ECE 2521: Analysis of Stochastic Processes

Lecture 7

Department of Electrical and Computer Engineering
University of Pittsburgh

October, 27th 2021

Azime Can-Cimino

One Function of Two Random Variables

Let $W = g(X, Y)$ be a function of RVs X and Y

- *Discrete Random Variables*
  - If X and Y are discrete RVs, then W will also be a discrete random variable characterized by a PMF $p_W(w)$
  - The PMF $p_W(w)$ can be obtained by adding the values of $p_{X,Y}(x, y)$ corresponding to x and y pairs for which $g(x, y) = w$: $p_W(w) = \sum_{\{(x,y)|g(x,y)=w\}} p_{X,Y}(x, y)$
- *Continuous Random Variables*
  - If X and Y are continuous RVs and $g(X, Y)$ is a continuous function, then $W = g(X, Y)$ is also a continuous RV
  - To find the PDF $f_W(w)$ of W first find CDF $F_W(w)$ and then take its derivative:
    $F_W(w) = \text{Prob}(W \leq w) = \iint_{g(x,y)\leq w} f_{X,Y}(x, y)dxdy$

Example

- Let X and Y be any continuous random variables

(1) Determine the PDF of $Z = X + Y$

(2) What if X and Y are independent?

(3) Consider the case when X and Y are independent and uniformly distributed random variables:

$$f_X(x) = u(x) - u(x-1)$$

$$f_Y(y) = 0.5u(y) - 0.5u(y-2)$$

Calculate and plot the PDF of $Z = X + Y$.

Two Functions of Two Random Variables

- Let $g(X,Y)$ and $h(X,Y)$ be continuous and differentiable functions such that:
$$g(X,Y) = Z \quad \text{and} \quad h(X,Y) = W. \tag{1}$$
- For a given (z,w), (1) may have many solutions. Let $(x_1,y_1),\ \ldots,\ (x_n,y_n)$ represent these multiple solutions, such that $g(x_i,y_i) = z$ and $h(x_i,y_i) = w$ for $i = 1,\ \ldots,\ n$. Then:
$$f_{ZW}(z,w) = \sum_{i=1}^{n} \frac{1}{|J(x_i,y_i)|} f_{XY}(x_i,y_i)$$
where $x_i = g_I(z,w)$ and $y_i = h_I(z,w)$, and $|J(x_i,y_i)|$ is the determinant of the Jacobian of the transform given in (1) such that:
$$|J(x_i,y_i)| = \begin{vmatrix} \frac{\partial g}{\partial x} & \frac{\partial g}{\partial y} \\ \frac{\partial h}{\partial x} & \frac{\partial h}{\partial y} \end{vmatrix}$$

Example

- Let $Z = aX + bY$ and $W = cX + dY$ are two functions of random variables X and Y.
- The joint pdf of X and Y is given by $f_{XY}(x, y)$.
- Find the joint pdf of Z and W, $f_{ZW}(z, w)$

Bivariate Gaussian Random Variables

- Let X and Y be two Gaussian random variables with correlation coefficient $\rho_{XY} = \rho$, where $-1 \le \rho \le 1$
- Their joint probability density function (PDF) is completely characterized by the mean μ_X and standard deviation σ_X of random variable X, mean σ_Y and standard deviation σ_Y of random variable Y, and their correlation coefficient ρ:

$$f_{X,Y}(x,y) = \frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho^2}} \exp\left[\frac{\frac{(x-\mu_X)^2}{\sigma_X^2} - \frac{2\rho(x-\mu_X)(y-\mu_Y)}{\sigma_X\sigma_Y} + \frac{(y-\mu_Y)^2}{\sigma_Y^2}}{-2(1-\rho^2)}\right]$$

- If X and Y uncorrelated $\rho = 0$, their joint PDF becomes:

$$f_{X,Y}(x,y) = \frac{1}{2\pi\sigma_X\sigma_Y} \exp\left[-\frac{(x-\mu_X)^2}{2\sigma_X^2} - \frac{(y-\mu_Y)^2}{2\sigma_Y^2}\right] = f_X(x)f_Y(y)$$

- The above implies that uncorrelated Gaussian random variables are also independent.

