Capítulo III: Vetores Aleatórios Reais

Probabilidades e Aplicações
Licenciatura em Ciências da Computação
Universidade do Minho
Ano Letivo 2023/2024

1. Definição e função distribuição de um vector aleatório real

Em muitas situações práticas, é frequente haver interesse em associar aos resultados decorrentes de uma experiência aleatória **mais do que uma** característica numérica. Em tais situações, interessa estudar, em simultâneo, várias variáveis aleatórias reais e a existência de relações entre elas.

Neste caso, a cada resultado da experiência será associado um **vector de números reais**.

Definição [Vetor aleatório real ($\overrightarrow{ve}.a.r.$) e respetivas margens]

Sejam $n \in \mathbb{N}$ e $(\Omega, \mathcal{A}, P)$ o espaço de probabilidade associado a uma experiência aleatória. Chama-se vector aleatório real (abrevia-se por $\overrightarrow{ve}.a.r.$) de dimensão n a todo o n-uplo $(X_1, X_2, \ldots, X_n)$ formado por n v.a.r.'s definidas sobre $(\Omega, \mathcal{A}, P)$.

As v.a.r's

$$X_1, X_2, \ldots, X_n$$

são designadas de margens (ou marginais) do $\overrightarrow{ve}.a.r.$ $(X_1, X_2, \ldots, X_n)$.

1. Definição e função distribuição de um vector aleatório real

Definições [σ-álgebra de Borel sobre $\mathbb{R}^n$ e lei de probabilidade de um $\overrightarrow{ve}.a.r.$]

i) Chamamos σ-álgebra de Borel sobre $\mathbb{R}^n$, denota-se por $\mathcal{B}_n(\mathbb{R})$, à σ-álgebra gerada pela seguinte família, $\mathcal{S}$, de subconjuntos de $\mathbb{R}^n$

$$\mathcal{S} = \{B_1 \times B_2 \times \ldots \times B_n : B_i \in \mathcal{B}(\mathbb{R}), i = 1, \ldots, n\},$$

onde $\mathcal{B}(\mathbb{R})$ denota a σ-álgebra de Borel sobre $\mathbb{R}$.

ii) Seja $(X_1, X_2, \ldots, X_n)$ um $\overrightarrow{ve}.a.r.$ definido sobre $(\Omega, \mathcal{A}, P)$ espaço de probabilidade. Chama-se lei de probabilidade do $\overrightarrow{ve}.a.r.$

$$(X_1, X_2, \ldots, X_n),$$

denota-se por $P_{(X_1,X_2,\ldots,X_n)}$, à medida de probabilidade sobre $(\mathbb{R}^n, \mathcal{B}_n(\mathbb{R}))$ definida por:

$$\begin{aligned} B \in \mathcal{B}_n(\mathbb{R}),\ P_{(X_1,X_2,\ldots,X_n)}(B) &= P((X_1, X_2, \ldots, X_n)^{-1}(B)) \\ &\equiv P\left(\{\omega \in \Omega : (X_1, X_2, \ldots, X_n)(\omega) \in B\}\right) \\ &\equiv P((X_1, X_2, \ldots, X_n) \in B) \end{aligned}$$

1. Definição e função distribuição de um vector aleatório real

Nota: É possível mostrar que

$$(X_1, X_2, \ldots, X_n) : \Omega \to \mathbb{R}^n$$

é um $\overrightarrow{ve}.a.r.$ se e só se

$$\forall\, B \in \mathcal{B}_n(\mathbb{R}), \quad (X_1, X_2, \ldots, X_n)^{-1}(B) \in \mathcal{A},$$

com $(\Omega, \mathcal{A}, P)$ o espaço de probabilidade.

1. Definição e função distribuição de um vector aleatório real

Tal como acontece com as v.a.r.'s, a lei de probabilidade de um $\overrightarrow{ve}.a.r.$ é caracterizada pela respectiva função de distribuição, que vamos definir de seguida.

Definição [Função de distribuição de um $\overrightarrow{ve}.a.r.$]

Seja $(X_1, X_2, \ldots, X_n)$ um $\overrightarrow{ve}.a.r.$ definido sobre o espaço de probabilidade $(\Omega, \mathcal{A}, P)$. Chama-se função de distribuição de $(X_1, X_2, \ldots, X_n)$ à função de distribuição da respectiva lei de probabilidade $P_{(X_1,X_2,\ldots,X_n)}$, i.e., à função $F : \mathbb{R}^n \to [0, 1]$ definida por

$$\begin{aligned} F((c_1, c_2, \ldots, c_n)) &= P_{(X_1,X_2,\ldots,X_n)}\left(]-\infty, c_1]\times]-\infty, c_2] \times \ldots \times]-\infty, c_n]\right) \\ &\equiv P(X_1 \leq c_1, X_2 \leq c_2, \ldots, X_n \leq c_n) \end{aligned}$$

