



Estimação e Filtragem Estocástica
Programa de Pós-Graduação em Engenharia de Sistemas Eletrônicos e de Automação
Departamento de Engenharia Elétrica
Universidade de Brasília

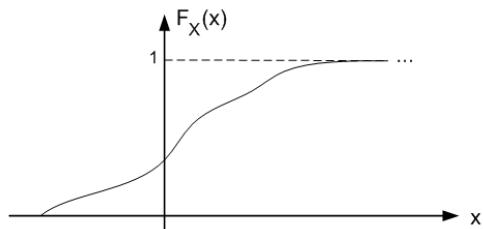
FUNDAMENTOS DE PROBABILIDADE E ESTATÍSTICA

Geovany A. Borges
gaborges@ene.unb.br

Variáveis aleatórias

■ Variáveis aleatórias contínuas

- Função de distribuição: $F_X(x) = \Pr\{X(\omega) \leq x\}$



- Propriedades de $F_X(x)$:
Monotônica não-decrescente: se $x_1 < x_2$, então $\{X(\omega) \leq x_1\} \subset \{X(\omega) \leq x_2\}$ e assim

$$F_X(x_1) \leq F_X(x_2)$$

Variáveis aleatórias

Definição 1 (Variável Aleatória (v.a.)):

Uma função real $X(\cdot)$ definida no espaço de probabilidades Ω é dita uma variável aleatória (v.a.) se para todo número $x \in \mathfrak{R}$, a inequação

$$X(\omega) \leq x$$

representa um conjunto (eventos) cuja probabilidade de ocorrência é definida.

Tem-se ainda que

$$\begin{aligned} \lim_{x \rightarrow -\infty} F_X(x) &= 0 \\ \lim_{x \rightarrow \infty} F_X(x) &= 1 \end{aligned}$$

Se $x_1 < x_2$ então $\{X(\omega) \leq x_2\} = \{X(\omega) \leq x_1\} \cup \{x_1 < X(\omega) \leq x_2\}$, que são conjuntos disjuntos. Assim

$$\Pr\{X(\omega) \leq x_2\} = \Pr\{X(\omega) \leq x_1\} + \Pr\{x_1 < X(\omega) \leq x_2\}$$

levando a

$$\Pr\{x_1 < X(\omega) \leq x_2\} = \Pr\{X(\omega) \leq x_2\} - \Pr\{X(\omega) \leq x_1\}.$$

- Função de Densidade de Probabilidade (FDP):

$$p_X(x) = \frac{dF_X(x)}{dx}$$

Como $F_X(x)$ é monotônica não-decrescente, então $p_X(x) \geq 0$.
E ainda

$$F_X(x) = \int_{-\infty}^x p_X(\xi) d\xi$$

com

$$F_X(\infty) = \int_{-\infty}^{\infty} p_X(\xi) d\xi = 1.$$

Tem-se ainda que

$$p_X(-\infty) = p_X(\infty) = 0.$$

Considerando o intervalo $x \leq X(\omega) \leq x + \Delta x$, tem-se

$$\begin{aligned} \Pr\{x \leq X(\omega) \leq x + \Delta x\} &= \int_x^{x+\Delta x} p_X(\xi) d\xi \\ &\approx p_X(x) \Delta x \end{aligned} \quad (1)$$

se Δx suficientemente pequeno.

Caracterização de variáveis aleatórias

■ Expectância (esperança, média ou primeiro momento)

- Caso contínuo:

$$E\{X(w)\} = \int_{-\infty}^{\infty} x p_X(x) dx \quad (4)$$

- Caso discreto:

$$E\{X(w)\} = \sum_{\xi \in \Omega} \xi m_X(\xi) \quad (5)$$

- Propriedades:

- Se $y = f(x)$ com $f(\cdot)$ sendo uma função escalar monotônica no intervalo $x_1 \leq x \leq x_2$. Tem-se que

$$\Pr\{y_1 \leq y \leq y_2\} = \Pr\{x_1 \leq x \leq x_2\}. \quad (6)$$

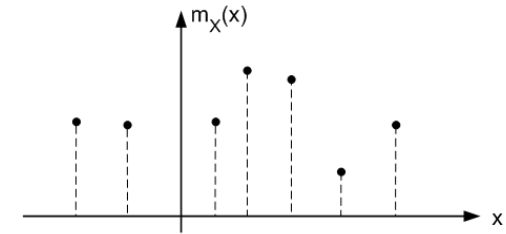
Variáveis aleatórias

■ Variáveis aleatórias discretas (número contável de valores)

- Função de massa: $m_X(x) > 0$ tal que

$$F_X(x) = \Pr\{X(\omega) \leq x\} \quad (2)$$

$$= \sum_{\xi \leq x} m_X(\xi) \quad (3)$$



A partir de então, pode-se demonstrar que

$$p_Y(y) = p_X(f^{-1}(y)) \left| \frac{dy}{dx} \right|_{x=f^{-1}(y)}^{-1} \quad (7)$$

Assim

$$E\{Y\} = \int_{-\infty}^{\infty} y p_Y(y) dy = \int_{-\infty}^{\infty} f(x) p_X(x) dx \quad (8)$$

- Sendo $X = c$, com c sendo uma constante, ou seja, $p_X(x) = \delta(x - c)$, então $E\{X\} = c$.
- Sendo $Y = a \cdot g_1(X) + b \cdot g_2(X)$, com a e b sendo constantes, então

$$E\{Y\} = a \cdot E\{g_1(X)\} + b \cdot E\{g_2(X)\} \quad (9)$$

Caracterização de variáveis aleatórias

- n -ésimo momento (momento de ordem n)

$$E\{X^n\} = \int_{-\infty}^{\infty} x^n p_X(x) dx$$

- n -ésimo momento central (momento central de ordem n)

$$E\{(X - E\{X\})^n\} = \int_{-\infty}^{\infty} (x - E\{X\})^n p_X(x) dx$$

Caso especial: para $n = 2$ o momento central é chamado de variância

$$\begin{aligned} var\{X\} &= E\{(X - E\{X\})^2\} \\ &= E\{X^2\} - E\{X\}^2 \end{aligned}$$

Se $p_X(x)$ for monomodal e simétrica com relação à média $E\{X\}$, então $med\{X\} = mod\{X\} = E\{X\}$.

Define-se também $\sigma\{X\} = \sqrt{var\{X\}}$ como desvio padrão da v.a. X .

- Mediana

$$med\{X\} = \arg_{x_{med}} F_X\{x_{med}\} = \frac{1}{2}$$

- Moda: máximos locais da FDP $p_X(x)$, ou seja $mod\{X\}$ deve satisfazer a ambas as relações abaixo:

$$\begin{aligned} \left. \frac{\partial p_X(x)}{\partial x} \right|_{x=mod\{X\}} &= 0 \\ \left. \frac{\partial^2 p_X(x)}{\partial x^2} \right|_{x=mod\{X\}} &< 0 \end{aligned}$$

Se $p_X(x)$ tiver apenas uma moda, diz-se que ela é *monomodal*. Uma FDP com duas ou mais modas é dita *multimodal*.

Distribuições unidimensionais discretas

- Distribuição de Bernoulli

Nesta distribuição $\Omega = \{0, 1\}$ representa a ocorrência ou não de um evento. A função de massa é dada por

$$m_X(x) = \begin{cases} 1 - p, & x = 0 \\ p, & x = 1 \end{cases} \quad (10)$$

em que p é a probabilidade de ocorrência do evento ($x = 1$).

- Distribuição Binomial

Uma v.a. Y com distribuição Binomial representa a soma de n v.a.s X_k de Bernoulli com parâmetro p

$$Y = \sum_{k=1}^n X_k. \quad (11)$$

Sendo assim, $\Omega = \{0, 1, 2, \dots, n\}$ e

$$m_Y(k) = \binom{n}{k} p^k (1-p)^{n-k}$$

com

$$\binom{n}{k} = \frac{n!}{k!(n-k)!}$$

Usada na contagem de eventos.