Conditional Gaussian PDF

- If X and Y are bivariate Gaussian random variables, the conditional PDF of X given $Y = y$ is:

$$f_{X|Y=y}(x) = \frac{1}{\sigma_X\sqrt{2\pi(1-\rho^2)}}\exp\left[\frac{\left(x-\mu_X-\rho\frac{\sigma_X}{\sigma_Y}(y-\mu_Y)\right)^2}{2\sigma_X^2(1-\rho^2)}\right]$$

- The conditional mean of random variable X given $Y = y$ is:

$$E[X|Y=y] = \mu_X + \rho\frac{\sigma_X}{\sigma_Y}(y-\mu_Y)$$

- The corresponding conditional variance of X is:

$$\text{Var}(X|Y=y) = \sigma_X^2(1-\rho^2).$$

Exercise 1

Rectangular to Polar coordinate transformation

- $X, Y \sim \mathcal{N}[0, 1]$ are independent jointly Gaussian random variables
- $R = \sqrt{X^2 + Y^2}$ such that $r = g(x, y) = \sqrt{x^2 + y^2}$
- $\Phi = tan^{-1}\left(\frac{Y}{X}\right)$ such that $\phi = h(x, y) = tan^{-1}\left(\frac{y}{x}\right)$
- Find PDFs of R and Φ.

Exercise 2

- Let $Z = max(X, Y)$ and $W = min(X, Y)$.
- Determine the PDFs $f_Z(z)$ and $f_W(w)$:

$$
\begin{aligned}
z &= max(x, y) = \begin{cases} x & \text{if } x > y \\ y & \text{if } x \leq y \end{cases} \\
w &= min(x, y) = \begin{cases} y & \text{if } x > y \\ x & \text{if } x \leq y \end{cases}
\end{aligned}
$$

Probability Models of Multiple Random Variables

- In Chapter 6 we introduce the probability measures for multiple random variables
- A *vector random variable* X is a function that assigns a vector of real numbers to each outcome ξ in S, the sample space of the random experiment:

$$\mathrm{X} = \left[\begin{array}{ccc} X_1 & \dots & X_n \end{array} \right]^T : S \to \mathbb{R}^n$$

- The probability models of n random variables are the generalization of the probability models of two random variables.

Probability Models of Multiple Random Variables

- A **random vector** is a column vector $\mathrm{X} = \begin{bmatrix} X_1 & \ldots & X_n \end{bmatrix}^T$, where each X_i is a random variable: when $n = 1$ a random vector reduces to a random variable
- A **sample value of a random vector** is a column vector $\mathrm{x} = \begin{bmatrix} x_1 & \ldots & x_n \end{bmatrix}^T$, where each x_i is a sample value of the random variable X_i
- **Random vector probability functions**:
  (a) The CDF of a random vector X is
  $$F_{\mathrm{X}}(\mathrm{x}) = F_{X_1,\ldots,X_n}(x_1, \ldots, x_n)$$
  (b) The PMF of a discrete random vector X is
  $$p_{\mathrm{X}}(\mathrm{x}) = p_{X_1,\ldots,X_n}(x_1, \ldots, x_n)$$
  (c) The PDF of a continuous random vector X is
  $$f_{\mathrm{X}}(\mathrm{x}) = f_{X_1,\ldots,X_n}(x_1, \ldots, x_n)$$

Multivariate Joint CDF

- The joint CDF of random variables $X_1, \ldots, X_n$ is

$$F_{\mathrm{X}}(x_1, \ldots, x_n) = P(X_1 \leq x_1, \ldots, X_n \leq x_n)$$

- The joint CDF is defined for discrete, continuous, and mixed type random variables
- **Properties**
  (1) $0 \leq F_{\mathrm{X}}(\mathrm{x}) \leq 1$.
  (2) $F_{\mathrm{X}}(x_1, \ldots, x_n)$ is nondecreasing on all x_i for $i = 1, \ldots, n$.
  (3) $\lim_{x_1 \to -\infty, \ldots, x_n \to -\infty} F_{\mathrm{X}}(x_1, \ldots, x_n) = 0$.
  (4) $\lim_{x_1 \to \infty, \ldots, x_n \to \infty} F_{\mathrm{X}}(x_1, \ldots, x_n) = 1$.
  (5) Joint CDF for $X_1, \ldots, X_{n-1}$ is given by $F_{X_1, \ldots, X_n}(x_1, \ldots, x_{n-1}, \infty)$.

