Stochastic Signal Processing

Lecture 4 - Multiple Random Variables

August, 2024

What will be presented today?

- Conditional Probability for 2 RV case
- Extension of the probability concepts defined for 2 RVs to multiple RVs
- Function of multiple RVs and their probabilistic representation
- Linear Transformation of Random Vectors
- Whitening the data

Conditional Probability I

Conditional Probability: The probability that Y is in A given that we know $X = x$ is

$$P[Y \text{ in } A | X = x] = \frac{P[Y \text{ in } A, X = x]}{P[X = x]} \text{ for } P[X = x] > 0$$

- When X and Y are discrete
  1. The conditional PMF of Y given $X = x$ is defined by

     $$p_Y(y|x) = P[Y = y | X = x] = \frac{P[X = x, Y = y]}{P[X = x]} = \frac{p_{X,Y}(x, y)}{p_X(x)}.$$

     for x such that $P[X = x] > 0$.
  2. The probability of an event A given $X = x_k$ is

     $$P[Y \text{ in } A | X = x_k] = \sum_{y_j \text{ in } A} p_Y(y_j | x_k).$$

  3. When X and Y are independent, what will be $p_Y(y_j|x_k)$?

Conditional Probability II

- When X is discrete and Y is continuous

1. The conditional CDF of Y given $X = x_k$ is:

$$F_Y(y|x_k) = \frac{P[Y \leq y, X = x_k]}{P[X = x_k]} \text{ for } P[X = x_k] > 0.$$

2. The conditional PDF of Y given $X = x_k$, if the derivative exists is, is

$$f_Y(y|x_k) = \frac{d}{dy} F_Y(y|x_k).$$

3. The probability of an event A given $X = x_k$ is

$$P[Y \text{ in } A | X = x_k] = \int_{y \text{ in } A} f_Y(y|x_k) dy.$$

4. When X and Y are independent, what will be $F_Y(y|x)$ and $f_Y(y|x)$?

Conditional Probability III

- When both X and Y are continuous:
  1. The conditional CDF of Y given $X = x$ is:

$$F_Y(y|x) = lim_{h \to 0} F_Y(y|x < X \leq x + h) = \frac{\int_{-\infty}^{y} f_{X,Y}(x, y')dy'}{f_X(x)}$$

  2. The conditional PDF is obtained by taking the derivative of $F_Y(y|x)$ with respect to y:

$$f_Y[y|x] = \frac{d}{dy} F_Y(y|x) = \frac{f_{X,Y}(x, y)}{f_X(x)}.$$

  3. The probability of an event A given $X = x$ is

$$P[Y \text{ in } A|X = x] = \int_{y \text{ in } A} f_Y(y|x)dy.$$

  4. What happens when both the RVs are independent?

Conditional Probability IV

- Theorem of total probability in terms of PMF and PDF:
  - Discrete RV X case:

$$P[Y \text{ in } A] = \sum_{all\ \ x_k} P[Y \text{ in } A | X = x_k] p_X(x_k).$$

  - Continuous RV X case:

$$P[Y \text{ in } A] = \int_{-\infty}^{\infty} P[Y \text{ in } A | X = x] f_X(x) dx.$$

Conditional Expectation

- The conditional expectation of Y given $X = x$ is defined by

$$E[Y|X] = \int_{-\infty}^{\infty} y f_Y(y|x) dy.$$

- For discrete RVs, we have

$$E[Y|x] = \sum_{y_j} y_j p_Y(y_j|x).$$

- The conditional expectation $E[Y|X]$ can be seen as a function of X. Hence, it makes sense to view it as a RV $g(X) = E[Y|X]$. The expectation of $g(X)$ can be taken:

$$E[g(X)] = E[E[Y|X]].$$

- Prove that: $E[Y] = E[E[Y|X]]$.

