2 Variáveis Aleatórias e Distribuições de Probabilidade Discretas

2.1 Introdução

Exemplos ilustrativos

1. ξ: lançamento de duas moedas equilibradas, uma a seguir à outra, anotando as faces que ficam voltadas para cima.

   $\Omega = \{(C, C), (C, \overline{C}), (\overline{C}, C), (\overline{C}, \overline{C})\}$, com C :"saída de cara".

   Por exemplo, pode definir-se a **variável aleatória** (**v.a.**) que representa o **número de caras** obtidas no lançamento das duas moedas.

2. ξ: lançamento de dois dados, anotando os números de pontos das faces que ficam voltadas para cima.

   $\Omega = \{(i, j), i, j = 1, 2, 3, 4, 5, 6\}$.

   Por exemplo, pode definir-se a **v.a.** que representa a **soma dos pontos** obtidos.

3. ξ: medição da altura de uma pessoa, escolhida ao acaso.

   Ω é o conjunto de todas as alturas atribuíveis a uma pessoa.

   Pode definir-se a **v.a.** que representa a **altura de uma pessoa**.

Embora da experimentação nem sempre resultem valores numéricos, o conceito de **variável aleatória permite associar valores numéricos (números ou intervalos reais) aos elementos de Ω (sejam estes numéricos ou não)**, relação esta com grande interesse prático.

Definição
Seja Ω um espaço de resultados associado a uma experiência aleatória. Designa-se por **variável aleatória** (v.a.) uma função $\mathbf{X}$ de domínio Ω e contradomínio em $\mathbb{R}$.

$$\mathbf{X} : \Omega \longrightarrow \mathbb{R}$$

$$w \longrightarrow X(w)$$

- $\forall A \subseteq \Omega, \;\; X(A) = \{X(w) : w \in A\}$ (imagem de A por X)
- $\forall E \subseteq \mathbb{R}$, $X^{-1}(E) = \{w \in \Omega : X(w) \in E\}$(imagem inversa de E por X)

  Note que se $w \in X^{-1}(E) \Rightarrow X(w) \in E$.

Assim, uma v.a. é uma aplicação que estabelece uma relação entre subconjuntos de Ω e subconjuntos de $\mathbb{R}$ e **vice-versa**! Pode, assim, ser introduzida em $\mathbb{R}$ uma **probabilidade**, denotada por P_X, que permite calcular probabilidades das v.a.'s variarem em subconjuntos reais

$$P_X(E) = P(\underbrace{X^{-1}(E)}_{\subseteq \Omega}), \;\forall E \subseteq \mathbb{R}.$$

A P_X é usual chamar-se lei de probabilidade da v.a. X.
Em termos de notação, $P_X(E) = P(X^{-1}(E)) = P(X \in E)$

2.2 Variáveis Aleatórias Discretas

Uma **v.a.** X diz-se **discreta** se $\exists S \subset \mathbb{R}$: S é finito ou infinito numerável tal que $P_X(S) = P(X \in S) = 1$.
Ao menor subconjunto de $\mathbb{R}$ que tem probabilidade 1 chamamos *Suporte* de X.

Chama-se **função de probabilidade** de X à função $p : \mathbb{R} \longrightarrow [0, 1]$ definida por

$$p(x) = P(X = x), \;\forall x \in \mathbb{R}.$$

(1) $0 \leq p(x) \leq 1, \forall x \in \mathbb{R}$

(2) $\displaystyle\sum_x p(x) = 1.$

- Sendo S o suporte de X, $\mathbb{R} = S \cup \overline{S}$. Então $p(x) > 0$ se $x \in S$, e $p(x) = 0$ se $x \in \overline{S}$.

- O conhecimento da função de probabilidade de X, p, implica o conhecimento da lei de probabilidade de X, P_X, e por isso p também é designada por lei de probabilidade de X.
- Identificamos as v.a.'s por letras maiúsculas (X) e as suas concretizações por letras minúsculas (x).

