Universidade Federal do Rio Grande do Sul
Escola de Engenharia
Departamento de Engenharia Elétrica
Programa de Pós-Graduação em Engenharia Elétrica
ELE00071-Tópicos Especiais em Automação e Controle II

Probabilidade e Variáveis Aleatórias

Prof. Walter Fetter Lages

4 de outubro de 2004

1 Introdução

Intuitivamente, tem-se uma noção do que seja um sinal aleatório ou com ruído. Este tipo de sinal não pode ser descrito adequadamente por funções matemáticas explícitas, como senoides, exponenciais, etc. A sua descrição deve ser feita de forma probabilística.

Por outro lado, as fontes de ruído em um sistema geralmente não podem ser eliminadas por completo, de forma que mesmo após eliminarem-se todas as possíveis causas de ruído ainda restará uma parcela de ruído que terá que ser eliminada por filtragem. Se a faixa de frequências do ruído for suficientemente separada da faixa de frequências do sinal, pode-se utilizar filtros convencionais, passa-baixas, passa-altas, passa-faixa ou rejeita-faixa para eliminar o ruído. No entanto, em algumas situações o ruído encontra-se na faixa de frequências do sinal de interesse. Nestes casos é necessário a utilização de filtro estocásticos. Para o desenvolvimento destes filtros é necessário um entendimento de forma quantitativa do ruído, que pelas suas características deve ser descrito de forma probabilística.

2 Probabilidade Intuitiva

Intuitivamente, a definição de probabilidade é feita considerando-se todas os possíveis resultados de um experimento e a probabilidade de ocorrência de um evento particular, A é definida como

$$P(A) = \frac{\text{Possibilidade de resultado com o evento A}}{\text{Total de resultados Possíveis}}$$

Este resultado pode ser estendido para uma interpretação estatística de probabilidade como sendo a frequência relativa de ocorrência do evento.

3 Probabilidade Axiomática

As noções intuitivas de probabilidade permitem tratar problemas relativamente simples, em especial quando tem-se igualdade de condições para todos os eventos. No entanto, frequenetemente deseja-se tratar situações onde alguns eventos não são "honestos". Adicionalmente, em alguns casos não se pode enumerar todos os possíveis resultados de um experimento. A formulação axiomática da teoria da probabilidade simplifica o tratamento nestes casos. Esta formulação é baseada em três axiomas. A apresentação destes axiomas requer algumas definições:

Espaço amostral é o conjunto de todos os possíveis resultados de um experimento. O espaço amostral é denotado por S.

Elementos ou pontos no espaço amostral são os resultados individuais de um experimento. O conjunto de elementos do espaço amostral é denotado por $\{s_1, s_2, s_3, \ldots\}$. Elementos são mutuamente exclusivos ou disjuntos. O número de pontos no espaço amostral pode ser:

- **finito** quando o espaço amostral é discreto e finito
- **infinito contável** quando o espaço amostral é discreto e infinito
- **infinito incontável** quando o espaço amostral é contínuo

evento é um subconjunto de S. Será denotado por letras maiúsculas. Eventualmente serão consideradas operações de união, intersecção e complemento de eventos.

ocorrência do evento A se dá quando ocorre algum ponto em A.

O espaço amostral é denotado por S e o seu conjunto de elementos por $\{s_1, s_2, s_3, \ldots\}$.

3.1 Axiomas da Probabilidade

Sejam S o espaço amostral e A qualquer evento[1] definido em S.Tem-se:

Axioma 1 $P(A) \geq 0$

[1]Note-se que a probabilidade é associada aos eventos e não aos pontos do espaço amostral. A diferenciação entre ponto e evento é especialmente importante quando o espaço amostral é infinito

Axioma 2 $P(S) = 1$

Sejam também $A_1, A_2, A_3, \ldots$ eventos mutuamente exclusivos (disjuntos) em S. Tem-se

Axioma 3 $P(A_1 \cup A_2 \cup A_3 \cup \ldots) = P(A_1) + P(A_2) + P(A_3) + \ldots$

3.2 Espaço de Probabilidade

A associação de um espaço amostral, um conjunto de eventos neste espaço e a atribuição de probabilidades de cada evento definem um *espaço de probabilidade*.

Exemplo 1 *Considere o lançamento de dois dados. Supondo que se está interessado apenas na soma dos números da face superior dos dados, pode-se definir o espaço amostral como*

$$S = 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12$$

O conjunto de possíveis eventos pode ser definido como sendo todos os possíveis subconjuntos de S, incluindo o conjunto vazio e o próprio S. A atribuição de probabilidade aos evento pode ser feita conforme a tabela 1[2].

Tabela 1: Probabilidades para o Lançamento de Dois Dados

Soma dos Dados	Probabilidade Atribuída
2	1/36
3	2/36
4	3/36
5	4/36
6	5/36
7	6/36
8	5/36
9	4/36
10	3/36
11	2/36
12	1/36

Tem-se, portanto, um espaço de probabilidade adequadamente definido sobre o qual pode-se fazer diversas inferências

[2]Esta tabela não atribui probabilidade a todos os eventos do espaço amostral, mas a probabilidade dos demais eventos pode ser computada a partir dos eventos relacionados na tabela.

1. *Qual a probabilidade de obter-se um 7 ou um 11?*

$$P(\textit{7 ou 11}) = P(7 \cup 11) = \frac{6}{36} + \frac{2}{36} = \frac{2}{9}$$

2. *Qual a probabilidade de não obter-se 2, 3 ou 12?*

$$\begin{aligned} P(\textit{não obter 2, 3, ou 12}) &= P(\overline{2 \cup 3 \cup 12}) \\ &= P(4 \cup 5 \cup 6 \cup 7 \cup 8 \cup 9 \cup 10 \cup 11) \\ &= \frac{3+4+5+6+5+4+3+2}{36} \end{aligned}$$

3. *Qual a probabilidade de obter-se dois 4s? Este evento não faz parte do espaço amostral, que foi definido como sendo as possíveis somas dos dois dados.*

3.3 Probabilidade Conjunta

Além das operações de união e complemento, a operação de intersecção também é útil. A intersecção de dois eventos A e B é o evento contento pontos comuns a A e B, como pode ser visto no diagrama de Venn da figura 1.

Da geometria do diagrama de Venn tem-se

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

A probabilidade $P(A \cap B)$ é denominada probabilidade conjunta de A e B e representa a probabilidade de ocorrência de ambos os eventos.

Exemplo 2 *Retornando ao exemplo 1, define-se o evento A como a obtenção de 4, 5, 6 ou 7 e o evento B como a obtenção de 7, 8 , 9, 10 ou 11.*

1. *Qual a probabilidade do evento A e B?*

$$P(A \textit{ e } B) = P(A \cap B) = P(7) = \frac{1}{6}$$

2. *Qual a probabilidade do evento A ou B?*

$$P(A \textit{ ou } B) = P(A \cup B) = P(A) + P(B) - P(A \cap B) = \frac{18}{36} + \frac{20}{36} - \frac{6}{36} = \frac{8}{9}$$

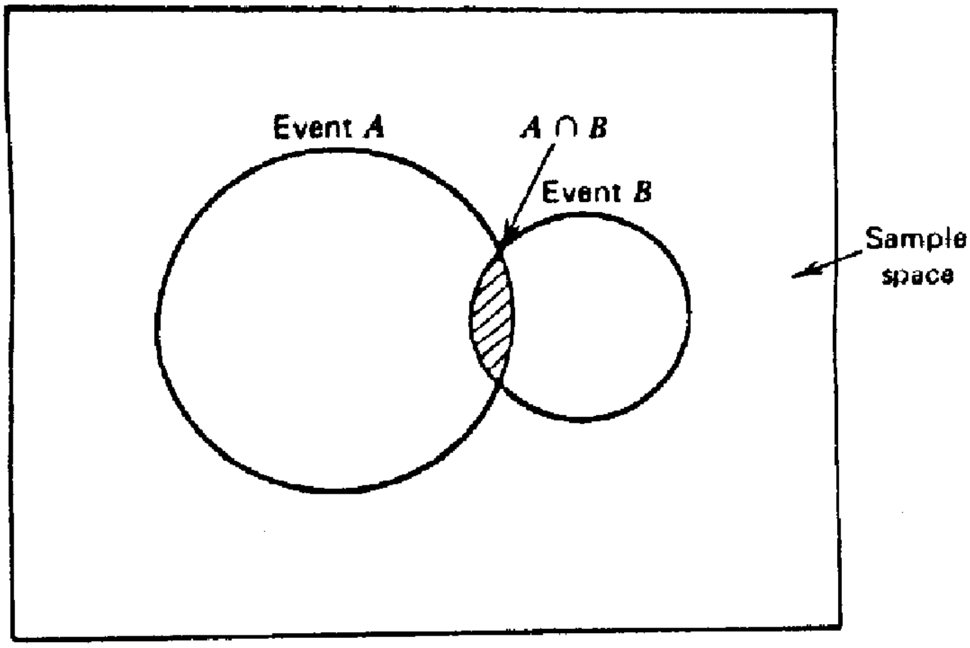

Figura 1: Diagrama de Venn para dois eventos não disjuntos.