■ Distribuição Uniforme:

Uma v.a. discreta uniforme X possui espaço amostral $\Omega = \{a, a+1, \dots, b-1, b\}$ tal que $b \geq a$ são parâmetros que definem um intervalo e

$$m_X(k) = \begin{cases} \frac{1}{b-a+1}, & k = a, \dots, b \\ 0, & \text{caso contrário.} \end{cases}$$

Usada quando os elementos de Ω forem equiprováveis.

■ Distribuição Normal (ou Gaussiana)

Uma v.a. X de distribuição Gaussiana com parâmetros μ e σ^2 possui $\Omega = \mathbb{R}$ e FDP

$$p_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp \left[-\frac{1}{2} \frac{(x-\mu)^2}{\sigma^2} \right]. \quad (15)$$

Notação: $X \sim N(\mu, \sigma^2)$

Tem-se ainda que

$$\begin{aligned} E\{X\} &= \mu, \\ E\{(X - E\{X\})^2\} &= \sigma^2. \end{aligned}$$

A função de distribuição de X é dada por

$$F_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^x \exp \left[-\frac{1}{2} \frac{(u-\mu)^2}{\sigma^2} \right] du = G \left(\frac{x-\mu}{\sigma} \right)$$

Distribuições unidimensionais contínuas

■ Distribuição Uniforme

Uma v.a. X de distribuição uniforme com parâmetros $b \geq a$ possui $\Omega = [a, b]$ e FDP

$$p_X(x) = \begin{cases} \frac{1}{b-a}, & a \leq x \leq b \\ 0, & \text{caso contrário.} \end{cases} \quad (12)$$

Notação: $X \sim U(a, b)$.

Tem-se ainda que

$$F_X(x) = \begin{cases} 0, & x < a \\ \frac{x-a}{b-a}, & a \leq x < b \\ 1, & x \geq b \end{cases} \quad (13)$$

$$E\{X\} = \frac{b-a}{2} \quad (14)$$

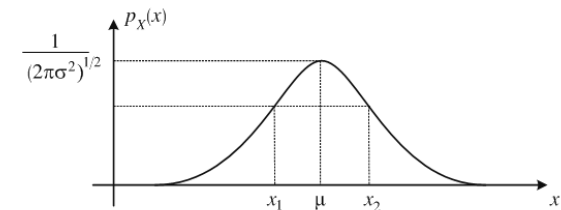
em que $G(\cdot)$ é denominada de função de Gauss sendo dada por

$$G(u) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^u \exp \left[-\frac{v^2}{2} \right] dv$$

Deve ser observado que sendo $G(\infty) = F_X(\infty) = 1$, então

$$\int_{-\infty}^{\infty} \exp \left[-\frac{v^2}{2} \right] dv = \sqrt{2\pi} \cdot G(\infty) = \sqrt{2\pi}$$

- Interpretação gráfica da FDP: monomodal e simétrica com relação à média.



Exemplo 1 (Transformação afim de uma v.a. Gaussiana):

Se $X \sim N(\mu_X, \sigma_X^2)$, qual seria a distribuição da v.a. $Y = aX + b$, sendo $a \neq 0$ e b constantes reais? Solução: aplicar a eq. (7).

Tem-se ainda que

$$\begin{aligned} E\{X\} &= n \\ \text{var}\{X\} &= 2n \end{aligned}$$

Exemplo 2 (Transformação quadrática de uma v.a. Gaussiana):

Se $V \sim N(0, 1)$, é uma v.a. de distribuição normal, então $W = V^2$ é uma v.a. de distribuição χ_1^2 .

Distribuições unidimensionais contínuas

■ Distribuição χ_n^2 (Qui-quadrado com n graus de liberdade)

Uma v.a. X de distribuição χ_n^2 com parâmetro $n \geq 1$ inteiro possui $\Omega = [0, \infty)$ e FDP

$$p_X(x) = \begin{cases} \frac{x^{n/2-1}}{2^{n/2}\Gamma(n/2)} \exp(-x/2), & x \geq 0 \\ 0 & \text{caso contrário} \end{cases} \quad (16)$$

com

$$\Gamma(u) = \int_0^\infty x^{u-1} \exp(-x) dx,$$

$u > 0$, sendo função Gama. Esta função possui as seguintes propriedades:

$$\Gamma(u) = (u-1)\Gamma(u-1) = (u-1)!$$

se u é um inteiro e $\Gamma(1/2) = \sqrt{\pi}$.

Notação: $X \sim \chi_n^2$

Variáveis aleatórias de distribuição conjunta

■ Se as v.a.s $X_1, X_2, \dots, X_n$ estão definidas no mesmo espaço Ω , então elas são ditas de distribuição conjunta com distribuição dada por

$$\begin{aligned} F_{X_1, X_2, \dots, X_n}(x_1, x_2, \dots, x_n) &= \Pr\{X_1 \leq x_1, X_2 \leq x_2, \dots, X_n \leq x_n\} \\ &= \int_{-\infty}^{x_1} \int_{-\infty}^{x_2} \cdots \int_{-\infty}^{x_n} p_{X_1, X_2, \dots, X_n}(u_1, u_2, \dots, u_n) du_1 du_2 \cdots du_n \end{aligned}$$

com $p_{X_1, X_2, \dots, X_n}(x_1, x_2, \dots, x_n)$ sendo a FDP conjunta tal que

$$p_{X_1, X_2, \dots, X_n}(x_1, x_2, \dots, x_n) = \frac{\partial^n F_{X_1, X_2, \dots, X_n}(x_1, x_2, \dots, x_n)}{\partial x_1 \partial x_2 \cdots \partial x_n} \quad (17)$$

■ Propriedades de $F_{X_1, X_2, \dots, X_n}(x_1, x_2, \dots, x_n)$:

$$F_{X_1, \dots, X_i, \dots, X_n}(x_1, \dots, x_i = -\infty, \dots, x_n) = 0$$

uma vez que $\Pr\{X_i \leq -\infty\} = 0$.

$$F_{X_1, \dots, X_n}(-\infty, \dots, -\infty) = 0$$

$$F_{X_1, \dots, X_n}(\infty, \dots, \infty) = 1$$

$$p_{X_1, X_2, \dots, X_n}(x_1, x_2, \dots, x_n) \geq 0$$

$$\Pr\{x_a < X_1 \leq x_b, X_2 \leq x_2, \dots, X_n \leq x_n\} = F_{X_1, X_2, \dots, X_n}(x_b, x_2, \dots, x_n) - F_{X_1, X_2, \dots, X_n}(x_a, x_2, \dots, x_n)$$

$$\Pr\{x_1 < X_1 \leq x_1 + \Delta x_1, \dots, x_n < X_n \leq x_n + \Delta x_n\} \approx p_{X_1, X_2, \dots, X_n}(x_1, x_2, \dots, x_n) \cdot \Delta x_1 \cdots \Delta x_n$$

■ Expectância marginal $E\{X_k\}$, com $1 \leq k \leq n$: partindo de

$$E\{X_k\} = \int_{-\infty}^{\infty} x_k \cdot p_{X_k}(x_k) dx_k$$

Com

$$\begin{aligned} p_{X_k}(x_k) &= \frac{dF_{X_k}(x_k)}{dx_k} = \frac{\partial}{\partial x_k} F_{X_1, \dots, X_{k-1}, X_k, X_{k+1}, \dots, X_n}(\infty, \dots, x_k, \dots, \infty) \\ &= \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} p_{X_1, \dots, X_{k-1}, X_k, X_{k+1}, \dots, X_n}(u_1, \dots, u_{k-1}, x_k, u_{k+1}, \dots, u_n) \times \\ &\quad du_1, \dots, du_{k-1}, du_{k+1}, \dots, du_n \end{aligned}$$

Variáveis aleatórias de distribuição conjunta

■ Função de distribuição conjunta de um subconjunto de v.a.s: sendo $m < n$

$$\begin{aligned} F_{X_1, \dots, X_m}(x_1, \dots, x_m) &= F_{X_1, \dots, X_m, \dots, X_n}(x_1, \dots, x_m, \infty, \dots, \infty) \\ &= \underbrace{\int_{-\infty}^{x_1} \cdots \int_{-\infty}^{x_m} \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} p_{X_1, X_2, \dots, X_n}(u_1, u_2, \dots, u_n) du_1 du_2 \cdots du_n}_{n} \end{aligned}$$

$F_{X_1, \dots, X_m}(x_1, \dots, x_m)$ é chamada de distribuição marginal das v.a.s $X_1, \dots, X_m$.

