Example (Gaussian Signalling over a Gaussian Channel)

\[ N(0, P) \sim X \xrightarrow{+} y = x + z \quad \text{Assume } X \text{ and } z \text{ are independent.} \]

Find \( g(y) \) that minimize the MSE

\[ E[(x - g(y))^2] \]

we already know that \( g^*(y) = E[X|Y=y] \)

Recall that:

\[ f_{X|Y}(y|x) = \frac{1}{\sqrt{2\pi N}} e^{-\frac{(x-y)^2}{2N}} \quad \text{and} \quad f_Y(y) = \frac{1}{\sqrt{2\pi N}} e^{-\frac{y^2}{2N}} \]

Hence:

\[ f_{X|Y}(x|y) = \frac{f_{X,Y}(x,y)}{f_Y(y)} = \frac{1}{\sqrt{2\pi P}} e^{-\frac{(x-(P+P_N)y)^2}{2P}} \]

In other word \( X|Y \sim N \left( \frac{P}{P+N} y + \frac{P_N}{P+N} \mu, \frac{P}{P+N} \right) \)

Thus, \( g^*(y) = \frac{P}{P+N} y + \frac{P_N}{P+N} \mu \)

\[ \to \mu \quad \text{as } N \to \infty \]

\[ \to y \quad \text{as } P \to 0^+ \]

Furthermore the resulting MSE is:

\[ E \left[ \text{Var}(X|Y) \right] = E \left[ \frac{PN}{P+N} \right] = \frac{PN}{P+N} < P \]

Random Vectors

Let \( X_1, X_2, \ldots, X_n \) be Random Variables on the same probability space, we define a random vector as:

\[ \mathbf{X} = \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_n \end{bmatrix} \]

the random vector \( \mathbf{X} \) is completely specified by its joint cdf:

\[ F_{\mathbf{X}}(a) = P \left( X_1 \leq a_1, X_2 \leq a_2, \ldots, X_n \leq a_n \right) \]

if \( \mathbf{X} \) is continuous (i.e. \( F_{\mathbf{X}}(a) \) is continuous) then it suffices to know its pdf

\[ f_{\mathbf{X}}(a) = \frac{\partial^n}{\partial x_1 \partial x_2 \cdots \partial x_n} F_{\mathbf{X}}(a) = f_{X_1, X_2, \ldots, X_n}(a_1, a_2, \ldots, a_n) \]
If $X$ is discrete, then it suffices to know its pmf
$$P_X(x) = P\left(\{ X_1=x_1, X_2=x_2, \ldots, X_n=x_n \}\right)$$

A marginal cdf (pdf/pmf) is the joint cdf (pdf/pmf) of a (proper) subset of random variables.

For example,
$$f_{X_1}(x_1), f_{X_2}(x_2), f_{X_3}(x_3), f_{X_1,X_3}(x_1,x_3), f_{X_2,X_3}(x_2,x_3), f_{X_1,X_2,X_3}(x_1,x_2,x_3)$$
are marginals of $f_{X_1,X_2,X_3}(x_1,x_2,x_3)$
we can obtain marginals from the joint in the usual way.

For example:
$$F_{X_1}(x_1) = \lim_{y_2,y_3 \to \infty} F_{X_1,X_2,X_3}(x_1,y_2,y_3)$$
$$f_{X_1|X_2,X_3}(x_1|x_2,x_3) = \frac{f_{X_1,X_2,X_3}(x_1,x_2,x_3)}{f_{X_2,X_3}(x_2,x_3)}$$

The Conditional cdf (pdf/pmf) can be also defined in usual way.

For example:
$$f_{X_1|X_2,X_3}(x_1|x_2,x_3) = \frac{f_{X_1,X_2,X_3}(x_1,x_2,x_3)}{f_{X_2,X_3}(x_2,x_3)}$$
$$P_{X_1,X_2,X_3}(x_1,x_2,x_3) = \frac{P_{X_1,X_2,X_3}(x_1,x_2,x_3)}{P_{X_1}(x_1)}$$

Chain Rule:
$$f_X(x) = f_{X_1}(x_1) f_{X_2|X_1}(x_2|x_1) \cdots f_{X_n|X_1,\ldots,X_{n-1}}(x_n|x_1,\ldots,x_{n-1})$$

Independence

$X_1, X_2, \ldots, X_n$ are mutually independent if
$$f_X(x) = f_{X_1}(x_1) f_{X_2}(x_2) \cdots f_{X_n}(x_n)$$
if further $X_1, \ldots, X_n$ have the same marginal distribution
then they called independent and identically distributed (i.i.d.)