Chama-se **função distribuição de** X à função $F:\, \mathbb{R} \longrightarrow [0,1]$ definida por

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

Se X tem suporte S então

$$F(x) = \underbrace{\sum_{\mathbf{a}\, \in S \cap]-\infty, x]}}_{\text{“valores do suporte de X e que são inferiores ou iguais a } x\text{”}} P(X = \mathbf{a}).$$

Propriedades da função distribuição de uma v.a. discreta

(1) $\forall x \in \mathbb{R},\ 0 \leq F(x) \leq 1$ (F é uma função limitada)

(2) $\forall a, b \in \mathbb{R}$, se $a < b \Rightarrow F(a) \leq F(b)$ (F é não decrescente)

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

(4) $\forall a \in \mathbb{R},\ \lim_{x \to a^+} F(x) = F(a)$ (F é contínua à direita)

(5) F admite um número finito ou infinito numerável de pontos de descontinuidade (pontos do Suporte de X)

(6) A cada v.a. X corresponde uma e uma só função distribuição

(7) Para $\forall\, a, b \in \mathbb{R}$,

$$\begin{aligned}
P(a < X < b) &= P(X < b) - P(X \leq a) \\
&= \lim_{x \to b^-} F(x) - F(a); \\
P(a \leq X \leq b) &= F(b) - F(a) - P(X = a); \\
P(a \leq X < b) &= F(b) - P(X = b) - F(a) + P(X = a)
\end{aligned}$$

Independência de Variáveis Aleatórias[a]

- As variáveis aleatórias X e Y dizem-se independentes se e só se

$$\forall x, y \in \mathbb{R},\ P(X \leq x \cap Y \leq y) = P(X \leq x) \times P(Y \leq y).$$

- As n $(n \geq 2)$ variáveis aleatórias $X_1, X_2, ..., X_n$ dizem-se independentes se e só se

$$\forall x_1, x_2, ..., x_n \in \mathbb{R}, P(X_1 \leq x_1 \cap X_2 \leq x_2 \cap ... \cap X_n \leq x_n) = \prod_{i=1}^{n} P(X_i \leq x_i).$$

[a] Voltaremos a este conceito mais tarde. Nesta fase, recorde independência entre acontecimentos (pág. 5).

Parâmetros de localização e de dispersão

Seja X uma v.a. discreta de suporte S. A **esperança matemática de** X (ou **valor médio de** X ou **valor esperado de** X), caso exista, é definida por

$$E(X) = \sum_{x \in S} x P(X = x) \quad .$$

$-$A existência do valor médio de uma v.a. X depende da convergência da série anterior.
$-$Se a v.a. X é discreta com suporte finito, então $E(X)$ existe sempre!
$-$O valor médio de uma v.a. é um parâmetro de localização (tendência central).

- Se $E(X) = 0$ diz-se que a v.a. X é centrada.

Propriedades da esperança matemática

1. Seja $X = c$, $c \in \mathbb{R}$. $E(X) = E(c) = c$.

2. Seja X uma v.a. discreta de suporte S e $g : \mathbb{R} \to \mathbb{R}$ é tal que $g(X)$ é uma v.a. discreta. Tem-se
$$E[\,g(X)\,] = \sum_{x \in S} g(x) P(X = x).$$
   - Seja $Y = g(X) = aX + b$, $\forall a, \, b \in \mathbb{R}$. Se $E(X)$ existe então $E(Y) = aE(X) + b$.
   - Seja $k \in \mathbb{N}$ (arbitrariamente fixo) e $g(X) = X^k$.
   $E(X^k) = \sum_{x \in S} x^k P(X = x)$ (*momento de ordem* k de X).