Se $n = 2$, então tem-se uma distribuição Bivariável

Se $n > 2$, então tem-se uma distribuição Multivariável

Portanto

$$\begin{aligned} E\{X_k\} &= \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} x_k \cdot p_{X_1, \dots, X_k, \dots, X_n}(u_1, \dots, x_k, \dots, u_n) du_1, \dots, dx_k, \dots, du_n \\ &= \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} u_k \cdot p_{X_1, \dots, X_n}(u_1, \dots, u_n) du_1, \dots, du_n \end{aligned}$$

■ Independência: duas v.a.s de distribuição conjunta X_i e X_j , $i \neq j$, são ditas independentes se os eventos $\{X_i \leq x_i\}$ e $\{X_j \leq x_j\}$ forem independentes. Isto leva a

$$F_{X_i, X_j}(x_i, x_j) = F_{X_i}(x_i) F_{X_j}(x_j) \quad (18)$$

e em consequência

$$p_{X_i, X_j}(x_i, x_j) = \frac{\partial^2 F_{X_i, X_j}(x_i, x_j)}{\partial x_i \partial x_j} = p_{X_i}(x_i) p_{X_j}(x_j) \quad (19)$$

Variáveis aleatórias de distribuição conjunta

Exemplo 3 (Distribuição conjunta bivariável uniforme):

Considerando X e Y v.a.s distribuídas conjuntamente tal que

$$p_{X,Y}(x,y) = \begin{cases} \frac{1}{T_X T_Y}, & \text{no espaço } [0, T_X] \times [0, T_Y] \\ 0, & \text{caso contrário} \end{cases}$$

determinar as densidades marginais de X e Y e verificar se estas v.a.s são independentes.

Exercício 1 (Independência de v.a.s Gaussianas):

Considerando X e Y v.a.s gaussianas distribuídas conjuntamente tal que

$$p_{X,Y}(x,y) = c \exp \left\{ \frac{1}{18} [36x - y^2 - 9x^2 - 36] \right\}$$

determinar as condições necessárias para que X e Y sejam estatisticamente independentes.

Variáveis aleatórias de distribuição conjunta

- Expectância de uma função aleatória multivariável: sendo $Y = f(X_1, \dots, X_n)$, então

$$E\{Y\} = \int_{-\infty}^{\infty} y \cdot p_Y(y) dy \quad (22)$$

$$= \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} f(x_1, \dots, x_n) p_{X_1, \dots, X_n}(x_1, \dots, x_n) dx_1 \dots dx_n \quad (23)$$

- Covariância: se X_i e X_j , $i \neq j$, são duas v.a.s, define-se a covariância entre elas como

$$\begin{aligned} cov\{X_i, X_j\} &= E\{(X_i - E\{X_i\}) \cdot (X_j - E\{X_j\})\} \\ &= E\{X_i \cdot X_j\} - E\{X_i\}E\{X_j\} \end{aligned} \quad (24)$$

Variáveis aleatórias de distribuição conjunta

- V.A.s mutuamente independentes: as v.a.s $X_1, \dots, X_n$ são mutuamente independentes se

$$F_{X_1, \dots, X_n}(x_1, \dots, x_n) = \prod_{i=1}^n F_{X_i}(x_i) \quad (20)$$

e em consequência

$$p_{X_1, \dots, X_n}(x_1, \dots, x_n) = \prod_{i=1}^n p_{X_i}(x_i) \quad (21)$$

De forma similar, $X_1, \dots, X_i$ e $X_{i+1}, \dots, X_n$ são independentes se

$$\begin{aligned} p_{X_1, \dots, X_i, X_{i+1}, \dots, X_n}(x_1, \dots, x_i, x_{i+1}, \dots, x_n) = \\ p_{X_1, \dots, X_i}(x_1, \dots, x_i) \cdot p_{X_{i+1}, \dots, X_n}(x_{i+1}, \dots, x_n) \end{aligned}$$

com $cov\{X_i, X_i\} = var\{X_i\}$.

Intuitivamente, se $cov\{X_i, X_j\} \neq 0$ existe alguma relação (linear ou não) entre elas.

- Coeficiente de correlação:

$$\rho(X_i, X_j) = \frac{cov\{X_i, X_j\}}{\sigma\{X_i\}\sigma\{X_j\}} \quad (25)$$

tal que $-1 \leq \rho(X_i, X_j) \leq 1$.

Exemplo 4 (Covariância de v.a.s independentes):

Se X_i e X_j são independentes, demonstrar que

$$\text{cov}\{X_i, X_j\} = E\{X_i \cdot X_j\} - E\{X_i\}E\{X_j\} = 0 \quad (26)$$

o que significa que X_i e X_j são não-correlacionadas e portanto

$$E\{X_i \cdot X_j\} = E\{X_i\}E\{X_j\}$$

E ainda, $\rho(X_i, X_j) = 0$ se $\rho\{X_i\}\rho\{X_j\} \neq 0$.

■ Matriz de covariâncias de X :

$$\begin{aligned} \mathbf{P}_X &= E\{(X - E\{X\}) \cdot (X - E\{X\})^T\} \\ &= E\left\{\begin{bmatrix} X_1 - E\{X_1\} \\ X_2 - E\{X_2\} \\ \vdots \\ X_n - E\{X_n\} \end{bmatrix} \cdot [\bullet]^T\right\} \end{aligned}$$

Portanto,

$$\mathbf{P}_X = \begin{bmatrix} \text{var}\{X_1\} & \text{cov}\{X_1, X_2\} & \cdots & \text{cov}\{X_1, X_n\} \\ \text{cov}\{X_2, X_1\} & \text{var}\{X_2\} & \cdots & \text{cov}\{X_2, X_n\} \\ \vdots & \vdots & \ddots & \vdots \\ \text{cov}\{X_n, X_1\} & \text{cov}\{X_n, X_2\} & \cdots & \text{var}\{X_n\} \end{bmatrix}$$

sendo simétrica uma vez que $\text{cov}\{X_i, X_j\} = \text{cov}\{X_j, X_i\}$ e positiva semidefinida pois $\mathbf{u}^T \mathbf{P}_X \mathbf{u} \geq 0$ para todo $\mathbf{u} \neq \mathbf{0}$.

Vetores de variáveis aleatórias

■ Um vetor aleatório é formado a partir de v.a.s $X_1, X_2, \dots, X_n$ de distribuição conjunta, e dado por

$$X = \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_n \end{bmatrix} = [X_1 \quad X_2 \quad \cdots \quad X_n]^T$$

simplicando a notação, a PDF associada é dada por

$$p_{X_1, X_2, \dots, X_n}(x_1, x_2, \dots, x_n) = p_X(\mathbf{x})$$

em que $\mathbf{x} = [x_1 \quad x_2 \quad \cdots \quad x_n]^T$ é um vetor e $\int p_X(\mathbf{x}) d\mathbf{x} = 1$

Observa-se que $p_X(\mathbf{x})$ é um escalar. De forma similar

$$E\{X\} = [E\{X_1\} \quad E\{X_2\} \quad \cdots \quad E\{X_n\}]^T$$

Vetores de variáveis aleatórias

■ Distribuição Gaussiana Multivariada

$$p_X(\mathbf{x}) = \frac{1}{(2\pi)^{n/2} |\mathbf{P}_X|^{1/2}} \exp \left\{ -\frac{1}{2} (\mathbf{x} - E\{X\})^T \cdot \mathbf{P}_X^{-1} \cdot (\mathbf{x} - E\{X\}) \right\}$$

com $|\mathbf{P}_X|$ sendo o determinante de $\mathbf{P}_X$.