Multivariate Joint PMF

- The joint PMF of discrete random variables $X_1, \ldots, X_n$:

$$p_{\mathrm{X}}(\mathrm{x}) = p_{X_1,\ldots,X_n}(x_1, \ldots, x_n) = \text{Prob}\left[X_1 = x_1, \ldots, X_n = x_n\right]$$

- Satisfies the axioms of probability:
  (a) Non-negativity: $p_{X_1,\ldots,X_n}(x_1, \ldots, x_n) \geq 0$
  (b) Normalization: $\sum_{x_1} \ldots \sum_{x_n} p_{X_1,\ldots,X_n}(x_1, \ldots, x_n) = 1$

- Probability of an event A is given by:

$$\mathrm{P}\left[A\right] = \sum_{(x_1,\ldots,x_n)\in A} \ldots \sum p_{X_1,\ldots,X_n}(x_1, \ldots, x_n) \qquad X_1, \ldots, X_n \text{ discrete}$$

Multivariate Joint PMF

- Marginal PMFs:

$$\begin{aligned} p_{X_1,\ldots,X_{n-1}}(x_1,\ldots,x_{n-1}) &= \sum_{x_n} p_{X_1,\ldots,X_n}(x_1,\ldots,x_n) \\ p_{X_1}(x_1) &= \sum_{x_2}\cdots\sum_{x_n} p_{X_1,\ldots,X_n}(x_1,\ldots,x_n) \end{aligned}$$

- Conditional PMFs:

$$p_{X_n}(x_n|x_1,\ \ldots,\ x_{n-1}) = \frac{p_{X_1,\ldots,X_n}(x_1,\ldots,x_n)}{p_{X_1,\ldots,X_{n-1}}(x_1,\ldots,x_{n-1})}$$

- Recursively, we can obtain:

$$p_{X_1,\ldots,X_n}(x_1,\ldots,x_n) = p_{X_n}(x_n|x_1,\ \ldots,\ x_{n-1}) \\ p_{X_{n-1}}(x_{n-1}|x_1,\ \ldots,\ x_{n-2})\cdots p_{X_2}(x_2|x_1)p_{X_1}(x_1)$$

Multivariate Joint PDF

- The joint PDF of continuous random variables $X_1, \ldots, X_n$ is denoted by $f_{\mathrm{X}}(\mathrm{x}) = f_{X_1,\ldots,X_n}(x_1, \ldots, x_n)$, where:

$$\text{Prob}\left[a_1 \leq X_1 \leq b_1, \ldots, a_n \leq X_n \leq b_n\right] =$$

$$\int_{a_n}^{b_n} \cdots \int_{a_1}^{b_1} f_{X_1,\ldots,X_n}(x_1, \ldots, x_n) dx_1 \ldots dx_n$$

- Satisfies the axioms of probability:
  (a) Non-negativity: $f_{X_1,\ldots,X_n}(x_1, \ldots, x_n) \geq 0$
  (b) Normalization: $\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f_{X_1,\ldots,X_n}(x_1, \ldots, x_n) dx_1 \ldots dx_n = 1$
- Probability of an event A is given by:

$$\mathrm{P}\left[A\right] = \underset{(x_1,\ldots,x_n)\in A}{\int \cdots \int} f_{X_1,\ldots,X_n}(x_1, \ldots, x_n) dx_1 \ldots dx_n \qquad X_1, \ldots, X_n \text{ continuous}$$

Multivariate Joint PDF

- Marginal PDFs:

$$f_{X_1,\ldots,X_{n-1}}(x_1,\ldots,x_{n-1}) = \int_{-\infty}^{\infty} f_{X_1,\ldots,X_n}(x_1,\ldots,x_n)\ dx_n$$

$$f_{X_1}(x_1) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f_{X_1,\ldots,X_n}(x_1,\ldots,x_n)\ dx_2 \ldots dx_n$$

- Conditional PDFs

$$f_{X_n}(x_n|x_1,\ \ldots,\ x_{n-1}) = \frac{f_{X_1,\ldots,X_n}(x_1,\ldots,x_n)}{f_{X_1,\ldots,X_{n-1}}(x_1,\ldots,x_{n-1})}$$

Multivariate Joint PDF

- Then recursively, we can obtain:

$$f_{X_1,\ldots,X_n}(x_1,\ldots,x_n) = f_{X_n}(x_n|x_1,\ \ldots,\ x_{n-1}) \cdot f_{X_{n-1}}(x_{n-1}|x_1,\ \ldots,\ x_{n-2}) \cdots f_{X_2}(x_2|x_1) f_{X_1}(x_1)$$

Note:

$$F_{X_1,\ldots,X_n}(x_1,\ldots,x_n) = \int_{-\infty}^{x_n} \cdots \int_{-\infty}^{x_1} f_{X_1,\ldots,X_n}(x_1,\ldots,x_n) dx_1 \ldots dx_n$$

- Therefore:

$$f_{X_1,\ldots,X_n}(x_1,\ldots,x_n) = \frac{\partial^n}{\partial x_1 \ldots \partial x_n} F_{X_1,\ldots,X_n}(x_1,\ldots,x_n)$$

Example

- Random variables $X_1, \ldots, X_n$ have joint PDF:

$$f_{X_1,\ldots,X_n}(x_1, \ldots, x_n) = \begin{cases} 1 & 0 \le x_i \le 1, \; i = 1, \ldots, n \\ 0 & \text{otherwise} \end{cases}$$

- Let A denote the event that $\max_i X_i \le \frac{1}{2}$
- Find $P[A]$.