Nota: Entre v.a.r.'s, é usual a "," substituir a "$\cap$". Assim,

$$P(X_1 \leq c_1, X_2 \leq c_2, \ldots, X_n \leq c_n) \equiv P(X_1 \leq c_1 \cap X_2 \leq c_2 \cap \ldots \cap X_n \leq c_n).$$

Definição [$\overrightarrow{ve}.a.r.$ discreto e respetivo contradomínio]

Sejam $(\Omega, \mathcal{A}, P)$ um espaço de probabilidade e $(X_1, X_2, \ldots, X_n)$ um $\overrightarrow{ve}.a.r.$ com lei de probabilidade $P_{(X_1,X_2,\ldots,X_n)}$. Este vector diz-se discreto se a lei $P_{(X_1,X_2,\ldots,X_n)}$ for uma lei discreta sobre $(\mathbb{R}^n, \mathcal{B}_n(\mathbb{R}))$, i.e., se existir um elemento M de $\mathcal{B}_n(\mathbb{R})$, finito ou infinito numerável, tal que

$$P_{(X_1,X_2,\ldots,X_n)}(M) = 1.$$

Ao menor conjunto M que satisfaz esta condição chamamos contradomínio de $(X_1, X_2, \ldots, X_n)$ e denotamos por $C_{(X_1,X_2,\ldots,X_n)}$.

2. Vector aleatório real discreto

Se $(X_1, \ldots, X_n)$ é um $\overrightarrow{ve}.a.r.$ discreto, de contradomínio C, então a sua lei de probabilidade, $P_{(X_1,\ldots,X_n)}$, é caracterizada por uma função

$$f : \mathbb{R}^n \to [0, 1],$$

designada de *função de probabilidade conjunta de* $(X_1, \ldots, X_n)$, definida por

$$f((a_1, \ldots, a_n)) = \begin{cases} P_{(X_1,\ldots,X_n)}(\{(a_1, \ldots, a_n)\}) & se \quad (a_1, \ldots, a_n) \in C \\ 0 & se \quad c.c. \end{cases}$$

Nota: Observe que $P_{(X_1,\ldots,X_n)}(\{(a_1, \ldots, a_n)\})$ representa

$$P\left((X_1, \ldots, X_n) = (a_1, \ldots, a_n)\right)$$

que se reduz a

$$P(X_1 = a_1, \ldots, X_n = a_n).$$

2. Vector aleatório real discreto

Se $(X_1, \ldots, X_n)$ é um $\overrightarrow{ve}.a.r.$ discreto, de contradomínio C e função de probabilidade conjunta $f : \mathbb{R}^n \to [0,1]$, então, para todo $B \in \mathcal{B}_n(\mathbb{R})$, tem-se

$$P((X_1, X_2, \ldots, X_n) \in B) = \sum_{(a_1, a_2, \ldots, a_n) \in \mathbb{R}^n : (a_1, a_2, \ldots, a_n) \in B \cap C} f((a_1, a_2, \ldots, a_n)).$$

2. Vector aleatório real discreto

$\rightsquigarrow$ Como obter as leis de probabilidade das margens de um $\overrightarrow{ve}.a.r.$ discreto?
Seja $(X_1, X_2, \ldots, X_n)$ um $\overrightarrow{ve}.a.r.$ discreto, de contradomínio C e função de probabilidade conjunta f. Então as v.a.r.'s $X_1, X_2, \ldots, X_n$ são discretas e, para a i-ésima v.a.r. marginal, X_i, tem-se

- o contradomínio:

$$C_{X_i} = \{a \in \mathbb{R} : (x_1, x_2, \ldots, x_{i-1}, a, x_{i+1}, \ldots, x_n) \in C\}$$

- a função de probabilidade: $f_{X_i} : \mathbb{R} \to [0, 1]$, para $a \in C_{X_i}$,

$$f_{X_i}(a) = \sum_{x_1 \in C_{X_1}, \ldots, x_{i-1} \in C_{X_{i-1}}, x_{i+1} \in C_{X_{i+1}}, \ldots, x_n \in C_{X_n}} f((x_1, \ldots, x_{i-1}, a, x_{i+1}, \ldots, x_n))$$

Nota: No caso dos pares aleatórios ($n = 2$), é usual apresentar a função de probabilidade conjunta através de uma tabela de dupla entrada. Tal tabela permite aceder muito facilmente às funções de probabilidade das variáveis marginais.