- Notação: $X \sim N(E\{X\}, \mathbf{P}_X)$
- Se os elementos de X forem não-correlacionados, ou seja, $\text{cov}\{X_i, X_j\} = 0$, $i \neq j$, então pode-se mostrar que

$$p_X(\mathbf{x}) = \prod_{i=1}^n p_{X_i}(x_i) = \prod_{i=1}^n \frac{1}{(2\pi)^{1/2} \text{var}\{X_i\}^{1/2}} \exp \left\{ -\frac{1}{2} \frac{(x_i - E\{X_i\})^2}{\text{var}\{X_i\}} \right\}$$

uma vez que $|\mathbf{P}_X| = \prod_{i=1}^n \text{var}\{X_i\}$ e

$$\mathbf{P}_X^{-1} = \begin{bmatrix} \frac{1}{\text{var}\{X_1\}} & 0 & \cdots & 0 \\ 0 & \frac{1}{\text{var}\{X_2\}} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \frac{1}{\text{var}\{X_n\}} \end{bmatrix}$$

Neste caso, como X é formado por elementos independentes e de mesma distribuição (embora de diferentes parâmetros $E\{X_i\}$ e $\text{var}\{X_i\}$), diz-se que X é formado por elementos i.i.d. (independentes e identicamente distribuídos).

- Pode-se verificar que, para $\mathbf{A}$ e $\mathbf{b}$ sendo contantes tais que $\dim\{\mathbf{A}\} = m \times n$ e $\dim\{\mathbf{b}\} = m \times 1$ e X sendo uma v.a. de distribuição multivariada, então a transformação afin

$$Y = \mathbf{A}X + \mathbf{b} \quad (27)$$

resulta em

$$E\{Y\} = \mathbf{A}E\{X\} + \mathbf{b} \quad (28)$$

$$\mathbf{P}_Y = \mathbf{A}\mathbf{P}_X\mathbf{A}^T \quad (29)$$

é χ_n^2 . Para verificar este resultado, considera-se a v.a.

$$U = \mathbf{P}_X^{-1/2}(X - E\{X\}) \quad (32)$$

tal que $\mathbf{P}_X^{-1/2} = (\mathbf{P}_X^{-1})^{1/2} = (\mathbf{P}_X^{1/2})^{-1}$, uma vez que $\mathbf{P}_X = \mathbf{P}_X^{1/2}\mathbf{P}_X^{1/2}$ é simétrica e positiva semidefinida, e $\mathbf{P}_X^{1/2}$ é sua matriz raiz-quadrada (também simétrica). Pode-se verificar que

$$\begin{aligned} E\{U\} &= \mathbf{0} \\ \mathbf{P}_U &= \mathbf{I}_n \end{aligned}$$

significando que os elementos de U são i.i.d. Gaussianos de média nula e variância unitária. Por outro lado, Q pode ser decomposta em

$$Q = U^T U = \sum_{i=1}^n U_i^2 \quad (33)$$

que é a definição de uma v.a. de distribuição χ_n^2 .

No mais, se X é Gaussiana, então Y também o será. Para comprovar esta última afirmação, faz-se necessário determinar $p_Y(\mathbf{y})$.

- Simulação de uma v.a. Gaussiana multivariável: usando o resultado acima, se $X \sim N(\mathbf{0}, \mathbf{I}_n)$, com $\mathbf{I}_n$ sendo a matriz identidade de dimensão n , então $Y \sim N(\mathbf{b}, \mathbf{A}\mathbf{A}^T)$. Isto significa que, se for desejado obter $Y \sim N(E\{Y\}, \mathbf{P}_Y)$, então basta escolher

$$\mathbf{b} = E\{Y\}, \quad (30)$$

$$\mathbf{A}\mathbf{A}^T = \mathbf{P}_Y, \quad (31)$$

Para determinar a matriz $\mathbf{A}$ tal que eq.(31) seja resolvida, usa-se a fatoração de Cholesky. Pela fatoração de Cholesky, obtém-se a matriz triangular superior $\mathbf{R}$ tal que $\mathbf{P}_Y = \mathbf{R}^T \mathbf{R}$. Portanto, tem-se $\mathbf{A} = \mathbf{R}^T$.

- A variável aleatória

$$Q = (X - E\{X\})^T \cdot \mathbf{P}_X^{-1} \cdot (X - E\{X\})$$

- A função

$$q = (\mathbf{x} - E\{X\})^T \cdot \mathbf{P}_X^{-1} \cdot (\mathbf{x} - E\{X\})$$

é comumente usada para testar a hipótese do vetor $\mathbf{x}$ pertencer a uma distribuição $N(E\{X\}, \mathbf{P}_X)$, considerada verdadeira se

$$q \leq \chi_n^2(1 - r)$$

com $1 - r$ sendo o percentual da distribuição χ_n^2 . Geralmente escolhe-se $1 - r = 0,95$. Consultando a tabela χ_n^2 , montada para valores de $1 - r$, obtém-se [1]

$$(1 - r) = 0,95 \text{ com } n = 1 \implies \chi_1^2(1 - r) = 3,841$$

$$(1 - r) = 0,95 \text{ com } n = 2 \implies \chi_2^2(1 - r) = 5,991$$

$$(1 - r) = 0,95 \text{ com } n = 3 \implies \chi_3^2(1 - r) = 7,815$$

Vetores de variáveis aleatórias

Exemplo 5 (Casamento probabilístico de padrões):

Existe uma classe de problemas de reconhecimento de padrões que se apresentam da seguinte forma [2]. Dispõe-se de K vetores de características $\mathbf{x}_1, \dots, \mathbf{x}_K$, cada qual assumido de distribuição Gaussiana de média $E\{X_k\}$ e matriz de covariâncias $\mathbf{P}_{X_k}$. Deseja-se verificar qual destes vetores corresponde (com maior probabilidade) a um vetor de referência $\mathbf{y}$ suposto $N(E\{Y\}, \mathbf{P}_Y)$. Neste problema, encontrar a boa correspondência significa determinar o par $\{\mathbf{x}_{k^*}, \mathbf{y}\}$ tal que a distância (probabilística) $\mathbf{x}_{k^*}$ e $\mathbf{y}$ seja a menor.

Solução:

Para todo $1 \leq k \leq K$, determinar o vetor de diferença

$$\boldsymbol{\varepsilon}_k = \mathbf{y} - \mathbf{x}_k$$

para o qual $E\{\boldsymbol{\varepsilon}_k\} = E\{Y\} - E\{X_k\}$ e $\mathbf{P}_{\boldsymbol{\varepsilon}_k} = \mathbf{P}_Y + \mathbf{P}_{X_k}$. Pode-se observar que $\sqrt{\boldsymbol{\varepsilon}_k^T \boldsymbol{\varepsilon}_k}$ é a distância euclidiana entre $\mathbf{y}$ e $\mathbf{x}_k$. No entanto, esta métrica de distância

Vetores de variáveis aleatórias

■ Distribuição Gaussiana Bivariada

- Considera-se o caso de uma distribuição Gaussiana com $n = 2$:

$$\begin{aligned} X &= \begin{bmatrix} X_1 & X_2 \end{bmatrix}^T \\ E\{X\} &= \begin{bmatrix} E\{X_1\} & E\{X_2\} \end{bmatrix}^T \\ \mathbf{P}_X &= \begin{bmatrix} \sigma_{X_1}^2 & \sigma_{X_1 X_2} \\ \sigma_{X_1 X_2} & \sigma_{X_2}^2 \end{bmatrix} \end{aligned}$$

em que $\text{cov}\{X_1, X_2\} = \text{cov}\{X_2, X_1\} = \sigma_{X_1 X_2} = \rho \sigma_{X_1} \sigma_{X_2}$. Como $\mathbf{P}_X$ é simétrica, ela pode ser decomposta em

$$\mathbf{P}_X = e_1 \mathbf{v}_1 \mathbf{v}_1^T + e_2 \mathbf{v}_2 \mathbf{v}_2^T$$

com e_1 e e_2 sendo os autovalores de $\mathbf{P}_X$ e $\mathbf{v}_1$ e $\mathbf{v}_2$ sendo seus autovetores.

não considera as incertezas devido às distribuições.

Para todo $1 \leq k \leq K$, determinar a medida

$$q_k = (\boldsymbol{\varepsilon}_k - E\{\boldsymbol{\varepsilon}_k\})^T \cdot \mathbf{P}_{\boldsymbol{\varepsilon}_k}^{-1} \cdot (\boldsymbol{\varepsilon}_k - E\{\boldsymbol{\varepsilon}_k\})$$

que segue χ_n^2 . É comum também assumir que a correspondência é verdadeira, e assim $E\{Y\} = E\{X_k\}$, resultando em

$$q_k = (\mathbf{y} - \mathbf{x}_k)^T \cdot (\mathbf{P}_Y + \mathbf{P}_{X_k})^{-1} \cdot (\mathbf{y} - \mathbf{x}_k)$$

e neste caso, esta medida é chamada de Distância de Mahalanobis.

