Random vectors and estimation theory

\[ X = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \text{, where } X_i \text{ are RVs / RANDOM VECTOR} \]
\[ Y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} \]

Two important matrices associated with RV's

\[
\begin{align*}
\text{Correlation matrix: } & E[X X^T] \\
\text{Cross-correlation: } & E[X T] \\
\text{Covariance matrix: } & E[(X-E(X))(X^T-E(X)^T)] \\
\text{Cross-covariance matrix: } & E[(X-E(X))(Y^T-E(Y)^T)]
\end{align*}
\]

Claim: Correlation and covariance matrices are symmetric, positive semi-definite matrices. If \( K \) is positive semi-definite, then there exists a zero-mean random vector \( X \) with \( K \) as its correlation matrix.

Proof: If \( K \) is a correlation matrix, then
\[
K = E[X X^T] \text{ for some RV } X
\]
\[
q^T K q = q^T E[X X^T] q = E[(q^T X)(X^T q)] = E[Q^T X^T X Q] > 0
\]
for non-random, \( q \in \mathbb{R}^n \).

For the second claim, suppose that \( K \) is positive semi-definite, then \( K \) has non-negative eigenvalues \( \lambda_1, \ldots, \lambda_n \) and the orthonormal matrix \( U \) formed from the eigenvector basis.

Let \( Y_i \) be independent RVs, s.t. \( E[Y_i] = 0, E[Y_i^2] = \lambda_i \)

\[ Y = \begin{bmatrix} Y_1 \\ \vdots \\ Y_n \end{bmatrix} \]

Then, \( \text{Cov}(Y, Y) = \Lambda = \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_n \end{bmatrix} \)

Define \( X = U Y \)

Then, \( E[X] = U E[Y] = 0 \) and

\[
\text{Cov}(X, X) = \text{Cov}(U Y, U Y) = E[U Y Y^T U^T] = U E[Y Y^T] U^T = U \Lambda U^T = K
\]

One more definition:

Characteristic function \( \phi_X(u) = E[\exp(ju^T X)] \)
Orthogonality principle

Let us start with the case of scalar RVs.
Let \( X \) be a scalar RV that we cannot observe directly. We can try to "estimate" \( X \) either via repeated experiments (say, \( X' \)), or via the observation of some other RV \( Y \).

Say, having complicated formula, do not want to drag RV in it, prefer to have a constant \( b \) to be a "proxy" for \( X \).

What is the best choice for \( b \)?

Intuitively, mean should be a "good" proxy. Formally, if we seek to find \( b \) that minimizes MSE = mean square error \( E[(X-b)^2] \), we get

\[
+ 2E[(X-E[X])(E[X]-b)] + (E[X]-b)^2 = \text{var}(X) + (E[X]-b)^2
\]

\( b^* = E[X] \) minimizes the error, which in that case equals the variance.

Key to the simple proof was:

\[
E[(X-E[X])] = 0 \quad \text{for any estimator} \]

\[
i.e. \ E[(X-b^*)] = 0 \quad \forall \ b \in \mathbb{R}, \ \text{const.}
\]

error of optimal estimator

Formally:

Recall that \( X, Y \) are said to be uncorrelated if \( E[XY] = E[X]E[Y] \),

\( X, Y \) are orthogonal if

\( E[XY] = 0 \), \( X \perp Y \)

Above orthogonality principle asserts that \( (X-E[X]) \perp \mathbf{1} \in \mathbb{C} \in \mathbb{C} \)

\( L^2(S, \mathcal{F}, P) \) is set of all RV's on given PS will have second moment (why finite - will use MSE)

\( \forall \in L^2(S, \mathcal{F}, P) \) s.t.,

a) \( \gamma \) is linear class: If \( Z_1, Z_2 \in \gamma \), then

\( a_1 Z_1 + a_2 Z_2 \in \gamma \), \( \forall \) const. \( a_1, a_2 \)

b) \( \gamma \) is closed in HS sense: \( Z_1, Z_2, \ldots \in \gamma \), \( Z_n \to Z \in \gamma \)

\( \Rightarrow Z \in \gamma \)
Let \( Y \) be as above.

a) \( \exists \hat{X} \in \mathcal{D} \) such that \( E[(X - \hat{X})^2] \leq E[(X - Z)^2] \quad \forall Z \in \mathcal{D} \) \( \hat{X} \) is unique.