4 Probabilidade Condicional

Sejam dois experimentos A e B e sejam A_1, A_2, $\ldots$, A_m eventos disjuntos associados com o experimento A. Similarmente, sejam B_1, B_2, $\ldots$, B_n eventos disjuntos associados com o experimento B. Tem-se, portanto a matriz de probabilidade conjunta mostrada na tabela 2. Obviamente, somando as linhas tem-se a probabilidade de um evento particular no experimento A independentemente dos resultados do experimento B. Similarmente, somando-se as colunas resulta $P(B_1)$, $P(B_2)$ e assim sucessivamente. Como os eventos são disjuntos, a soma das probabilidades marginais é 1.

A tabela 2 mostra a frequência relativa de ocorrência dos diversos eventos em um conjunto dado um evento em particular do outro conjunto. Por exemplo, a linha 1 lista $P(A_1 \cap B_1)$, $P(A_1 \cap B_2)$, $\ldots$, $P(A_1 \cap B_n)$. Como nenhuma outra entrada na tabela envolve $A1$, esta linha da tabela mostra a distribuição relativa dos eventos B_1, B_2, $\ldots$, B_n dado que A_1 ocorreu. No entanto, o conjunto de números desta linha não é uma distribuição de probabilidade válida, pois a soma não é 1, mas sim $P(A_1)$. Pode-se, porém, renormalizar todas as colunas da linha, dividindo-se por $P(A_1)$. O conjunto de números resultante será $P(A_1 \cap B_1)/P(A_1)$, $P(A_1 \cap B_2)/P(A_1)$, $\ldots$, $P(A_1 \cap B_n)/P(A_1)$, a soma será 1 e a distribuição relativa corresponde à frequência relativa de ocorrência de B_1,

$B_2, \ldots, B_n$ dado que A_1 ocorreu.

Tabela 2: Matriz de Probabilidades Conjuntas

	Evento B_1	Evento B_2	$\cdots$	Evento B_n	Prob. marginal
Evento A_1	$P(A_1 \cap B_1)$	$P(A_1 \cap B_2)$	$\cdots$	$P(A_1 \cap B_n)$	$P(A_1)$
Evento A_2	$P(A_2 \cap B_1)$	$P(A_2 \cap B_2)$	$\cdots$	$P(A_2 \cap B_n)$	$P(A_2)$
$\vdots$	$\vdots$	$\vdots$		$\vdots$	$\vdots$
Evento A_m	$P(A_m \cap B_1)$	$P(A_m \cap B_2)$	$\cdots$	$P(A_m \cap B_n)$	$P(A_m)$
Prob. marginal	$P(B_1)$	$P(B_2)$	$\cdots$	$P(B_n)$	Soma $= 1$

A *probabilidade condicional* de B_i dado A_j é definida como

$$P(B_i|A_j) = \frac{P(A_j \cap B_i)}{P(A_j)} \tag{1}$$

Similarmente, a probabilidade condicional de A_j dado B_j é definida como

$$P(A_j|B_i) = \frac{P(A_j \cap B_i)}{P(B_i)} \tag{2}$$

Teorema 1 (Teorema de Bayes)

$$P(A_j|B_i) = \frac{P(B_i|A_j)P(A_j)}{P(B_i)} \tag{3}$$

Prova 1 *Combinando-se as expressões (1) e (2).*

5 Independência

Dois eventos são ditos independentes se a ocorrência de um não afeta o outro. Formalmente, dois eventos são ditos independentes se

$$P(A \cap B) = P(A)P(B) \tag{4}$$

Também deve ser evidente da expressão (4) e da definição de probabilidade condicional (1) e (2) que se A e B são independentes

$$\left.\begin{array}{l} P(A|B) = P(A) \\ P(B|A) = P(B) \end{array}\right\} \text{somente para } A \text{ e } B \text{ independentes}$$

6 Variáveis Aleatórias

Para controle, se está tipicamente interessado em sinais como tensão, torque, distância, que possuem um significado físico. Nestes casos, a chance de ocorrência está associada a números reais e não a "coisas"como faces de dados. Uma *variável aleatória* é um número $x(\zeta)$ atribuído a cada resultado ζ de um experimento. Assim, uma variável aleatória é uma função cujo domínio é o conjunto S de resultados do experimento.

7 Funções Distribuição e Densidade de Probabilidade

Quando o espaço amostral consiste em um número finito de elementos, a atribuição de probabilidade pode ser feita diretamente com base nos elementos do espaço amostral, de acordo com a possibilidade de ocorrência. Estas probabilidades transferem-se diretamente para os eventos equivalentes no espaço da variável aleatória.

Tendo-se variáveis aleatórias discretas, o conjunto discreto de probabilidades associado é denominado *distribuição de massa de probabilidade* ou *distribuição de probabilidade*.

No caso de variáveis aleatórias contínuas, o espaço amostral correspondente terá infinitos pontos e portanto não pode-se atribuir probabilidades diretamente aos pontos do espaço amostral. Isto tem que ser feito para eventos definidos.

Considere o jogo de girar um ponteiro montado sobre um cartão circular e livre para girar sobre o seu centro.

Seja X uma variável aleatória contínua correspondente à posição angular do ponteiro ao parar, o que pode ser qualquer ângulo entre 0 e 2π radianos. Portanto, a probabilidade de exatamente qualquer posição em particular é zero. Logo, atribui-se probabilidade ao evento do ponteiro parar dentro de uma certa faixa angular, por exemplo entre 0 e θ radianos. Se todas as posições são igualmente prováveis é razoável atribuir as probabilidades da seguinte maneira:

$$P(X \leq \theta) = \begin{cases} 0, & \theta < 0 \\ \frac{1}{2\pi}\theta, & 0 \leq \theta \leq 2\pi \\ 1, & \theta > 2\pi \end{cases}$$

Esta função é denominada *função de distribuição acumulada* ou *função de distribuição de probabilidade* e descreve a atribuição de probabilidade. Especificamente, a função de distribuição de probabilidade associada com a variável aleatória X é definida como

$$F_X(\theta) = P(X \leq \theta)$$

onde θ é um parâmetro representando a realização de X.

É claro, a partir da definição que uma função de distribuição de probabilidade tem as seguintes propriedades:

1. $F_X(\theta) \to 0$, quando $\theta \to -\infty$
2. $F_X(\theta) \to 1$, quando $\theta \to \infty$
3. $F_X(\theta)$ é uma função não decrescente de θ

A informação contida na função de distribuição pode ser apresentada na forma diferencial. Especificamente, seja $f_X(\theta)$ definida como

$$f_X(\theta) = \frac{d}{d\theta} F_X(\theta)$$

A função $f_X(\theta)$ é conhecida como *função de densidade de probabilidade* associada com a variável aleatória X. Das propriedades da função de distribuição, é óbvio que a função densidade tem as seguintes propriedades

1. $f_X(\theta)$ é não negativa
2. $$\int_{-\infty}^{\infty} f_X(\theta) d\theta = 1$$

Também deve ser aparente que a área abaixo da função de densidade representa a probabilidade de X estar entre θ_1 e θ_2.