O par $\{\mathbf{x}_{k^*}, \mathbf{y}\}$ de menor distância estatística é aquele cujo índice k^* satisfaz a

$$k^* = \arg_k \min q_k$$

e a correspondência será aceita se $q_{k^*} \leq \chi_n^2(1 - r)$.

- FDP conjunta: para um dado $\mathbf{x} = [x_1, x_2]^T$

$$p_X(\mathbf{x}) = \frac{1}{2\pi \sigma_{X_1} \sigma_{X_2} \sqrt{1 - \rho^2}} \exp \left\{ -\frac{1}{2} q \right\}$$

com

$$\begin{aligned} q &= \frac{1}{1 - \rho^2} \left\{ \left(\frac{x_1 - E\{X_1\}}{\sigma_{X_1}} \right)^2 \right. \\ &\quad \left. - 2\rho \left(\frac{x_1 - E\{X_1\}}{\sigma_{X_1}} \right) \left(\frac{x_2 - E\{X_2\}}{\sigma_{X_2}} \right) + \left(\frac{x_2 - E\{X_2\}}{\sigma_{X_2}} \right)^2 \right\} \end{aligned}$$

Fazendo $u_1 = (x_1 - E\{X_1\})/\sigma_{X_1}$ e $u_2 = (x_2 - E\{X_2\})/\sigma_{X_2}$, q pode ser re-escrita como

$$q = \mathbf{u}^T \mathbf{P}_u^{-1} \mathbf{u} = \frac{1}{1 - \rho^2} \{u_1^2 - 2u_1 u_2 \rho + u_2^2\}$$

$$\text{em que } \mathbf{u} = [u_1, u_2]^T \text{ e } \mathbf{P}_u = \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}.$$

- Representação gráfica.

No plano $x_1 \times x_2$, uma forma de representar graficamente a distribuição bivariada é por meio da curva de nível correspondentes a uma densidade constante, ou seja, $p_X(\mathbf{x}) = \text{const.}$ Isto implica em

$$q = \frac{1}{1 - \rho^2} \{u_1^2 - 2u_1u_2\rho + u_2^2\} = c \quad (34)$$

com c sendo uma constante cujo significado ficará logo claro.

Dado que $|\rho| \leq 1$, a eq. (34) corresponde a uma elipse com os seguintes parâmetros:

– Centro:

$$E\{X\} = [E\{X_1\}, E\{X_2\}]^T \quad (35)$$

– Ângulo com relação ao eixo x_1 :

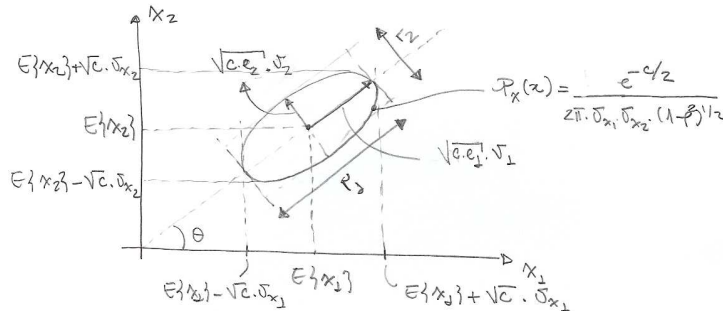
$$\theta = \frac{1}{2} \arctan \left(\frac{2\rho\sigma_{X_1}\sigma_{X_2}}{\sigma_{X_1}^2 - \sigma_{X_2}^2} \right)$$

– Eixos principais : r_1 (menor) e r_2 (maior)

$$r_1^2 = \frac{c\sigma_{X_1}^2\sigma_{X_2}^2(1 - \rho^2)}{\sigma_{X_2}^2 \cos^2(\theta) - 2\rho\sigma_{X_1}\sigma_{X_2} \sin(\theta) \cos(\theta) + \sigma_{X_1}^2 \sin^2(\theta)} \quad (36)$$

$$r_2^2 = \frac{c\sigma_{X_1}^2\sigma_{X_2}^2(1 - \rho^2)}{\sigma_{X_2}^2 \sin^2(\theta) + 2\rho\sigma_{X_1}\sigma_{X_2} \sin(\theta) \cos(\theta) + \sigma_{X_1}^2 \cos^2(\theta)} \quad (37)$$

Assim, no plano $x_1 \times x_2$ tem-se abaixo um exemplo com $\rho > 0$.



Esta elipse é comumente chamada de “Elipse de incerteza $\sqrt{c}$ -sigma”, cujo interior incorpora uma determinada porcentagem p da distribuição de X .

Pode-se demonstrar que

$$p = 1 - \frac{1}{\sqrt{e^c}}. \quad (38)$$

No caso especial de $\sqrt{c} = 3$, tem-se a chamada elipse 3-sigma, que incorpora $p = 98,9981\%$ da distribuição de X .

Verifique você mesmo: se $n = 3$, então $p_X(\mathbf{x}) = \text{const.}$ implica em uma elipsóide.

Exemplo 6 (Elipse 3-sigma de uma distribuição gaussiana bivariada):

Simulação *gaussiana_bivariada* realizada considerando $E\{X_1\} = 5$, $E\{X_2\} = 2$ e diferentes valores de $\mathbf{P}_X$.

Probabilidade condicional de variáveis aleatórias

■ Considerando A e B eventos, a probabilidade condicional é dada por

$$\Pr\{A|B\} = \frac{\Pr\{A \cap B\}}{\Pr\{B\}}$$

Se associarmos estes eventos a variáveis aleatórias X e Y de distribuição contínua, pode-se considerar

$$A = \{X \leq x\}, \quad B = \{y < Y < y + \Delta y\},$$

com $\Delta y > 0$. Então, pode-se mostrar que

$$\begin{aligned} F_{X|Y}(x|y) &= \lim_{\Delta y \rightarrow 0} \Pr\{X \leq x | y < Y < y + \Delta y\} \\ &= \frac{\partial F_{X,Y}(x, y)}{\partial y} \bigg/ \frac{dF_Y(y)}{dy} = \frac{\int_{-\infty}^x p_{X,Y}(u, y) du}{p_Y(y)}. \end{aligned} \quad (39)$$

■ Expectância condicional:

$$E\{X|Y\} \triangleq \int_{-\infty}^{\infty} x p_{X|Y}(x|y) dx \quad (43)$$

Como $E\{X|Y\}$ e $p_{X|Y}(x|y)$ são funções da v.a. Y , elas são portanto v.a.s.

■ Propriedades:

$$p_{X|Y}(x|y) \geq 0 \quad (44)$$

$$F_{X|Y}(\infty|y) = \int_{-\infty}^{\infty} p_{X|Y}(x|y) dx = 1 \quad (45)$$

$$p_X(x) = \int_{-\infty}^{\infty} p_{X,Y}(x, y) dy = \int_{-\infty}^{\infty} p_{X|Y}(x|y) p_Y(y) dy \quad (46)$$

$$= E_Y\{p_{X|Y}(x|y)\} \quad (47)$$

■ FDP condicional

$$\begin{aligned} p_{X|Y}(x|y) &\triangleq \frac{dF_{X|Y}(x|y)}{dx} \\ &= \frac{p_{X,Y}(x, y)}{p_Y(y)} \\ &= \frac{p_{X,Y}(x, y)}{\int_{-\infty}^{\infty} p_{X,Y}(x, y) dy} \end{aligned} \quad (40)$$

No caso discreto:

$$m_{X|Y}(x, y) = \frac{m_{X,Y}(x, y)}{m_Y(y)} \quad (41)$$

■ Regra de Bayes para FDPs de variáveis aleatórias:

$$p_{X|Y}(x|y) = \frac{p_{Y|X}(y|x) p_X(x)}{p_Y(y)} \quad (42)$$

$$E\{X\} = \int_{-\infty}^{\infty} x p_X(x) dx = \int_{-\infty}^{\infty} x \int_{-\infty}^{\infty} p_{X|Y}(x|y) p_Y(y) dy dx \quad (48)$$

$$= \int_{-\infty}^{\infty} \left[\int_{-\infty}^{\infty} x p_{X|Y}(x|y) dx \right] p_Y(y) dy = E_Y\{E\{X|Y\}\}$$

$$E\{g(Y)|Y\} = \int_{-\infty}^{\infty} g(Y) p_{X|Y}(x|y) dx \quad (49)$$

$$= g(Y) \int_{-\infty}^{\infty} p_{X|Y}(x|y) dx = g(Y) \quad (50)$$

com $g(Y)$ sendo uma função escalar, ou seja, $g(Y) : \mathbb{R}^m \rightarrow \mathbb{R}$.