b) \( W = Z \)
\( \implies \)
\( \forall Z \in \mathcal{D} \), \( (X - W) \perp Z \)

\[ \text{c) MVESE} = E[(X - \hat{X})^2] = E[X^2] - E[(\hat{X})^2] \]

**Proof**

b) Assume that \( W \) is \( (X-W) \perp Z \) \( \forall Z \in \mathcal{D} \) (linearity) hence \( (X-W) \perp (W-Z) \)
\[ \text{i.e.} \quad E[(X-W)^2] \]
Assume has smallest MVESE

Assume has smallest MVESE

Show orthogonality

Expanding

\[ E[(X-(\hat{X}-cZ))^2] = E[(X-\hat{X}-cZ)^2] = E[(X-\hat{X})^2] - E[2(X-\hat{X})cZ] \]
\[ + E[c^2Z^2] \]

From **

\[ 0 \leq E[c^2Z^2] - 2cE[(X-\hat{X})Z] \]
\[ \frac{d(c)}{dc} \]

\[ f'(c) = 0 \implies \]
\[ \left( E[c^2Z^2] - 2cE[(X-\hat{X})Z] \right) = 0 \]
\[ \left( 2cE[Z^2] - 2E[(X-\hat{X})Z] \right) = 0 \]
\[ \implies E[(X-\hat{X})Z] = 0 \]


\[ \text{Easier} \]
Summary of properties of projections
Denote the projection of $X$ onto $Y$ as $\Pi_Y(X)$

1. $\Pi_Y(a_1X_1 + a_2X_2) = a_1\Pi_Y(X_1) + a_2\Pi_Y(X_2)$

2. Both $\nu_1, \nu_2$ are closed, $\nu_1 \subseteq \nu_2$

   $$\Pi_{\nu_2}(x) = \Pi_{\nu_1}(\Pi_{\nu_1}(x))$$

   $$E[(X - \Pi_{\nu_2}(x))^2] = E[(X - \Pi_{\nu_1}(x))^2] + E[(\Pi_{\nu_1}(x) - \Pi_{\nu_1}(x))^2]$$

   and

   $$E[(X - \Pi_{\nu_2}(x))^2] > E[(X - \Pi_{\nu_1}(x))^2]$$

3. Both $\nu_1, \nu_2$ are closed, $\nu_1 \perp \nu_2$ 

   $$\forall \nu_1 \in \nu_1, \forall \nu_2 \in \nu_2, \quad E[\nu_1\nu_2] = 0$$

Let $\nu = \nu_1 \cup \nu_2 = \{ x_1 + x_2 : x_1 \in \nu_1, \exists x_2 \in \nu_2 \}$

$$\Pi_\nu(x) = \Pi_{\nu_1}(x) + \Pi_{\nu_2}(x)$$

$$E[(X - \Pi_\nu(x))^2] = E[X^2] - E[(\Pi_{\nu_1}(x))^2] - E[(\Pi_{\nu_2}(x))^2]$$

The more important case: Estimating $X$ based on observations of another variable $Y$. Estimator = function $f(Y) = f(Y)$

- All functions $f$
- Linear functions $f$

Let $Y = f(g(Y)) : g: \mathbb{R}^m \to \mathbb{R}^n, \quad E[\hat{g}(Y)\hat{g}(Y)] < \infty$

Assume that $X, Y$ have a well-defined joint distribution, and in particular, a joint pdf

$$E[(X - g(Y))^2] = \int_{\mathbb{R}^m} E[(X - g(Y))^2 | Y = y] f_Y(y) dy$$

where

$$E[(X - g(Y))^2 | Y = y] = \int_{\mathbb{R}^m} (X - g(y))^2 f_{X|Y}(x) dx$$

Optimal MSE estimator $E[XY]$

Orthogonality principle $E[(X - E[XY])g(Y)] = 0$
Since \( E[\|X-g(Y)\|^2] = \sum_{i=1}^{m} E[(x_i-g_i(Y))^2] \), MLE estimation analysis carries over from the scalar to the vector case in the obvious way - through componentwise estimation.

\[
S^Y(x) = E\{X|Y\} = \begin{bmatrix}
E(Y_1|x)
\vdots
E(Y_m|x)
\end{bmatrix}
\]