General Random Vectors

- Let $X_1, X_2, \ldots, X_n$ be the components of an $n-$dimensional vector $\boldsymbol{X}$

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

- Each event A involving $\boldsymbol{X} = (X_1, \ldots, X_n)$ has a corresponding region in n-dimensional real space $\mathbb{R}^n$.
- Product form events: Let A_k be a one dimensional event that involves X_k only, then product form event A can be written as

$$A = \{X_1 \text{ in } A_1\} \cap \{X_2 \text{ in } A_2\} \cap \ldots \cap \{X_n \text{ in } A_n\}.$$

- $P[A] = P[\boldsymbol{X} \in A] = P[\{X_1 \text{ in } A_1\} \cap \{X_2 \text{ in } A_2\} \cap \ldots \cap \{X_n \text{ in } A_n\}]$.

Joint CDF of Random Vector

- The joint cumulative CDF of a random vector $\boldsymbol{X} = [X_1, X_2, \ldots, X_n]^T$ is defined as

$$F_X(\boldsymbol{x}) = F_{X_1,X_2,\ldots,X_n}(x_1, \ldots, x_n) = P[X_1 \leq x_1, X_2 \leq x_2, \ldots, X_n \leq x_n].$$

- The joint CDF is valid for all (discrete, continuous and mixed random vectors).
- Marginal CDF (e.g., CDF of X_1)

$$F_{X_1}(x) = F_{X_1,X_2,\ldots,X_n}(x_1, \infty, \ldots, \infty)$$

- Marginal CDF (e.g., Joint CDF of X_1 and X_2)

$$F_{X_1,X_2}(x_1, x_2) = F_{X_1,X_2,\ldots,X_n}(x_1, x_2, \infty, \ldots, \infty)$$

Joint PMFs of Discrete Random Vectors I

- The joint PMF of a random vector $X_1, X_2, \ldots, X_n$ is defined as

$$p_X(\boldsymbol{x}) = p_{X_1,X_2,\ldots,X_n}(x_1, \ldots, x_n) = P[X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n].$$

- The probability of n-dimensional event A is found by summing the PMF over all points in the event:

$$P[\boldsymbol{X} \text{ in } A] = \sum \cdots \sum_{\boldsymbol{x} \text{ in } A} p_{X_1,\ldots,X_n}(x_1, \ldots, x_n).$$

- Marginal PMF (e.g., one-dimensional PMF of X_j)

$$p_{X_j}(x_j) = \sum_{x_1} \cdots \sum_{x_{j-1}} \sum_{x_{j+1}} \cdots \sum_{x_n} p_{X_1,X_2,\ldots,X_n}(x_1, x_2, \ldots, x_n)$$

Joint PMFs of Discrete Random Vectors II

- Conditional PMF:

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

if $p_{X_1,X_2,\ldots,X_{n-1}}(x_1,x_2,\ldots,x_{n-1}) > 0$.

$$\begin{aligned} p_{X_1,X_2,\ldots,X_n}(x_1,x_2,\ldots,x_n) = {} & p_{X_n}(x_n|x_1,\ldots,x_n-1) \\ & \times p_{X_{n-1}}(x_{n-1}|x_1,x_2,\ldots,x_{n-2}) \\ & \times \ldots \times \\ & p_{X_2}(x_2|x_1) \times p_{X_1}(x_1), \end{aligned}$$

Joint Probability Density Functions I

- Random vectors $X_1, X_2, \ldots, X_n$ are jointly continuous if the probability of any n-dimensional event A is given by a n-dimensional integral of the joint PDF, $f_{X_1,X_2,\ldots,X_n}(x_1, x_2, \ldots, x_n)$:

$$P[\boldsymbol{X} \text{ in } A] = \int \cdots \int_{\boldsymbol{x} \in A} f_{X_1,X_2,\ldots,X_n}(x_1', x_2', \ldots, x_n') dx_1' dx_2' \ldots dx_n'.$$

- The joint CDF is obtained from

$$F_{\boldsymbol{X}}(\boldsymbol{x}) = F_{X_1,X_2,\ldots,X_n}(x_1,x_2,\ldots,x_n) = \int_{-\infty}^{x_1} \ldots \int_{-\infty}^{x_n} f_{X_1,X_2,\ldots,X_n}(x_1',x_2',\ldots,x_n') dx_2' dx_2' \ldots dx_n'.$$

- Marginal PDF (e.g., one-dimensional PDF for X_1) is

$$f_{X_1}(x_1) = \int_{-\infty}^{\infty} \ldots \int_{-\infty}^{\infty} f_{X_1,X_2,\ldots,X_n}(x_1, x_2', \ldots, x_n') dx_1' \ldots dx_n'.$$

What will be the marginal PDF for $(X_1, X_2 \ldots X_{n-1})$?

