c_1 X_1 + c_2 X_2) = \sum_{x_1} \sum_{x_2} (c_1 x_1 + c_2 x_2) f(x_1, x_2) =$$
$$= c_1 \sum_{x_1} \sum_{x_2} x_1 f(x_1, x_2) + c_2 \sum_{x_1} \sum_{x_2} x_2 f(x_1, x_2) =$$
$$= c_1 E(X_1) + c_2 E(X_2)$$

2) Generalização de **(1)**

$$E(c_1 X_1 + c_2 X_2 + \ldots + c_p X_p) =$$
$$= c_1 E(X_1) + c_2 E(X_2) + \ldots + c_p E(X_p)$$

Variância duma combinação linear:

1)
$$V(c_1 X_1 + c_2 X_2) =$$
$$= c_1^2 V(X_1) + c_2^2 V(X_2) + 2 c_1 c_2 \operatorname{cov}(X_1, X_2)$$

Demonstração:

$$V(c_1 X_1 + c_2 X_2) =$$
$$= E\left[(c_1 X_1 + c_2 X_2)^2 \right] - \left[E(c_1 X_1 + c_2 X_2) \right]^2 =$$
$$= E\left(c_1^2 X_1^2 + c_2^2 X_2^2 + 2 c_1 c_2 X_1 X_2 \right) - (c_1 \mu_1 + c_2 \mu_2)^2 =$$
$$= c_1^2 E(X_1^2) + c_2^2 E(X_2^2) + 2 c_1 c_2 E(X_1 X_2) -$$
$$- c_1^2 \mu_1^2 - c_2^2 \mu_2^2 - 2 c_1 c_2 \mu_1 \mu_2 =$$
$$= c_1^2 \left[E(X_1^2) - \mu_1^2 \right] + c_2^2 \left[E(X_2^2) - \mu_2^2 \right] +$$
$$+ 2 c_1 c_2 \left[E(X_1 X_2) - \mu_1 \mu_2 \right]$$

c.q.d.

2) Generalização de **(1)**

$$V\left(c_1X_1+\ldots+c_pX_p\right)=\sum_{i=1}^{p}c_i^2V\left(X_i\right)+$$
$$+2\sum_{i=1}^{p}\sum_{j=i+1}^{p}c_ic_j\operatorname{cov}\left(X_i,X_j\right)$$

3) Caso particular de 2)

Se $\operatorname{cov}\left(X_i,X_j\right)=0,\quad \forall_{i\neq j}$, ou seja, se as v.a. forem não correlacionadas duas a duas tem-se

$$V\left(c_1X_1+\ldots+c_pX_p\right)=\sum_{i=1}^{p}c_i^2V\left(X_i\right)$$

Nota: o mesmo acontece se as v.a. forem independentes duas a duas, uma vez que, como já se viu, nesse caso as covariâncias duas a duas são nulas.

Casos especiais de somas de variáveis aleatórias:

I) Propriedade reprodutiva da distribuição binomial

a) Se $X_1\sim Bin\left(n_1,p\right)$, $X_2\sim Bin\left(n_2,p\right)$ e X_1 e X_2 forem independentes, então

$$X_1+X_2\sim Bin\left(n_1+n_2,p\right)$$

(X_1+X_2 representa o número de sucessos em n_1+n_2 provas de Bernoulli independentes com P(sucesso) constante e igual a p)

b) (Generalização) Se $X_i\sim Bin\left(n_i,p\right)$, $i=1,\ldots,k$, independentes, então

$$X_1+\ldots+X_k\sim Bin\left(\sum_{i=1}^{k}n_i,p\right)$$

Caso particular: $X_i\sim Ber(p)=Bin(1,p)$, $i=1,\ldots,n$ independentes, então $X_1+\ldots+X_n\sim Bin(n,p)$

II) Propriedade reprodutiva da distribuição de Poisson

a) Se $X_1\sim Poisson\left(\lambda_1\right)$, $X_2\sim Poisson\left(\lambda_2\right)$ e X_1 e X_2 forem independentes, então

$$X_1+X_2\sim Poisson\left(\lambda_1+\lambda_2\right)$$

(Este caso pode ser visto como um limite do caso **I)a)**)

b) (Generalização) Se $X_i\sim Poisson\left(\lambda_i\right)$, $i=1,\ldots,k$, independentes, então

$$X_1+\ldots+X_k\sim Poisson\left(\sum_{i=1}^{k}\lambda_i\right)$$

III) Propriedade reprodutiva da distribuição Normal

Se $X_1\sim N\left(\mu_1,\sigma_1^2\right)$, $X_2\sim N\left(\mu_2,\sigma_2^2\right)$ e X_1 e X_2 forem independentes, então

$$c_1X_1+c_2X_2\sim N\left(c_1\mu_1+c_2\mu_2,c_1^2\sigma_1^2+c_2^2\sigma_2^2\right)$$