8 Esperança, Média e Função Característica

média amostral

$$\bar{X} = \frac{X_1 + X_2 + \cdots + X_N}{N}$$

Valor Esperado

$$\text{Valor esperado de } X = E(X) = \sum_{i=1}^{n} p_i x_i$$

$$\text{Valor esperado de } X = E(X) = \int_{-\infty}^{\infty} x f_X(x) dx$$

Valor Esperado de uma Função

$$E(g(X)) = \sum_{i=1}^{n} p_i g(x_i)$$

$$E(g(X)) = \int_{-\infty}^{\infty} g(x) f_X(x) dx$$

k-ésimo momento $E(X^k)$

$$E(X^k) = \int_{-\infty}^{\infty} x^k f_X(x) dx$$

Segundo momento

$$E(X^2) = \int_{-\infty}^{\infty} x^2 f_X(x) dx$$

Variância segundo momento em torno da média

$$\sigma_X^2 = E((X - E(x)^2)$$

$$\sigma_X^2 = E[X^2 - 2XE(X) + E(X)^2] = E(X^2) - E(X)^2$$

Desvio padrão $\sigma_X = \sqrt{\sigma_X^2}$

Função Característica

$$\Psi_x(\omega) = \int_{-\infty}^{\infty} f_X(x) e^{j\omega x} dx$$

Exemplo 3 *Seja X uniformemente distribuído no intervalo $(0, 2\pi)$. Encontrar a média, variância e desvio padrão de X.*

Tem-se, portanto a seguinte função de densidade de probabilidade:

$$f_X(x) = \begin{cases} \frac{1}{2\pi}, & 0 \leq x < 2\pi \\ 0, & \textit{caso contrário} \end{cases}$$

$$E(X) = \int_0^{2\pi} x \frac{1}{2\pi} dx = \left[\frac{1}{2\pi} \frac{x^2}{2} \right]_0^{2\pi} = \pi$$

$$\sigma_X^2 = \int_0^{2\pi} x^2 \frac{1}{2\pi} dx - \pi^2 = \frac{4}{3}\pi^2 - \pi^2 = \frac{1}{3}\pi^2$$

$$\sigma_X = \sqrt{\sigma_X^2} = \sqrt{\frac{1}{3}\pi^2} = \frac{1}{\sqrt{3}}\pi$$

Exemplo 4 *Mostre que a função característica pode ser utilizada para calcular os momentos de X.*

Os momentos de X podem ser escritos como

$$\begin{aligned} E(X) &= \int_{-\infty}^{\infty} x f_X(x) dx \\ E(X^2) &= \int_{-\infty}^{\infty} x^2 f_X(x) dx \\ &\vdots \\ &\textit{etc.} \end{aligned}$$

As derivadas de $\Psi_X(\omega)$ calculadas em $\omega = 0$ são:

$$\begin{aligned} \left.\frac{d\Psi_x(\omega)}{d\omega}\right|_{\omega=0} &= \left.\int_{-\infty}^{\infty} jx f_X(x) e^{j\omega x} dx\right|_{\omega=0} = \int_{-\infty}^{\infty} jx f_X(x) dx \\ \left.\frac{d^2\Psi_x(\omega)}{d\omega^2}\right|_{\omega=0} &= \left.\int_{-\infty}^{\infty} (jx)^2 f_X(x) e^{j\omega x} dx\right|_{\omega=0} = \int_{-\infty}^{\infty} j^2 x^2 f_X(x) dx \\ &\vdots \\ &\textit{etc.} \end{aligned}$$

Portanto,

$$\begin{aligned} E(X) &= \left.\frac{1}{j}\frac{d\psi_x}{d\omega}\right|_{w=0} \\ E(X^2) &= \left.\frac{1}{j^2}\frac{d^2\psi_x}{d\omega^2}\right|_{w=0} \\ &\vdots \\ &\textit{etc.} \end{aligned}$$

9 Variáveis Aleatórias Gaussianas

Uma variável aleatória X é denominada *normal* ou *Gaussiana*, se sua função densidade de probabilidade é

$$f_X(x) = \frac{1}{\sqrt{2\pi}\sigma_X} e^{-\frac{1}{2\sigma_X^2}(x-m_X)^2}$$

onde os parâmetros m_X e σ_X são a média e a variância da variável aleatória.

Note que uma função de densidade Gaussiana é completamente especificada através da sua média e da sua variância. Assim, é usual escrever-se $X \sim N(m_X, \sigma_X^2)$ para denotar que X é uma variável aleatória Gaussiana de média m_X e variância σ_X^2. As figuras 2 e 3 apresentam esboços das funções densidade e distribuição Gaussianas, respectivamente. Infelizmente, a função distribuição Gaussiana não pode ser computada de forma fechada.

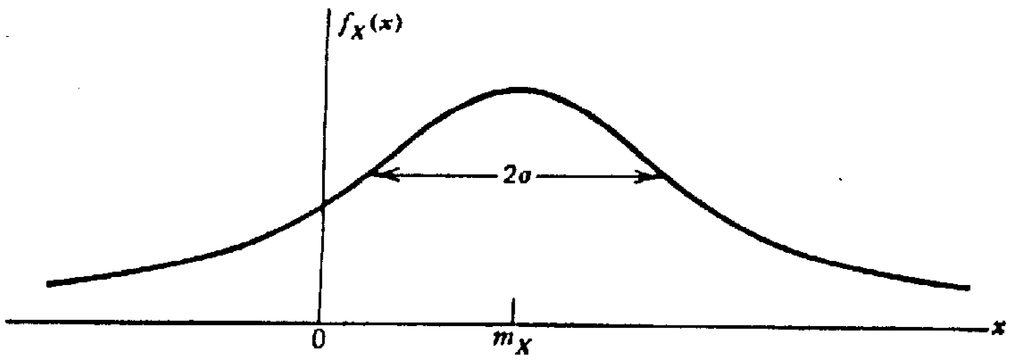

Figura 2: Função densidade de probabilidade Gaussiana.

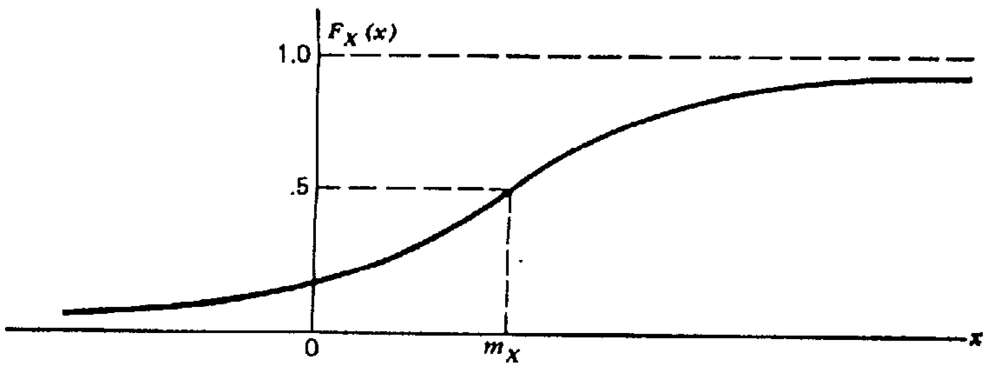

Figura 3: Função distribuição de probabilidade Gaussiana.

Exemplo 5 *Seja uma variável aleatória* $X \sim N(1, 4)$. *Deseja-se*

1. *O valor da função densidade no seu pico;*

2. *A probabilidade de que* $X \geq 2$*;*

3. A probabilidade de que $0 \leq X \leq 2$.

Deve-se ter em mente que as tabelas para as funções densidade e distribuição de probabilidade são normalizadas para média zero e variância unitária. Ou seja, a função tabelada é

$$f_X(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}$$

A função densidade da variável X *é*

$$f_X(x) = \frac{1}{\sqrt{2\pi} \cdot 2} e^{-\frac{1}{2\cdot 4}(x-1)^2}$$

Obviamente o pico ocorre em $x = 1$ *e o seu valor é* $1/2\sqrt{2\pi} \approx 0.199$. *Este valor poderia ser obtido da tabela notando o valor de* $f_X(x)$ *para* $x = 0$ *e dividindo-se pelo desvio padrão.*

$$\begin{aligned} P(X \geq 2) &= \int_2^\infty \frac{1}{\sqrt{2\pi} \cdot 2} e^{-\frac{1}{2\cdot 4}(x-1)^2} dx \\ &= 1 - \int_{-\infty}^2 \frac{1}{\sqrt{2\pi} \cdot 2} e^{-\frac{1}{2\cdot 4}(x-1)^2} dx \end{aligned}$$

$$P(X \geq 2) = 1 - \int_{-\infty}^{1/2} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}v^2} dv$$

$$P(X \geq 2) = 1 - 0.691462 = 0.308538$$

$$P(0 \leq X \leq 2) = \int_0^2 \frac{1}{\sqrt{2\pi} \cdot 2} e^{-\frac{1}{2\cdot 4}(x-1)^2} dx$$

Fazendo-se $v = \frac{x-1}{2}$, *tem-se*

$$P(0 \leq X \leq 2) = \int_{-1/2}^{1/2} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}v^2} dv$$

que devido a simetria da curva pode ser escrita como

$$\begin{aligned} P(0 \leq X \leq 2) &= 2\int_0^{1/2} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}v^2} dv \\ &= 2\left(\int_{-\infty}^{1/2} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}v^2} dv - 0.5\right) \\ &= 2\,(0.691462 - 0.5) = 0.382924 \end{aligned}$$

10 Variáveis Aleatórias Múltiplas

No estudo de controle estocástico frequentemente serão tratadas diversas variáveis aleatórias e seus relacionamentos mútuos. As várias relações probabilísticas serão apresentadas aqui para o caso bivariável. A extensão para o caso de três ou mais variáveis é direta e não será especificamente discutida.

