

2 Sinais Aleatórios em Tempo Contínuo. Parte I: Espaço de Probabilidade e Variáveis Aleatórias.

Na Figura 2.1 está representado um modelo simplificado de um sistema de comunicação. No Capítulo 1 justificou-se a opção por um modelo aleatório para representar uma fonte de informação. Embora a discussão se tenha focado no caso de fontes digitais, as razões apresentadas mantêm-se válidas quando a fonte de informação gera sinais analógicos em tempo contínuo. Assim sendo, utilizaremos um **modelo estocástico** (ou aleatório) para representar a classe de sinais passíveis de serem gerados por uma fonte de informação analógica. Em geral, o sinal transmitido resulta de uma ou mais transformações realizadas sobre o sinal fonte, aqui representadas de forma integrada num único bloco genericamente designado por emissor. Embora o sinal transmitido seja em geral diferente do sinal fonte, as transformações envolvidas preservam a informação original, e têm por objectivo adaptar a transmissão ao canal de comunicação e combater o efeito do ruído.

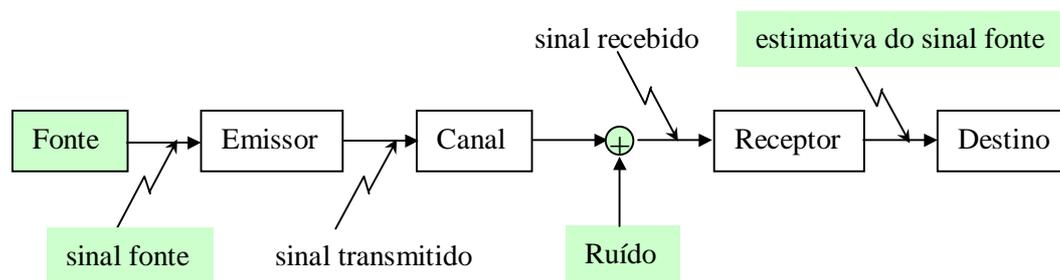


Figura 2.1: Modelo simplificado de um sistema de comunicação

Por natureza própria dos fenómenos físicos envolvidos o mecanismo de geração do ruído é aleatório, pelo que o modelo mais adequado para o representar é também do tipo estocástico. De acordo com o modelo da Figura 2.1, o sinal recebido à entrada do receptor é uma versão do sinal transmitido, eventualmente distorcida pelo canal de comunicação, e corrompida por uma amostra de ruído aditivo:

$$\begin{aligned} x_{\text{rec}}(t) &= C_t [x_{\text{tra}}(\tau \in T_C)] + n(t) \\ x_{\text{tra}}(t) &= E_t [x(\tau \in T_E)] \end{aligned} \quad (2.1)$$

onde C_t e E_t representam as transformações, em geral não instantâneas, realizadas pelo canal e pelo emissor sobre os respectivos sinais de entrada. x , x_{tra} e x_{rec} são o sinal fonte, o sinal transmitido e o sinal recebido, respectivamente, e n é uma amostra do ruído. Note-se que o sinal disponível à entrada do receptor é desconhecido pois nem o sinal fonte nem o ruído são conhecidos. Embora o canal, isto é, a transformação C_t , possa também ser desconhecido, iremos admitir que dispomos de um modelo adequado para o representar. Tendo em conta a falta de conhecimento prévio sobre um conjunto de componentes importantes do sistema, torna-se óbvio que é impossível o receptor fornecer ao destinatário uma réplica exacta do sinal fonte. O problema consiste então em desenhar o emissor e o receptor de modo a que este último possa, a partir do sinal recebido, gerar a melhor réplica possível do sinal fonte. Para tal, e também para avaliar a qualidade do sistema, isto é, o grau de semelhança entre o sinal fonte e a respectiva estimativa, torna-se necessário estudar o modelo de representação da fonte e do ruído.

2.1 Introdução aos Processos Estocásticos

O conceito de processo estocástico constitui uma extensão da noção de variável aleatória e permite modelar uma classe de sinais cujo comportamento ao longo do tempo é não determinístico. A interpretação deste modelo pode fazer-se com base na Figura 2.2. A ideia básica é a de que cada sinal ou função amostra daquela classe ocorre de acordo com os resultados de um modelo experimental probabilístico. Com efeito, um processo estocástico $X(t)$ não é mais do que um conjunto de sinais $x(t; \xi_i)$, designados por funções amostra, onde ξ_i é um dos resultados elementares de um fenómeno físico completamente caracterizado pelo conjunto \mathcal{S} de todos os resultados experimentais directos. Antes de avançar mais na descrição deste modelo, convém recordar as noções de **espaço de probabilidade** e de **variável aleatória**.

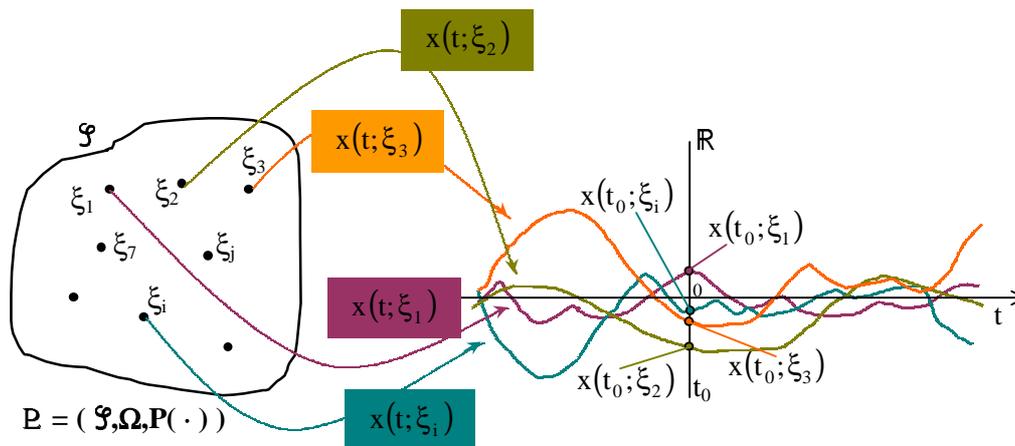


Figura 2.2: Modelação de um Processo Estocástico

2.1.1 Espaço de Probabilidade

Consideremos o conjunto \mathcal{S} formado pelos elementos $\xi_i \in \mathcal{S}$, os quais simbolizam os resultados elementares de uma experiência. Por exemplo, ξ_i pode representar a ocorrência de uma das faces resultante do lançamento de um dado ou o intervalo de tempo que decorre entre duas chamadas telefónicas consecutivas. \mathcal{S} é portanto um **modelo experimental**. Qualquer subconjunto de \mathcal{S} é um acontecimento: a ocorrência de uma face par corresponde à ocorrência das alternativas face_2 , face_4 ou face_6 . Neste contexto, o conjunto vazio \emptyset corresponde ao acontecimento impossível e \mathcal{S} é o acontecimento certo. O conjunto de todos os subconjuntos de \mathcal{S} constitui o **espaço de amostras** Ω . Note-se que os elementos de Ω são a união de acontecimentos elementares de \mathcal{S} os quais, por definição, são mutuamente exclusivos. Recorde-se que dois acontecimentos A e B são **mutuamente exclusivos** sse $A \cap B = \emptyset$. Para completar este modelo probabilístico é necessário atribuir uma medida de probabilidade a todos os acontecimentos de Ω . A probabilidade é uma função dos elementos de Ω que verifica os axiomas:

$$\begin{aligned}
 \text{(i)} \quad & P(A) \geq 0 \\
 \text{(ii)} \quad & P(\mathcal{S}) = 1 \\
 \text{(iii)} \quad & P(A \cup B) = P(A) + P(B) \text{ se } A \cap B = \emptyset
 \end{aligned}
 \tag{2.2}$$

O triplete $(\mathcal{F}, \Omega, P(\cdot))$ é designado por **espaço de probabilidade**. Note-se que o espaço de amostras Ω , sendo o conjunto formado por todos os subconjuntos de \mathcal{F} , satisfaz as seguintes propriedades:

$$\begin{aligned} \text{(i)} \quad & \mathcal{F} \in \Omega \\ \text{(ii)} \quad & \text{se } A \in \Omega \text{ então } A^c \in \Omega \\ \text{(iii)} \quad & \forall i, A_i \in \Omega \text{ então } \cup_i A_i \in \Omega \end{aligned} \quad (2.3)$$