3. Se $E(X)$ existe então $|E(X)| \leq E(|X|)$.

4. Sejam X_1, $X_2, ..., X_n$ $(n \in \mathbb{N})$ v.a.'s todas definidas sobre o mesmo espaço Ω e tais que $E(X_i)$ existe para $i = 1, 2, ..., n$. Então
$$E\left[\sum_{i=1}^{n} a_i X_i \, + \, b\right] = \sum_{i=1}^{n} a_i E(X_i) \, + \, b,$$
para $\forall \, a_1, a_2, ..., a_n, \, b \in \mathbb{R}$. **(Linearidade da esperança)**

   - $E\left[\sum_{i=1}^{n} X_i\right] = \sum_{i=1}^{n} E(X_i).$

Seja X uma v.a. discreta de suporte S e esperança $E(X) = \mu$.
A **variância de X**, denotada por $V(X)$, $Var(X)$ ou σ^2, é definida por

$$\begin{aligned} V(X) &= E\left[(X-\mu)^2\right] \\ &= \sum_{x \in S} (x-\mu)^2 P(X=x) \end{aligned}$$

O **desvio-padrão de X** é a raiz quadrada positiva da variância de X, e denota-se por $\sigma(X) = \sigma = \sqrt{V(X)}$.

A variância e o desvio padrão de uma v.a. X são ambos parâmetros de dispersão; no entanto, o desvio padrão é mais usado, uma vez que vem expresso nas mesmas unidades de X.

Propriedades da variância

1. Seja X uma v.a. tal que $E(X)$ e $E(X^2)$ existem. Então

$$V(X) = E(X^2) - E^2(X).$$

2. Se $X = c,\ c \in \mathbb{R}$, então $V(X) = 0$.

3. Seja X uma v.a. tal que $E(X)$ e $E(X^2)$ existem e $a,\ b \in \mathbb{R}$. Então

$$V(aX+b) = a^2 V(X).$$

   - $V(X+b) = V(X)$;
   - $V(aX) = a^2 V(X)$;
   - $V(-X) = V(X)$.

Outros parâmetros: Mediana e Quantis

- Chama-se **mediana** de X ao número real M_d tal que

$$F(M_d) = F(M_d^+) \geq 0.5 \quad \wedge \quad F(M_d^-) \leq 0.5$$

A mediana é um parâmetro de localização, alternativo (ou complementar) ao valor esperado.

- Seja $p \in]0,1[$. Chama-se **quantil de ordem p de X** ao número real x_p tal que

$$F(x_p^+) \geq p \quad \wedge \quad F(x_p^-) \leq p$$

2.3 Distribuições Especiais Discretas

Nesta secção vamos focar em algumas famílias de distribuições discretas que modelam um vasto leque de situações.

1- Considere um lote com 5 bolas, das quais 2 são brancas e 3 são vermelhas. São extraídas 3 bolas do lote, com reposição. Qual é a probabilidade de saírem 2 bolas brancas? E qual é a probabilidade de saírem, no máximo, 2 bolas brancas?

2- Um avião comercial tem 4 motores independentes e num voo a probabilidade de cada motor funcionar, sem avarias, é de 0.9. Qual a probabilidade do avião fazer uma viagem segura, se para isso precisar de pelo menos 2 motores a funcionar corretamente?

3- Um material radioativo emite um certo tipo de partículas a uma taxa média de duas por milissegundo, e segue uma distribuição de *Poisson*. Qual é a probabilidade de:

serem emitidas 3 partículas num milissegundo?

serem emitidas 5 partículas em dois milissegundos?

serem emitidas pelo menos 3 partículas em dois milissegundos?

2.3.1 Distribuição de Bernoulli. Distribuição Binomial

Considere uma determinada experiência aleatória ξ, em que pode ocorrer o acontecimento A com $P(A) = p$, $p \in]0,1[$, ou o seu complementar $\overline{A}$. Considere agora a variável aleatória X_1:

$$X_1 = \begin{cases} 1 & \text{se} \quad \text{A ocorre} \\ 0 & \text{se} \quad \text{A não ocorre} \end{cases} .$$

Então X_1 é uma v.a. discreta de Suporte $S = \{0,1\}$ e função de probabilidade

$$p(x) = \begin{cases} 1-p & \text{se} \quad x = 0 \\ p & \text{se} \quad x = 1 \\ 0 & \text{se} \quad x \in \overline{S} \end{cases} = \begin{cases} p^x\,(1-p)^{1-x} & \text{se} \quad x \in \{0,1\} \\ 0 & \text{se} \quad x \in \overline{S} \end{cases} .$$

- A função (lei) de probabilidade assim definida é denominada por **lei de Bernoulli de parâmetro p**, denotada por $\mathcal{B}(\mathbf{p})$ (ou **Bernoulli**$(\mathbf{p})$).