10.1 Variáveis Aleatórias Discretas

Sejam duas variáveis aleatórias *discretas* X e Y. Define-se a *distribuição de probabilidade conjunta* como

$$p_{XY}(x_i, y_j) = P(X = x_i \text{ e } Y = y_j)$$

Tal como no caso de eventos (vide seção 4), a distribuição conjunta de X e Y pode ser considerada como uma matriz de probabilidades bi-dimensional, com cada elemento representando a probabilidade de ocorrência de uma combinação particular de X e Y. A soma dos números da matriz deve ser unitária, assim como as somas das colunas ou das linhas resulta na probabilidade marginal, como no caso dos eventos.

De forma similar à seção 4, pode-se escrever as seguintes relações:

Probabilidade marginal (incondicional)

$$p_X(x_i) = \sum_j p_{XY}(x_i, y_j)$$

$$p_Y(y_j) = \sum_i p_{XY}(x_i, y_j)$$

Probabilidade condicional

$$p_{X|Y} = \frac{p_{XY}}{P_Y} \tag{5}$$

$$p_{Y|X} = \frac{p_{XY}}{P_X} \tag{6}$$

Teorema de Bayes

$$p_{X|Y} = \frac{p_{Y|X} p_X}{P_Y}$$

As variáveis aleatórias discretas X e Y são definidas como sendo *estatisticamente independentes* se

$$p_{XY}(x_i, y_j) = p_X(x_i) p_Y(y_j)$$

para todos possíveis x_i e y_i.

10.2 Variáveis Aleatórias Contínuas

Tal como no caso monovariável, a descrição da variável deve ser feita em termos de uma função de distribuição acumulada ou de uma função de densidade.

Sejam X e Y variáveis aleatórias contínuas. A *função de distribuição conjunta acumulada* é definida como

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

Obviamente, F_{XY} possui as seguintes propriedades:

1. $F_{XY}(-\infty, -\infty) = 0$
2. $F_{XY}(\infty, \infty) = 1$
3. F_{XY} é não decrescente em x e y

A *função de densidade conjunta* de variáveis aleatórias contínuas é dada por

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

Note-se que a relação integral entre a função de distribuição acumulada e a função de densidade existente para o caso monovariável também existe para o caso multivariável. Assim, a probabilidade de uma realização conjunta de X e Y estar dentro de uma certa região R no plano xy é dada por

$$P(X \text{ e } Y \text{ estarem dentro de } R) = \int \int_R f_{XY}(x, y) dx dy$$

Se a região R for um retângulo diferencial (vide figura 4), a probabilidade de X e Y estarem dentro do retângulo será:

$$P(x_0 \leq X \leq x_0 + dx \text{ e } y_0 \leq Y \leq y_0 + dy) = f_{XY}(x, y) dx dy$$

As densidades marginais ou incondicionais são obtidas de forma semelhante ao caso discreto, substituindo-se o somatório pela integral, tem-se portanto:

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

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

As expressões (5) e (6) para probabilidades condicionais discretas podem ser aplicadas para funções densidade para regiões diferenciais. Considerando-se a região diferencial mostrada na figura 4, pode-se obter as seguintes relações:

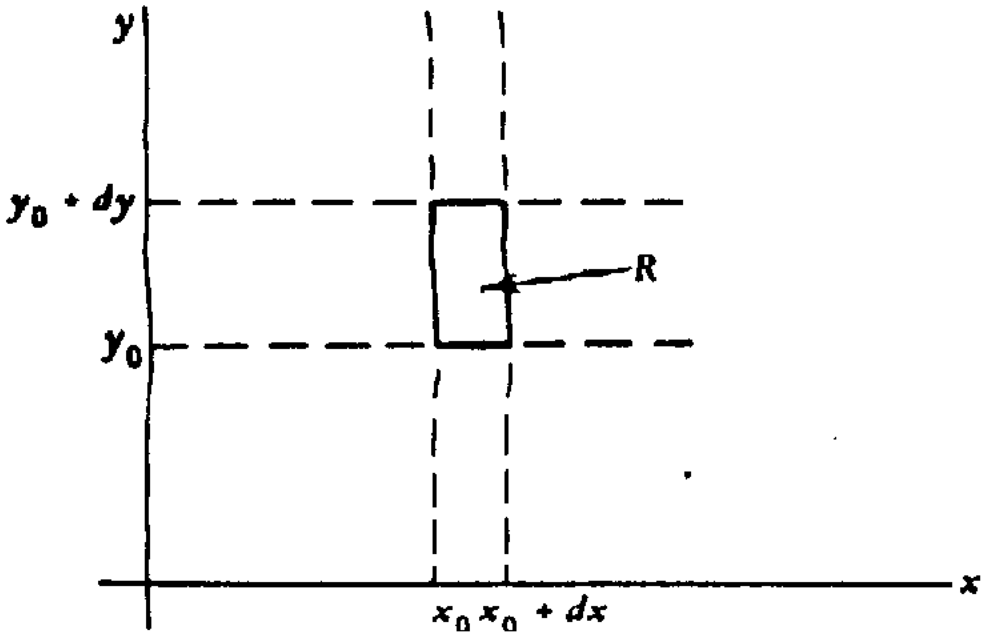

Figura 4: Região diferencial R no plano xy.

$$P(X \text{ está na faixa } dx | Y \text{ está na faixa } dy) = \frac{f_{XY}(x_0, y_0)dxdy}{f_Y(y_0)dy}$$

Cancelando-se os dy's e considerando-se que "Y está na faixa dy"é aproximadamente o mesmo que "Y é igual a y_0", tem-se que

$$P(x_0 \leq X x_0 + dx | Y = y_0) = \left(\frac{f_{XY}(x_0, y_0)}{f_Y(y_0)} \right) dx$$

O lado direito desta expressão possui todas as características de uma função de densidade e a sua interpretação está no lado esquerdo da expressão. Assim, define-se *densidade condicional* como[3]:

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

Analogamente, tem-se

[3] A dependência de $f_{X|Y}$ de y é omitida, para enfatizar que esta é uma função de densidade em x, já que y aparece apenas como um parâmetro determinístico

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

E consequentemente a expressão do teorema de Bayes surge diretamente:

$$f_{X|Y}(x) = \frac{f_{Y|X}(y)f_X(x)}{f_Y(y)}$$

Similarmente, X e Y serão estatisticamente independentes se

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

11 Correlação, Covariância e Ortogonalidade

A esperança do produto de duas variáveis aleatórias X e Y é dada por

$$E(XY) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xyf_{XY}(x,y)dxdy \tag{7}$$

Caso as variáveis X e Y sejam independentes, a expressão (7) reduz-se, através do teorema de Bayes, à

$$E(XY) = \int_{-\infty}^{\infty} xf_X(x)dx \int_{-\infty}^{\infty} yf_Y(y)dy = E(X)E(Y) \tag{8}$$

Quando X e Y possuem a propriedade da expressão (8) diz-se que elas são *descorrelacionadas*. Obviamente, quando X e Y são independentes, elas também são descorrelacionadas. No entanto, o inverso não é verdadeiro, a não ser em casos especiais.

Se $E(XY) = 0$, X e Y são ditas serem *ortogonais*.