Example

- Random variables $X_1, \ldots, X_n$ have joint PDF:

$$f_{X_1,\ldots,X_n}(x_1, \ldots, x_n) = \begin{cases} 1 & 0 \le x_i \le 1, \; i = 1, \ldots, n \\ 0 & \text{otherwise} \end{cases}$$

- Let A denote the event that $\max_i X_i \le \frac{1}{2}$
- Find $P[A]$.

Solution The maximum of n numbers is less than $\frac{1}{2}$ if and only if each of the n numbers is less than $\frac{1}{2}$; therefore

$$\begin{aligned} P[A] &= P\left[\max_i X_i \le \frac{1}{2}\right] = P\left[X_1 \le \frac{1}{2}, \ldots, X_n \le \frac{1}{2}\right] \\ &= \int_0^{\frac{1}{2}} \cdots \int_0^{\frac{1}{2}} 1 \; dx_1 \ldots dx_n = \frac{1}{2^n} \end{aligned}$$

Independence

- $X_1, \ldots, X_n$ are **independent** if for all $x_1, \ldots, x_n$:

$$\begin{aligned} p_{X_1,\ldots,X_n}(x_1, \ldots, x_n) &= p_{X_1}(x_1)p_{X_2}(x_2)\ldots p_{X_n}(x_n) \\ f_{X_1,\ldots,X_n}(x_1, \ldots, x_n) &= f_{X_1}(x_1)f_{X_2}(x_2)\ldots f_{X_n}(x_n) \end{aligned}$$

- $X_1, \ldots, X_n$ are **Independent Identically Distributed (i.i.d)** if for all $x_1, \ldots, x_n$:

$$\begin{aligned} p_{X_1,\ldots,X_n}(x_1, \ldots, x_n) &= p_X(x_1)p_X(x_2)\ldots p_X(x_n) \\ f_{X_1,\ldots,X_n}(x_1, \ldots, x_n) &= f_X(x_1)f_X(x_2)\ldots f_X(x_n) \end{aligned}$$

Example

- The random variables X_1,X_2 and X_3 have the joint Gaussian PDF:

$$f_{X_1,X_2,X_3}(x_1,x_2,x_3) = \frac{e^{-(x_1^2+x_2^2-\sqrt{2}x_1x_2+\frac{1}{2}x_3^2)}}{2\pi\sqrt{\pi}}$$

- Find the marginal PDFs $f_{X_1,X_3}(x_1,x_3)$, $f_{X_1}(x_1)$ and $f_{X_3}(x_3)$.

Example - Solution

- The marginal PDF for the pair X_1 and X_3 is found by integrating the joint PDF over X_2:

$$f_{X_1,X_3}(x_1,x_3) = \frac{e^{-x_3^2/2}}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \frac{e^{-(x_2^2-\sqrt{2}x_1x_2+\frac{x_1^2}{2}+\frac{x_1^2}{2})}}{\pi\sqrt{2}} dx_2 = \frac{e^{-x_3^2/2}}{\sqrt{2\pi}} \frac{e^{-x_1^2/2}}{\sqrt{2\pi}}$$

- Marginal PDF for X_1 is found by integrating $f_{X_1,X_3}(x_1,x_3)$ over X_3:

$$f_{X_1}(x_1) = \frac{e^{-x_1^2/2}}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \frac{e^{-x_3^2/2}}{\sqrt{2\pi}} dx_3 = \frac{e^{-x_1^2/2}}{\sqrt{2\pi}}$$

- Marginal PDF for X_3 is found by integrating $f_{X_1,X_3}(x_1,x_3)$ over X_1:

$$f_{X_3}(x_3) = \frac{e^{-x_3^2/2}}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \frac{e^{-x_1^2/2}}{\sqrt{2\pi}} dx_1 = \frac{e^{-x_3^2/2}}{\sqrt{2\pi}}$$

Note: $f_{X_1,X_3}(x_1,x_3) = f_{X_1}(x_1) f_{X_3}(x_3)$, therefore X_1 and X_3 are independent zero-mean, unit variance Gaussian random variables.