- Diz-se que a v.a. **$\mathbf{X_1}$ segue uma lei de Bernoulli** (ou **$\mathbf{X_1}$ tem distribuição de Bernoulli**) **de parâmetro p**, e escreve-se simbolicamente $\mathbf{X_1} \sim \mathcal{B}(\mathbf{p})$.

A distribuição de Bernoulli está associada a experiências aleatórias com apenas dois resultados possíveis, usualmente denominados por **sucesso** e **insucesso**, com probabilidade de ocorrer sucesso p (o seu parâmetro!) e insucesso $q = 1 - p$. Uma experiência aleatória assim caracterizada chama-se **prova (ou tentativa) de Bernoulli**.

- Se $X_1 \sim \mathcal{B}(p)$ então

$$E(X_1) = p$$
$$V(X_1) = p(1-p)$$

Suponha agora que a experiência ξ é repetida (realizada) **n** vezes, com $\mathbf{n > 1}$, sempre nas mesmas condições. Note que os resultados das sucessivas repetições (A ou $\overline{A}$) são independentes, e $P(A) = p$ em qualquer repetição. Considere agora a variável aleatória

X = número de vezes que ocorre A, nas n realizações da experiência aleatória

- O suporte de X é $S = \{0, 1, ..., n\}$

- A probabilidade (por exemplo) de ocorrer exatamente x vezes o acontecimento A, **e de seguida** $n - x$ vezes $\overline{A}$, é dada por

$$P(\underbrace{A \cap ... \cap A}_{x \ vezes} \cap \underbrace{\overline{A} \cap ... \cap \overline{A}}_{n-x \ vezes}) = \underbrace{P(A)...P(A)}_{x \ vezes} \underbrace{P(\overline{A})...P(\overline{A})}_{n-x \ vezes} = p^x (1-p)^{n-x}$$

- O número de maneiras distintas de ocorrer x vezes o acontecimento A, nas n realizações, é $C_x^n = \begin{pmatrix} n \\ x \end{pmatrix} = \dfrac{n!}{(n-x)!x!}$

- A probabilidade de ocorrer exatamente x vezes o acontecimento A, nas n realizações, é dada por

$$P(X = x) = \begin{pmatrix} n \\ x \end{pmatrix} p^x (1-p)^{n-x}, \ x \in \{0, 1, ..., n\}.$$

X é uma v.a. discreta de Suporte $S = \{0, 1, ..., n\}$ e função de probabilidade dada por

$$p(x) = P(X = x) = \begin{cases} \begin{pmatrix} n \\ x \end{pmatrix} p^x (1-p)^{n-x} & \text{se} \quad x \in \{0, 1, ..., n\} \\ 0 & \text{se} \quad x \in \overline{S} \end{cases}.$$

- $P(X = x) > 0, \forall x \in \{0, 1, ..., n\}$
- $\displaystyle\sum_{x=0}^{n} P(X = x) = 1$

Diz-se que a v.a. **X segue uma lei Binomial** (ou X **tem distribuição Binomial**) **de parâmetros n** e **p**, e escreve-se simbolicamente $\mathbf{X \sim \mathcal{B}(n, p)}$.

- Se $X \sim \mathcal{B}(n, p)$ então $E(X) = np$
  $V(X) = np(1-p)$

- Se $X \sim \mathcal{B}(n, p)$ então $Y = n - X \sim \mathcal{B}(n, 1-p)$ (Interpretação de Y ao cuidado do aluno!)