Joint Probability Density Functions II

- Conditional PDF

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

if $f_{X_1,X_2,\ldots,X_{n-1}}(x_1,x_2,\ldots,x_{n-1})>0$.

Repeated applications of the above yields:

$$\begin{aligned}f_{X_1,X_2,\ldots,X_n}(x_1,x_2,\ldots,x_n)=&f_{X_n}(x_n|x_1,\ldots,x_n-1)\\&\times f_{X_{n-1}}(x_{n-1}|x_1,x_2,\ldots,x_{n-2})\\&\times\ldots\times\\&f_{X_2}(x_2|x_1)\times f_{X_1}(x_1),\end{aligned}$$

Question: Suppose that $X_1,X_2,\ldots,X_n$ are independent RVs. Could you write the corresponding PMF, CDF and PDF of the vector RVs in simplified form?

Expected Value of Scalar Function of Random Vector

The expected value of a scalar function $Z = g(\boldsymbol{X}) = g(X_1, \ldots, X_n)$ of a vector RV $\boldsymbol{X}$ is given by:

$$E(Z) = \int_{-\infty}^{\infty} \ldots \int_{-\infty}^{\infty} g(x_1, \ldots, x_n) f_{\boldsymbol{X}}(x_1, \ldots, x_n) dx_1 \ldots dx_n.$$

$$E(Z) = \sum_{x_1} \ldots \sum_{x_n} g(x_1, \ldots, x_n) p_{\boldsymbol{X}}(x_1, \ldots, x_n).$$

- Let $g(\boldsymbol{X})$ be the sum of the functions of $\boldsymbol{X}$), then:

$$E[g_1(\boldsymbol{X}) + \ldots + g_n(\boldsymbol{X})] = E[g_1(\boldsymbol{X})] + \ldots + E[g_n(\boldsymbol{X})].$$

- Problem: Let $Z = X + Y$. Find $E[Z]$. If we extend the summation over n RVs: $Z = X_1 + X_2 + \ldots + X_n$, what will be $E[Z]$?

Single Function of Several RVs

Let a RV $Z = g(X_1, X_2, \ldots, X_n)$ be defined as a function of several RVs. The CDF of Z is found by first finding the equivalent event of $\{Z \leq z\}$ i.e., the set $R_z = \{\boldsymbol{x} = (x_1, x_2, \ldots, x_n)$ such that $g(\boldsymbol{x}) \leq z\}$, then

$$
\begin{aligned}
F_Z(z) &= P[\boldsymbol{X} \text{ in } R_z] \\
&= \int \cdots \int_{\boldsymbol{x} \text{ in } R_z} f_{X_1, X_2, \ldots, X_n (x_1', \ldots, x_n') dx_1' \ldots dx_n'}
\end{aligned}
$$

How to find the PDF of Z?

For the transformation of discrete RVs i.e., $Z = g(X, Y)$, the PMF of Z can be expressed in terms of the joint PMF of X and Y as follows:

$$
p_Z[z_k] = \sum \sum_{\{(i,j): z_k = g(x_i, y_j)\}} p_{X,Y}[x_i, y_j] \tag{1}
$$

Example: Let Z=X+Y; Find the PDF of Z in terms of joint PDF of X and Y. What happens when X and Y are independent?

PDF of Linear Transformations I

Goal: To find the joint PDF of transformed RVs $\boldsymbol{Z} = (Z_1, \dots Z_n)$, obtained from RVs $\boldsymbol{X} = (X_1, \dots X_n)$ through $\boldsymbol{Z} = A\boldsymbol{X}$; A denotes an $n \times n$ invertible matrix.