Functions of Random Vectors

- Let $\mathrm{X} = \begin{bmatrix} X_1 & \ldots & X_n \end{bmatrix}^T$ and $Y = g(\mathrm{X})$; that is $g : \mathbb{R}^n \to \mathbb{R}$ with $\mathrm{X} \in \mathbb{R}^n$ and $Y \in \mathbb{R}$. Then:

$$F_Y(y) = Prob(g(\mathrm{X}) \leq y) = Prob(X \in R_Y)$$

where $R_Y = \{\mathrm{x} : g(\mathrm{x}) \leq y\}$.

Transformations of Random Vectors

- Consider the random vector: $\mathrm{X} = \begin{bmatrix} X_1 & \dots & X_n \end{bmatrix}^T$
- Let $\mathrm{Y} = \mathrm{g}(\mathrm{X}) = [g_1(\mathrm{X}) \cdots g_n(\mathrm{X})]^T$ such that $\mathrm{g} : \mathbb{R}^n \rightarrow \mathbb{R}^n$ with $\mathrm{X} \in \mathbb{R}^n$ and $\mathrm{Y} \in \mathbb{R}^n$
- If $\mathrm{X} = \begin{bmatrix} X_1 & \dots & X_n \end{bmatrix}^T = \mathrm{g}^{-1}(\mathrm{Y}) = [g_1^{-1}(\mathrm{Y}) \cdots g_n^{1}(\mathrm{Y})]^T$, we can compute $f_{\mathrm{Y}}(\mathrm{y})$ as:

$$f_{\mathrm{Y}}(\mathrm{y}) = \frac{f_{\mathrm{X}}(\mathrm{g}^{-1}(\mathrm{y}))}{|J(x_1, \dots, x_n)|}$$

where $|J(x_1, \dots, x_n)|$ is the determinant of the Jacobian:

$$J(x_1, \dots, x_n) = \begin{bmatrix} \frac{\partial g_1(\mathrm{x})}{\partial x_1} & \cdots & \frac{\partial g_1(\mathrm{x})}{\partial x_n} \\ \vdots & & \vdots \\ \frac{\partial g_n(\mathrm{x})}{\partial x_1} & \cdots & \frac{\partial g_n(\mathrm{x})}{\partial x_n} \end{bmatrix}$$

Special Case (Linear Transformation)

- Let $\mathrm{X} = \begin{bmatrix} X_1 & \ldots & X_n \end{bmatrix}^T$ and $\mathrm{Y} = \mathrm{g}(\mathrm{X}) = \boldsymbol{A}\mathrm{X} + \mathrm{b}$ where $\boldsymbol{A}$ is an invertible $n \times n$ matrix and b is an $n \times 1$ vector
- Then $\mathrm{X} = \boldsymbol{A}^{-1}(\mathrm{Y} - \mathrm{b})$ and:

$$f_{\mathrm{Y}}(\mathrm{y}) = \frac{f_{\mathrm{X}}(\boldsymbol{A}^{-1}(\mathrm{Y} - \mathrm{b}))}{|\boldsymbol{A}|}$$

Expected Values of Random Vectors

- Let $\mathrm{X} = \begin{bmatrix} X_1 & \dots & X_n \end{bmatrix}^T$ and $Y = g(\mathrm{X}) = g(X_1, \dots, X_n)$, then the expected value of Y is:

$$E[Y] = \begin{cases} \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} g(\mathrm{X}) f_{\mathrm{X}}(x_1, \dots, x_n) \mathrm{d}x_1 \cdots \mathrm{d}x_n & \mathrm{X} \text{ is jointly continuous} \\ \sum_{x_1} \cdots \sum_{x_n} g(\mathrm{X}) p_{\mathrm{X}}(x_1, \dots, x_n) \mathrm{d}x_1 \cdots \mathrm{d}x_n & \mathrm{X} \text{ is jointly discrete} \end{cases}$$

Mean Vector

- Let $\mathrm{X} = \begin{bmatrix} X_1 & \dots & X_n \end{bmatrix}^T$, then the expected value of X - also the mean vector μ_{X} - is defined as:

$$\mu_{\mathrm{X}} = E[\mathrm{X}] = \begin{bmatrix} E[X_1] \\ \vdots \\ E[X_n] \end{bmatrix} = \begin{bmatrix} \mu_1 \\ \vdots \\ \mu_n \end{bmatrix}$$

- In general if $\mathrm{Y} = \mathrm{g}(\mathrm{X}) = [g_1(\mathrm{X}) \cdots g_n(\mathrm{X})]^T$, then the expected value of Yis computed as:

$$E[\mathrm{g}(\mathrm{X})] = \begin{bmatrix} E[g_1(\mathrm{X})] \\ \vdots \\ E[g_n(\mathrm{X})] \end{bmatrix}$$