A lei Binomial descreve a contagem do número de *sucessos* em n repetições, independentes, de uma prova de Bernoulli com probabilidade de sucesso p. Note que se X_i representar o número de sucessos (0 ou 1) obtidos na prova i, então $X_i \sim \mathcal{B}(p)$, $i = 1, 2, ..., n$, e o número de sucessos obtidos nas n provas é a v.a.

$$X = \sum_{i=1}^{n} X_i \sim \mathcal{B}(n, p).$$

Uma das aplicações mais interessantes da lei Binomial é a contagem do número de elementos de determinado tipo numa amostra de dimensão n, quando selecionados com reposição.

Nota A família de distribuições Binomial depende apenas dos parâmetros n e p. Esta está implementada na maioria das máquinas de calcular, assim como tabelada nos livros da área. Em particular, nas aulas de ME podem (ou devem, em falta de alternativa) ser consultadas as tabelas da disciplina onde constam as probabilidades

$$F(x) = P(X \leq x) = \sum_{k=0}^{x} \binom{n}{k} p^k (1-p)^{n-k} \text{ para } x \in \{0, 1, ..., n\}.$$

2.3.2 Distribuição Hipergeométrica

Suponha que tem um número finito de N objetos, dos quais M são de um tipo (*sucesso*) e os restantes $N - M$ de outro tipo (*insucesso*). Considere a experiência aleatória
ξ: extração, ao acaso, de um objeto (e regista-se se é sucesso ou não). Suponha agora que esta mesma experiência é realizada n vezes, isto é, extração sucessiva, e sem reposição, de n dos N objetos (note que os resultados das sucessivas experiências não são independentes). Defina-se a v.a.

X = número de objetos do tipo sucesso obtidos nas n extrações

- $x \leq n$ e $x \leq M \iff x \leq min(n, M)$

  $n - x \leq n$ e $n - x \leq N - M \iff x \geq 0$ e $x \geq n - (N - M)$

  $\iff x \geq max(0, n - (N - M))$

  Então

  $$S = \{\max(0, n - (N - M)), 1, \ldots, \min(n, M)\}$$

- O número de casos favoráveis à saída de x sucessos, nas n extrações, é $\binom{M}{x}\binom{N-M}{n-x}$

- O número de casos possíveis nas n extrações, é $\binom{N}{n}$

- A probabilidade de obter exatamente x sucessos, nas n extrações, é dada por

$$P(X=x)=\frac{\begin{pmatrix} M \\ x \end{pmatrix}\begin{pmatrix} N-M \\ n-x \end{pmatrix}}{\begin{pmatrix} N \\ n \end{pmatrix}}, \quad \text{para } x\in S$$

X é então uma v.a. discreta de Suporte $S=\{\max(0,n-(N-M)),1,\ldots,\min(n,M)\}$ e função de probabilidade dada por

$$p(x)=P(X=x)=\begin{cases} \dfrac{\begin{pmatrix} M \\ x \end{pmatrix}\begin{pmatrix} N-M \\ n-x \end{pmatrix}}{\begin{pmatrix} N \\ n \end{pmatrix}} & \text{se} \quad x\in S \\ 0 & \text{se} \quad x\in \overline{S} \end{cases}.$$

Diz-se que a v.a. **X segue a lei Hipergeométrica** (ou **X tem distribuição Hipergeométrica**) **de parâmetros n**, **N** e **M**, e escreve-se simbolicamente $\mathbf{X}\sim\mathcal{H}(\mathbf{n},\mathbf{N},\mathbf{M})$.