A *covariância* de X e Y é definida como:

$$\sigma_{XY} = E\left((X - m_x)(Y - m_y)\right) \tag{9}$$

O *coeficiente de correlação* de X e Y é definido como

$$\rho = \frac{\sigma_{XY}}{\sqrt{\sigma_X^2}\sqrt{\sigma_Y^2}} = \frac{\sigma_{XY}}{\sigma_X\sigma_Y} \tag{10}$$

O coeficiente de correlação é uma medida normalizada ($-1 \leq \rho \leq 1$) do grau de correlação entre duas variáveis aleatórias[4].

[4]Note que se $X = Y$, então $\rho = 1$; se $X = -Y$, então $\rho = -1$; se X e Y são descorrelacionadas, então $\rho = 0$.

12 Soma de Variáveis Aleatórias Independentes

Sejam X e Y duas variáveis aleatórias independentes com funções densidade de probabilidade $f_X(x)$ e $f_Y(y)$, respectivamente. Seja Z outra variável aleatória tal que $Z = X + Y$.

Seja z uma realização de Z com valor fixo. Todas as possíveis realizações de X e Y satisfazem $x + y = z$ e o lugar geométrico destes pontos no plano xy é uma reta, como mostrado na figura 5.

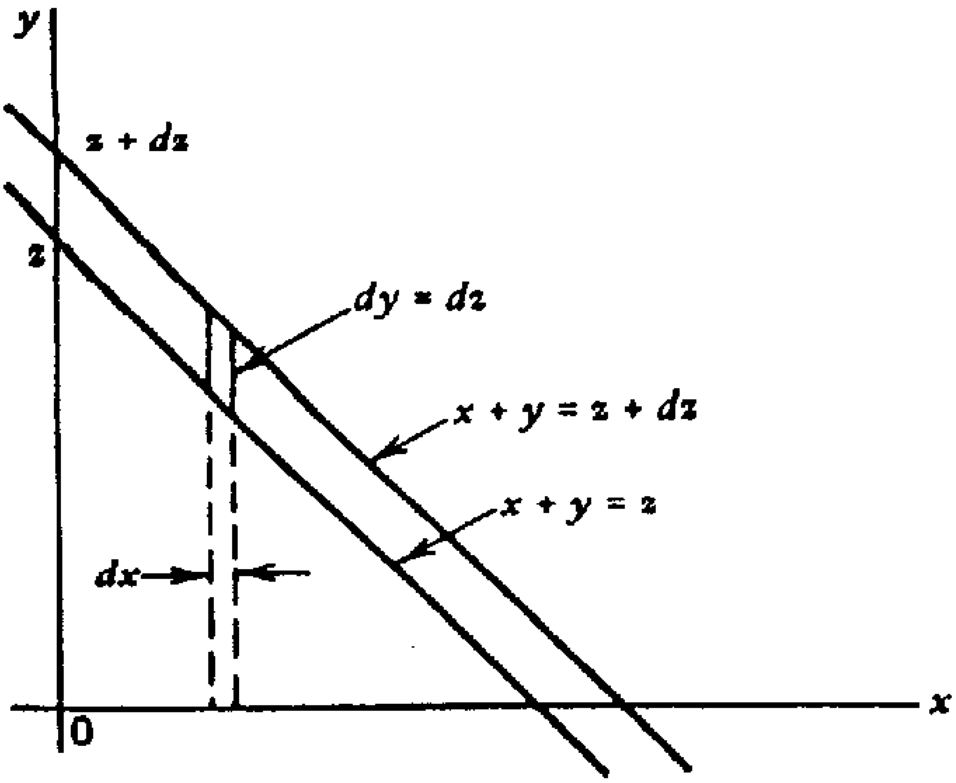

Figura 5: Faixa diferencial para dedução de $f_Z(z)$.

Considere, agora, uma perturbação incremental de z para $z + dz$ e o correspondente lugar geométrico no plano xy das realizações de X e Y que resultam $z + dz$, que também é uma reta mostrada na figura 5.

Pode-se perceber que todos os x e y dentro da faixa diferencial entre as duas retas mapeiam-se em pontos entre z e $z + dz$ no espaço z. Logo,

$$P(z \leq Z \leq z + dz) \quad = \quad P(x \text{ e } y \text{ estejam na faixa diferencial})$$

$$= \int\int_{\substack{\text{faixa}\\\text{diferencial}}} f_X(x) f_Y(y) dx dy$$

No entanto, dentro da faixa diferencial $y = z - x$, e como a largura da faixa é diferencial, a integral dupla pode ser reduzida á uma integral simples. Escolhendo-se x como a variável de integração e notando-se que $dy = dz$ tem-se

$$P(z \leq Z \leq z + dz) = \left[\int_{-\infty}^{\infty} f_X(x) f_Y(z - x) dx\right] dz$$

e o valor entre colchetes é a função de densidade de probabilidade de Z, portanto

$$f_Z(z) = \int_{-\infty}^{\infty} f_X(x) f_Y(z - x) dx \tag{11}$$

A integral na expressão (11) é uma integral de convolução. Assim, da teoria da transformada de Fourier pode-se escrever

$$\mathcal{F}[f_Z] = \mathcal{F}[f_X] \cdot \mathcal{F}[f_Y] \tag{12}$$

Exemplo 6 (Teorema do Limite Central) *Sejam X, Y e V três variáveis aleatórias independentes com funções de densidade retangulares idênticas, como mostrado na figura 6.*

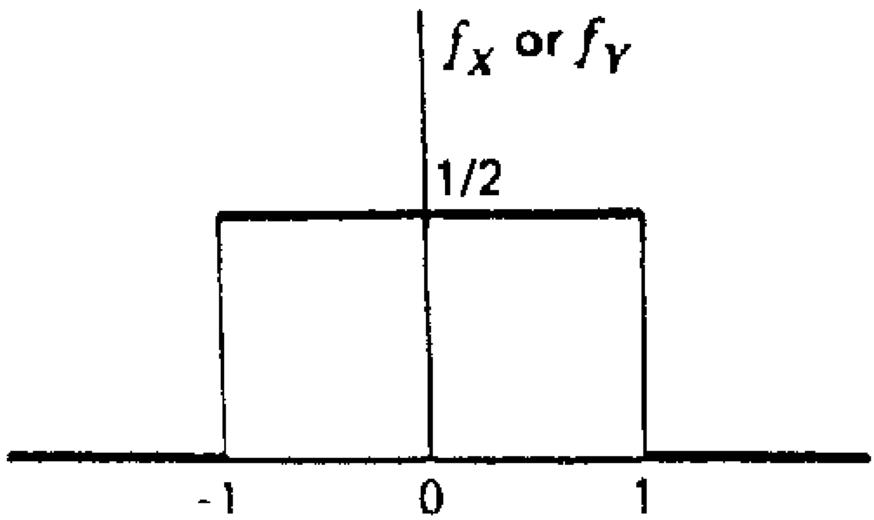

Figura 6: Função de densidade de probabilidade para X, Y e V.

A função de densidade de probabilidade de $Z = X + Y$ pode ser obtida através da convolução de duas funções como a mostrada na figura 6, cujo resultado é mostrado na figura 7

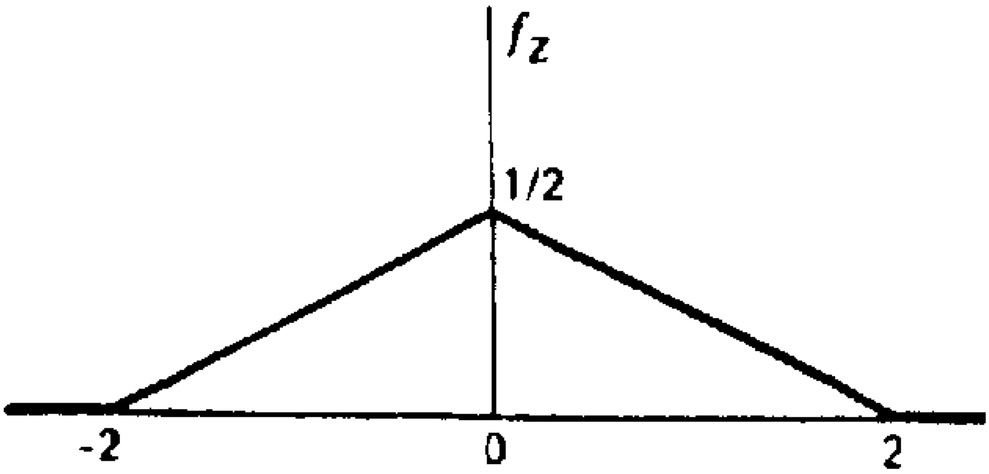

Figura 7: Função de densidade de probabilidade para Z.