A partir de (2.2) e (2.3) é possível derivar algumas propriedades adicionais da medida de probabilidade $P(\cdot)$, tais como¹:

$$1. \quad P(A^c) = 1 - P(A) \quad (2.4)$$

$$2. \quad P(\emptyset) = 0 \quad (2.5)$$

$$3. \quad P(A \cup B) = P(A) + P(B) - P(A \cap B) \quad (2.6)$$

$$4. \quad A \subset B \Rightarrow P(A) \leq P(B) \quad (2.7)$$

2.1.1.1 Probabilidade Condicional e Independência Estatística

Def. 2.1: Dado um acontecimento M tal que $P(M) \neq 0$, a probabilidade de ocorrência do acontecimento A condicionada na certeza da ocorrência de M é definida por

$$P(A|M) = \frac{P(A \cap M)}{P(M)}. \quad (2.8)$$

Suponhamos que A e M são acontecimentos pertencentes ao espaço de amostras Ω associado ao espaço de probabilidade $\mathbb{P} = (\mathcal{F}, \Omega, P(\cdot))$. Admitamos ainda que em N repetições da experiência representada pelo modelado experimental \mathcal{F} os acontecimentos M e $A \cap M$ ocorrem N_M e $N_{A \cap M}$ vezes, respectivamente. É sabido que para um valor de N suficientemente elevado para que N_M e $N_{A \cap M}$ tomem também valores muito elevados se verifica

$$P(M) \cong \frac{N_M}{N} \quad \text{e} \quad P(A \cap M) \cong \frac{N_{A \cap M}}{N},$$

isto é,

$$\frac{P(A \cap M)}{P(M)} \cong \frac{N_{A \cap M}}{N_M}.$$

Note-se que o acontecimento $A \cap M$ ocorre sse A e M ocorrerem em simultâneo; $N_{A \cap M}$ conta assim o número de vezes que, na série de ocorrências de M , o acontecimento A também ocorre. Este raciocínio explica, ainda que de forma empírica, a definição de probabilidade condicional expressa em (2.8). A probabilidade condicional goza das seguintes propriedades:

$$1. \quad P(A|M) \geq 0 \quad (2.9)$$

$$2. \quad P(\mathcal{F}|M) = 1 \quad (2.10)$$

$$3. \quad A \cap B = \emptyset \Rightarrow P(A \cup B|M) = P(A|M) + P(B|M) \quad (2.11)$$

¹ Por ser relativamente simples, deixa-se como exercício a demonstração destas propriedades.

A propriedade (2.9) resulta directamente do facto de $A \cap M$ ser um acontecimento em Ω . De facto, tendo em conta que a medida de probabilidade é não negativa, $P(A \cap M) \geq 0$, e a quantidade definida em (2.8) é não negativa. Como $\mathcal{S} \cap M = M$, a igualdade (2.10) é imediata. Usando (2.8), temos

$$P(A \cup B | M) = \frac{P((A \cup B) \cap M)}{P(M)} = \frac{P((A \cap M) \cup (B \cap M))}{P(M)}.$$

Sendo A e B mutuamente exclusivos, o mesmo acontece com $A \cap M$ e $B \cap M$. Assim,

$$\frac{P((A \cap M) \cup (B \cap M))}{P(M)} = \frac{P(A \cap M)}{P(M)} + \frac{P(B \cap M)}{P(M)}$$

e, tendo em conta (2.8), obtém-se (2.11).

Def. 2.2: Dois acontecimentos A e B são estatisticamente independentes sse

$$P(A \cap B) = P(A)P(B). \quad (2.12)$$

Desta relação e de (2.8) conclui-se ainda que se dois acontecimentos A e B são estatisticamente independentes, então

$$\begin{aligned} P(A|B) &= P(A) \\ P(B|A) &= P(B) \end{aligned} \quad (2.13)$$

2.1.1.2 Probabilidade Total

Consideremos o conjunto \mathcal{S} de todos os acontecimentos elementares associados a um determinado modelo experimental. Suponhamos que, tal como se mostra na Figura 2.3, se define a partição $[A_1, A_2, \dots, A_M]$ de \mathcal{S} .

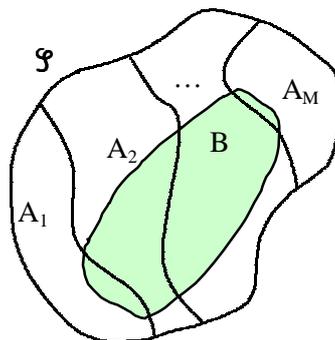


Figura 2.3: Partição do conjunto \mathcal{S}

Seja B um acontecimento qualquer definido em \mathcal{S} , isto é, $B \in \Omega$. Como $\mathcal{S} = A_1 \cup A_2 \cup \dots \cup A_M$, então

$$\begin{aligned} B &= B \cap (A_1 \cup A_2 \cup \dots \cup A_M) \\ &= (B \cap A_1) \cup (B \cap A_2) \cup \dots \cup (B \cap A_M) \end{aligned}$$

Como se vê na Figura 2.3, os acontecimentos A_1, \dots, A_M são mutuamente exclusivos e portanto também com os acontecimentos $(B \cap A_1), \dots, (B \cap A_M)$ são mutuamente exclusivos. Logo

$$P(B) = P(B \cap A_1) + P(B \cap A_2) + \dots + P(B \cap A_M),$$

ou, atendendo a (2.8),

$$P(B) = \sum_{m=1}^M P(B | A_m) P(A_m). \quad (2.14)$$

Este resultado é conhecido por **Teorema da Probabilidade Total** e permite calcular a probabilidade de um acontecimento B se as probabilidades condicionais $P(B|A_m)$ e *a priori* $P(A_m)$ forem conhecidas.

2.1.1.3 Teorema de Bayes

Suponhamos que os acontecimentos B e A_m , $m=1,2,\dots,M$, verificam o teorema da probabilidade total. Então, o **Teorema de Bayes** diz que as probabilidades *a posteriori* $P(A_m|B)$, $m=1,2,\dots,M$, se exprimem em termos das probabilidades *a priori* $P(A_m)$, $m=1,2,\dots,M$, do seguinte modo

$$P(A_m | B) = \frac{P(B | A_m) P(A_m)}{\sum_{m=1}^M P(B | A_m) P(A_m)}. \quad (2.15)$$

Este resultado deriva directamente da definição de probabilidade condicional (2.8) e do teorema da probabilidade total (2.14) e, como veremos mais tarde, desempenha um papel muito importante em muitos problemas de engenharia, em particular, no desenho de receptores em sistemas de comunicações digitais.

2.1.2 Variáveis Aleatórias

Uma variável aleatória é uma função $X(\xi)$ cujo domínio é o conjunto \mathcal{S} de resultados experimentais elementares ξ , sendo o conjunto dos números reais o respectivo contradomínio². Formalmente, a especificação de uma variável aleatória assenta no espaço de probabilidade $\mathcal{P}=(\mathcal{S},\Omega,P(\cdot))$, isto é, no conjunto \mathcal{S} de acontecimentos elementares, no espaço de amostras Ω , e na medida de probabilidade $P(\cdot)$ definida para cada elemento de Ω . Para ilustrar a construção do modelo de uma variável aleatória consideremos a Figura 2.4. Como se pode verificar, a cada elemento de \mathcal{S} faz-se corresponder um e um só número real. No entanto, pode acontecer que a mais de um elemento de \mathcal{S} corresponda um único valor de \mathbb{R} .

² Uma variável aleatória complexa define-se por $X(\xi) = X_r(\xi) + jX_i(\xi)$, onde $X_r(\xi)$ e $X_i(\xi)$ são variáveis aleatórias reais e $j = \sqrt{-1}$ é a unidade imaginária.