- Se $X\sim\mathcal{H}(n,N,M)$ então:

$$E(X)=n\frac{M}{N};\ V(X)=n\frac{M}{N}\left(1-\frac{M}{N}\right)\left(\frac{N-n}{N-1}\right).$$

Tal como no caso da Binomial, a distribuição Hipergeométrica é usada na contagem do número de sucessos que ocorrem em n realizações de uma experiência aleatória, cada uma com apenas dois resultados possíveis (sucesso ou insucesso). No entanto, a Hipergeométrica é usada quando a probabilidade do sucesso é alterada em cada realização. Uma das aplicações mais interessantes destas distribuições é na área da Amostragem, onde uma amostra de dimensão n é selecionada numa população de dimensão N. Uma nova leitura dos parâmetros da Hipergeométrica, (n,N,M), é

n: dimensão da amostra
N: dimensão da população (finita)
M: número de sucessos na população

Prova-se, no entanto, que em certas condições (n é muito pequeno relativamente a N), a distribuição $\mathcal{H}(n,N,M)$ pode ser aproximada pela distribuição $\mathcal{B}(n,p)$, com $p=\frac{M}{N}$. Na prática,

$$\text{se } \frac{n}{N}\leq 0.1 \text{ então } \mathcal{H}(n,N,M) \underset{\downarrow}{\dot\sim} \mathcal{B}\left(n,\tfrac{M}{N}\right)$$

"*aproximadamente (no limite)*"

2.3.3 Distribuição de Poisson

Seja λ um número real positivo. Diz-se que uma v.a. **X** segue a **lei de** *Poisson* **de parâmetro** λ, e escreve-se simbolicamente $\mathbf{X} \sim \mathcal{P}(\lambda)$, se X for discreta de suporte $S = \mathbb{N}_0$, com função de probabilidade

$$p(x) = P(X = x) = \begin{cases} e^{-\lambda}\dfrac{\lambda^x}{x!} & \text{se} \quad x \in \mathbb{N}_0 \\ 0 & \text{se} \quad x \notin \mathbb{N}_0 \end{cases}.$$

- $P(X = x) > 0, \forall x \in \mathbb{N}_0$
- $\displaystyle\sum_{x=0}^{+\infty} P(X = x) = 1$
- $E(X) = V(X) = \lambda$

[**Aditividade da Poisson**]
Sejam $X_1, X_2, ...X_k$, v.a.'s independentes tais que $X_i \sim \mathcal{P}(\lambda_i)$, $i = 1, ..., k$. Então,

$$Y = \sum_{i=1}^{k} X_i \sim \mathcal{P}(\lambda_1 + \lambda_2 + ... + \lambda_k)$$

Concluir que uma certa v.a. segue a lei de Poisson (ao contrário das leis Binomial e Hipergeométrica) não é imediato, e sai fora do âmbito do nosso curso. No entanto, esta distribuição é aplicada na contagem de eventos independentes que ocorrem durante um dado intervalo de tempo ou numa dada região espacial. Esta distribuição também é conhecida pela distribuição dos acontecimentos raros, pois a probabilidade de ocorrer mais do que um evento num intervalo muito pequeno é nula.
Prova-se que a distribuição $\mathcal{B}(n, p)$, com p muito pequeno e n suficientemente grande, pode ser aproximada pela distribuição $\mathcal{P}(\lambda)$, com $\lambda = np$. Na prática,

$$\text{se} \quad p \leq 0.1 \quad \text{então} \quad \mathcal{B}(n, p) \mathrel{\dot\sim} \mathcal{P}(np)$$

2.4 Variáveis Aleatórias Bidimensionais

Exemplo ilustrativo

Lançamento de dois dados, anotando os números de pontos das faces que ficam voltadas para cima.

$\Omega = \{(i, j) : i, j \in \{1, 2, 3, 4, 5, 6\}\}$.

Seja $\mathbf{X} = (X_1, X_2)$ o vetor aleatório em que X_1 e X_2 representam, respetivamente, a **soma** e a **diferença** dos pontos obtidos.