Para $W = X + Y + V$, a função densidade de probabilidade, mostrada na figura 8, pode ser obtida pela convolução das funções mostradas nas figuras 6 e 7.

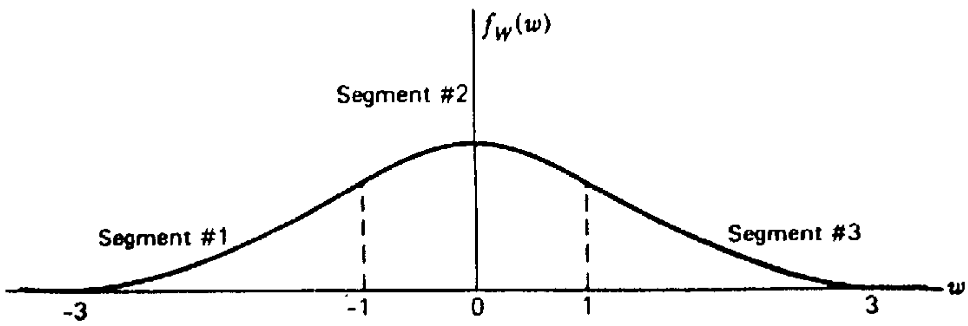

Figura 8: Função de densidade de probabilidade para W.

Pode-se facilmente perceber a semelhança entre esta curva e a curva de uma densidade normal com média zero. Para o somatório de quatro variáveis aleatórias com densidade como mostrado na figura 6, o resultado seria uma curva *formada por segmentos de cúbicas de -4 a $+4$, cuja aparência se assemelhará mais ainda com a curva de uma densidade normal, e assim por diante. Cada convolução adicional resultará uma curva mais próxima da curva normal.*

A generalização deste resultado leva à conclusão de que a superposição de variáveis aleatórias independentes tente à curva normal, independentemente da distribuição de cada variável aleatória contribuindo para a soma. Este resultado é conhecido como Teorema do Limite Central.

Em aplicações de engenharia, tipicamente o ruído é devido à superposição de pequenas contribuição de muitos fenômenos. Assim, tem-se um bom motivo para assumir que o ruído possui uma distribuição normal.

O teorema do limite central explica também o interesse exagerado em variáveis aleatórias normais, pois elas são uma ocorrência bastante comum na natureza.

13 Transformações de Variáveis Aleatórias

Na análise de sistemas é comum a utilização de transformações matemáticas que mapeiam um conjunto de variáveis (por exemplo entradas) em outro conjunto de variáveis (saídas). Considere a relação entrada-saída descrita por

$$y = g(x) \tag{13}$$

onde x é uma realização da variável aleatória de entrada X e y é a correspondente realização da variável aleatória de saída Y.

Assumindo-se que $g(x)$ é um mapeamento um-para-um para todos os valores possíveis de x, a relação (13) pode ser invertida:

$$x = h(y) \tag{14}$$

As probabilidades de que X e Y estejam dentro de regiões diferenciais correspondentes devem ser iguais, ou seja:

$$P(x \leq X \leq x + dx) = P(y \leq Y \leq y + dy)$$

ou

$$\int_x^{x+dx} f_X(u)du = \begin{cases} \int_y^{y+dy} f_Y(u)du & \text{para } dy \text{ positivo} \\ -\int_y^{y+dy} f_Y(u)du & \text{para } dy \text{ negativo} \end{cases} \tag{15}$$

Note que a expressão (15) expõe uma das particularidades deste problema. Se dx leva a dy negativo, a integral de f_Y deve ser de $y + dy$ a y, de forma a manter uma probabilidade positiva.

Assumindo-se dx positivo, tem-se que o equivalente diferencial de (15) é

$$f_X(x)dx = f_Y(y)|dy|$$

Além disso, x está restrito a ser igual a $h(y)$. Tem-se portanto

$$f_Y(y) = \left|\frac{dx}{dy}\right| f_X(h(y)) \tag{16}$$

ou ainda

$$f_Y(y) = \left|\frac{dh(y)}{dy}\right| f_X(h(y))$$

Exemplo 7 *Considere uma entrada $X \sim N(0, \sigma_X^2)$ e obtenha a função densidade de saída para as seguintes transformações*

1. $y = Kx$ *(K=constante)*
2. $y = x^3$

1. $y = Kx$ *(K=constante)*

$$x = h(y) = \frac{1}{K}y$$

$$\left|\frac{dh(y)}{dy}\right| = \left|\frac{1}{K}\right|$$

De (16) pode-se obter a expressão para f_Y:

$$f_Y(y) = \frac{1}{K}\frac{1}{\sqrt{2\pi}\sigma_X}exp\left[-\frac{\left(\frac{y}{K}\right)^2}{2\sigma_X^2}\right]$$

ou reescrevendo na forma normalizada:

$$f_Y(y) = \frac{1}{\sqrt{2\pi(K\sigma_X)^2}}exp\left[-\frac{y^2}{2(K\sigma_X)^2}\right]$$

Portanto, pode-se concluir que a transformação de uma variável aleatória normal com média zero por um fator de escala resulta outra variável aleatória normal com uma alteração correspondente no seu desvio padrão[5].

[5] O desvio padrão é equivalente à amplitude da variável aleatória

Em outras palavras, a "normalidade"da variável aleatória é preservada em uma transformação linear[6].

2. $y = x^3$

   Invertendo-se a transformação tem-se

$$x = h(y) = \sqrt[3]{y}$$

   A derivada de x *é:*

$$\frac{dh(y)}{dy} = \frac{1}{3}y^{-2/3}$$

   Note que $y^{2/3}$ *pode ser escrito como* $(y^{1/3})^2$*, logo* $y^{2/3}$ *é sempre positivo para* $y^{1/3}$ *real. A função densidade para* Y *é portanto*

$$f_Y(y) = \frac{1}{3y^{2/3}} \frac{1}{\sqrt{2\pi}\sigma_X} e^{-(y^{1/3})^2/2\sigma_X}$$

   Este é um exemplo de uma transformação não linear que converte uma variável aleatória normal para uma forma não gaussiana.

14 Função de Densidade Normal Multivariável

Nas seções 9 e 10 foram abordados as funções densidade para os casos monovariável e bivariável. Nesta seção será obtida uma forma geral para funções de densidade normais n-dimensionais.

Considere um conjunto de n variáveis aleatórias gaussianas $X_1, X_2, \ldots, X_n$, que pode ser escrito na forma de um vetor de variáveis aleatórias:

$$X = \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_n \end{bmatrix}$$

Em geral, os componentes de X podem ser correlacionados e ter médias $m_1, m_2, \ldots, m_n$ diferentes de zero. Portanto, define-se o vetor de médias:

[6]Pode-se também provar que a soma de duas variáveis aleatórias normais é uma variável aleatória normal. Vide exercício 5

$$m = \begin{bmatrix} m_1 \\ m_2 \\ \vdots \\ m_n \end{bmatrix}$$

Similarmente, o conjunto de realizações $x_1, x_2, \ldots, x_n$ de $X_1, X_2, \ldots, X_n$ pode ser escrito na forma de vetor:

$$x = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$$

A *matriz de covariância* para a variável X é definida como

$$C = \begin{bmatrix} E\left[(X_1 - m_1)^2\right] & E\left[(X_1 - m_1)(X_2 - m_2)\right] & \cdots \\ E\left[(X_2 - m_2)(X_1 - m_1)\right] & & \\ \vdots & \ddots & \\ & & E\left[(X_n - m_n)^2\right] \end{bmatrix}$$

Os termos na diagonal principal de C são as variâncias das variáveis, e os termos fora da diagonal são as covariâncias.