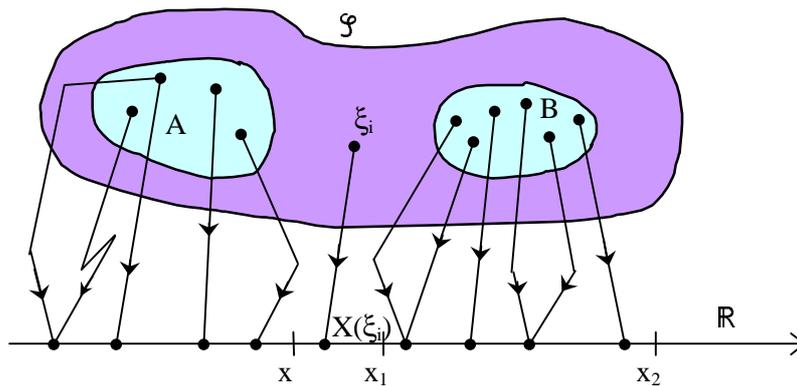


Figura 2.4: Modelação de um Variável Aleatória Real

Note-se que qualquer acontecimento em Ω é representado por um intervalo em \mathbb{R} . Por exemplo, os acontecimentos A e B são representados pelos intervalos

$$A = \{X \leq x\} \quad \text{e} \quad B = \{x_1 < X \leq x_2\},$$

onde x , x_1 e x_2 são números reais. No exemplo da Figura 2.4 o contradomínio da variável aleatória X é constituído por um conjunto contável de números reais. Quando assim é diz-se que a variável aleatória é discreta. Uma variável aleatória X é contínua se o seu contradomínio for uma união de intervalos em \mathbb{R} . Finalmente, se o contradomínio for a união de intervalos de \mathbb{R} com um conjunto contável de números reais, a variável aleatória é do tipo misto. A especificação de uma variável aleatória completa-se recorrendo à medida de probabilidade $P(\cdot)$ associada ao espaço de probabilidade \mathcal{E} .

2.1.2.1 Função de Distribuição Cumulativa

Def. 2.3: A função de distribuição cumulativa de uma variável aleatória X é dada por³:

$$F_X(x) = P(X \leq x). \quad (2.16) \quad \square$$

A função de distribuição goza das seguintes propriedades⁴:

- $$\begin{aligned} F_X(-\infty) &= P(X \leq -\infty) = 0 \\ F_X(+\infty) &= P(X \leq +\infty) = 1 \end{aligned} \quad (2.17)$$

- $$F_X(\cdot)$$
 é monotónica crescente, isto é,

$$\text{se } x_1 < x_2 \quad \text{então } F_X(x_1) \leq F_X(x_2). \quad (2.18)$$

³ A função de distribuição será sempre representada por F ; o subíndice em letra maiúscula designa a variável aleatória a que se refere F .

⁴ Por ser relativamente simples, deixa-se como exercício a demonstração destas propriedades.

$$3. \quad P(X > x) = 1 - F_X(x). \quad (2.19)$$

$$4. \quad P(x_1 < X \leq x_2) = F_X(x_2) - F_X(x_1). \quad (2.20)$$

A função de distribuição pode ser contínua ou descontínua conforme o tipo de variável aleatória a que se refere. Consideremos uma variável aleatória do tipo misto cujo contradomínio é o conjunto $C_D =]-\infty, x_0[\cup]x_0, +\infty[$, e admitamos que $P(x_0) = p_0$. A Figura 2.5 ilustra esta situação. A diferença

$$p_0 = F_X(x_0^+) - F_X(x_0^-)$$

é o salto de descontinuidade de F_X no ponto $x = x_0$. A partir da figura podemos concluir que $P(X \leq x_0) = F_X(x_0^+)$, ou seja, $F_X(x_0) = F_X(x_0^+)$.

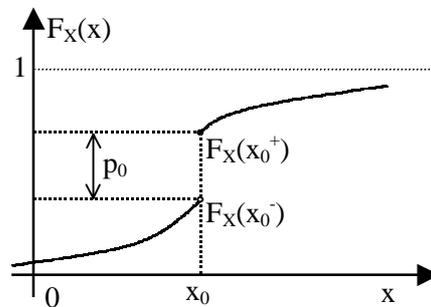


Figura 2.5: Distribuição de probabilidade de uma variável aleatória mista

Como os acontecimentos $\{X < x_0\}$ e $\{X = x_0\}$ são mutuamente exclusivos, temos

$$P(X \leq x_0) = P(X < x_0) + P(X = x_0),$$

isto é,

$$\begin{aligned} P(X = x_0) &= F_X(x_0) - F_X(x_0^-) = p_0 \\ P(X < x_0) &= F_X(x_0^-) \end{aligned} .$$

Partindo desta discussão podemos inferir que, nos casos das variáveis aleatórias discretas e contínuas, as respectivas funções de distribuição são funções em escada no primeiro caso e contínuas no segundo, como se mostra na Figura 2.6.

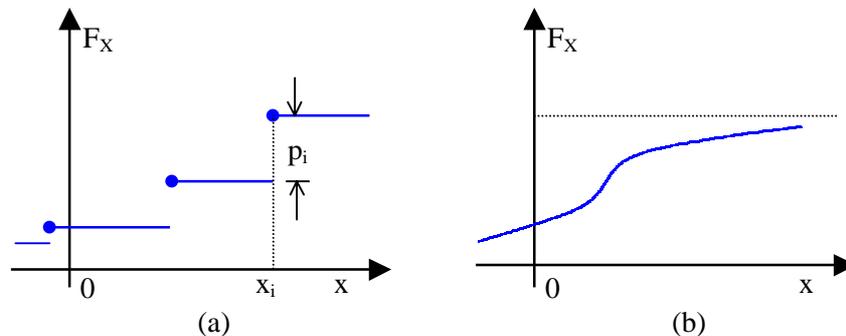


Figura 2.6: Funções de distribuição: (a) variável discreta, (b) variável contínua

2.1.2.2 Função Densidade de Probabilidade

Para além da função de distribuição, a função densidade de probabilidade pode ser usada para caracterizar uma variável aleatória de forma equivalente.

Def. 2.4: Seja X uma variável aleatória contínua com função de distribuição $F_X(x)$. A função densidade de probabilidade de X é definida por⁵:

$$f_X(x) = \frac{dF_X(x)}{dx}. \quad (2.21)$$

A função densidade de probabilidade goza das seguintes propriedades⁶:

1. Uma vez que F_X é não decrescente, ver (2.18), então

$$f_X \geq 0. \quad (2.22)$$

2. De (2.20) resulta

$$P(x_1 < X \leq x_2) = \int_{x_1}^{x_2} f_X(\mu) d\mu. \quad (2.23)$$

- 3.

$$F_X(x) = \int_{-\infty}^x f_X(\mu) d\mu. \quad (2.24)$$

- 4.

$$1 = F_X(+\infty) = \int_{-\infty}^{+\infty} f_X(\mu) d\mu. \quad (2.25)$$

Graficamente, a função densidade de probabilidade de uma variável aleatória contínua tem o aspecto genérico que se mostra na Figura 2.7. Como se vê, a função densidade de probabilidade tende para zero quando x vai para $\pm\infty$, e pode ser multimodal, isto é, apresentar diversos máximos.

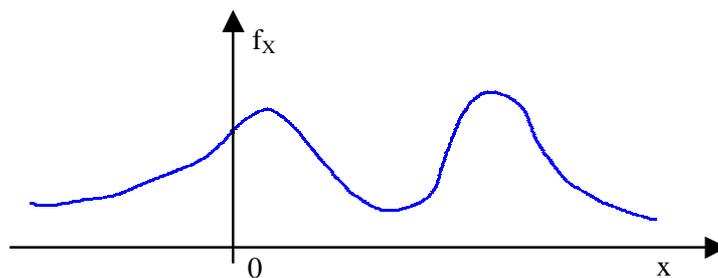


Figura 2.7: Densidade de probabilidade de uma variável aleatória contínua

⁵ A função densidade de probabilidade será sempre representada por f ; o subíndice em letra maiúscula designa a variável aleatória a que se refere f .

⁶ Por ser relativamente simples, deixa-se como exercício a demonstração destas propriedades.