■ Matriz de covariâncias condicional

$$\mathbf{P}_{X|Y} = E\{(X - E\{X\}) \cdot (X - E\{X\})^T | Y\} \quad (51)$$

Probabilidade condicional de variáveis aleatórias

Exercício 2 (Condicional de distribuição exponencial):

Considere X, Y v.a.s cuja distribuição conjunta seja

$$p_{X,Y}(x, y) = \begin{cases} e^{-y} & 0 < x < y < \infty \\ 0 & \text{caso contrário} \end{cases}.$$

Determine: (i) a fdp marginal de X , (ii) a fdp condicional de Y dado $X = x$ e (iii) a esperança e variância de Y dado $X = x$.

Exercício 3 (Distribuição bivariada gaussiana):

Considere X, Y v.a.s Gaussianas cuja distribuição conjunta seja

$$p_{X,Y}(x, y) = \frac{1}{2\pi} \exp \left\{ -\frac{(2x^2 - 6xy + 5y^2)}{2} \right\}.$$

Determine $p_Y(y)$ e $p_{X|Y}(x|y)$. Sugestão: complete os quadrados da função $f(x, y) = (2x^2 - 6xy + 5y^2)$ de forma a obter $f(x, y) = g(x, y) + h(y)$.

Estatística amostral

Estimação da média, matriz de covariâncias e FDP de uma v.a. X a partir de M amostras $\mathbf{x}_1, \dots, \mathbf{x}_M$.

■ Média amostral:

$$\bar{\mathbf{x}} = \frac{1}{M} \sum_{i=1}^M \mathbf{x}_i \quad (52)$$

Esta estimativa é não-polarizada pois $E\{\bar{\mathbf{x}}\} = E\{X\}$ e consistente uma vez que $\mathbf{P}_{\bar{\mathbf{x}}} \rightarrow 0$ com $M \rightarrow \infty$.

■ Covariância amostral:

$$\bar{\mathbf{P}}_X = \frac{1}{M} \sum_{i=1}^M (\mathbf{x}_i - \bar{\mathbf{x}})(\mathbf{x}_i - \bar{\mathbf{x}})^T \quad (53)$$

Exercício 4 (Distribuição bivariada gaussiana):

Considere X, Y vetores de v.a.s Gaussianas dos quais se obtém

$$Z = X + \mathbf{g}(Y).$$

em que $\mathbf{g}(\cdot)$ é uma transformação não-linear. Determine $p_{Z|Y}(\mathbf{z}|\mathbf{y})$, $E\{Z|Y\}$ e $E\{(Z - E\{Z|Y\})(Z - E\{Z|Y\})^T|Y\}$.

Esta estimativa é polarizada pois $E\{\bar{\mathbf{P}}_X\} \neq \mathbf{P}_X$. Pode-se mostrar que

$$\bar{\mathbf{P}}_X = \frac{1}{M-1} \sum_{i=1}^M (\mathbf{x}_i - \bar{\mathbf{x}})(\mathbf{x}_i - \bar{\mathbf{x}})^T \quad (54)$$

é uma estimativa não polarizada (chamada de matriz de covariâncias amostral não-polarizada).

■ FDP a partir de histograma

- Caso escalar: dadas amostras $x_1, \dots, x_M$, obtém-se um histograma, que nada mais é que um gráfico do número de amostras pertencentes a intervalos. Os intervalos possuem espaçamentos Δx iguais entre os limites $x_{\min} = \min(x_1, \dots, x_M)$ e $x_{\max} = \max(x_1, \dots, x_M)$:

$$x_{\min}, x_{\min} + \Delta x, x_{\min} + 2\Delta x, \dots, x_{k-1}, x_k, \dots, x_{\max} - \Delta x, x_{\max}.$$

com o k -ésimo intervalo sendo dado por $[x_{k-1}, x_k]$. Denomina-se n_k o número

de amostras neste intervalo. Assim, tem-se que

$$\begin{aligned}\Pr\{x_{k-1} \leq X < x_k\} &\approx \frac{n_k}{M} \\ &\approx p_X(x_{k-1})\Delta x\end{aligned}$$

Portanto, pode-se obter uma estimativa para a PDF em $p_X(x_{k-1})$ através de

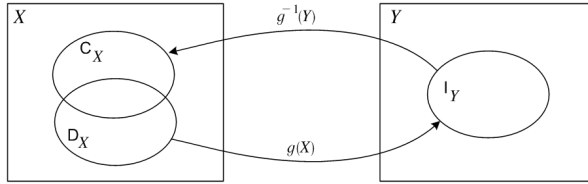
$$p_X(x_{k-1}) = \frac{n_k}{M\Delta x}$$

Deve-se observar que esta é uma forma não paramétrica da PDF $p_X(x)$.

- Caso bivariável: dadas amostras $\mathbf{x}_1, \dots, \mathbf{x}_M$, obtém-se um histograma bidimensional, que é um gráfico do número de amostras pertencentes a intervalos em duas dimensões. Os intervalos possuem espaçamento $\Delta \mathbf{x} = [\Delta x_1, \Delta x_2]^T$ igual entre os limites $x_{d,\min} = \min(x_{d,1}, \dots, x_{d,M})$ e $x_{d,\max} = \max(x_{d,1}, \dots, x_{d,M})$, para $d = 1, 2$ (no caso $N = 2$ é a dimensão da v.a. amostrada). O histograma é portanto um gráfico de contagens de amostras

Funções de variáveis aleatórias

- Conceitos básicos: seja $Y = g(X)$ tal que $g(X) : \mathcal{R}^n \rightarrow \mathcal{R}^m$ é um mapeamento de domínio D_X e imagem I_Y . O problema consiste em determinar a distribuição de Y a partir da distribuição de X .



Este problema resolve-se a partir de

$$\Pr\{I_Y\} = \Pr\{C_X\}$$

em que C_X é o contra-domínio de $g(X)$, determinado a partir de I_Y e $g^{-1}(X)$ (função inversa).

inseridas nos intervalos bidimensionais

$$I_{i,j} = [x_{1,i-1}, x_{1,i}] \times [x_{2,j-1}, x_{2,j}]$$

para os quais i e j os índices do intervalo nas dimensões 1 e 2, respectivamente. Denomina-se $n_{I_{i,j}}$ o número de amostras em cada intervalo $I_{i,j}$. Assim, tem-se que

$$\begin{aligned}\Pr\{\{x_{1,k-1} \leq X_1 < x_{1,k}\}, \{x_{2,k-1} \leq X_2 < x_{2,k}\}\} &\approx \frac{n_{I_{i,j}}}{M} \\ &\approx p_X(\mathbf{x}_{ij} = [x_{1,i-1} \ x_{2,j-1}]^T) \Delta x_1 \Delta x_2\end{aligned}$$

Portanto, pode-se obter uma estimativa para a PDF em $p_X(\mathbf{x}_{ij})$ através de

$$p_X(\mathbf{x}_{ij}) = \frac{n_{I_{i,j}}}{M\Delta x_1\Delta x_2}$$

- Caso multi-variável: extensão para N variáveis a partir do caso bivariável.