Exemplo/Exercício da Folha Prática 9:
Considere a experiência aleatória que consiste em efetuar 2 lançamentos consecutivos de um dado equilibrado. Seja (X, Y) o par aleatório em que X representa o número de ases e Y representa o número de faces par obtidas nos dois lançamentos.

a) Determine a função de probabilidade conjunta do par (X, Y).
b) Calcule $P(Y < X)$.
c) Identifique as funções de probabilidade das margens.

3. Vector aleatório real absolutamente contínuo

Definição [$\overrightarrow{ve}.a.r.$ difuso]

Um $\overrightarrow{ve}.a.r.$ $(X_1, X_2, \ldots, X_n)$ diz-se difuso se a sua lei de probabilidade for uma lei de probabilidade difusa sobre $(\mathbb{R}^n, \mathcal{B}_n(\mathbb{R})$, i.e., se

$$P_{(X_1,X_2,\ldots,X_n)}(\{(a_1, a_2, \ldots, a_n)\}) = 0, \quad \forall (a_1, a_2, \ldots, a_n) \in \mathbb{R}^n. \tag{1}$$

Obs.: Note que a igualdade (1) é equivalente a escrever que

$$P((X_1, X_2, \ldots, X_n) = (a_1, a_2, \ldots, a_n)) = 0$$

ou ainda que

$$P(X_1 = a_1, X_2 = a_2, \ldots, X_n = a_n) = 0,$$

para todo o $(a_1, a_2, \ldots, a_n) \in \mathbb{R}^n$.

Tal como no caso univariado, aqui vamos estudar apenas as leis de probabilidade sobre $(\mathbb{R}^n, \mathcal{B}_n(\mathbb{R})$ difusas que são absolutamente contínuas. Estas leis são agora caracterizadas através de uma função densidade de probabilidade sobre $\mathbb{R}^n$.

3. Vector aleatório real absolutamente contínuo

Definição [Função densidade de probabilidade sobre $\mathbb{R}^n$]

Uma função $f : \mathbb{R}^n \to \mathbb{R}$ diz-se uma função densidade de probabilidade sobre $\mathbb{R}^n$ se satisfaz as seguintes condições:

i) $f((x_1, x_2, \ldots, x_n)) \geq 0$, para todo o $(x_1, x_2, \ldots, x_n) \in \mathbb{R}^n$,

ii) f é integrável em $\mathbb{R}^n$ e

$$\int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} f((x_1, x_2, \ldots, x_n)) dx_1 dx_2 \ldots dx_n = 1.$$

3. Vector aleatório real absolutamente contínuo

Definição [$\overrightarrow{ve}.a.r.$ absolutamente contínuo]

Um $\overrightarrow{ve}.a.r.$ $(X_1, X_2, \ldots X_n)$ diz-se absolutamente contínuo se a sua lei de probabilidade é uma lei de probabilidade sobre $(\mathbb{R}^n, \mathcal{B}_n(\mathbb{R}))$ absolutamente contínua, i.e., se existir uma função $f : \mathbb{R}^n \to \mathbb{R}$, função densidade de probabilidade sobre $\mathbb{R}^n$, tal que, para todo o

$$a_1, a_2, \ldots, a_n, b_1, b_2, \ldots, b_n, \text{ com } a_i < b_i, i = 1, \ldots n,$$

se tem

$$P(a_1 < X_1 < b_1, a_2 < X_2 < b_2, \ldots, a_n < X_n < b_n) = \\ = \int_{a_1}^{b_1} \left[\int_{a_2}^{b_2} \ldots \left[\int_{a_n}^{b_n} f((x_1, x_2, \ldots, x_n)) dx_n \right] \ldots dx_2 \right] dx_1.$$

A função f é designada de função densidade de probabilidade conjunta do $\overrightarrow{ve}.a.r.$ $(X_1, X_2, \ldots X_n)$.

3. Vector aleatório real absolutamente contínuo

De um modo geral, se $(X_1, X_2, \ldots, X_n)$ é um $\overrightarrow{ve}.a.r.$ absolutamente contínuo, com função densidade de probabilidade conjunta $f : \mathbb{R}^n \to \mathbb{R}$, tem-se, para todo $B \in \mathcal{B}_n(\mathbb{R})$,

$$P((X_1, X_2, \ldots, X_n) \in B) = \int \underset{B}{\int} \ldots \int f((x_1, x_2, \ldots, x_n)) dx_1 dx_2 \ldots dx_n.$$

3. Vector aleatório real absolutamente contínuo

$\rightsquigarrow$ Como obter as leis de probabilidade das margens de um $\overrightarrow{ve}.a.r.$ absolutamente contínuo?