2.1.2.2.1 Densidade de Probabilidade de uma Variável Aleatória Discreta

No caso de variáveis aleatórias discretas, a função de distribuição não é contínua e portanto (2.21) em Def. 2.1 não tem sentido. Uma variável aleatória discreta X toma valores x_k com probabilidade $P(X=x_k) = p_k = F_X(x_k^+) - F_X(x_k^-)$, pelo que podemos usar a notação

$$f_X(x_k) = p_k, \quad k = 1, 2, \dots \quad (2.26)$$

que especifica a densidade de probabilidade pontual. Neste contexto, f_X não é a derivada de F_X , sendo definida pelas amplitudes dos saltos de descontinuidade da função de distribuição. Graficamente, f_X representa-se como se ilustra na Figura 2.8.

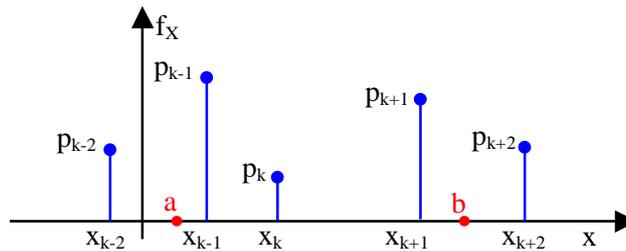


Figura 2.8: Densidade de probabilidade de uma variável aleatória discreta

Neste contexto, as propriedades da função densidade de probabilidade pontual equivalentes a (2.23), (2.24) e (2.25) são⁷:

1. Sejam x_i e x_j os valores de X imediatamente superior a a e inferior ou igual a b , respectivamente. Ver exemplo da Figura 2.8. Então

$$P(a < X \leq b) = \sum_{k=i}^j p_k = \sum_{k=i}^j f_X(x_k). \quad (2.27)$$

2. Seja x_j o valor de X imediatamente inferior ou igual a a . Então

$$F_X(a) = \sum_{k=-\infty}^j f_X(x_k). \quad (2.28)$$

- 3.

$$F_X(+\infty) = 1 = \sum_{k=-\infty}^{+\infty} f_X(x_k). \quad (2.29)$$

Alternativamente, podemos estender o espaço onde se definem as funções densidade de probabilidade definidas por (2.21) de modo a incluir a distribuição $\delta(\cdot)$ de Dirac

$$\int_a^b \delta(x - c) dx = \begin{cases} 1 & \text{se } a \leq c \leq b \\ 0 & \text{caso contrário.} \end{cases} \quad (2.30)$$

⁷ A propriedade (2.22) mantém-se.

Deste modo, a densidade de probabilidade pontual de uma variável aleatória discreta é generalizável de modo a suportar a dedição (2.21), e escreve-se na forma

$$f_X(x) = \sum_{k=-\infty}^{+\infty} p_k \delta(x - x_k). \quad (2.31)$$

Usando esta definição em (2.23), (2.24) e (2.25) e tendo em conta (2.30), recuperam-se facilmente as expressões (2.27), (2.28) e (2.29), respectivamente.

2.1.2.3 Operador Valor Expectável

Considere-se a variável aleatória X caracterizada pela densidade de probabilidade f_X e a transformação

$$Y = g(X) \quad (2.32)$$

a partir da qual se define uma nova variável aleatória Y .

Def. 2.5: O valor expectável da variável aleatória Y é definido por

$$E\{Y\} = \int_{-\infty}^{+\infty} g(\mu) f_X(\mu) d\mu. \quad (2.33)$$

Note-se que se, em particular, a transformação $g(\cdot)$ for a identidade, isto é, $Y = X$, então de (2.33) resulta

$$E\{X\} = \int_{-\infty}^{+\infty} \mu f_X(\mu) d\mu, \quad (2.34)$$

ou seja, o valor expectável da variável aleatória X . Pode também mostrar-se, como veremos mais adiante, que o **operador valor expectável é linear**, isto é,

$$Y = \alpha_1 Y_1 + \alpha_2 Y_2 \Rightarrow E\{Y\} = \alpha_1 E\{Y_1\} + \alpha_2 E\{Y_2\}. \quad (2.35)$$

2.1.2.3.1 Momentos

Def. 2.6: Fazendo em (2.32) $Y = X^n$, e usando (2.33), obtém-se o momento de ordem n da variável aleatória X :

$$m_X^{(n)} = E\{X^n\} = \int_{-\infty}^{+\infty} \mu^n f_X(\mu) d\mu. \quad (2.36)$$

O momento de primeira ordem, abreviadamente representado por m_X , é o valor expectável de X já introduzido em (2.34). O momento de segunda ordem, $m_X^{(2)} = E\{X^2\}$, é normalmente designado por correlação.

2.1.2.3.2 Momentos Centrados

Def. 2.7: Fazendo em (2.32) $Y = (X - m_X)^n$ e usando (2.33), obtém-se o momento centrado de ordem n da variável aleatória X :

$$\mu_X^{(n)} = E\{(X - m_X)^n\} = \int_{-\infty}^{+\infty} (v - m_X)^n f_X(v) dv. \quad (2.37)$$

O momento centrado de primeira ordem é nulo. O momento centrado de segunda ordem é designado por variância: $\sigma_X^2 = \mu_X^{(2)}$.

É fácil verificar que a variância verifica a igualdade

$$\sigma_X^2 = m_X^{(2)} - m_X^2, \quad (2.38)$$

ou seja, é a diferença entre a correlação e quadrado do valor expectável. Portanto, no caso em que X tem valor expectável (média) nulo, a variância coincide com a correlação.

2.1.2.3.3 Desigualdade de Chebyshev

Na caracterização estatística de uma variável aleatória é frequente fazer-se uso de um outro parâmetro, relacionado com a variância, e designado por **desvio padrão**:

$$\sigma_X = (\sigma_X^2)^{1/2} = (m_X^{(2)} - m_X^2)^{1/2}. \quad (2.39)$$

Para se entender que tipo de estatística é medida pelo desvio padrão consideremos uma variável aleatória qualquer Z e um número real arbitrário $\varepsilon > 0$ infinitesimal. Qualquer que seja $a > 0$,

$$E\{Z^2\} = \int_{-\infty}^{+\infty} \mu^2 f_Z(\mu) d\mu \geq \int_{-\infty}^{-a} \mu^2 f_Z(\mu) d\mu + \int_{a-\varepsilon}^{+\infty} \mu^2 f_Z(\mu) d\mu;$$

em qualquer das parcelas do termo mais à direita da relação anterior, o factor μ^2 toma, em qualquer das duas integrandas, valores não inferiores a a^2 . Portanto, podemos ainda escrever

$$E\{Z^2\} \geq a^2 \left[\int_{-\infty}^{-a} f_Z(\mu) d\mu + \int_{a-\varepsilon}^{+\infty} f_Z(\mu) d\mu \right] \xrightarrow{\varepsilon \rightarrow 0} a^2 P(|Z| \geq a),$$

ou ainda

$$P(|Z| \geq a) \leq \frac{E\{Z^2\}}{a^2}. \quad (2.40)$$

Fazendo em (2.40) $Z = X - m_X$ e $a = k\sigma_X$, $k \in \mathbb{Z}^+$, obtém-se a **desigualdade de Chebyshev**, a qual se pode escrever em qualquer das formas alternativas seguintes:

$$P(|X - m_X| \geq k\sigma_X) \leq \frac{1}{k^2}; \quad (2.41)$$

$$P(|X - m_X| < k\sigma_X) > 1 - \frac{1}{k^2}. \quad (2.42)$$

A desigualdade de Chebyshev permite afirmar que, independentemente da função f_X , a probabilidade de X tomar valores fora de um intervalo centrado em torno do respectivo valor expectável e com comprimento igual a $2k$ desvios padrão é sempre não superior a $1/k^2$. Um valor pequeno do desvio padrão significa um pequeno espalhamento dos valores mais prováveis de X em torno do valor médio. Por exemplo, para $k = 2$, qualquer variável aleatória X toma valores entre $m_X - 2\sigma_X$ e $m_X + 2\sigma_X$ com probabilidade superior a 0.75.

2.1.2.4 Exemplos de Variáveis Aleatórias

Neste parágrafo iremos dar exemplo de algumas variáveis aleatórias de grande interesse para o desenho e análise de sistemas de telecomunicações.