Exemplo 7 (Transformação linear):

Considere $Y = aX + b$ com a, b constantes, e X e Y v.a.s unidimensionais. Sendo esta função monotônica, pode-se verificar que

$$F_Y(y) = \begin{cases} F_X\left(\frac{y-b}{a}\right) & , a > 0 \\ 1 - F_X\left(\frac{y-b}{a}\right) & , a < 0 \\ 1(y-b) & , a = 0 \end{cases} \quad p_Y(y) = \begin{cases} p_X\left(\frac{y-b}{a}\right)\frac{1}{|a|} & , a > 0 \\ -p_X\left(\frac{y-b}{a}\right)\frac{1}{|a|} & , a < 0 \\ \delta(y-b) & , a = 0 \end{cases}$$

em que $1(u)$ é a função degrau em $u = 0$, e $\delta(u)$ é a função impulso em $u = 0$. Se $a \neq 0$, $p_Y(y)$ pode ser escrito como

$$p_Y(y) = p_X\left(\frac{y-b}{a}\right)\frac{1}{|a|}$$

que é o mesmo resultado obtido pela eq. (7).

Exemplo 8 (Transformação quadrática):

Considere $Y = X^2$ com X e Y v.a.s unidimensionais. Pode-se verificar que C_X somente existe para $Y \geq 0$. Neste caso,

$$I_Y = \{0 \leq Y \leq y\} \implies C_X = \{-\sqrt{y} \leq X \leq \sqrt{y}\}$$

Portanto,

$$\begin{aligned} F_Y(y) &= \begin{cases} F_X(\sqrt{y}) - F_X(-\sqrt{y}) & , y \geq 0 \\ 0 & , \text{caso contrário.} \end{cases} \\ p_Y(y) &= \begin{cases} \frac{1}{2\sqrt{y}} [p_X(\sqrt{y}) + p_X(-\sqrt{y})] & , y \geq 0 \\ 0 & , \text{caso contrário.} \end{cases} \end{aligned}$$

■ Determinação direta da FDP

- Caso vetorial com $\mathbf{g}(X)$ monotônica, $\mathbf{g}(X) : \mathbb{R}^n \rightarrow \mathbb{R}^n$. (atenção: X e Y são de mesma dimensão)

$$\begin{aligned} p_Y(\mathbf{y}) &= p_X(g^{-1}(\mathbf{y})) |\det(\mathbf{J}^{-1})| \\ &= \frac{p_X(g^{-1}(\mathbf{y}))}{|\det(\mathbf{J})|} \end{aligned} \quad (57)$$

com $\det(\mathbf{J}^{-1}) = 1/\det(\mathbf{J})$ e $\mathbf{J}$ sendo a Jacobiana da transformação, de dimensão $n \times n$, dada por

$$\mathbf{J} = \left[\begin{array}{ccc} \frac{\partial g_1(\mathbf{x})}{\partial x_1} & \dots & \frac{\partial g_1(\mathbf{x})}{\partial x_n} \\ \vdots & & \vdots \\ \frac{\partial g_n(\mathbf{x})}{\partial x_1} & \dots & \frac{\partial g_n(\mathbf{x})}{\partial x_n} \end{array} \right]_{\mathbf{x}=\mathbf{g}^{-1}(\mathbf{y})} \quad (58)$$

■ Determinação direta da FDP

- Caso escalar: $g^{-1}(Y)$ pode ter um ou mais mapeamentos.

$$p_Y(y) = \frac{p_X(x_1)}{|g'(x_1)|} + \dots + \frac{p_X(x_n)}{|g'(x_n)|} \quad (55)$$

$$= \sum_{i=1}^n p_X(x_i) \left| \frac{dg(y)}{dx} \right|_{x=x_i}^{-1} \quad (56)$$

em que $x_1, \dots, x_n$ são as soluções para a variável x em $g^{-1}(y) = x$.

Exercício 5 (Transformação escalar raiz-quadrada):

Considere $Y = X^{1/2}$ com X e Y v.a.s unidimensionais. Sendo $X \sim N(\mu, \sigma_X^2)$ com parâmetros μ e σ^2 tais que $\Pr\{X \leq 0\} \approx 0$. Determinar $p_Y(y)$.

em que as funções $g_i(\mathbf{x})$ são obtidas de $\mathbf{g}(\mathbf{x})$

$$\mathbf{g}(\mathbf{x}) = \begin{bmatrix} g_1(\mathbf{x}) \\ \vdots \\ g_n(\mathbf{x}) \end{bmatrix}. \quad (59)$$

Exemplo 9 (Conversão coordenadas cartesianas em coordenadas polares):

Considere um ponto P de coordenadas cartesianas (X, Y) tal que $X \sim N(x_0, \sigma^2)$ e $Y \sim N(y_0, \sigma^2)$ são independentes. Determinar a fdp conjunta das coordenadas polares de P. Quais as distribuições marginais de tais coordenadas?

Propagação de incertezas

- Definição do problema: determinar a média e a matriz de covariâncias da v.a.

$$Y = \mathbf{g}(X)$$

em que X é uma v.a. de média $E\{X\}$ e matriz de covariâncias $\mathbf{P}_X$. Nesse caso, $\mathbf{g}(X) : \mathbb{R}^n \rightarrow \mathbb{R}^l$ em que não necessariamente $l = n$.

- Solução fechada para $\mathbf{g}(X) = \mathbf{A}X + \mathbf{b}$:

$$\begin{aligned} E\{Y\} &= \mathbf{A}E\{X\} + \mathbf{b} \\ \mathbf{P}_Y &= \mathbf{A}\mathbf{P}_X\mathbf{A}^T. \end{aligned}$$

Observa-se que nada é necessário ser afirmado sobre a distribuição de Y ou de X .

- Aproximação de primeira ordem (linearização):
Se $\mathbf{g}(\mathbf{x})$ for suficientemente suave em torno do suporte da distribuição de X , pode-se mostrar que sendo X gaussiano, Y será aproximadamente gaussiano. Para demonstrar isto, considera-se a seguinte parametrização de $X \sim N(E\{X\}, \mathbf{P}_X)$:

$$X = E\{X\} + \Delta_X$$

com $\Delta_X \sim N(\mathbf{0}, \mathbf{P}_X)$. Por expansão em série de Taylor, truncada em primeira ordem,

$$\begin{aligned} Y &= E\{Y\} + \Delta_Y = \\ &= \mathbf{g}(E\{X\} + \Delta_X) \approx \mathbf{g}(E\{X\}) + \left. \frac{\partial \mathbf{g}(\mathbf{u})}{\partial \mathbf{u}} \right|_{\mathbf{u}=E\{X\}} \Delta_X \end{aligned}$$

Então, Y é aproximadamente obtido por uma transformação afim sobre a v.a.

Propagação de incertezas

- Aproximações para $\mathbf{g}(X)$ sendo uma função não-linear

- Simulação de Monte Carlo:
Gera-se S amostras $\{\mathbf{x}_i\}_{i=1}^S$, das quais obtém-se $\mathbf{y}_i = \mathbf{g}(\mathbf{x}_i)$. Assim, estimativas para a média e matriz de covariâncias de Y são dadas por

$$E\{Y\} \approx \bar{\mathbf{y}} = \frac{1}{S} \sum_{i=1}^S \mathbf{y}_i \quad (60)$$

$$\mathbf{P}_Y \approx \bar{\mathbf{P}}_Y = \frac{1}{S-1} \sum_{i=1}^S (\mathbf{y}_i - \bar{\mathbf{y}})(\mathbf{y}_i - \bar{\mathbf{y}})^T \quad (61)$$

gaussiana Δ_X . Pode-se então obter

$$\begin{aligned} E\{Y\} &\approx \mathbf{g}(E\{X\}) \\ \mathbf{P}_Y &\approx \left(\left. \frac{\partial \mathbf{g}(\mathbf{u})}{\partial \mathbf{u}} \right|_{\mathbf{u}=E\{X\}} \right) \mathbf{P}_X \left(\left. \frac{\partial \mathbf{g}(\mathbf{u})}{\partial \mathbf{u}} \right|_{\mathbf{u}=E\{X\}} \right)^T \end{aligned}$$

- Propagação implícita:
Técnica proposta em [3] para estimação de matriz de covariâncias de parâmetros estimados por minimização de função de custo. A ser detalhado mais à frente no curso.

- *Unscented Transform* (UT):

Forma simplificada da simulação de Monte Carlo em que são geradas amostras consideradas relevantes de X .