- **Linear transformation of two RVs**:

$$\begin{bmatrix} V \\ W \end{bmatrix} = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \begin{bmatrix} X \\ Y \end{bmatrix}$$

  The above matrix A is considered to have an inverse meaning that there is one-to-one correspondence between each pair of (x, y) and (v, w)

$$\begin{bmatrix} x \\ y \end{bmatrix} = A^{-1} \begin{bmatrix} v \\ w \end{bmatrix}$$

PDF of Linear Transformations II

Equivalent Events: $f_{X,Y}dxdy = f_{V,W}dP$ where dP is the area of the parallelogram. The joint PDF of V and W is thus

$$f_{V,W} = \frac{f_{X,Y}}{\left|\frac{dP}{dxdy}\right|}$$

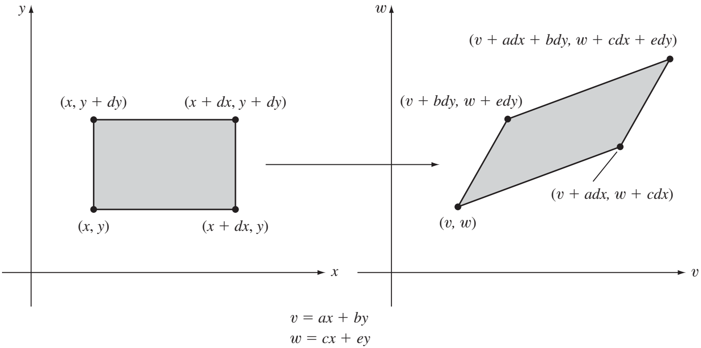
It can be shown that

$$\left|\frac{dP}{dxdy}\right| = |ae - bc| = |A|,$$

where A is the determinant of A.
The joint PDF of $\boldsymbol{Z}$ is then

$$f_{\boldsymbol{Z}}(\boldsymbol{z}) = \frac{f_{\boldsymbol{x}}(A^{-1}\boldsymbol{z})}{|A|}.$$

PDF of Linear Transformations III



Source: A. Garcia (2008) Probability, Statistics and Random Processes for Electrical Engineering. New york:Pearson

PDF of Linear Transformations IV

- **Extension to n-dimensional vector $\boldsymbol{Z}$**: $\boldsymbol{Z} = A\boldsymbol{X}$ where A is $n \times n$ invertible matrix. The joint PDF of Z is

$$f_{\boldsymbol{Z}}(\mathbf{z}) = \frac{f_{\boldsymbol{X}}(\boldsymbol{x})}{|A|}\bigg|_{\boldsymbol{x}=A^{-1}\boldsymbol{z}}$$

- Problem (Example 5.45): Let X and Y be the jointly Gaussian RVs and let V and W be obtained from X and Y by

$$\begin{bmatrix} V \\ W \end{bmatrix} = \begin{bmatrix} 1/\sqrt{2} & 1/\sqrt{2} \\ -1/\sqrt{2} & -1/\sqrt{2} \end{bmatrix} \begin{bmatrix} X \\ Y \end{bmatrix}$$

Find the joint PDF of V and W.

Expected Values of Vector RVs I

- For $\boldsymbol{X} = (X_1, \ldots, X_n)$, the mean vector is defined as the column vector of the expected values of the components:

$$\boldsymbol{m_X} = E[\boldsymbol{X}] = \begin{bmatrix} E[X_1] \\ E[X_2] \\ \vdots \\ E[X_n] \end{bmatrix} \tag{2}$$

- The correlation matrix is nxn symmetric and is given by

$$\boldsymbol{R_X} = E[\boldsymbol{X}\boldsymbol{X}^T] = \begin{pmatrix} E[X_1^2] & E[X_1X_2] & \ldots & E[X_1X_n] \\ E[X_2X_1] & E[X_2^2] & \ldots & E[X_2X_n] \\ \vdots & \vdots & \ldots & \vdots \\ E[X_nX_1] & E[X_nX_2] & \ldots & E[X_n^2] \end{pmatrix}$$

What would be the covariance matrix $\boldsymbol{K_X}$ and what would be its relation to $\boldsymbol{R_X}$?