1. Bernoulli

Variável aleatória discreta cujo contradomínio é o conjunto $C_D = \{0,1\}$ com distribuição de probabilidade

$$\begin{aligned} P(0) &= p \\ P(1) &= 1 - p \end{aligned} \quad \text{com } 0 \leq p \leq 1. \quad (2.43)$$

Esta variável aleatória é usada para modelar fontes binárias e a ocorrência de erros de transmissão num sistema de comunicações digitais.

2. Binomial

Variável aleatória discreta cujo contradomínio é o conjunto $C_D = \mathbb{Z}_0^+$ dos inteiros não negativos e que representa, por exemplo, o número de 0's que ocorrem numa sequência de n ocorrências de Bernoulli. A respectiva densidade de probabilidade pontual é da forma

$$P(X = k) = \begin{cases} \binom{n}{k} p^k (1-p)^{n-k} & 0 \leq k \leq n \\ 0 & k > n \end{cases} \quad (2.44)$$

A distribuição binomial serve, por exemplo, para modelar o número total de símbolos recebidos com erro numa sequência de n símbolos estatisticamente independentes, sendo p a probabilidade de erro por símbolo.

3. Uniforme

Variável aleatória contínua cuja função densidade de probabilidade é definida como

$$f_X(x) = \begin{cases} \frac{1}{b-a} & a \leq x \leq b \\ 0 & \text{caso contrário} \end{cases} \quad (2.45)$$

A distribuição uniforme pode ser usada para representar a fase de uma senoide num intervalo de comprimento 2π .

4. Laplace

Variável aleatória contínua cuja densidade de probabilidade é definida por

$$f_X(x) = \frac{\alpha}{2} \exp(-\alpha|x|) \quad -\infty < x < +\infty \quad (2.46)$$

Neste caso X pode ser usada para modelar a amplitude de um sinal de voz.

5. Gaussiana ou Normal

A variável aleatória gaussiana é de grande importância na análise do desempenho de sistemas de comunicações pois o ruído térmico, uma das fontes de ruído mais típicas neste tipo de sistemas, tem uma distribuição de amplitude gaussiana. A respectiva densidade de probabilidade é dada por

$$f_X(x) = \frac{1}{\sqrt{2\pi}\sigma_X} \exp\left(-\frac{(x - m_X)^2}{2\sigma_X^2}\right) \quad -\infty < x < +\infty \quad (2.47)$$

onde m_X e σ_X são a média e o desvio padrão de X, respectivamente.

2.1.2.5 Transformação de Uma Variável Aleatória

Consideremos uma variável aleatória X com densidade de probabilidade f_X e a transformação

$$Y = g(X), \quad (2.32)$$

onde g é uma função conhecida para a qual existe a transformação inversa

$$x = g^{-1}(y). \quad (2.48)$$

Em geral, g pode ser não monotónica pelo que a solução da equação $y = g(x)$ dada por (2.48) não é única. O problema que iremos abordar é o de calcular a densidade de probabilidade f_Y dadas a transformação (2.32) e a densidade f_X . Para simplificar, começaremos por assumir que g é uma função monotónica, isto é, a solução dada por (2.48) é única. Consideremos a ilustração deste problema apresentada na Figura 2.9. Note-se que ao acontecimento $]x_0, x_0 + dx_0]$ definido sobre o espaço de amostras da variável X corresponde o mesmo acontecimento $]y_0, y_0 + dy_0]$, este definido sobre o espaço de amostras de Y, e determinado pela transformação g. Porque se trata do mesmo acontecimento, podemos concluir que

$$P(x_0 < X \leq x_0 + dx_0) = P(y_0 < Y \leq y_0 + dy_0);$$

Admitindo que dx_0 e dy_0 são infinitesimais, as probabilidades da igualdade anterior, dadas pelas áreas sombreadas na Figura 2.9, podem ser aproximadas por

$$f_X(x_0)|dx_0| \cong f_Y(y_0)|dy_0|,$$

ou seja,

$$f_Y(y_0) \equiv \frac{f_X(x_0)}{\left| \frac{dy_0}{dx_0} \right|}. \quad (2.49)$$

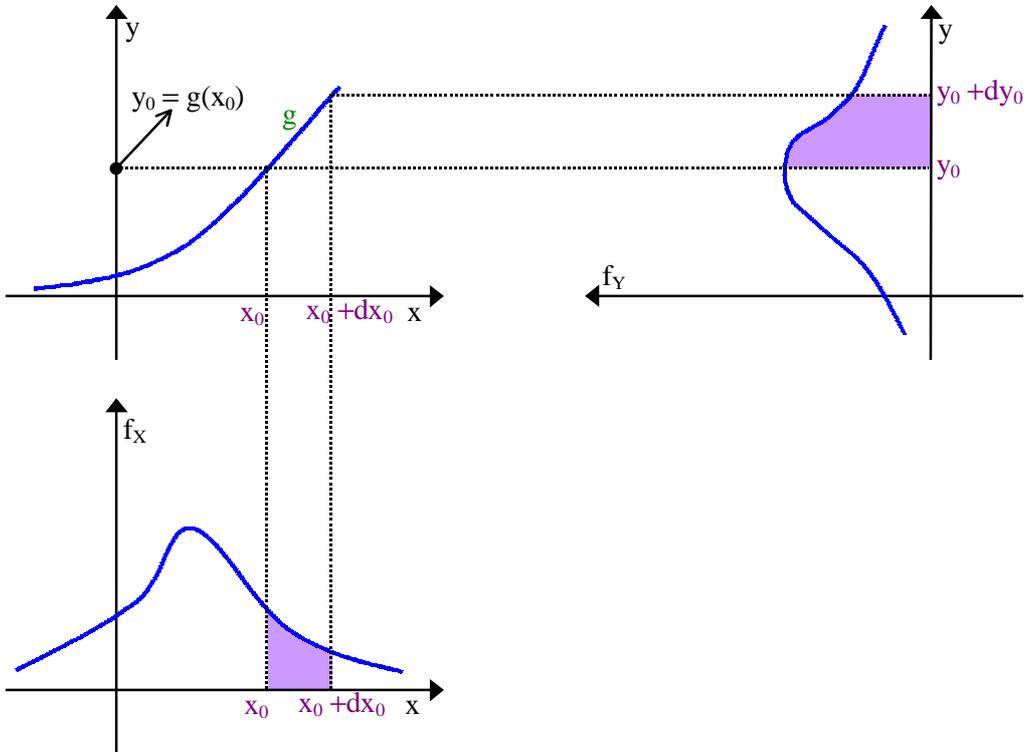


Figura 2.9: Transformação de uma variável aleatória

Uma vez que $x_0 = g^{-1}(y_0)$, no limite, quando $dx_0 \rightarrow 0$, a aproximação (2.49) converge para a solução exacta

$$\forall (x_0, y_0): y_0 = g(x_0) \Rightarrow f_Y(y_0) = \frac{f_X(g^{-1}(y_0))}{|g'(g^{-1}(y_0))|}, \quad (2.50)$$

onde $g' = \frac{dg(x)}{dx}$.

Exemplo 2.1: Consideremos a transformação $Y = m + \alpha X$ onde X é uma variável gaussiana de média nula e variância unitária. Pretendemos conhecer a densidade de probabilidade de Y . Usando (2.47) podemos escrever

$$f_X(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right). \quad (2.51)$$

Por outro lado, $g'(x) = \alpha$ e $g^{-1}(y) = \alpha^{-1}(y - m)$. Substituindo em (2.50) obtemos

$$f_Y(y) = \frac{1}{\sqrt{2\pi}|\alpha|} \exp\left(-\frac{(y-m)^2}{2\alpha^2}\right),$$

isto é, uma gaussiana de média m e variância α^2 .

Exemplo 2.2: Suponhamos a transformação $Y = |X|$, onde X é também gaussiana de média nula e variância unitária como no exemplo anterior (ver (2.51)). Neste exemplo, a transformação é não monotónica pelo que (2.50) não pode ser aplicada directamente. No entanto, se nos socorrermos da Figura 2.10 verificamos que um acontecimento A_y definido sobre o espaço de amostras de Y é o mesmo que a união de dois acontecimentos mutuamente exclusivos B_{-y} e B_y definidos no espaço de amostras de X . Portanto, a probabilidade associada a A_y é a soma das probabilidades associadas a B_{-y} e B_y . Isto quer dizer que a solução (2.50) tem no caso presente duas parcelas correspondentes às duas soluções da equação $y = |x|$.