Em suma, de acordo com [4], gera-se $S = 2n + 1$ amostras $\{\mathbf{x}_i\}_{i=0}^{S-1}$ (ou também denominados de pontos-sigma) segundo as regras abaixo:

$$\mathbf{x}_i = \begin{cases} E\{X\}, & \text{para } i = 0 \\ E\{X\} + \left(\sqrt{\frac{n}{1-w_0}} \mathbf{P}_X\right)_i & \text{para } i = 1, \dots, n \\ E\{X\} - \left(\sqrt{\frac{n}{1-w_0}} \mathbf{P}_X\right)_{i-n} & \text{para } i = n+1, \dots, 2n \end{cases}$$

em que $(\cdot)_i$ é um operador que corresponde à i -ésima coluna da matriz argumento. Deve ser observado que $\sqrt{\cdot}$ é a raiz quadrada matricial, sendo que em alguns casos usa-se também a decomposição de Cholesky.

Associado a cada ponto-sigma tem-se um peso $w_i = \frac{1-w_0}{2n}$ para $i \neq 0$, e w_0 é um parâmetro de entrada. Supondo-se que a distribuição de X seja Gaussiana, de acordo com [4] recomenda-se $w_0 = \frac{1}{3}$.

A partir dos ponto-sigma obtém-se $\mathbf{y}_i = \mathbf{g}(\mathbf{x}_i)$. As estimativas para a média e matriz de covariâncias de Y são dadas por

$$E\{Y\} \approx \bar{\mathbf{y}} = \sum_{i=0}^{S-1} w_i \mathbf{y}_i \quad (62)$$

$$\mathbf{P}_Y \approx \bar{\mathbf{P}}_Y = \sum_{i=0}^{S-1} w_i (\mathbf{y}_i - \bar{\mathbf{y}})(\mathbf{y}_i - \bar{\mathbf{y}})^T \quad (63)$$

Exemplo 10 (Propagação em um caso de distribuição gaussiana bivariada):

Simulação *propagacao_incertezas* para avaliar a escolha entre raiz quadrada matricial e decomposição de Cholesky na propagação pela UT com diferentes valores de $\mathbf{P}_X$.

Geração de números aleatórios

■ Gerador de números aleatórios: algoritmo que gera amostras

$$x_1, x_2, x_3, \dots$$

de uma v.a. X a partir de p_X ou de F_X .

■ Geradores de distribuições uniformes

- Fórmula genérica de geradores lineares congruentes:

$$x_i = (a \cdot x_{i-1} + b) \bmod (K + 1)$$

que gera números com as seguintes propriedades:

- I. Valores gerados entre 0 e K .
- II. K é um número primo muito grande
- III. x_i é uma sequência periódica, com periodicidade $K_0 \leq K$

IV. x_0 é chamado de semente do gerador

V. a , b e K devem ser bem escolhidos

- Alguns geradores:

- Park & Miller:

$$a = 7^5 \quad b = 0 \quad K = 2^{31} - 2$$

- Gerador com $K_0 \approx 2^{30}$

$$a = 1103515245 \quad b = 12345 \quad K = 2^{32} - 1$$

- Gerador $U(0, 1)$:

$$x_i = \left(\frac{a \cdot x_{i-1} + b}{K + 1} \right) \bmod(1)$$

de onde verifica-se que os geradores funcionam se a sequência discreta

$$x_i = \frac{a}{K + 1} \cdot x_{i-1} + \frac{b}{K + 1}$$

for instável.

■ Geradores de distribuições não-uniformes

- Método da inversa generalizada

Sendo $F_X(x)$ uma função de distribuição, define-se a inversa generalizada como sendo

$$F_X^{-1}(u) = \inf \{x; F_X(x) \geq u\}$$

Portanto, para se obter uma amostra x_i da distribuição F_X , utiliza-se o seguinte algoritmo:

- I. Gera-se $u \sim U(0, 1)$
 - II. Obtem-se uma amostra $x_i = F_X^{-1}(u)$
- Gerador normal usando o teorema do limite central
- Pelo teorema do limite central, a soma

$$Y = \sum_{i=1}^m X_i$$

com os X_i sendo i.i.d., tende a uma distribuição Gaussiana quando $m \rightarrow \infty$. A

média e variância de Y são dados por

$$\begin{aligned} E\{Y\} &= \bar{Y} = \sum_{i=1}^m E\{X_i\} \\ E\{(Y - E\{Y\})^2\} &= \sigma_Y^2 = \sum_{i=1}^m \sigma_{X_i}^2 \end{aligned}$$

Sendo $X_i \sim U(0, 1)$, obtém-se $\bar{Y} = m/2$ e $\sigma_Y^2 = m/12$. Com $m \rightarrow \infty$, pelo teorema do limite central obtém-se

$$Y \sim N\left(\frac{m}{2}, \frac{m}{12}\right)$$

Assim, se for de interesse gerar $Z \sim N(0, 1)$ basta fazer

$$Z = \frac{(Y - m/2)}{\sqrt{m/12}}$$

que se torna fácil de resolver se $m = 12$.

- Método da mistura

Se a FDP de X puder ser escrita sob a forma

$$p_X(x) = \sum_{i=1}^m \omega_i p_i(x)$$

com $p_i(x)$ sendo FDP da qual pode-se gerar amostras e

$$\sum_{i=1}^m \omega_i = 1$$

de forma a satisfazer $F_X(\infty) = 1$. De posse deste modelo, uma amostra de $p_X(x)$ é obtida da distribuição $p_i(x)$ com probabilidade ω_i . Ou seja, segue-se o seguinte algoritmo

- I. Gera-se $u \sim U(0, 1)$
- II. Determina-se o índice j que satisfaz a

$$s_{i-1} \leq u \leq s_i$$

com

$$s_i = \sum_{j=1}^i \omega_j$$

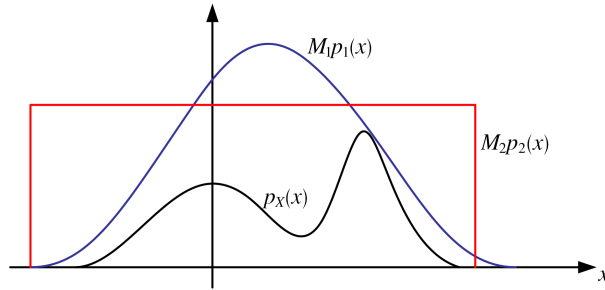
III. Como resultado, obtém-se uma amostra de $p_i(x)$

Não seria a mesma coisa se fossem gerados $x_i \sim p_i(x)$ e a amostra final dada por $\sum_{i=1}^m \omega_i x_i$?

- Método da amostragem por rejeição (rejection sampling)
Considera-se o caso de $p_X(\cdot)$ ser uma fdp da qual se deseja obter amostras. Seja $p_A(\cdot)$ uma fdp auxiliar da qual obtém-se amostras mais facilmente, e uma constante M tal que

$$p_X(x) \leq M p_A(x)$$

Por exemplo, sendo p_1 gaussiana e p_2 uniforme:



Referências

- [1] MEYER, P. *Probabilidade - aplicações à estatística*. Segunda edição. [S.l.]: Livros Técnicos e Científicos (LTC), 2003.
- [2] DUDA, R. O.; HART, P. E.; STORK, D. G. *Pattern Classification*. 2nd. ed. [S.l.]: Wiley Interscience, 2000.
- [3] HARALICK, R. Propagating covariances in computer vision. In: *International Conference on Pattern Recognition*. [S.l.: s.n.], 1994. p. 493–498.
- [4] JULIER, S. J.; UHLMANN, J. Unscented filtering and nonlinear estimation. *Proceedings of the IEEE*, v. 92, n. 3, p. 401–422, 2004.

Observa-se que as áreas sobre as curvas se relacionam por

$$\frac{\text{Área}\{M_1 p_1(x)\}}{\text{Área}\{p_X(x)\}} = M_1$$

Algoritmo de amostragem:

- I. Gera-se $x \sim p_A$ e $u \sim U(0, 1)$
- II. Se

$$\frac{p_X(x)}{M p_A(x)} \geq u$$

então x deve ser aceito como amostra de $p_X(\cdot)$. Caso contrário, retornar ao passo 1.

Exemplo 11 (Simulação de números aleatórios):

Ver simulação *geradores_aleatorios*.