Conditional independence

$X_1$ and $X_2$ are said to be conditionally independent given $X_3$ if:
$$f_{X_1,X_2|X_3}(x_1,x_2|x_3) = f_{X_1|X_3}(x_1|x_3) f_{X_2|X_3}(x_2|x_3) \quad \text{for all } x_1,x_2,x_3$$
Equivalently:
$$f_{X_1,X_2,X_3}(x_1,x_2,x_3) = f_{X_1|X_3}(x_1|x_3) f_{X_2|X_3}(x_2|x_3) \quad \text{for all } x_1,x_2,x_3$$
Remark: Neither independence nor conditional independence implies the other.

Example) $X_1, X_2$ are i.i.d. $\text{Bern} \left( \frac{1}{2} \right)$

and $X_3 = X_1 \oplus X_2$. Then $X_1$ and $X_3$ independent.

But $X_1$ and $X_3$ are not conditionally independent given $X_2$.

Example) Let $p \in (0,1]$ let $X_1, X_2, \ldots, X_n | \{P=p\} \sim \text{Bern}(p)$

By definition, $X_1, X_2, \ldots, X_n$ are conditionally independent given $p$.

But $X_1, X_2, \ldots, X_n$ are not independent.

To see this, consider

$$P(\{X_1=1\}) = P(\{X_2=1\}) = \int_0^1 P \, dp = \frac{1}{2}$$

$$P(\{X_1=1, X_2=1\}) = \int_0^1 P^2 \, dp = \frac{1}{3} \neq \left( \frac{1}{2} \right)^2$$

Mean and Covariance Matrix

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

The covariance matrix of $X$ is $\text{Cov}(X) = K_X = E_{X} = \begin{bmatrix} \cdots & \text{Cov}(X_i, X_j) & \cdots \end{bmatrix}$

for $n=2$

$$K_X = \begin{pmatrix} \text{Var}(X_1) & \text{Cov}(X_1, X_2) \\ \text{Cov}(X_2, X_1) & \text{Var}(X_2) \end{pmatrix} = \begin{pmatrix} \sigma_{X_1}^2 & \rho_{X_1, X_2} \sigma_{X_1} \sigma_{X_2} \\ \rho_{X_1, X_2} \sigma_{X_1} \sigma_{X_2} & \sigma_{X_2}^2 \end{pmatrix}$$

Properties of Covariance Matrix $K_X$:

1. $K_X$ is symmetric

2. $K_X$ is positive semi-definite (non-negative definite)

Namely for any column vector $a$, the quadratic form $a^T K_X a \geq 0 \forall$ any $a$.

To see this, note that:

$$K_X = E \left[ \begin{pmatrix} X_1 - E(X_1) \\ X_2 - E(X_2) \\ \vdots \\ X_n - E(X_n) \end{pmatrix} \begin{pmatrix} X_1 - E(X_1) & X_2 - E(X_2) & \cdots & X_n - E(X_n) \end{pmatrix} \right]$$
why? The (i,j) element is
\[ E\left[(X_i - E(X_i))(X_j - E(X_j))\right] = \text{COV}(X_i, X_j) \]

By linearity of expectation:
\[
\alpha^T K x \alpha = \alpha^T E \left[ (X - E(X))(X - E(X))^T \right] \alpha
\]
\[
= E \left[ \alpha^T (X - E(X))(X - E(X))^T \alpha \right]
\]
\[
= E \left[ (\alpha^T (X - E(X))^2 \right] \geq 0
\]

(3) Conversely, any symmetric positive semidefinite matrix \( K \) is a covariance matrix of some random vector.

Example: which of the following can be covariance matrix.

1) \[
\begin{bmatrix}
1 & 0 & 0 \\
0 & 1 & 0 \\
0 & 0 & 1
\end{bmatrix}
\]
   Yes

2) \[
\begin{bmatrix}
1 & 2 & 1 \\
2 & 1 & 1 \\
1 & 1 & 1
\end{bmatrix}
\]
   No \[ |2,1| = -1 \leq 0 \]

3) \[
\begin{bmatrix}
1 & 0 & 1 \\
0 & 1 & 2 \\
1 & 3 & 0
\end{bmatrix}
\]
   No \[ Not \ symmetric. \]

4) \[
\begin{bmatrix}
1 & 1 & 1 \\
1 & 1 & 1 \\
1 & 1 & 1
\end{bmatrix}
\]
   No \[ Variance \ can't \ be \ negative. \]

5) \[
\begin{bmatrix}
1 & 1 & 1 \\
1 & 2 & 1 \\
1 & 1 & 3
\end{bmatrix}
\]
   Yes \[ Play \]