Se $(X_1, X_2, \ldots, X_n)$ é um $\overrightarrow{ve}.a.r.$ absolutamente contínuo, com função densidade de probabilidade conjunta $f : \mathbb{R}^n \to \mathbb{R}$, então as margens $X_1, X_2, \ldots, X_n$ são v.a.r.'s absolutamente contínuas e, para a i-ésima v.a.r. marginal, X_i, a função densidade de probabilidade $f_{X_i} : \mathbb{R} \to \mathbb{R}$ é dada por

$$f_{X_i}(x) = \int_{-\infty}^{+\infty} \ldots \int_{-\infty}^{+\infty} f((x_1, \ldots, x_{i-1}, x, x_{i+1}, \ldots, x_n))dx_1 \ldots dx_{i-1}dx_{i+1} \ldots dx_n,$$

para $i = 1, \ldots, n$.

3. Vector aleatório real absolutamente contínuo

Exemplo/Exercício da Folha Prática 9:

Seja (X, Y) um par aleatório com função densidade de probabilidade conjunta dada por

$$f((x,y)) = \begin{cases} e^{-(x+y)} & se \quad x > 0,\ y > 0 \\ 0 & se \quad c.c. \end{cases}$$

Comece por mostrar que se trata, de facto, de uma função densidade de probabilidade sobre $\mathbb{R}^n$.

a) Determine as funções densidade de probabilidade das margens.

b) Calcule $P(X \leq Y)$.

c) Calcule, para todo o $u \in \mathbb{R}$, $P(X + Y \leq u)$. Use o resultado obtido para calcular $P(1 < X + Y \leq 2)$.

4. Variáveis aleatórias independentes

Definição [Independência de v.a.r.'s]

Sejam $X_1, X_2, \ldots, X_n$ v.a.r.'s, todas definidas sobre o mesmo espaço de probabilidade $(\Omega, \mathcal{A}, P)$, com leis de probabilidade $P_{X_1}, P_{X_2}, \ldots, P_{X_n}$, respectivamente.

$X_1, X_2, \ldots, X_n$ dizem-se v.a.r.'s independentes se, para todos os $B_1, B_2, \ldots, B_n \in \mathcal{B}(\mathbb{R})$,

$$P_{(X_1, X_2, \ldots, X_n)}(B_1 \times B_2 \times \ldots \times B_n) = P_{X_1}(B_1) P_{X_2}(B_2) \ldots P_{X_n}(B_n), \qquad (2)$$

com $P_{(X_1, X_2, \ldots, X_n)}$ a lei de probabilidade do $\vec{ve}.a.r.$ $(X_1, X_2, \ldots, X_n)$.

Nota: Observe que a condição (2) pode ser escrita do seguinte modo:

$$P(X_1 \in B_1, X_2 \in B_2, \ldots, X_n \in B_n) = P(X_1 \in B_1) P(X_2 \in B_2) \ldots P(X_n \in B_n).$$

4. Variáveis aleatórias independentes

Teorema

Sejam $X_1, X_2, \ldots, X_n$ v.a.r.'s independentes e $g_1, g_2, \ldots, g_n$ funções tais que, para $i = 1, \ldots, n$,

- $g_i : D_i \subseteq \mathbb{R} \to \mathbb{R}$,
- $g_i \circ X_i$ é uma v.a.r..

Então $g_1 \circ X_1, g_2 \circ X_2, \ldots, g_n \circ X_n$ também são v.a.r.'s independentes.

[Sem demonstração] (Ver Lopes & Gonçalves)

Exemplos: Da utilização deste teorema resulta, por exemplo, que:

- Se X e Y são v.a.r.'s independentes então, para todo $k \in \mathbb{N}$, as v.a.r.'s X^k e Y^k também são independentes.
- Se X e Y são v.a.r.'s independentes então, se $Y > 0$, as v.a.r.'s e^X e $\log(Y)$ também são independentes.

4. Variáveis aleatórias independentes

$\rightsquigarrow$ Condições necessárias e suficientes para independência de v.a.r.'s:

Provar, por definição, que $X_1, X_2, \ldots, X_n$ são v.a.r.'s independentes é extremamente difícil. Precisamos assim de ter "formas alternativas"para o fazer.

1) $X_1, X_2, \ldots, X_n$ são v.a.r.'s independentes se e só se a função de distribuição conjunta, F, do $\overrightarrow{ve}.a.r.$ $(X_1, X_2, \ldots, X_n)$ verifica a seguinte condição:

$$\forall (c_1, c_2, \ldots, c_n) \in \mathbb{R}^n, \quad F((c_1, c_2, \ldots, c_n)) = \prod_{i=1}^{n} F_{X_i}(c_i),$$

onde F_{X_i} representa a função de distribuição de X_i, $i = 1, \ldots, n$.
[Demonstração] $(\Rightarrow)$ é consequência imediata de definição de independência de v.a.r.'s; ver Lopes & Gonçalves para $(\Leftarrow)$.

