Example (Gaussian Signalling over a Gaussian channel)

\[ Z \sim N(0, N) \]

\[ X \rightarrow \mathcal{B} \rightarrow Y = X + Z \]

Assume that \( X \perp \perp Z \sim N(\mu, P) \)

Find \( g^*(y) \) that minimizes the MSE \( E[(X - g(y))^2] \)

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

Recall that

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

Hence,

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

\[ = \frac{1}{\sqrt{2\pi(P+N)}} \exp \left( -\frac{1}{2\left(\frac{P}{P+N}\right)} (x - \left( \frac{P}{P+N} y + \frac{N}{P+N} \mu \right))^2 \right) \]

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

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

as \( N \to \infty \)

Observation

Prior belief

\( y \) as \( P \to \infty \)

Furthermore, the resulting MSE is

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

Let $X_1, X_2, \ldots, X_n$ be random variables on the same probability space.

We define a random vector as

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

The random vector $X$ is completely specified by its joint cdf

$$
F_X(x) = P(\{X_1 \leq x_1, X_2 \leq x_2, \ldots, X_n \leq x_n\})
$$

- If $X$ is continuous (i.e., $F_X(x)$ is continuous), then it suffices to know its pdf $f_X(x) = \frac{\partial^n}{\partial x_1 \ldots \partial x_n} F_X(x) = f_{X_1, \ldots, X_n}(x_1, x_2, \ldots, x_n)$
- If $X$ is discrete, then it suffices to know its pmf $P_X(x) = p(\{X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n\})$

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

For example,

$$
f_{X_1}(x_1), f_{X_2}(x_2), f_{X_3}(x_3), f_{X_1,X_2}(x_1,x_2), f_{X_2,X_3}(x_2,x_3), f_{X_3,X_1}(x_3,x_1)
$$

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,

$$
F_{X_1}(x_1) = \lim_{x_2,x_3 \to \infty} F_{X_1,X_2,X_3}(x_1,x_2,x_3)
$$

$$
f_{X_1,X_2}(x_1,x_2) = \int f_{X_1,X_2,X_3}(x_1,x_2,x_3) dx_3.
$$

The conditional cdf (pdf/pmf) can be also defined in the usual way. For example,

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

$$
P_{X_1,X_2|X_3}(x_1,x_2|x_3) = \frac{P_{X_1,X_2,X_3}(x_1,x_2,x_3)}{P_X(x_1)}
$$

Chain rule

$$
f_Z(z) = f_{X_1}(x_1) f_{X_2|X_1}(x_2|x_1) \cdots f_{X_n|X_{n-1},\ldots,X_1}(x_n|x_1,\ldots,x_{n-1}).$$
Independence

\( X_1, X_2, \ldots, X_n \) are (mutually) independent if

\[ f_{X}(x) = f_{X_1}(x_1) f_{X_2}(x_2) \cdots f_{X_n}(x_n) \quad \text{for all } x \]

If further \( X_1, \ldots, X_n \) have the same marginal distribution, then they are called independent and identically distributed (i.i.d.)

Conditional independence

\( X_1 \) and \( X_3 \) are said to be conditionally independent given \( X_2 \) if

\[ f_{X_1, X_3 | X_2}(x_1, x_3 | x_2) = f_{X_1 | X_2}(x_1 | x_2) f_{X_3 | X_2}(x_3 | x_2) \quad \text{for all } x_1, x_2, x_3. \]

Equivalently,

\[ f_{X_3 | X_1, X_2}(x_3 | x_1, x_2) = f_{X_3 | X_2}(x_3 | x_2) \quad \text{for all } x_1, x_2, x_3. \]

Remark: Neither independence nor conditional independence implies the other.

Example(1) \( X_1, X_2 \) are i.i.d. Bern \((\frac{1}{2})\) and \( X_3 := X_1 \oplus X_2. \)

Then \( X_1 \) and \( X_3 \) are independent.  
But \( X_1 \) and \( X_3 \) are not conditionally independent given \( X_2. \)

(a) Let \( p \sim U[0, 1]. \) Let \( X_1, X_2, \ldots, X_n, \mid \) \( p \sim \) 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

\[ P(\{X_i = 1\}) = P(\{X_i = 0\}) = \int_0^1 p \, dp = \frac{1}{2} \quad \text{(or by symmetry)} \]

\[ P(\{X_i = 1, X_i = 1\}) = \int_0^1 p^2 \, dp = \frac{1}{3} \neq \left( \frac{1}{2} \right)^2. \]

Mean and Covariance matrix

The mean (vector) of the random vector \( X \) is \( \mu(X) = \begin{pmatrix} \mu(X_1) \\ \mu(X_2) \end{pmatrix}. \)

The covariance matrix of \( X \) is

\[ \text{Cov}(X) = \Sigma_X = \begin{pmatrix} \text{Cov}(X_1, X_1) & \text{Cov}(X_1, X_2) \\ \text{Cov}(X_2, X_1) & \text{Cov}(X_2, X_2) \end{pmatrix} \]

For \( n=2, \)

\[ \Sigma_X = \begin{pmatrix} \text{Var}(X_1) & \text{Cov}(X_1, X_2) \\ \text{Cov}(X_2, X_1) & \text{Var}(X_2) \end{pmatrix} = \begin{pmatrix} \sigma_{x_1}^2 & \rho_{x_1, x_2} \sigma_{x_1} \sigma_{x_2} \\ \rho_{x_1, x_2} \sigma_{x_1} \sigma_{x_2} & \sigma_{x_2}^2 \end{pmatrix}. \]
Properties of covariance matrix $K_x$

1. $K_x$ is symmetric
2. $K_x$ is positive semidefinite (nonnegative definite), namely, for any vector $a$, the quadratic form $a^T K_x a \geq 0$.

To see this, note that

$$K_x = E \begin{bmatrix}
(x-Ex_1) \\
(x-Ex_2) \\
\vdots \\
(x-Ex_n)
\end{bmatrix}
\begin{bmatrix}
(x-Ex_1) \\
(x-Ex_2) \\
\vdots \\
(x-Ex_n)
\end{bmatrix}^T$$

Why? $(k_x)_{ij} = E[(x_i-Ex_i)(x_j-Ex_j)] = Cov(x_i, x_j)$

By linearity of expectation,

$$a^T K_x a = a^T E[(x-Ex)(x-Ex)^T] a$$
$$= E[a^T(x-Ex)(x-Ex)^T a]$$
$$= E[(a^T(x-Ex))^2] \geq 0$$

3. Conversely, any positive semidefinite matrix $K$ is a covariance matrix of some random vector.

Which of the following can be a covariance matrix?

1. $\begin{bmatrix}
0 & 0 & 0 \\
0 & 1 & 0 \\
0 & 0 & 1
\end{bmatrix}$ Yes.

2. $\begin{bmatrix}
2 & 1 \\
1 & 0
\end{bmatrix}$ No.

3. $\begin{bmatrix}
0 & 1 \\
1 & 2
\end{bmatrix}$ No.

4. $\begin{bmatrix}
-1 & 1 & 1 \\
1 & -1 & 1 \\
1 & 1 & -1
\end{bmatrix}$ No.

5. $\begin{bmatrix}
1 & 2 & 1 \\
2 & 1 & 3 \\
1 & 3 & 1
\end{bmatrix}$ Yes.