As variáveis aleatórias $X_1, X_2, \ldots, X_n$ são ditas *conjuntamente normais* ou *conjuntamente gaussianas* se a sua função de densidade de probabilidade conjunta é dada por

$$f_X(x) = \frac{1}{(2\pi)^{n/2}|C|^{1/2}} exp\left\{-\frac{1}{2}\left[(x-m)^T C^{-1}(x-m)\right]\right\} \tag{17}$$

Note que $f_X(x)$ definida pela expressão (17) é escalar e que C^{-1} deve existir para que $f_X(x)$ esteja adequadamente definida. Obviamente, a expressão (17) reduz-se à forma normal padrão para o caso monovariável. Para o caso bivariável pode-se escrever F_X explicitamente em termos de x_1 e x_2:

$$X = \begin{bmatrix} X_1 \\ X_2 \end{bmatrix} \quad x = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \quad m = \begin{bmatrix} m_1 \\ m_2 \end{bmatrix}$$

e

$$C = \begin{bmatrix} E\left[(X_1 - m_1)^2\right] & E\left[(X_1 - m_1)(X_2 - m_2)\right] \\ E\left[(X_2 - m_2)(X_1 - m_1)\right] & E\left[(X_2 - m_2)^2\right] \end{bmatrix} = \begin{bmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{bmatrix}$$

O determinante e a inversa de C são dados por

$$|C| = \begin{vmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{vmatrix} = (1-\rho^2)\sigma_1^2\sigma_2^2$$

$$C^{-1} = \begin{bmatrix} \frac{\sigma_2^2}{|C|} & -\frac{\rho\sigma_1\sigma_2}{|C|} \\ -\frac{\rho\sigma_1\sigma_2}{|C|} & \frac{\sigma_1^2}{|C|} \end{bmatrix} = \begin{bmatrix} \frac{1}{(1-\rho^2)\sigma_1^2} & -\frac{\rho}{(1-\rho^2)\sigma_1\sigma_2} \\ -\frac{\rho}{(1-\rho^2)\sigma_1\sigma_2} & \frac{1}{(1-\rho^2)\sigma_2^2} \end{bmatrix}$$

e portanto

$$\begin{aligned} f_{X_1X_2}(x_1,x_2) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} exp\left\{-\frac{1}{2(1-\rho^2)}\left[\frac{(x_1-m_1)^2}{\sigma_1^2}\right.\right. \\ \left.\left. -\frac{2\rho(x_1-m_1)(x_2-m_2)}{\sigma_1\sigma_2} + \frac{(x_2-m_2)^2}{\sigma_2^2}\right]\right\} \end{aligned} \quad (18)$$

Um esboço de $f_{X_1X_2}(x_1,x_2)$ é mostrado na figura 9. A densidade bivariável normal é uma superfície suave no plano x_1, x_2 com pico diretamente acima do o ponto (m_1, m_2). Contornos de igual altura na superfície $f_{X_1X_2}(x_1,x_2)$ projetam-se como elipses no plano x_1, x_2 (mostrada na figura 9 para um coeficiente de correlação positivo). Pontos na elipse representam combinações igualmente prováveis de x_1 e x_2. Se $\rho = 0$ tem-se o caso de X_1 e X_2 descorrelacionados e as elipses tem seus eixos principal e secundário paralelos aos eixos x_1 e x_2. Se $\sigma_1 = \sigma_2$ (mantendo-se $\rho = 0$) as elipses degeneram para círculos. Por outro lado, quando $|\rho|$ tende à unidade, as elipses tornam-se mais excêntricas.

O caso descorrelacionado é de especial interesse e neste caso $f_{X_1X_2}(x_1,x_2)$ reduz-se à

$$\begin{aligned} f_{X_1X_2}(x_1,x_2) &= \frac{1}{2\pi\sigma_1\sigma_2} exp\left\{-\frac{1}{2}\left[\frac{(x_1-m_1)^2}{\sigma_1^2} + \frac{(x_2-m_2)^2}{\sigma_2^2}\right]\right\} \\ &= \frac{1}{\sqrt{2\pi}\sigma_1} e^{-(x_1-m_1)^2/2\sigma_1^2} \frac{1}{\sqrt{2\pi}\sigma_2} e^{-(x_2-m_2)^2/2\sigma_2^2} \end{aligned} \quad (19)$$

Portanto, duas variáveis aleatórias normais que são descorrelacionadas são também estatisticamente independentes. Pode-se também verificar facilmente que esta propriedade é mantida para qualquer número de variáveis aleatórias normais não correlacionadas. Note que, em geral, correlação zero não implica independência estatística. No entanto, no caso gaussiano, implica.

Esta interpretação geométrica apresentada aqui para o caso bivariável pode ser generalizada para três ou mais variáveis.

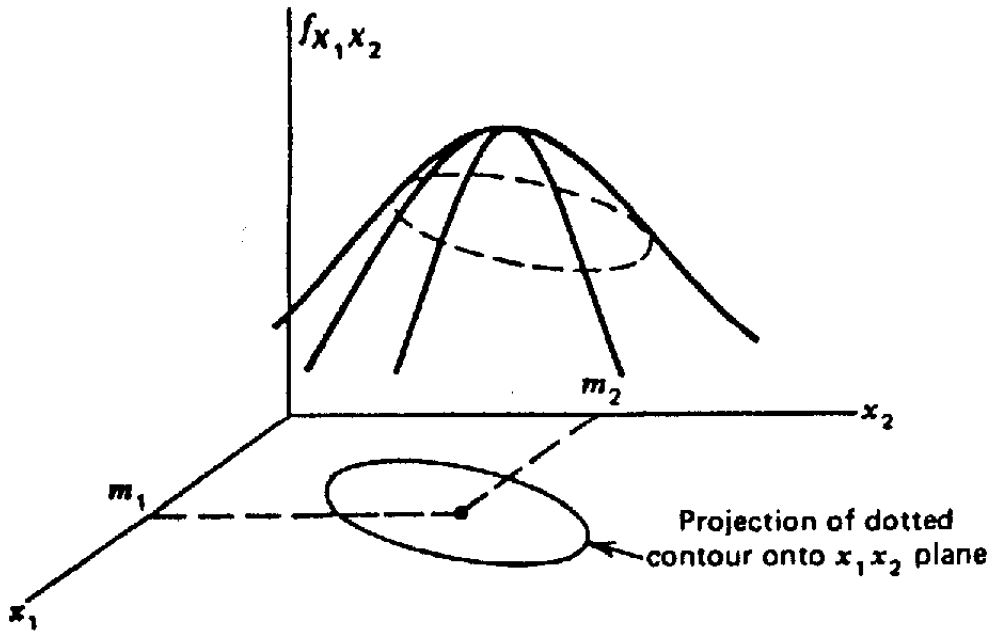

Figura 9: Distribuição bivariável.

15 Propriedades de Variáveis Aleatórias Gaussianas sujeitas à Transformações Lineares

A função densidade para variáveis aleatórias conjuntamente normais $X_1, X_2, \ldots, X_n$ é dada por

$$f_X(x) = \frac{1}{(2\pi)^{n/2}|C_X|^{1/2}} exp\left\{-\frac{1}{2}\left[(x - m_X)^T C_X^{-1}(x - m_X)\right]\right\} \tag{20}$$

Definindo-se um conjunto de variáveis aleatórias $Y_1, Y_2, \ldots, Y_n$ linearmente relacionadas com $X_1, X_2, \ldots, X_n$ através da expressão

$$y = Ax + b \tag{21}$$

onde b é um vetor constante e A é uma matriz quadrada não singular, tem-se que a função densidade para Y será dada por uma generalização de (16)

$$f_Y(y) = f_X(h(y))\,|J_h(y)| \tag{22}$$

Invertendo-se a relação (21) obtém=se

$$x = A^{-1}y - A^{-1}b \tag{23}$$

com

$$A^{-1} = \begin{bmatrix} d_{11} & d_{12} & \cdots & d_{1n} \\ d_{21} & d_{22} & \cdots & d_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ d_{n1} & d_{n2} & \cdots & d_{nn} \end{bmatrix}$$

e portanto

$$\begin{aligned} x_1 &= (d_{11}y_1 + d_{12}y_2 + \cdots) - (d_{11}b_1 + d_{12}b_2 + \cdots) \\ x_2 &= (d_{21}y_1 + d_{22}y_2 + \cdots) - (d_{21}b_1 + d_{22}b_2 + \cdots) \\ x_3 &= (d_{31}y_1 + d_{32}y_2 + \cdots) - (d_{31}b_1 + d_{32}b_2 + \cdots) \\ \vdots &= \vdots \end{aligned} \tag{24}$$

O Jacobiano da transformação é então

$$J_h(y) = \begin{bmatrix} \frac{\partial x_1}{\partial y_1} & \frac{\partial x_2}{\partial y_1} & \cdots \\ \frac{\partial x_1}{\partial y_2} & \frac{\partial x_2}{\partial y_2} & \cdots \\ \vdots & \ddots & \vdots \end{bmatrix} \tag{25}$$

$$|J_h(y)| = \left| \text{Det} \begin{bmatrix} d_{11} & d_{21} & \cdots \\ d_{21} & d_{22} & \cdots \\ \vdots & \ddots & \vdots \end{bmatrix} \right| = \left| \text{Det} \left(A^{-1} \right)^T \right| = \left| \text{Det} \left(A^{-1} \right) \right| \tag{26}$$