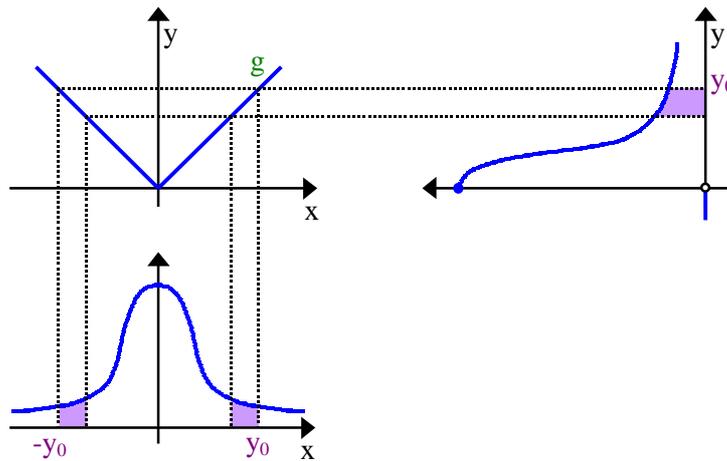


Figura 2.10: Valor absoluto de uma gaussiana

Como em valor absoluto a derivada da transformação é unitária, teremos

$$f_Y(y) = \frac{1}{|-1|\sqrt{2\pi}} \exp\left(-\frac{(-y)^2}{2}\right) + \frac{1}{|+1|\sqrt{2\pi}} \exp\left(-\frac{(+y)^2}{2}\right) \quad y \geq 0,$$

ou seja,

$$f_Y(y) = \begin{cases} \sqrt{\frac{2}{\pi}} \exp\left(-\frac{y^2}{2}\right) & y \geq 0 \\ 0 & y < 0 \end{cases} \quad (2.52)$$

A partir deste exemplo podemos generalizar (2.50) para o caso em que a transformação g é não monotónica.

Facto 2.1: Sejam x_1, \dots, x_N as soluções da equação $y = g(x)$, onde g é uma transformação arbitrária. Sendo $Y = g(X)$, onde X tem densidade de probabilidade f_X , então a densidade de probabilidade de Y vem dada por

$$f_Y(y) = \sum_{n=1}^N \frac{f_X(x_n)}{|g'(x_n)|} \Big|_{x_n=g^{-1}(y)}. \quad (2.53)$$

2.1.3 Distribuições Conjuntas e Condicionais

Os conceitos introduzidos na secção anterior podem ser generalizados para o caso de vectores aleatórios, isto é, cujos componentes são variáveis aleatórias. Por questões de simplicidade, iremos conduzir a apresentação para o caso de vectores bidimensionais.

2.1.3.1 Função de Distribuição Conjunta

Consideremos duas variáveis aleatórias X e Y definidas sobre o conjunto \mathcal{S} associado a um determinado modelo experimental. Naturalmente, as propriedades estatísticas de cada uma das variáveis aleatórias quando consideradas isoladamente são completamente determinadas pelas respectivas funções de distribuição F_X e F_Y . No entanto, o mesmo não acontece quando se consideram as propriedades conjuntas de X e Y . Em particular, o conhecimento de F_X e de F_Y não é em geral suficiente para calcular a probabilidade $P\{(X, Y) \in D\}$ de que o par (X, Y) tome valores numa região $D \subseteq \mathbb{R}^2$.

Def. 2.8- A função de distribuição conjunta das variáveis aleatórias X e Y é definida como

$$F_{XY}(x, y) = P(X \leq x, Y \leq y), \quad (2.54)$$

onde $\{X \leq x, Y \leq y\} = \{X \leq x\} \cap \{Y \leq y\}$.

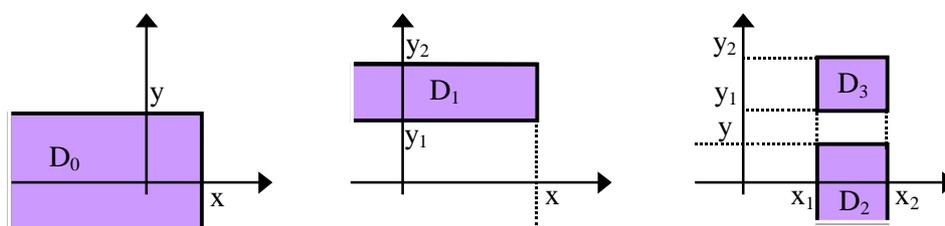


Figura 2.11: Acontecimentos em \mathbb{R}^2

Uma vez que $F_{XY}(x, y)$ representa a probabilidade de o par (X, Y) tomar valores na região D_0 representada na Figura 2.11, é fácil verificar a partir de (2.54) que as probabilidades de ocorrência (X, Y) nas regiões D_1 , D_2 , e D_3 são, respectivamente,

$$P(X \leq x, y_1 < Y \leq y_2) = F_{XY}(x, y_2) - F_{XY}(x, y_1); \quad (2.55)$$

$$P(x_1 < X \leq x_2, Y \leq y) = F_{XY}(x_2, y) - F_{XY}(x_1, y); \quad (2.56)$$

$$P(x_1 < X \leq x_2, y_1 < Y \leq y) = F_{XY}(x_2, y_2) - F_{XY}(x_1, y_2) - F_{XY}(x_2, y_1) + F_{XY}(x_1, y_1). \quad (2.57)$$

2.1.3.2 Densidade de Probabilidade Conjunta

Def. 2.9- A função densidade de probabilidade conjunta das variáveis aleatórias X e Y é definida como

$$f_{XY}(x, y) = \frac{\partial^2 F_{XY}(x, y)}{\partial x \partial y}. \quad (2.58) \square$$

Desta definição resulta que

$$P\{(X, Y) \in D\} = \iint_D f_{XY}(\mu, \nu) d\mu d\nu \quad (2.59)$$

e, em particular,

$$F_{XY}(x, y) = \int_{-\infty}^x \int_{-\infty}^y f_{XY}(\mu, \nu) d\mu d\nu. \quad (2.60)$$

Exemplo 2.3: Consideremos a região D_3 definida na Figura 2.11; fazendo uso de (2.59) e de (2.58) obtém-se sucessivamente:

$$\begin{aligned} P\{(X, Y) \in D_3\} &= \int_{x_1}^{x_2} \int_{y_1}^{y_2} f_{XY}(\mu, \nu) d\mu d\nu \\ &= \int_{x_1}^{x_2} \frac{d}{d\mu} [F_{XY}(\mu, y_2) - F_{XY}(\mu, y_1)] d\mu \\ &= F_{XY}(x_2, y_2) - F_{XY}(x_1, y_2) - F_{XY}(x_2, y_1) + F_{XY}(x_1, y_1) \end{aligned} \quad \square$$

Note-se finalmente que

$$f_{XY}(x, y) \geq 0 \quad (2.61)$$

e que

$$F_{XY}(+\infty, +\infty) = \iint_{\mathbb{R}^2} f_{XY}(\mu, \nu) d\mu d\nu = 1. \quad (2.62)$$