Seja Ω o espaço de resultados associado a uma experiência aleatória. Designa-se por *variável aleatória bidimensional* (ou *vetor aleatório de dimensão dois*) uma função $\mathbf{X}$ de domínio Ω e contradomínio em $\mathbb{R}^2$.

$$\mathbf{X} : \Omega \longrightarrow \mathbb{R}^2$$

$$w \longrightarrow \mathbf{X}(w) = (X_1(w), X_2(w))$$

Notas

− Uma variável aleatória bidimensional é um vetor cujas componentes são variáveis aleatórias unidimensionais.

− Associado a uma experiência aleatória, pode interessar o estudo de k, com $k \geq 2$, medidas/características. Surge o conceito de variável (vetor) aleatória(o) k-dimensional, com

$$\mathbf{X}(w) = (X_1(w), X_2(w), ..., X_k(w)) \in \mathbb{R}^k.$$

Por uma questão de clareza e simplicidade na exposição dos conceitos, no nosso estudo vamos considerar apenas variáveis aleatórias bidimensionais. No entanto, todos os conceitos aqui apresentados generalizam-se facilmente para o caso k-dimensional.

− Os vetores aleatórios podem ser discretos (todas as componentes do vetor são v.a.'s discretas), contínuos (todas as componentes do vetor são v.a.'s contínuas) ou mistos. Embora neste curso se analise apenas o caso discreto, os conceitos aqui apresentados generalizam-se para o caso contínuo.

Um vetor aleatório (X, Y) diz-se discreto se existir um conjunto finito ou infinito numerável S tal que[a]

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

e tal que

$$P((X, Y) \in S) = \sum_{(x,y)\in S} P(X = x, Y = y) = 1.$$

[a] $P(X = x, Y = y) = P(X = x \cap Y = y)$

A **função de probabilidade conjunta** do par (X, Y) é dada por

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

$$\downarrow$$

"probabilidade de, simultaneamente,
X tomar o valor x e Y tomar o valor y"

- $0 \leq p_{XY}(x, y) \leq 1, \forall (x, y) \in \mathbb{R}^2$
- $\displaystyle\sum_{(x,y)\in\mathbb{R}^2} p_{XY}(x, y) = \sum_{y\in\mathbb{R}} \sum_{x\in\mathbb{R}} p_{XY}(x, y) = 1.$

A *função distribuição conjunta* do par de v.a.'s (X, Y) é dada por

$$F_{XY}(x, y) = P(X \leq x \cap Y \leq y), \; \forall (x, y) \in \mathbb{R}^2.$$

Se (X, Y) tem suporte $S = S_X \times S_Y$ então

$$F_{XY}(x, y) = \sum_{(\mathbf{a},\mathbf{b}) \in S \cap]-\infty,x]\times]-\infty,y]} P(X = \mathbf{a} \cap Y = \mathbf{b}).$$

$$= \sum_{\mathbf{a} \in S_X \cap]-\infty,x]} \; \sum_{\mathbf{b} \in S_Y \cap]-\infty,y]} P(X = \mathbf{a} \cap Y = \mathbf{b}).$$

Seja (X, Y) um vetor aleatório discreto com função de probabilidade conjunta p_{XY}.

- A **função de probabilidade marginal de** X é a função dada por

$$p_X(x) = P(X = x) = \sum_{\mathbf{y} \in \mathbb{R}} p_{XY}(x, \mathbf{y}) = \sum_{\mathbf{y} \in \mathbb{R}} P(X = x \cap Y = \mathbf{y}), \ \forall x \in \mathbb{R}.$$

- A **função de probabilidade marginal de** Y é a função dada por

$$p_Y(y) = P(Y = y) = \sum_{\mathbf{x} \in \mathbb{R}} p_{XY}(\mathbf{x}, y) = \sum_{\mathbf{x} \in \mathbb{R}} P(X = \mathbf{x} \cap Y = y), \ \forall y \in \mathbb{R}.$$

(As funções de probabilidade marginais são funções de probabilidade de v.a.'s unidimensionais.)