Substituindo-se (23) e (26) em (22) tem-se

$$f_Y(y) = \frac{|\text{Det}\,(A^{-1})|}{(2\pi)^{n/2}\,|C_X|^{1/2}} exp \left\{ -\frac{1}{2} \left[\left(A^{-1}y - A^{-1}b - m_X \right)^T C_X^{-1} \left(A^{-1}y - A^{-1}b - m_X \right) \right] \right\} \tag{27}$$

A média de Y pode ser calculada tomando-se a esperança de ambos os lados da transformação, portanto

$$m_y = Am_x + b$$

O argumento da exponencial da expressão (27) pode então ser escrito como

$$
\begin{aligned}
- & \frac{1}{2}\left[\left(A^{-1}y - A^{-1}b - A^{-1}Am_X\right)^T C_X^{-1}\left(A^{-1}y - A^{-1}b - A^{-1}Am_X\right)\right] \\
= & -\frac{1}{2}\left[(y - m_Y)^T \left(A^{-1}\right)^T C_X^{-1}A^{-1}(y - m_Y)\right] \\
= & -\frac{1}{2}\left[(y - m_Y)^T \left(AC_XA^T\right)^{-1}(y - m_Y)\right] \qquad (28)
\end{aligned}
$$

Notando-se que

$$
\left|\text{Det}\left(A^{-1}\right)\right| = \frac{1}{|\text{Det}A|} = \frac{1}{|\text{Det}A|^{1/2}\left|\text{Det}A^T\right|^{1/2}}
$$

chega-se à

$$
f_Y(y) = \frac{1}{(2\pi)^{n/2}\left|AC_XA^T\right|^{1/2}} exp\left\{-\frac{1}{2}\left[(y - m_Y)^T \left(AC_XA^T\right)^{-1}(y - m_Y)\right]\right\}
$$

que utilizando-se a definição

$$
C_Y = AC_XA^T
$$

pode-se escrever na forma

$$
f_Y(y) = \frac{1}{(2\pi)^{n/2}\left|C_Y\right|^{1/2}} exp\left\{-\frac{1}{2}\left[(y - m_Y)^T C_Y^{-1}(y - m_Y)\right]\right\} \qquad (29)
$$

Portanto, f_Y também é normal e possui média e matriz de covariância dadas por $m_Y = Am_X + b$ e $C_Y = AC_XA^T$. Logo, a normalidade é preservada em uma transformação linear. Apenas a média e a variância são alteradas, a forma da função densidade permanece inalterada.

Uma transformação linear particularmente interessante é a que produz uma nova matriz de covariância SC_XS^T que é diagonal. Isto pode ser obtido através da transformação de similaridade, onde a matriz S é formada pelos autovetores de C_X. Neste caso a transformação produz um conjunto de variáveis aleatórias normais descorrelacionadas e portanto estatisticamente independentes. Esta transformação sempre existirá se C_X for positiva definida, que no caso de uma matriz de covariância implica todos os coeficientes de correlação serem, em módulo, menores que a unidade.

Resumo das propriedades de variáveis aleatórias normais múltiplas:

1. A função densidade de probabilidade de uma variável aleatória vetorial X é completamente definida através da média e da matriz de covariância de X.

2. A matriz de covariância de X é positiva definida se os módulos de todos os coeficientes de correlação forem menores do que a unidade.

3. Se variáveis aleatórias normais são descorrelacionadas, elas também são estatisticamente independentes.

4. Uma transformação linear de variáveis aleatórias normais leva à outro conjunto de variáveis aleatórias normais. Uma transformação descorrelacionadora sempre existirá se a matriz de covariância for positiva definida.

5. Se a função densidade conjunta para n variáveis aleatórias é normal, todas das densidades condicionais e marginais associadas com as n variáves também serão normais.

16 Limites, Convergência e Estimadores não Polarizados

Um estimador é dito não polarizado se

$$E(\text{estimador de } X) = E(X) \tag{30}$$

Considere uma sequência de variáveis aleatórias $Y_1, Y_2, \ldots, Y_n$. A sequência é dita *convergir em média* se

$$\lim_{n\to\infty} E\left[(Y_n - Y)^2\right] = 0 \tag{31}$$

A sequência *converge em probabilidade* para Y se

$$\lim_{n\to\infty} P\left(|Y_n - Y| \geq \epsilon\right) = 0 \tag{32}$$

onde ϵ é uma constante positiva arbitrariamente pequena.

A grosso modo, a convergência em média indica que a dispersão (variância) em torno do valor limite tende para zero no limite. Similarmente a convergência em probabilidade significa que um critério de precisão arbitrariamente pequeno pode ser obtido com probabilidade um quando $n \to \infty$. A convergência em média é um requisito mais severo do que a convergência em probabilidade. Pode se provado que a convergência em média garante a convergência em probabilidade, mas o contrário não é verdadeiro.

17 Exercícios

1. Um jogo de dados possui as seguintes regras: O jogador lança dois dados e aposta contra a banca. Se o resultado do primeiro lançamento for 7 ou 11 o jogador ganha imediatamente; se for 2, 3 ou 12, o jogador perde imediatamente. Caso o resultado seja outro número, o jogador lança os dados sucessivamente até o mesmo número aparecer novamente, quando ele ganha, ou até aparecer um 7, quando ele perde. Qual a probabilidade total de se ganhar neste jogo?

2. Qual o tamanho do espaço amostral dos seguintes experimentos:

   (a) Retirar uma carta de um baralho com 52 cartas.

   (b) Lançar dois dados.

   (c) Lançar dois dados e observar sua soma.

   (d) Arremessar um dardo.

3. Suponha um retificador de meia-onda cuja entrada é uma sinal gaussiano qualquer com média zero e variância σ_X^2. Faça um esboço das funções densidade e distribuição de probabilidade do sinal de saída.

4. Considere o arremesso de um dardo em um alvo descrito pelas coordenadas x e y. Após o jogador estar treinado, é razoável supor que os erros nas direções horizontal e vertical terão as mesmas características e serão independentes. Supondo que estes erros tenham uma distribuição normal, com desvio padrão σ e média 0 (as coordenadas do centro do alvo são 0,0), calcule a expressão da probabilidade do dardo atingir uma região de raio r em torno do centro do alvo.

5. Mostre que a soma de duas variáveis aleatórias $X \sim N(0, \sigma_X^2)$ e $Y \sim N(0, \sigma_Y^2)$ é uma variável aleatória $Z \sim N(0, \sigma_Z^2)$, onde $\sigma_Z^2 = \sigma_X^2 + \sigma_Y^2$.

6. Prove (no exemplo 6 é feita apenas uma demonstração e não uma prova) o teorema do limite central. Dica: Utilize a transformada de Fourier e a expansão em série da exponencial.

7. Seja $X \sim N(0, \sigma_X^2)$. Determine a função densidade de $y = x^2$.

8. Calcule a média e a variância da saída do retificador do item 3.

9. Sejam X e Y duas variáveis aleatórias com função de densidade de probabilidade conjunta definida como:

$$f_{XY}(x,y) = \begin{cases} 0.25, & -1 \le x \le 1 \text{ e } -1 \le y \le 1 \\ 0, & \text{caso contrário} \end{cases}$$

X e Y são estatisticamente independentes?

10. Sejam X e Y duas variáveis aleatórias independentes com funções de densidade de probabilidade dadas por:

$$f_X(x) = \frac{1}{2}e^{-|x|}$$

$$f_Y(y) = e^{-2|y|}$$

ache a função de densidade de probabilidade de $X + Y$.

11. A variável aleatória gaussiana:

$$X = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$$

é completamente descrita pela média e matriz de covariância dadas por

$$m_X = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$$

$$C_X = \begin{bmatrix} 4 & 1 \\ 1 & 1 \end{bmatrix}$$

Considere outra variável aleatória Y relacionada com X pela expressão

$$y = Ax + b$$

onde

$$A = \begin{bmatrix} 2 & 1 \\ 1 & -1 \end{bmatrix}$$

$$b = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

Ache a média e a matriz de covariância de Y.