2.1.3.3 Distribuições e Densidades Marginais

Relativamente ao par de variáveis aleatórias (X, Y) , diz-se que F_X (ou F_Y) e f_X (ou f_Y) são, respectivamente, a distribuição e a densidade de probabilidade marginal da variável X (ou Y). Consideremos a Figura 2.12(a). Naturalmente, $\{X \leq x\} = \{X \leq x, Y \leq +\infty\}$ é o conjunto de todos os pares (X, Y) à esquerda de L_x e, portanto,

$$F_X(x) = F_{XY}(x, +\infty). \quad (2.63)$$

Por outro lado, e relativamente à Figura 2.12(b), podemos observar que

$$f_X(x) dx = P\{x < X \leq x + dx\} = P\{X \in \Delta D\},$$

ou seja,

$$f_X(x)dx = \iint_{\Delta D} f_{XY}(x,v) dx dv = dx \int_{-\infty}^{+\infty} f_{XY}(x,v) dv$$

e, em conclusão,

$$f_X(x) = \int_{-\infty}^{+\infty} f_{XY}(x,v) dv. \quad (2.64)$$

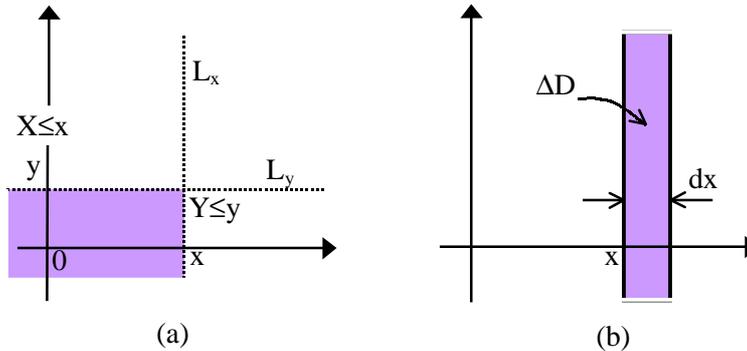


Figura 2.12: Distribuições marginais

De modo equivalente, podemos também escrever

$$F_Y(y) = F_{XY}(+\infty, y)$$

e

$$f_Y(y) = \int_{-\infty}^{+\infty} f_{XY}(\mu, y) d\mu.$$

2.1.3.4 Variáveis Aleatórias Estatisticamente Independentes

Def. 2.10- As variáveis aleatórias X e Y são estatisticamente independentes (ou simplesmente independentes) se os acontecimentos $\{X \leq x\}$ e $\{Y \leq y\}$ forem independentes, isto é, se

$$P\{X \leq x, Y \leq y\} = P\{X \leq x\}P\{Y \leq y\}. \quad (2.65)$$

Deste modo, se X e Y forem estatisticamente independentes, então a distribuição conjunta é o produto das distribuições marginais, ou seja,

$$\forall (x, y) \in \mathbb{R}^2: F_{XY}(x, y) = F_X(x)F_Y(y). \quad (2.66)$$

Atendendo a (2.58) e a (2.21), de (2.66) resulta

$$\forall (x, y) \in \mathbb{R}^2: f_{XY}(x, y) = f_X(x)f_Y(y), \quad (2.67)$$

o que significa que se X e Y forem independentes então a densidade de probabilidade conjunta é o produto das densidades marginais.

2.1.3.5 Média, Correlação e Covariância

Consideremos a variável aleatória $Z = g(X, Y)$, onde g é uma função $\mathbb{R}^2 \rightarrow \mathbb{R}$ conhecida. Seja ΔD_z a região de \mathbb{R}^2 onde $z \leq g(X, Y) \leq z + dz$; para cada dz corresponde uma região ΔD_z onde $g(X, Y) \cong z$ e

$$P\{z \leq Z \leq z + dz\} = P\{(X, Y) \in \Delta D_z\}.$$

À medida que dz cobre a recta real, as correspondentes regiões ΔD_z , disjuntas, cobrem o plano \mathbb{R}^2 e, portanto,

$$E\{Z\} = \int_{-\infty}^{+\infty} z f_Z(z) dz = \iint_{\mathbb{R}^2} g(x, y) f_{XY}(x, y) dx dy,$$

isto é,

$$E\{g(X; Y)\} = \iint_{\mathbb{R}^2} g(x, y) f_{XY}(x, y) dx dy. \quad (2.68)$$

Estamos agora em condições de demonstrar a **linearidade do operador valor expectável**. Com efeito, sendo $Y = \alpha_1 Y_1 + \alpha_2 Y_2$, podemos então sucessivamente escrever

$$\begin{aligned} E\{Y\} &= \iint_{\mathbb{R}^2} (\alpha_1 y_1 + \alpha_2 y_2) f_{Y_1 Y_2}(y_1, y_2) dy_1 dy_2 \\ &= \alpha_1 \iint_{\mathbb{R}^2} y_1 f_{Y_1 Y_2}(y_1, y_2) dy_1 dy_2 + \alpha_2 \iint_{\mathbb{R}^2} y_2 f_{Y_1 Y_2}(y_1, y_2) dy_1 dy_2 \\ &= \alpha_1 \int_{\mathbb{R}} y_1 \int_{\mathbb{R}} f_{Y_1 Y_2}(y_1, y_2) dy_2 dy_1 + \alpha_2 \int_{\mathbb{R}} y_2 \int_{\mathbb{R}} f_{Y_1 Y_2}(y_1, y_2) dy_1 dy_2 \\ &= \alpha_1 \int_{\mathbb{R}} y_1 f_{Y_1}(y_1) dy_1 + \alpha_2 \int_{\mathbb{R}} y_2 f_{Y_2}(y_2) dy_2 \\ &= \alpha_1 E\{Y_1\} + \alpha_2 E\{Y_2\} \end{aligned}$$

A **correlação** entre as variáveis aleatórias X e Y define-se por

$$R_{XY} = E\{XY\} \quad (2.69)$$

Sendo σ_X^2 e σ_Y^2 as variâncias de X e Y , respectivamente, então o **coeficiente de correlação** entre X e Y é

$$\rho_{XY} = \frac{C_{XY}}{\sigma_X \sigma_Y}, \quad (2.70)$$

onde

$$C_{XY} = E\{(X - m_X)(Y - m_Y)\} \quad (2.71)$$

é a **covariância** entre X e Y , e m_X e m_Y os respectivos valores expectáveis. Usando (2.71) e (2.69), obtém-se a seguinte relação:

$$C_{XY} = R_{XY} - m_X m_Y. \quad (2.72)$$

Def. 2.11- As variáveis aleatórias X e Y são **in correlacionadas sse**

$$C_{XY} = 0 \Leftrightarrow \rho_{XY} = 0 \Leftrightarrow R_{XY} = m_X m_Y. \quad (2.73) \quad \square$$

Def. 2.12- As variáveis aleatórias X e Y são ortogonais sse

$$R_{XY} = 0. \quad (2.74) \quad \square$$

Facto 2.2: Sejam X e Y variáveis aleatórias estatisticamente independentes. Então X e Y são variáveis aleatórias incorrelacionadas.

Se X e Y são estatisticamente independentes então $f_{XY}(x, y) = f_X(x)f_Y(y)$. Portanto, de (2.69) e (2.68) vem

$$\begin{aligned} R_{XY} &= \iint_{\mathbb{R}^2} xy f_{XY}(x, y) dx dy \\ &= \int_{\mathbb{R}} x f_X(x) dx \int_{\mathbb{R}} y f_Y(y) dy \\ &= m_X m_Y \end{aligned}$$

o que, de acordo com a Def. 2.11, significa que X e Y são incorrelacionadas. Sublinha-se que, em geral, duas variáveis aleatórias incorrelacionadas não são necessariamente estatisticamente independentes. \square

Exemplo 2.4: Variáveis aleatórias conjuntamente gaussianas. Sejam X e Y duas variáveis aleatórias conjuntamente gaussianas com médias m_X e m_Y e variâncias σ_X^2 e σ_Y^2 , respectivamente. Então, sendo ρ o coeficiente de correlação entre X e Y ,

$$f_{XY}(x, y) = \frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho^2}} \exp\left[-\frac{(x-m_X)^2}{2\sigma_X^2(1-\rho^2)} - \frac{(y-m_Y)^2}{2\sigma_Y^2(1-\rho^2)} + \frac{\rho(x-m_X)(y-m_Y)}{\sigma_X\sigma_Y(1-\rho^2)}\right] \quad (2.75)$$

Facto 2.3: A correlação entre duas variáveis aleatórias X e Y verifica a desigualdade de Schwarz

$$E^2\{XY\} \leq E\{X^2\}E\{Y^2\} \quad (2.76)$$

A igualdade é verificada quando X e Y são proporcionais, isto é, quando $Y = c_0 X$.