$\rightsquigarrow$ Condições necessárias e suficientes para independência de v.a.r.'s:

2) Sejam $X_1, X_2, \ldots, X_n$ v.a.r.'s discretas, cujos contradomínios são $C_{X_1}, C_{X_2}, \ldots, C_{X_n}$, respectivamente. Neste caso, $X_1, X_2, \ldots, X_n$ são v.a.r.'s independentes se e só

$$\underset{a_1 \in C_{X_1}, a_2 \in C_{X_2}, \ldots, a_n \in C_{X_n}}{\forall}, \quad P(X_1 = a_1, X_2 = a_2, \ldots, X_n = a_n) = \prod_{i=1}^{n} P(X_i = a_i). \tag{3}$$

[Sem demonstração] (Ver Lopes & Gonçalves)

Nota: Condição (3) equivale a dizer que a função de probabilidade conjunta do $\overrightarrow{ve}.a.r.$ $(X_1, X_2, \ldots, X_n)$ é igual ao produto das funções de probabilidade das v.a.r.'s marginais.

$\rightsquigarrow$ Condições necessárias e suficientes para independência de v.a.r.'s:

3) Uma condição necessária e suficiente para que um $\vec{ve}.a.r.$ absolutamente contínuo, $(X_1, X_2, \ldots, X_n)$, tenha as margens independentes é que a sua função densidade de probabilidade conjunta, $f_{(X_1,X_2,\ldots,X_n)}$, seja da forma

$$f_{(X_1,X_2,\ldots,X_n)}((x_1, x_2, \ldots, x_n)) = \prod_{i=1}^{n} f_{X_i}(x_i), \quad \forall (x_1, x_2, \ldots, x_n) \in \mathbb{R}^n,$$

onde $f_{X_1}, f_{X_2}, \ldots, f_{X_n}$ são funções densidade de probabilidade das margens $X_1, X_2, \ldots, X_n$, respectivamente.
[Sem demonstração] (Ver Lopes & Gonçalves)

Exemplo/Exercício: Para os pares aleatórios, (X, Y), trabalhados nos exemplos das Secções 2 e 3, diga se as variáveis marginais, X e Y, são independentes.

4. Variáveis aleatórias independentes

Antes de prosseguir com mais resultados relativos à independência de v.a.r.'s, vejamos a definição de esperança matemática (ou valor médio) de uma função de um $\overrightarrow{ve}.a.r.$.

Definição [Esperança matemática de uma função de um $\overrightarrow{ve}.a.r.$ discreto]

Sejam $(X_1, \ldots, X_n)$ um $\overrightarrow{ve}.a.r.$ e $\phi : \mathbb{R}^n \to \mathbb{R}$ uma função tal que $\phi((X_1, \ldots, X_n))$ é uma v.a.r..

i) Se $(X_1, \ldots, X_n)$ é um $\overrightarrow{ve}.a.r.$ discreto, de contradomínio C e função de probabilidade conjunta f, e tal que

$$\sum_{(x_1,\ldots,x_n)\in C} |\phi((x_1, \ldots, x_n))| f((x_1, \ldots, x_n)) < +\infty$$

então $E[\phi((X_1, \ldots, X_n))]$ existe e é dada por

$$E[\phi((X_1, \ldots, X_n))] = \sum_{(x_1,\ldots,x_n)\in C} \phi((x_1, \ldots, x_n)) f((x_1, \ldots, x_n)).$$

Definição [Esperança matemática de uma função de um $\overrightarrow{ve}.a.r.$ contínuo]

ii) Se $(X_1,\ldots,X_n)$ é um $\overrightarrow{ve}.a.r.$ absolutamente contínuo, com função densidade de probabilidade conjunta f, e tal que

$$\int_{-\infty}^{+\infty}\ldots\int_{-\infty}^{+\infty}|\phi((x_1,\ldots,x_n))|f((x_1,\ldots,x_n))dx_1\ldots dx_n<+\infty$$

então $E[\phi((X_1,\ldots,X_n))]$ existe e é dada por

$$E[\phi((X_1,\ldots,X_n))]=\int_{-\infty}^{+\infty}\ldots\int_{-\infty}^{+\infty}\phi((x_1,\ldots,x_n))f((x_1,\ldots,x_n))dx_1\ldots dx_n.$$

Observação:
A partir desta última definição, facilmente se conclui (TPC) que, se $(X_1, \ldots, X_n)$ é um $\overrightarrow{ve}.a.r.$ possuindo esperança matemática (i.e., $E[X_i]$ existe para todo $i = 1, \ldots, n$), então

$$E[a_1X_1 + a_2X_2 + \ldots + a_nX_n] = \sum_{i=1}^{n} a_iE[X_i],$$

para quaisquer constantes reais $a_1, a_2, \ldots, a_n$.