Covariance Matrix and Correlation Matrix

- The correlation matrix $\boldsymbol{R}_{\mathrm{X}} = E[\mathrm{X}\mathrm{X}^T]$:

$$\boldsymbol{R}_{\mathrm{X}} = \begin{bmatrix} E[X_1^2] & \cdots & E[X_1 X_n] \\ \vdots & & \vdots \\ E[X_n X_1] & \cdots & E[X_n^2] \end{bmatrix}$$

- The covariance matrix $\boldsymbol{K}_{\mathrm{X}} = E[(\mathrm{X} - \mu_{\mathrm{X}})(\mathrm{X} - \mu_{\mathrm{X}})^T]$:

$$\begin{aligned} \boldsymbol{K}_{\mathrm{X}} &= \begin{bmatrix} E[(X_1 - \mu_1)^2 & E[(X_1 - \mu_1)(X_2 - \mu_2) & \cdots & E[(X_1 - \mu_1)(X_n - \mu_n)] \\ E[(X_2 - \mu_2)(X_1 - \mu_1) & E[(X_2 - \mu_2)^2 & \cdots & E[(X_2 - \mu_2)(X_n - \mu_n)] \\ \vdots & \vdots & & \vdots \\ E[(X_n - \mu_n)(X_1 - \mu_1)] & E[(X_n - \mu_n)(X_2 - \mu_2)] & \cdots & E[(X_n - \mu_n)^2] \end{bmatrix} \\ &= \begin{bmatrix} \mathrm{Var}(X_1) & \mathrm{Cov}(X_1, X_2) & \cdots & \mathrm{Cov}(X_1, X_n)] \\ \vdots & \vdots & & \vdots \\ \mathrm{Cov}(X_n, X_1) & \mathrm{Cov}(X_n, X_2) & \cdots & \mathrm{Var}(X_n) \end{bmatrix} \end{aligned}$$

Note: $\boldsymbol{K}_{\mathrm{X}} = \boldsymbol{R}_{\mathrm{X}} - \boldsymbol{\mu}_{\mathrm{X}}\boldsymbol{\mu}_{\mathrm{X}}^T$.

Theorem

- For a linear transformation of a vector of random variables of the form $\mathrm{Y} = \boldsymbol{A}\mathrm{X} + \mathrm{b}$, the means of X and **Y** are related by:

$$\mu_{\mathrm{Y}} = \boldsymbol{A}\mu_{\mathrm{X}} + \mathrm{b}$$

- Also, the covariance matrices of X and Y are related by:

$$\boldsymbol{K}_{\mathrm{Y}} = \boldsymbol{A}\boldsymbol{K}_{\mathrm{X}}\boldsymbol{A}^T.$$

Remarks

- Both $\boldsymbol{R}_{\mathrm{X}}$ and $\boldsymbol{K}_{\mathrm{X}}$ are symmetric nonnegative definite $n \times n$ matrices.
- Recall from linear algebra that, if $\mathrm{u_i}$ for $i = 1, \ldots, n$ are eigenvectors with the corresponding eigenvalues λ_i with $\lambda_i \geq 0$ such that $\boldsymbol{K}_{\mathrm{X}}\mathrm{u}_i = \lambda_i \mathrm{u}_i$ and u_i's are orthogonal, then:

$$\boldsymbol{K}_{\mathrm{X}} = \boldsymbol{U}\Lambda\boldsymbol{U}^T$$

  where $\boldsymbol{U} = [\mathrm{u}_1 \cdots \mathrm{u}_n]$ is an orthogonal matrix with i^{th} eigenvector as the i^{th} column ($\boldsymbol{U}\boldsymbol{U}^T = \boldsymbol{I}$), and Λ is a diagonal matrix with i^{th} diagonal element as the i^{th} eigenvalue λ_i.
- Given $\mathrm{Y} = \boldsymbol{A}\mathrm{X}$, we can choose $\boldsymbol{A}$ such that Y has uncorrelated components: $\boldsymbol{A} = (\boldsymbol{U}\sqrt{\Lambda})^{-1}$ yields $\boldsymbol{K}_{\mathrm{Y}} = \boldsymbol{I}$.