Consideremos a seguinte igualdade onde c é uma constante arbitrária:

$$I(c) = E\{(Y - cX)^2\} = E\{Y^2\} - 2cE\{XY\} + c^2E\{X^2\}$$

Note-se que $I(c) \geq 0$ representa uma parábola. Portanto, a equação $I(c) = 0$ não pode ter mais do que uma solução real. Por outras palavras, o binómio discriminante da fórmula resolvente da equação algébrica do 2º grau deve verificar

$$4E^2\{XY\} - 4E\{X^2\}E\{Y^2\} \leq 0,$$

de onde (2.76) resulta imediatamente. Por outro lado, a solução real única obtém-se quando o binómio discriminante é nulo e vale precisamente $c_0 = Y/X$. \square

Facto 2.4: O coeficiente de correlação entre duas variáveis aleatórias X e Y verifica a desigualdade

$$|\rho_{XY}| \leq 1, \quad (2.77)$$

atingindo o valor máximo quando $Y = c_0 X$.

Este facto resulta directamente do Facto 2.3 e das definições (2.70) e (2.71). □

2.1.3.6 Funções de duas Variáveis Aleatórias

O problema que aqui consideramos é o de, dadas as variáveis aleatórias

$$\begin{aligned} Z &= g(X, Y) \\ W &= h(X, Y) \end{aligned} \quad (2.78)$$

e as transformações $g(\cdot, \cdot)$ e $h(\cdot, \cdot)$, exprimir a função densidade de probabilidade conjunta $f_{ZW}(\cdot, \cdot)$ em termos de $f_{XY}(\cdot, \cdot)$. Assume-se que as transformações (2.78) têm inversa. O resultado constitui uma generalização natural de (2.53) pelo que não será aqui demonstrado. Sejam (x_n, y_n) , $n = 1, 2, \dots, N$, as N soluções do sistema de equações (2.78) e

$$J(x, y) = \det \begin{bmatrix} \frac{\partial g(x, y)}{\partial x} & \frac{\partial g(x, y)}{\partial y} \\ \frac{\partial h(x, y)}{\partial x} & \frac{\partial h(x, y)}{\partial y} \end{bmatrix} \quad (2.79)$$

o respectivo Jacobeano. Então,

$$f_{ZW}(z, w) = \sum_{n=1}^N \frac{f_{XY}(x_n, y_n)}{|J(x_n, y_n)|}, \quad (2.80)$$

onde os pares (x_n, y_n) , $n = 1, 2, \dots, N$, aparecem expressos em termos de z e w .

Exemplo 2.5: Pretende-se determinar $f_Z(z)$, sabendo que $Z = X + Y$ e conhecendo $f_{XY}(x, y)$. Começamos por definir a variável aleatória $W = Y$, formando o par de transformações

$$\begin{cases} Z = X + Y \\ W = Y \end{cases}.$$

Qualquer que seja o par (z, w) , o sistema anterior tem apenas uma solução $(x, y) = (z - w, w)$. Por outro lado, $J(x, y) = 1$ e, portanto,

$$f_{ZW}(z, w) = f_{XY}(z - w, w). \quad (2.81)$$

Note-se que $f_Z(z)$ é uma das densidades marginais de $f_{ZW}(z, w)$, ou seja, atendendo a (2.64) e (2.81),

$$f_Z(z) = \int_{-\infty}^{+\infty} f_{XY}(z - w, w) dw. \quad (2.82)$$

No caso particular em que as variáveis aleatórias X e Y são estatisticamente independentes, $f_{XY}(\cdot, \cdot) = f_X(\cdot) f_Y(\cdot)$, e (2.82) toma a forma de um integral de convolução

$$f_Z(z) = \int_{-\infty}^{+\infty} f_X(z-w)f_Y(w)dw. \quad (2.83)$$

Facto 2.5: A função densidade de probabilidade da soma de duas variáveis aleatórias estatisticamente independentes é dada pelo integral de convolução das densidades de probabilidade das parcelas.

2.1.3.7 Teorema do Limite Central

Seja $Z = X_1 + X_2 + \dots + X_N$ a soma de N variáveis aleatórias $\{X_n\}_{n=1}^N$ estatisticamente independentes e com distribuições de probabilidade arbitrárias. O teorema do limite central diz que a variável aleatória Z tem uma distribuição de probabilidade que tende para a de uma gaussiana quando $N \rightarrow \infty$.

2.1.3.8 Distribuições Condicionais

Def. 2.13- A função de distribuição da variável aleatória X dada a ocorrência do acontecimento A é dada por

$$F_{X|A}(x|A) = P\{X \leq x | A\} = \frac{P\{X \leq x, A\}}{P\{A\}}. \quad (2.84)$$

Note-se que esta definição é coerente com as Defs. 2.3 e 2.1 de função de distribuição de uma variável aleatória e de probabilidade condicional, respectivamente.

Def. 2.14- A função densidade de probabilidade da variável aleatória X dada a ocorrência do acontecimento A é dada por

$$f_{X|A}(x|A) = \frac{dF_{X|A}(x|A)}{dx}. \quad (2.85)$$

Do teorema da probabilidade total, ver (2.14), e fazendo $B = \{X \leq x\}$, podemos escrever

$$F_X(x) = \sum_{m=1}^M F_{X|A_m}(x|A_m)P(A_m), \quad (2.86)$$

onde $\{A_m\}_{m=1}^M$ é uma partição arbitrária do conjunto \mathcal{S} e $P(A_m)$, $m=1,2,\dots,M$, são as respectivas probabilidades de ocorrência. De (2.86), (2.21) e (2.85) decorre ainda que

$$f_X(x) = \sum_{m=1}^M f_{X|A_m}(x|A_m)P(A_m). \quad (2.87)$$

A fórmula de Bayes pode também ser generalizada para o contexto das variáveis aleatórias. Se em

$$P(A|B) = \frac{P(B|A)}{P(B)}P(A)$$

fizemos $B = \{X \leq x\}$, então

$$P(A | X \leq x) = \frac{F_{X|A}(x | A)}{F_X(x)} P(A) \quad (2.88)$$

e, de modo análogo,

$$P(A | x_1 < X \leq x_2) = \frac{F_{X|A}(x_2 | A) - F_{X|A}(x_1 | A)}{F_X(x_2) - F_X(x_1)} P(A). \quad (2.89)$$

Usando o facto de $P(A | X = x) = \lim_{\Delta x \rightarrow 0} P(A | x < X \leq x + \Delta x)$, e fazendo $x_1 = x$ e $x_2 = x + \Delta x$ em (2.89), obtemos

$$P(A | X = x) = \frac{f_{X|A}(x | A)}{f_X(x)} P(A). \quad (2.90).$$

Multiplicando ambos os membros de (2.90) por $f_X(x)$ e integrando em x , vem

$$P(A) = \int_{-\infty}^{+\infty} P(A | X = x) f_X(x) dx, \quad (2.91)$$

pois a área limitada por $f_{X|A}(x | A)$ é obviamente unitária. Note-se que (2.91) constitui uma outra forma de apresentar o teorema da probabilidade total. Finalmente, de (2.90) e (2.91) obtém-se a correspondente fórmula de Bayes

$$f_{X|A}(x | A) = \frac{P(A | X = x) f_X(x)}{\int_{-\infty}^{+\infty} P(A | X = x) f_X(x) dx}. \quad (2.92)$$

Facto 2.6: A função de densidade de probabilidade conjunta de duas variáveis aleatórias X e Y é, em geral, dada por

$$f_{XY}(x, y) = f_X(x | Y = y) f_Y(y). \quad (2.93)$$

X e Y são estatisticamente independentes sse $f_X(x | Y = y) = f_X(x)$. Neste caso, de (2.93) obtém-se a relação (2.67).

Retomando (2.84) e fazendo $A = \{y < Y \leq y + \Delta y\}$, podemos escrever

$$P(X \leq x, y < Y \leq y + \Delta y) = F_{X|A}(x | y < Y \leq y + \Delta y) P(y < Y \leq y + \Delta y),$$

ou seja,

$$F_{XY}(x, y + \Delta y) - F_{XY}(x, y) = F_{X|A}(x | y < Y \leq y + \Delta y) [F_Y(y + \Delta y) - F_Y(y)].$$

Dividindo ambos os membros da igualdade anterior por Δy e tomando o limite, quando $\Delta y \rightarrow 0$, obtém-se

$$\frac{\partial F_{XY}(x, y)}{\partial y} = F_X(x | Y = y) \frac{dF_Y(y)}{dy},$$

resultando (2.93) após derivação de ambos os membros em ordem a x .