IV) Mudança de escala na distribuição exponencial

Se $X\sim Exp(\lambda)$ e $c>0$, então $Y=cX\sim Exp\left(\frac{\lambda}{c}\right)$

Demonstração: Seja para $y>0$

$$F_Y(y)=P(Y\leq y)=P(cX\leq y)=P\left(X\leq\frac{y}{c}\right)=$$
$$=F_X\left(\frac{y}{c}\right)=1-e^{-\frac{\lambda}{c}y}$$

uma vez que para $X\sim Exp(\lambda)$,

$$F_X(x)=1-e^{-\lambda x},x>0$$

5.5 Desigualdade de Chebychev*

Esta desigualdade permite relacionar probabilidades, relativas a uma qualquer v.a., discreta ou contínua, com os parâmetros μ e σ:

Proposição: Para qualquer v.a. X, com valor esperado μ e desvio padrão σ, verifica-se

$$P\left(|X-\mu| \geq c\sigma\right) \leq \frac{1}{c^2}$$

(só é útil para $c > 1$).

Exemplo:

c	$P\left(\|X-\mu\| \geq c\sigma\right)$ (qq v.a. X)
1.5	≤ 0.444
2	≤ 0.25
3	≤ 0.111
4	≤ 0.063

* Excepto para Probabilidades, Erros e Estatística

Estes valores são úteis se não conhecermos a distribuição da variável, mas podem ser muito pessimistas. Por exemplo para $X \sim N\left(\mu, \sigma^2\right)$, podem ser determinados exactamente e comparados com aqueles:

c	$P\left(\|X-\mu\| \geq c\sigma\right)$ $X \sim N\left(\mu,\sigma^2\right)$	$P\left(\|X-\mu\| \geq c\sigma\right)$ (qq v.a. X)
1.5	$= 0.1336$	≤ 0.444
2	$= 0.0456$	≤ 0.25
3	$= 0.0027$	≤ 0.111
4	$= 0.0001$	≤ 0.063

Demonstração: Considere-se a v.a. auxiliar Y, definida da seguinte forma

$$Y = \begin{cases} 1, & \text{se } |X-\mu| \geq c\sigma \\ 0, & \text{caso contrário} \end{cases}$$

Tem-se $E(Y) = 1 \times P(Y=1) = P\left(|X-\mu| \geq c\sigma\right)$,

por outro lado como $y = 0$ ou $y = 1$, e pela definição de Y,

$$(X-\mu)^2 \geq (X-\mu)^2 Y \ \text{ e } \ (X-\mu)^2 Y \geq c^2\sigma^2 Y$$

Logo

$$E(X-\mu)^2 \geq c^2\sigma^2 E(Y) \Leftrightarrow$$
$$\Leftrightarrow \sigma^2 \geq c^2\sigma^2 P\left(|X-\mu| \geq c\sigma\right) \Leftrightarrow$$
$$\Leftrightarrow P\left(|X-\mu| \geq c\sigma\right) \leq \frac{1}{c^2}$$

5.6 Teorema do Limite Central

Este teorema justifica a grande utilidade e importância da distribuição normal.

T.L.C.: Se para todo o inteiro positivo n, $X_1, \ldots, X_n$ forem v.a. independentes e identicamente distribuídas com valor esperado μ e variância σ^2, então para cada z, real, tem-se

$$\lim_{n\to\infty} P\left(\frac{\sum_{i=1}^{n} X_i - n\mu}{\sigma\sqrt{n}} \leq z\right) = \Phi(z)$$

Consequência: Para $X_1, \ldots, X_n$ nas mesmas condições e n grande verifica-se

$$P\left(a < \sum_{i=1}^{n} X_i < b\right) \cong \Phi\left(\frac{b-n\mu}{\sigma\sqrt{n}}\right) - \Phi\left(\frac{a-n\mu}{\sigma\sqrt{n}}\right)$$

Observações:

- A demonstração do teorema exige algumas ferramentas matemáticas avançadas.
- Como $E(X_i)=\mu$, $V(X_i)=\sigma^2$ e $X_1,\ldots,X_n$ são v.a. independentes tem-se
$$E\left(\sum_{i=1}^{n} X_i\right)=n\mu \qquad \text{e} \qquad V\left(\sum_{i=1}^{n} X_i\right)=n\sigma^2$$
pelo que
$$E\left(\frac{\sum_{i=1}^{n} X_i - n\mu}{\sigma\sqrt{n}}\right)=0 \quad V\left(\frac{\sum_{i=1}^{n} X_i - n\mu}{\sigma\sqrt{n}}\right)=1 \quad \forall_n$$
- As v.a. $X_1,\ldots,X_n$ podem ser discretas ou contínuas.
- A aproximação das distribuições Binomial e Poisson à distribuição normal (ver Capítulo 4), pode ser justificada por este teorema.
- Geralmente considera-se *n* grande se $n \geq 30$.

Exemplo: Suponha-se que ao adicionar números reais cada número é arredondado previamente para o inteiro mais próximo. Admita-se que os erros de arredondamento são v.a. independentes e identicamente distribuídas com distribuição uniforme contínua no intervalo [-0.5;0.5] (esta suposição é razoável se desconhecermos à partida tudo sobre os referidos números reais e admitirmos que também eles se distribuem uniformemente e independentemente nalgum intervalo)

a) Qual é a probabilidade de que, ao adicionar 1500 números, o valor absoluto do erro seja superior a 15?

Pode afirmar-se que o valor absoluto do erro está certamente compreendido entre 0 e 750, mas este valor é muito pessimista.

Considerem-se as v.a.

E_i - erro na parcela *i*

T - erro total $\qquad T=\sum_{i=1}^{1500} E_i$

Tem-se $E(E_i)=0$ e $V(E_i)=\frac{(0.5-(-0.5))^2}{12}=\frac{1}{12}$

$$E(T)=1500\times 0=0 \qquad V(T)=\frac{1500}{12}=125$$

Pelo T.L.C. conclui-se que $T=\sum_{i=1}^{1500} E_i \overset{a}{\sim} N(0,125)$

Logo

$$P(|T|>15)=1-P(|T|\leq 15)=1-P(-15\leq T\leq 15)=$$
$$=1-P\left(-\frac{15}{\sqrt{125}}\leq\frac{T}{\sqrt{125}}\leq\frac{15}{\sqrt{125}}\right)=$$
$$=1-P\left(-1.34\leq\frac{T}{\sqrt{125}}\leq 1.34\right)\cong$$
$$\cong 1-(\Phi(1.34)-\Phi(-1.34))=$$
$$=1-(0.9099-0.0901)=0.1802$$

b) Quantos números podem ser somados (*n*) para que

$$P\left(\left|\sum_{i=1}^{n} E_i\right|<10\right)\cong 0.9$$

$$T_n=\sum_{i=1}^{n} E_i \qquad E(T_n)=0 \qquad V(T_n)=\frac{n}{12}$$

Pelo T.L.C. conclui-se que $T_n \overset{a}{\sim} N\left(0,\frac{n}{12}\right)$

$$P(-10<T_n<10)\cong\Phi\left(\frac{10}{\sqrt{n/12}}\right)-\Phi\left(\frac{-10}{\sqrt{n/12}}\right)$$

Como $\Phi(-x)=1-\Phi(x)$

$$\Phi\left(\frac{10}{\sqrt{n/12}}\right)-\left(1-\Phi\left(\frac{10}{\sqrt{n/12}}\right)\right)=0.9\Leftrightarrow$$
$$\Leftrightarrow\Phi\left(\frac{10}{\sqrt{n/12}}\right)=0.95\Leftrightarrow\frac{10}{\sqrt{n/12}}=\Phi^{-1}(0.95)=1.645$$
$$\Leftrightarrow\sqrt{n/12}=6.079\Leftrightarrow n=12\times 6.079^2=443.45$$

Resposta: $n\cong 443$

Pode-se "observar" o T.L.C. em "acção":

Exemplo: Vamos realizar a experiência do exercício anterior. Para isso seleccionamos totalmente ao acaso e de modo independente n números reais num dado intervalo e fazemos a diferença para o inteiro mais próximo (isto é equivalente a seleccionar aleatoriamente n "erros" ao acaso no intervalo [-0.5;0.5]), em seguida somam-se os n erros. Temos então uma observação de T_n.

Repetimos este processo 1000 vezes (isto pode ser feito em computador).