Teorema

Seja $(X_1, \ldots, X_n)$ um $\overrightarrow{ve}.a.r.$ possuindo esperança matemática (i.e., $E[X_i]$ existe para todo $i = 1, \ldots, n$) e com margens independentes. Então

$$E\left[\prod_{i=1}^{n} X_i\right] = \prod_{i=1}^{n} E\left[X_i\right].$$

Demonstração: Vai ser feita apenas para o caso em que $(X_1, X_2, \ldots, X_n)$ é um $\overrightarrow{ve}.a.r.$ absolutamente contínuo (TPC: caso discreto).

Comecemos por provar que

$$E\left[\prod_{i=1}^{n} X_i\right]$$

existe.

Sejam f a função densidade de probabilidade conjunta do $\overrightarrow{ve}.a.r.$ e f_{X_i} a função densidade de probabilidade da v.a.r. X_i, $i = 1, \ldots, n$.

$$\begin{aligned}
E[|X_1 \ldots X_n|] &= \int_{-\infty}^{+\infty} \ldots \int_{-\infty}^{+\infty} |x_1 \ldots x_n| \, f((x_1, \ldots, x_n)) \, dx_1 \ldots dx_n \\
&= \int_{-\infty}^{+\infty} \ldots \int_{-\infty}^{+\infty} |x_1| \ldots |x_n| \, f_{X_1}(x_1) \ldots f_{X_n}(x_n) \, dx_1 \ldots dx_n \\
&= \underbrace{\int_{-\infty}^{+\infty} |x_1| f_{X_1}(x_1) dx_1}_{<+\infty} \ \ldots \ \underbrace{\int_{-\infty}^{+\infty} |x_n| f_{X_n}(x_n) dx_n}_{<+\infty} \\
&< +\infty.
\end{aligned}$$

E de modo análogo tem-se

$$\begin{aligned}
E[X_1 X_2 \ldots X_n] &= \int_{-\infty}^{+\infty} \ldots \int_{-\infty}^{+\infty} x_1 \ldots x_n \, f((x_1, \ldots, x_n)) \, dx_1 \ldots dx_n \\
&= \underbrace{\int_{-\infty}^{+\infty} x_1 f_{X_1}(x_1) dx_1}_{E[X_1]} \ \ldots \ \underbrace{\int_{-\infty}^{+\infty} x_n f_{X_n}(x_n) dx_n}_{E[X_n]}
\end{aligned}$$

c.q.d

5. Covariância e correlação

Recordar que, para caracterizar a dispersão de uma v.a.r. em torno do seu valor médio, fazemos uso de variância. Para um $\overrightarrow{ve}.a.r.$ a questão é mais complexa e é preciso introduzir um parâmetro que permita avaliar o tipo e o grau de dependência entre os pares de variáveis marginais. Para esse efeito, serão aqui apresentados os conceitos, e propriedades, da *covariância* e do *coeficiente de correlação*.

Definição [Covariância]

Sejam X e Y v.a.r.'s definidas sobre o mesmo espaço de probabilidade $(\Omega, \mathcal{A}, P)$ e ambas admitindo momento de segunda ordem. Chama-se covariância entre X e Y, denota-se por $Cov(X,Y)$, a

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

De certo modo, a covariância mede a dispersão do par aleatório (X,Y) relativamente ao par $(E[X], E[Y])$.

Observação: $Cov(X,X) = Var[X]$.

$\rightsquigarrow$ Propriedades da covariância:

Sejam $X, Y, X_1, X_2, \ldots, X_n$ v.a.r.'s, todas admitindo momento de segunda ordem.

1) $\forall a, b, c, d \in \mathbb{R}$, $Cov(aX + c, bY + d) = abCov(X, Y)$.

Demonstração:

$$\begin{aligned} Cov(aX + c, bY + d) &= E[(aX + c - E(aX + c))\,(bY + d - E(bY + d))] \\ &= E[a(X - E(X))\,b(Y - E(Y))] \\ &= ab\,E[(X - E(X))\,(Y - E(Y))] \\ &= ab\,Cov(X, Y) \end{aligned}$$

c.q.d.