Joint Moment Generating Functions of Random Vectors

- Let $\mathrm{X} = \begin{bmatrix} X_1 & \ldots & X_n \end{bmatrix}^T$, then the joint moment generating function of X is defined as:

$$\Phi_{\mathrm{X}}(\mathrm{s}) = \Phi_{X_1,\ldots,X_n}(s_1,\ldots,s_n) = E[e^{\mathrm{s}^T\mathrm{X}}] = E[e^{s_1X_1+\ldots+s_nX_n}]$$

where $\mathrm{s} = [s_1 \ldots s_n]^T$.
- The joint PDF can be obtained using the MGF of X:

$$f_{X_1,\ldots,X_n}(x_1,\ldots,x_n) = \frac{1}{(2\pi)^n}\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}\Phi_{\mathrm{X}}(\mathrm{s})e^{s_1X_1+,\ldots,s_nX_n}ds_1\cdots ds_n$$

- Recall that for $\mathrm{s} = j\omega = [j\omega_1 \ldots j\omega_n]$ we can compute the joint characteristic function of X
- If $X_1,\ldots,X_n$ are all independent, then:

$$\Phi_{\mathrm{X}}(\mathrm{s}) = \Phi_{X_1}(s_1)\cdots\Phi_{X_n}(s_n) = \prod_{i=1}^{n}\Phi_{X_i}(s_i)$$

Multivariate Gaussian Random Variables

- If a random vector $\mathrm{X} = \begin{bmatrix} X_1 & \dots & X_n \end{bmatrix}^T \in \mathbb{R}^n$ is said to follow a multivariate Gaussian distribution with mean μ_{X} and covariance $\boldsymbol{K}_{\mathrm{X}}$ (where $\boldsymbol{K}_{\mathrm{X}}$ is invertible), then

$$f_{\mathrm{X}}(\mathrm{X}) = (2\pi)^{-\frac{n}{2}} (\det \boldsymbol{K}_{\mathrm{X}})^{-\frac{1}{2}} \exp\left[-\frac{(\mathrm{X}-\mu_{\mathrm{X}})^T \boldsymbol{K}_{\mathrm{X}}^{-1} (\mathrm{X}-\mu_{\mathrm{X}})}{2}\right]$$

Properties

(1) Uncorrelated Gaussian random variables are independent. That is, if X and Y are jointly Gaussian and $E[(X-\mu_X)(Y-\mu_Y)]=0$, then X and Y are independent.

(2) If $\mathrm{X} \in \mathbb{R}^n$ follows a multivariate Gaussian distribution, then $\mathrm{Y} = \boldsymbol{A}\mathrm{X} + \mathrm{b}$ with $\boldsymbol{A}$ as an $n \times n$ matrix and b as an $n \times 1$ vector also follows a multivariate Gaussian distribution. That is $\mathrm{Y} \sim \mathcal{N}(\boldsymbol{A}\mathrm{X} + \mathrm{b}, \boldsymbol{A}^T \boldsymbol{K}_{\mathrm{X}} \boldsymbol{A})$

(3) All the marginal distributions are also Gaussian. That is, X_i for $i = 1, \ldots, n$ also follows a Gaussian distribution. That is $X_i \sim \mathcal{N}(\mu_i, \mathrm{Var}(X_i))$.

Properties

(4) If we denote $\mathrm{X} = \begin{bmatrix} \mathrm{X}_1 \\ \mathrm{X}_2 \end{bmatrix}$ such that $\mu_{\mathrm{X}} = \begin{bmatrix} \mu_{\mathrm{X}_1} \\ \mu_{\mathrm{X}_2} \end{bmatrix}$ and $\boldsymbol{K}_{\mathrm{X}} = \begin{bmatrix} \boldsymbol{K}_{\mathrm{X}_1} & \boldsymbol{K}_{\mathrm{X}_1,\mathrm{X}_2} \\ \boldsymbol{K}_{\mathrm{X}_2,\mathrm{X}_1} & \boldsymbol{K}_{\mathrm{X}_2} \end{bmatrix}$, then the conditional random variable $\mathrm{X}_1|\mathrm{X}_2$ also follows a Gaussian distribution such that $\mathrm{X}_1|\mathrm{X}_2 \sim$ $\mathcal{N}\left(\mu_{\mathrm{X}_1} + \boldsymbol{K}_{\mathrm{X}_1,\mathrm{X}_2}\boldsymbol{K}_{\mathrm{X}_2}^{-1}(\mathrm{x}_2 - \mu_{\mathrm{X}_2}), \boldsymbol{K}_{\mathrm{X}_1} - \boldsymbol{K}_{\mathrm{X}_1,\mathrm{X}_2}\boldsymbol{K}_{\mathrm{X}_2}^{-1}\boldsymbol{K}_{\mathrm{X}_2,\mathrm{X}_1}\right)$.

(5) The joint MGF of X: $\Phi_{\mathrm{X}} = \exp\left(\mathrm{s}^T\mu_{\mathrm{X}} + \frac{1}{2}\mathrm{s}^T\boldsymbol{K}_{\mathrm{X}}\mathrm{s}\right)$.