Linear Transformation of Random Vectors I

Consider the linear transformation of a random vector $\boldsymbol{X}$ such that

$$\boldsymbol{Y}=\boldsymbol{A}\boldsymbol{X}= \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \dots & \vdots \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{pmatrix} \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_n \end{bmatrix} \tag{3}$$

- The expected value of kth component of $\boldsymbol{Y}$ is

$$E[Y_k] = E\left[\sum_{j=1}^{n} a_{kj}X_j\right] = \sum_{j=1}^{n} a_{kj}E[X_j].$$

This yields

$$\boldsymbol{m_Y} = E[\boldsymbol{Y}] = \boldsymbol{A}\boldsymbol{m_X}\ .$$

Linear Transformation of Random Vectors II

- Let $\boldsymbol{K_X}$ and $\boldsymbol{K_Y}$ denote the covariance matrices of vector RVs $\boldsymbol{X}$ and $\boldsymbol{Y}$ respectively. The covariance matrix of $\boldsymbol{Y}$, in terms of $\boldsymbol{K_X}$, can be written as

$$\boldsymbol{K_Y} = \boldsymbol{A K_X A^T} .$$

- The cross-covariance matrix of two random vectors $\boldsymbol{Y}$ and $\boldsymbol{Y}$ is

$$\boldsymbol{K_{XY}} = \boldsymbol{E[(X - m_X)(Y - m_Y)^T]} = \boldsymbol{K_X A^T} .$$

- Example: What would be the covariance of $\boldsymbol{Y = AX}$ if $\boldsymbol{X}$ is an uncorrelated random vector i.e, for which $\boldsymbol{K_X = I}$?

Diagnolization of Covariance Matrix

Any covariance matrix $\boldsymbol{K_X}$ can be diagnolized via linear transformation.

- Let $\boldsymbol{X}$ be a random vector with covariance matrix $\boldsymbol{K_X}$. Can we perform/find a linear transformation of $\boldsymbol{X}$ i.e., $\boldsymbol{Y}=\boldsymbol{AX}$, such that $\boldsymbol{Y}$ has a diagonal covariance matrix (uncorrelated components).
- The transformation matrix is given by:

$$\boldsymbol{A}=\boldsymbol{P}^{\boldsymbol{T}}$$

  where the columns of $\boldsymbol{P}$ are eigenvectors of $\boldsymbol{K_X}$.
- The resulting covariance matrix $\boldsymbol{K_Y}$ is as follows:

$$\boldsymbol{K_Y}=\boldsymbol{P}^{\boldsymbol{T}}\boldsymbol{K_X}\boldsymbol{P}=\boldsymbol{D},$$

  where D is diagonal having the eigenvalues of $\boldsymbol{K_X}$ at its diagonal.

Generating RVs with arbitrary specified Covariance Matrix

- Let $\boldsymbol{Y}=\boldsymbol{A}^T\boldsymbol{X}$ where $\boldsymbol{X}$ is a uncorrelated RV with components that are zero mean and unit variance. Then $m_Y = Am_X = 0$ and

$$K_Y = A^T K_X A = A^T A$$

- Let P be the matrix whose columns are eigenvectors of K_Y and let Δ be the diagonal matrix of eigenvalues, then

$$P^T K_Y P = P^T P\Delta = \Delta$$

Pre-multiplying the above by P and post-multiplying by P^T and considering that P is orthogonal, we get

$$K_Y = P\Delta P^T = P\Delta^{1/2}\Delta^{1/2}P^T$$

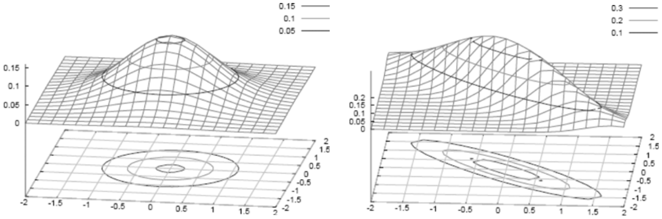
This leads to $A = (P\Delta^{1/2})^T$

Gaussian Distribution

For a D-dimensional vector $\boldsymbol{x}$, the vector Gaussian distribution is given by:

$$\mathcal{N}(\boldsymbol{x}\,|\boldsymbol{\mu},\boldsymbol{\Sigma}) = \frac{1}{(2\pi)^{D/2}|\boldsymbol{\Sigma}|^{1/2}}\, exp\left\{-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})\right\}$$

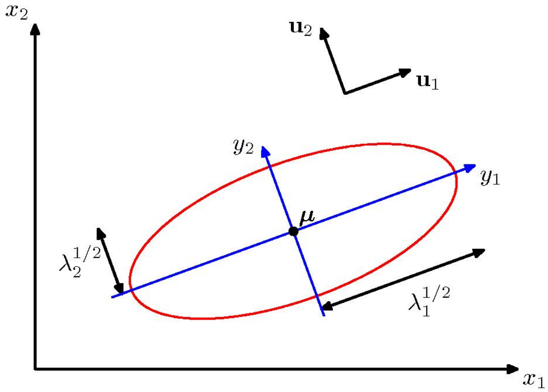
where $\boldsymbol{\mu} \in \mathbb{R}^D$ is a mean vector, $\boldsymbol{\Sigma} \in \mathbb{R}^{D\times D}$ is the covariance matrix and $|\cdot|$ denotes the determinant operator.



(a) (b)

Gaussian Distribution

The Gaussian distribution is constant on elliptical surfaces



where:

- λ_i is the i-th eigenvalue of $\boldsymbol{\Sigma}$
- $\boldsymbol{u}_i$ is the eigenvector of $\boldsymbol{\Sigma}$ associated to λ_i.

Conditional and marginal distributions of a Gaussian distribution:

Given a Gaussian distribution $\mathcal{N}(\boldsymbol{x}|\boldsymbol{\mu}, \boldsymbol{\Sigma})$ with:

$$\boldsymbol{x} = [\boldsymbol{x}_a, \boldsymbol{x}_b]^T, \quad \boldsymbol{\mu} = [\boldsymbol{\mu}_a, \boldsymbol{\mu}_b]^T$$

and

$$\boldsymbol{\Sigma} = \begin{pmatrix} \boldsymbol{\Sigma}_{aa} & \boldsymbol{\Sigma}_{ab} \\ \boldsymbol{\Sigma}_{ba} & \boldsymbol{\Sigma}_{bb} \end{pmatrix}$$

- The marginal distribution $p(\boldsymbol{x}_a)$ is a Gaussian distribution with parameters $(\boldsymbol{\mu}_a, \boldsymbol{\Sigma}_{aa})$ i.e., $p(\boldsymbol{x}_a) = \mathcal{N}(\boldsymbol{x}_a|\boldsymbol{\mu}_a, \boldsymbol{\Sigma}_{aa})$
- The conditional distribution $p(\boldsymbol{x}_a|\boldsymbol{x}_b)$ is a Gaussian distribution i.e., $p(\boldsymbol{x}_a|\boldsymbol{x}_b) = \mathcal{N}(\boldsymbol{x}|\boldsymbol{\mu}_{a|b}, \boldsymbol{\Sigma}_{a|b})$ with parameters:

$$\begin{aligned} \boldsymbol{\mu}_{a|b} &= \boldsymbol{\mu}_a + \boldsymbol{\Sigma}_{ab}\boldsymbol{\Sigma}_{bb}^{-1}(\boldsymbol{x}_b - \boldsymbol{\mu}_b) \\ \boldsymbol{\Sigma}_{a|b} &= \boldsymbol{\Sigma}_{aa} - \boldsymbol{\Sigma}_{ab}\boldsymbol{\Sigma}_{bb}^{-1}\boldsymbol{\Sigma}_{ba} \end{aligned}$$