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

Demonstração:

$$\begin{aligned} Cov(X,Y) &= E[(X - E(X))\,(Y - E(Y))] \\ &= E[XY - XE(Y) - YE(X) + E(X)E(Y)] \\ &= E[XY] - E[X]E(Y) - E[Y]E(X) + E(X)E(Y) \\ &= E[XY] - E[X]E[Y] \qquad \text{c.q.d.} \end{aligned}$$

Nota: Na dedução acima usamos o facto de $E[XY]$ existir. É possível mostrar (ver Lopes & Gonçalves) que, se X e Y admitem momento de segunda ordem, então $E[|XY|] < +\infty$, pelo que $E[XY]$ existe.

3) $Var\left[\sum_{i=1}^{n} X_i\right] = \sum_{i=1}^{n} Var[X_i] + 2\sum_{i=1}^{n}\sum_{j=i+1}^{n} Cov(X_i, X_j)$

Demonstração:

$$
\begin{aligned}
Var\left[\sum_{i=1}^{n} X_i\right] &= E\left[\left(\sum_{i=1}^{n} X_i - E\left[\sum_{i=1}^{n} X_i\right]\right)^2\right] = E\left[\left(\sum_{i=1}^{n}(X_i - E[X_i])\right)^2\right] \\
&= E\left[\sum_{i=1}^{n}(X_i - E[X_i])^2 + 2\sum_{i=1}^{n}\sum_{j=i+1}^{n}(X_i - E(X_i))(X_j - E(X_j))\right] \\
&= \sum_{i=1}^{n} E[(X_i - E(X_i))^2] + 2\sum_{i=1}^{n}\sum_{j=i+1}^{n} E\left[(X_i - E(X_i))(X_j - E(X_j))\right] \\
&= \sum_{i=1}^{n} Var[X_i] + 2\sum_{i=1}^{n}\sum_{j=i+1}^{n} Cov(X_i, X_j) \qquad \text{c.q.d}
\end{aligned}
$$

Observação: Se $X_1, X_2, \ldots, X_n$ são v.a.r.'s independentes, tem-se, para $i \neq j$, que $Cov(X_i, X_j) = 0$ e, consequentemente

$$Var\left[\sum_{i=1}^{n} X_i\right] = \sum_{i=1}^{n} Var[X_i].$$

A covariância entre duas v.a.r.'s permite definir um outro parâmetro, o coeficiente de correlação, que mede o grau e o tipo de dependência existente entre as v.a.r.'s.

Definição [Coeficiente de correlação]

Sejam X e Y v.a.r.'s, definidas sobre o mesmo espaço de probabilidade $(\Omega, \mathcal{A}, P)$, com variâncias σ_X^2 e σ_Y^2 estritamente positivas. Chama-se coeficiente de correlação entre X e Y, denota-se por $\rho(X,Y)$, ao número real

$$\rho(X,Y) = Cov\left(\frac{X - E(X)}{\sigma_X}, \frac{Y - E(Y)}{\sigma_Y}\right),$$

com σ_X e σ_Y o desvio-padrão de X e Y, respectivamente.

5. Covariância e correlação

⇝ Algumas observações sobre $\rho(X,Y)$:

1) Pelas propriedades da covariância é simples deduzir que

$$\rho(X,Y) = \frac{Cov(X,Y)}{\sigma_X\, \sigma_Y}.$$

2) É possível mostrar (ver Lopes & Gonçalves) que $|\rho(X,Y)| \leq 1$ e que

$$|\rho(X,Y)| = 1 \Leftrightarrow \underset{a,b,c\in\mathbb{R}:ab\neq 0}{\exists} \;:\; P(aX+bY+c=0) = 1.$$

Este resultado permite-nos concluir que o coeficiente de correlação quantifica a existência de uma **relação linear** entre X e Y.

3) Quando $\rho(X,Y)=0$ dizemos que X e Y não estão correlacionadas (significa que não existe uma relação linear entre X e Y).

4) Se X e Y são independentes tem-se, obviamente, $\rho(X,Y)=0$. No entanto, o recíproco é falso: podemos ter v.a.r.'s não correlacionadas mas que não são independentes.

Exemplo/Exercício: Calcular $cov(X,Y)$ e $\rho(X,Y)$ para as variáveis X e Y trabalhadas nos exemplos das Secções 2 e 3.

6. Algumas leis de probabilidade sobre $(\mathbb{R}^n, \mathcal{B}_n(\mathbb{R}))$ mais conhecidas

Entre as leis discretas multivariadas, a mais conhecida é a distribuição *Multinomial*, que generaliza a distribuição Binomial (univariada).

Entre as leis absolutamente contínuas multivariadas, a mais conhecida é a distribuição *Normal multivariada*, que generaliza a distribuição Normal (univariada).