Em seguida construímos o histograma das frequências relativas das 1000 observações de

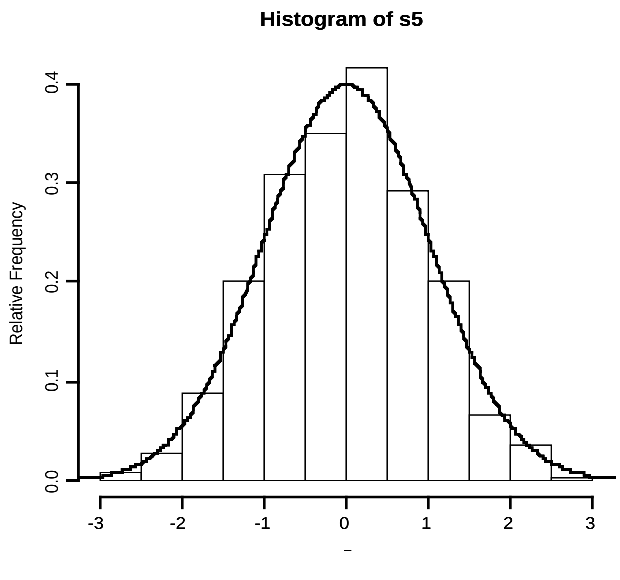
$$S_n = \frac{T_n - n\mu}{\sigma\sqrt{n}} = \frac{T_n}{\sqrt{n/12}}$$

e sobreposmo-lhe a densidade da distribuição $N(0,1)$

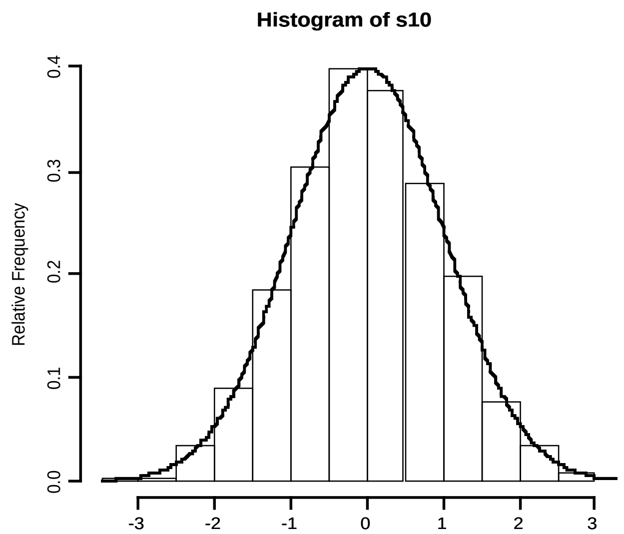
Resultados:

$n = 1$

$n = 5$



$n = 10$



A aproximação parece boa mesmo só com $n = 10$. (Uma explicação é que a distribuição de partida é simétrica e contínua)

Experimentemos com outra distribuição:

Exemplo: Seja $X_i \sim Exp(1)$, por exemplo, intervalo de tempo entre chegadas num processo de Poisson com taxa $\lambda = 1$. A v.a. $T_n = \sum_{i=1}^{n} X_i$ representa o intervalo de tempo até à n-ésima chegada. Vamos repetir o procedimento descrito no exemplo anterior. Ou seja, para cada n, obtemos 1000 observações de

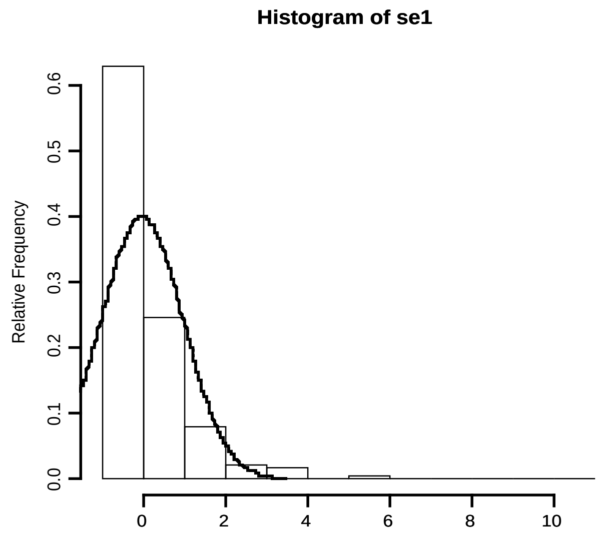
$$S_n = \frac{T_n - n\mu}{\sigma\sqrt{n}} = \frac{T_n - n}{\sqrt{n}}$$