- A **função de probabilidade condicionada de** Y **dado** $X = x$, com $p_X(x) > 0$, é a função de y dada por

$$p_{Y/X=x}(y) = P(Y = y/X = x) = \frac{p_{XY}(x, y)}{p_X(x)}, \ \forall y \in \mathbb{R}.$$

(de modo análogo se define função de probabilidade condicionada de X dado $Y = y$, com $p_Y(y) > 0$)

$p_{Y/X=x}$ é uma função de probabilidade.

Independência de Variáveis Aleatórias

Sejam X e Y duas variáveis aleatórias discretas. X e Y dizem-se independentes se e só se

$$\forall (x, y) \in \mathbb{R}^2, \, p_{XY}(x, y) = p_X(x) \times p_Y(y).$$

As n $(n \geq 2)$ variáveis aleatórias $X_1, X_2, ..., X_n$ dizem-se **independentes** se e só se

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

Parâmetros de Vetores Aleatórios

Seja (X, Y) um vetor aleatório tal que $E(X)$ e $E(Y)$ existem. Chama-se *esperança matemática* (*valor médio* ou *valor esperado*) do vetor (X, Y) ao vetor $(E(X), E(Y))$.

Seja (X, Y) um vetor aleatório discreto de suporte S e $g : \mathbb{R}^2 \to \mathbb{R}$ tal que $g(X, Y)$ é uma variável aleatória. Tem-se

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

Seja (X, Y) um vetor aleatório discreto tal que $E(X) = \mu_X$ e $E(Y) = \mu_Y$.

• A **Covariância entre** X **e** Y, denotada por $cov(X, Y)$, é definida por

$$cov(X, Y) = E[(X - \mu_X)(Y - \mu_Y)]\,.$$

$\triangleright$ $cov(X, Y) = \sum_{x\in S_X} \sum_{y\in S_Y} (x - \mu_X)(y - \mu_Y)P(X = x, Y = y) = E(XY) - E(X)E(Y)$

• O **coeficiente de correlação linear entre** X **e** Y, denotado por ρ_{XY}, é dado por

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

$\triangleright$ $-1 \leq \rho_{XY} \leq 1$

Algumas propriedades :

1. $V(X + Y) = V(X) + V(Y) + 2\,cov(X, Y)$
2. $V(X - Y) = V(X) + V(Y) - 2\,cov(X, Y)$
3. Se $X_1, X_2, ..., X_n$ $(n \in \mathbb{N})$ são v.a.'s **independentes** então $E\,(X_1\,X_2...X_n) = E(X_1)E(X_2)...E(X_n)$;

$$V\left[\sum_{i=1}^{n} a_i X_i\, +\, b\right] = \sum_{i=1}^{n} a_i^2 V(X_i),\ \forall\, a_1, a_2, ..., a_n,\, b \in \mathbb{R}.$$

Casos particulares:

i) $V\left[\sum_{i=1}^{n} X_i\right] = \sum_{i=1}^{n} V(X_i)$;

ii) $V(X_1 - X_2) = V(X_1) + V(X_2)$.

Nota: Se X e Y são independentes então $cov(X, Y) = 0$ (!), mas a implicação contrária já não é válida!

Exemplo

Uma fábrica possui duas linhas de produção de um certo motor. Considere o vetor aleatório (X, Y), onde

$X =$ número de motores produzidos, diariamente, na linha I

$Y =$ número de motores produzidos, diariamente, na linha II

A função de probabilidade conjunta de (X, Y) é dada na forma tabular, por

Y \ X	0	1	2
0	0	0.06	0.19
1	0.03	0.07	0.18
2	0.04	0.06	0.1
3	0.04	0.08	0.15

Funções de probabilidade marginais

x	0	1	2
$p_X(x)$	0.11	0.27	0.62

y	0	1	2	3
$p_Y(y)$	0.25	0.28	0.2	0.27

$E(X) = 1.51,\ \ V(X) = 0.4699,\ \ \sigma_X = 0.6855$

$E(Y) = 1.49,\ \ V(Y) = 1.2899,\ \ \sigma_Y = 1.1357$

(Comparar as duas linhas de produção.)