6. Algumas leis de probabilidade sobre $(\mathbb{R}^n, \mathcal{B}_n(\mathbb{R}))$ mais conhecidas

$\rightsquigarrow$ Lei Multinomial:

Considere uma experiência aleatória, ξ, na qual ocorrem, **em alternativa**, "sucesso de tipo 1", "sucesso de tipo 2", ..., "sucesso de tipo r-1" ou "insucesso" (ie., não ocorre qualquer tipo de sucesso), com probabilidades $p_1, p_2, \ldots, p_{r-1}$ e $p_r = 1 - (p_1 + \ldots + p_{r-1})$, respectivamente, sendo $r \geq 2$. Seja $(X_1, X_2, \ldots, X_{r-1})$ o $\overrightarrow{ve}.a.r.$ em que X_i é a v.a.r. que representa o número de vezes que ocorreu o sucesso de tipo i em n repetições independentes de ξ, $i = 1, \ldots, r-1$. A função de probabilidade conjunta deste $\overrightarrow{ve}.a.r.$ é dada por:

$$P(X_1 = n_1, X_2 = n_2, \ldots, X_{r-1} = n_{r-1}) = \frac{n!}{n_1! n_2! \ldots n_{r-1}! n_r!} p_1^{n_1} p_2^{n_2} \ldots p_{r-1}^{n_{r-1}} p_r^{n_r},$$

em que $n_i \in \mathbb{N}_0$, $i = 1, \ldots, n_r$ e, adicionalmente, satisfazem as condições

$$n_1 + n_2 + \ldots + n_{r-1} \leq n \quad \text{e} \quad n_r = n - (n_1 + n_2 + \ldots + n_{r-1}).$$

Nestas condições, diz-se que o $\overrightarrow{ve}.a.r.$ $(X_1, X_2, \ldots, X_{r-1})$ segue a lei Multinomial, com parâmetros n e $p_1, p_2, \ldots, p_{r-1}$, e abrevia-se por

$$(X_1, X_2, \ldots, X_{r-1}) \sim M(n; p_1, p_2, \ldots, p_{r-1}).$$

Exemplos/Exercícios da Folha Prática 9:

Numa lotaria com 10.000 bilhetes, numerados de 0000 a 9999, qual é a probabilidade de o primeiro prémio ir para um bilhete com exatamente dois dígitos ímpar e exatamente um zero?

$\rightsquigarrow$ Lei Normal Multivariada:

Diz-se que um $\overrightarrow{ve}.a.r.$ $(X_1, X_2, \ldots, X_p)$ absolutamente contínuo segue a lei Normal Multivariada, com parâmetros $\mathbf{u} = [u_1\, u_2\, \ldots\, u_p]^\top \in \mathbb{R}^p$ e $\Sigma = [\sigma_{ij}]_{i,j=1}^p$, se a função densidade de probabilidade conjunta do vetor é dada por

$$f(\mathbf{x}) = \frac{1}{(2\pi)^{p/2}\sqrt{det(\Sigma)}} \exp\left\{-\frac{1}{2}(\mathbf{x}-\mathbf{u})^\top \Sigma^{-1}(\mathbf{x}-\mathbf{u})\right\},$$

para todo o $\mathbf{x} = [x_1\, x_2\, \ldots\, x_p]^\top \in \mathbb{R}^p$, em que Σ é uma matriz real, quadrada de ordem p, invertível, simétrica e positiva definida. Abrevia-se por

$$(X_1, X_2, \ldots, X_p) \sim N_p(\mathbf{u}, \Sigma).$$

É possível mostrar que $\mathbf{u}$ é o vetor valor médio do $\overrightarrow{ve}.a.r.$ $(X_1, X_2, \ldots, X_p)$ e que Σ é a respectiva matriz das covariâncias, i.e,

$$\sigma_{ij} = Cov(X_i, X_j), \quad i, j \in \{1, \ldots, p\}.$$

Exemplo/Exercício da Folha Prática 9:
Sejam $X_1, X_2, \ldots, X_p$ v.a.r.'s independentes e tais que

$$X_i \sim N(u_i, \sigma_i^2),\ i = 1, \ldots, p.$$

Mostre que o $\overrightarrow{ve}.a.r.$ $(X_1,\, X_2,\, \ldots\, , X_p)$ é tal que

$$(X_1\, X_2\, \ldots\, X_p) \sim N_p(\mathbf{u}, \Sigma),$$

com $\mathbf{u} = [u_1\, u_2\, \ldots\, u_p]^\top$ e $\Sigma = Diag(\sigma_1^2, \sigma_2^2, \ldots, \sigma_p^2)$.