$(\mu = E(X_i) = 1 \qquad \sigma^2 = V(X_i) = 1)$

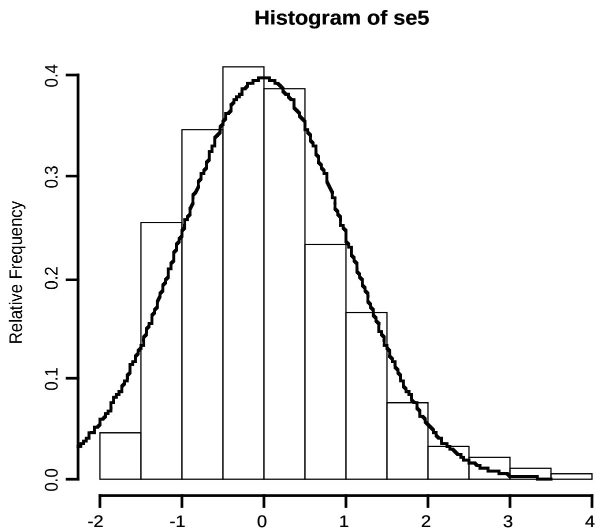
Em seguida construímos o histograma das frequências relativas das 1000 observações de S_n e sobreposmo-lhe a densidade da distribuição $N(0,1)$.

Resultados:

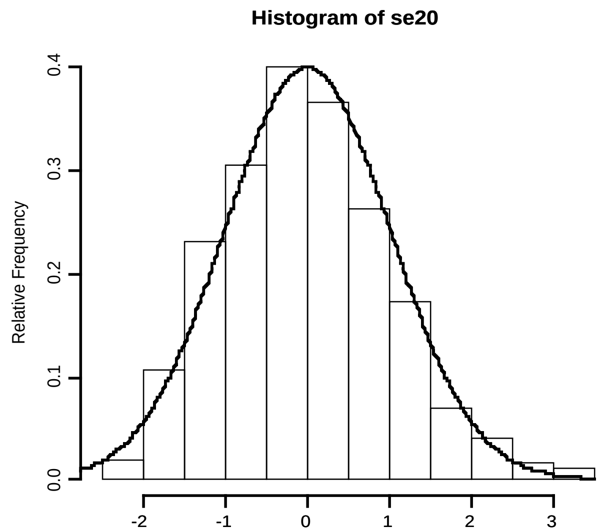
$n = 1$



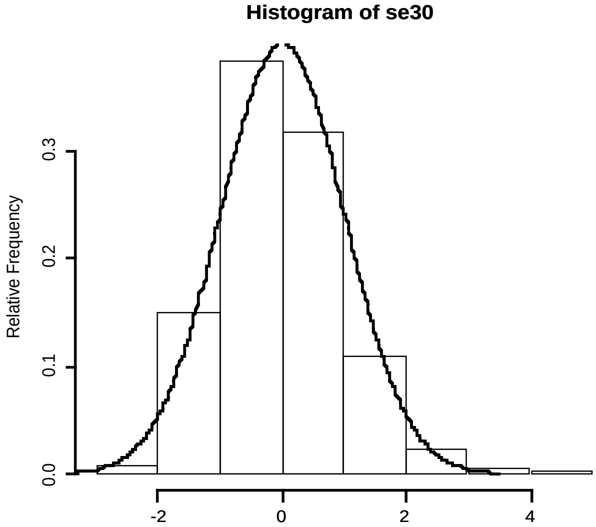
$n = 5$



$n = 20$

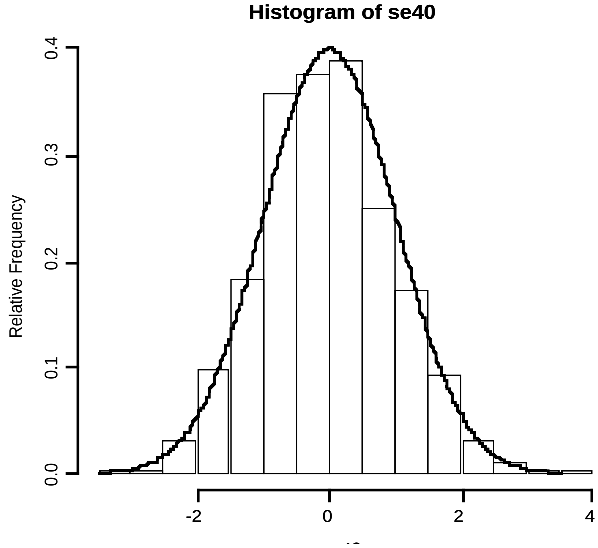


n = 30



©Ana Pires, IST, Outubro de 2000

n = 40



Aqui temos de ter *n* maior para que a aproximação comece a ser razoável.