Estimation versus Detection

- Main difference between estimation and detection problems involves how we measure success:

Detection: We might ask how often our guess is correct

Estimation: Common to measure an error between the true value and the estimated value.

- In detection problems, we are interesting in estimating a quantity that is discrete in nature:

Example 1: Radar systems: we are trying to decide whether or not a target is present based on observing radar returns

Example 2: Digital communication systems: we are trying to determine whether bits take on values of 0 or 1 based on samples of some receive signal

Maximum A-Posteriori (MAP) Estimator

- Assume X and Y are correlated to some degree
- Find the most probable input X given the observation $Y = y$

Discrete Find the value of x that maximizes the a posteriori probability $P[X = x|Y = y]$:

$$\hat{X}_{MAP} = \max_x P[X = x|Y = y]$$

Cont. $\hat{X}_{MAP} = \max_x f_{X|Y}(x|y)$

Maximum Likelihood (ML) Estimator

Discrete The a posteriori probability is given by:

$$P[X = x|Y = y] = \frac{P[Y = y|X = x]P[X = x]}{P[Y = y]}$$

- $P[Y = y]$ does not affect the optimization (ignore)
- The a priori probability $P[X = x]$ may not be known, and we can model it as a uniform distribution (constant)
- Select the estimator $\hat{X}_{ML}$ that maximizes $P[Y = y|X = x]$ as the maximum likelihood (ML) estimator of the observed value $Y = y$:

$$\hat{X}_{ML} = \max_x P[Y = y|X = x]$$

Cont. Similarly:

$$\hat{X}_{ML} = \max_x f_{Y|X}(Y|X)$$

Example

- Find the MAP and ML estimators of X in terms of the observations Y when X and Y are jointly Gaussian random variables with the following conditional PDFs:

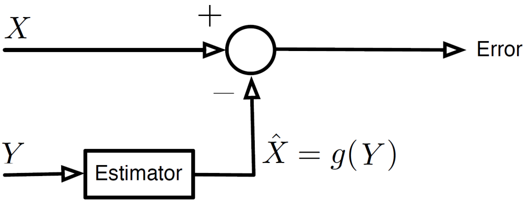
$$f_{X|Y} = \frac{e^{-\frac{1}{2(1-\rho^2)\sigma_X^2}\left(x-\rho\frac{\sigma_X}{\sigma_Y}(y-\mu_Y)-\mu_X\right)^2}}{\sqrt{2\pi\sigma_X^2(1-\rho^2)}}$$

$$f_{Y|X} = \frac{e^{-\frac{1}{2(1-\rho^2)\sigma_Y^2}\left(y-\rho\frac{\sigma_Y}{\sigma_X}(x-\mu_X)-\mu_Y\right)^2}}{\sqrt{2\pi\sigma_Y^2(1-\rho^2)}}$$

Estimation of Random Variables

- Estimating the parameters of one or more random variables (e.g. probabilities, means, variances, or covariances)
- Estimating the value of an inaccessible random variable X in terms of the observation of an accessible random variable Y:
  - Prediction Problems: predict future based on current and past observations
  - Interpolation Problems: given samples of a signal, we wish to interpolate to some in-between point in time
  - Filtering Problems: filter the noise out of a sequence of observations to provide the best estimate of the desired signal

Mean-Square Estimation (MSE)



- Assume X and Y are correlated to some degree
- If Y is observed, then estimate X so as to minimize the mean-square error:

$$e = E[(X - g(Y))^2]$$

Constant MSE

(a) Estimate the random variable X by a constant $\hat{X} = g(Y) = a$ so that the mean-square error is minimized.

(b) What is the mean-square error for this estimator?

Linear MSE

- Estimate X by a linear function $g(Y) = aY + b$ so that the mean-square error is minimized:

$$\min_{a,b} E[(X - aY - b)^2]$$

Step 1 We can apply the result from the previous example if we view the problem as estimating the random variable $(X - aY)$ with a constant b, such that:

$$b^* = E[X - aY] = E[X] - aE[Y]$$

Step 2 The minimization problem simplifies to one parameter a:

$$\min_{a} E[(X - E[X] - a(Y - E[Y]))^2]$$

such that $a^* = \dfrac{\text{Cov}(X,Y)}{\text{Var}(Y)}$

Linear MSE

- The linear estimate $g(Y) = aY + b$ of X is obtained:

$$\hat{X} = E[X] + \mathrm{Cov}(X, Y)\frac{Y - E[Y]}{\mathrm{Var}(Y)}$$

Note The linear mean-square estimator depends on second order moments: mean, variance and covariance.

- The minimum error of the linear MSE: $\epsilon_{MIN} = \mathrm{Var}(X)\left(1 - \rho^2\right)$.