**Linear estimators:**

\[
Y = \hat{Y} = a_0 + c_1Y_1 + c_2Y_2 + \cdots + c_mY_m \quad a_0, c_i \in \mathbb{R}
\]

\[
x = AY + b \quad e_i = x_i - A_i Y + b_i
\]

\[
E[\|x\|^2] < \infty \quad \Downarrow
\]

\[
E[\|x\|^2] < \infty
\]

1) \( e_i \perp \Delta \)
2) \( e_i \perp A_i \quad \forall i, \Delta \)

i) \( E[e_i] = 0 \Rightarrow E[x] = A E[Y] + b \Rightarrow b = E[x] - A E[Y] \)

ii) \( E[e_i Y_i^\Delta] = 0 \) which is equivalent to \( \text{cov} (e_i, Y_i^\Delta) = 0 \), since \( E[e_i Y_i^\Delta] = 0 \)

\[
\text{cov} (e, Y) = \text{cov} (x_i - A Y - b, Y) = \text{cov} (x_i, Y) - A \text{cov} (Y, Y) - 0 = 0
\]

\[
\Rightarrow \text{cov} (x_i, Y) = A \text{cov} (Y, Y)
\]

\[
A = \text{cov} (x_i, Y) \text{cov}^{-1} (Y, Y)
\]

Hence, the optimal estimator from the linear class is

\[
E [x|Y] = E [x] + A (Y - E[Y])
\]

\[
= E [x] + \text{cov}(x, Y) \text{cov}^{-1} (Y, Y) (Y - E[Y]) = 0
\]

\[
\text{cov} (e) = \text{cov} (x, x) = \text{cov} (x, x - g^Y(y)) = \text{cov} (x, x) - \text{cov} (x, g^Y(y))
\]

\[
= \text{cov} (x - E[X] - \text{cov}(x, y) \text{cov}^{-1} (Y, Y) (y - E[Y]), x) = \text{cov} (x) - \text{cov}(x, y) \text{cov}(y, y)^{-1} \text{cov}(y, x)
\]

**Estimation error:**

\[
\text{cov} (e) = \text{cov}(x) - \text{cov}(x, y) \text{cov}^{-1} (y, y) \text{cov}(y, x)
\]

**Examples**

Let \( Y = X + Z \) where the signal \( X \sim U[-1,1] \) and noise \( Z \sim N(0,1) \) are independent.

Find the MNSE for \( \text{sign}(X) \) (the sign function)

\[
g(y) = E [\text{sign}(X) | Y = y] = \int \text{sgn}(x) f_{X|Y}(x|y) dx
\]
\[
\begin{align*}
\frac{f_{X|Y}(x|y)}{f_Y(y)} &= \frac{f_{X,Y}(y|x)f_X(x)}{f_Y(y)} \\
&= \begin{cases} 
\frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{(x-y)^2}{2}} & \text{if } x \in [-1, 1] \\
0 & \text{otherwise}
\end{cases}
\end{align*}
\]

Since \(X, Y\) are independent \(Y|X = x_i \sim N(x_i, 1)\)

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

\[
= \frac{1}{\sqrt{2}} \left( Q(y-1) - Q(y+1) \right)
\]

Hence, \(g(y) = \frac{Q(y+1) - 2Q(y) + Q(y-1)}{Q(y-1) - Q(y+1)}\)

\[
\phi_X(u) = e^{\frac{-u^2}{2}}
\]

Claim: \(X_1, \ldots, X_n\) are independent Gaussian RVs, then any \(\sum a_i X_i\) is Gaussian.

Proof: suffices to show that claim is true for \(n = 2\), \(a_1 = a_2 = 1\)

\[
\begin{align*}
X &= X_1 + X_2 \\
X_1 &\sim N(\mu_1, \sigma_1^2) \\
X_2 &\sim N(\mu_2, \sigma_2^2)
\end{align*}
\]

\[
\phi_X(u) = E[e^{-iuX}] = E[e^{-iuX_1}]E[e^{-iuX_2}] = e^{-\frac{u^2\sigma_1^2}{2}} e^{-\frac{u^2\sigma_2^2}{2}} e^{-u^2\mu_1 u} e^{-u^2\mu_2 u}
\]

**Definition:** \((X_i : i \in I)\) is jointly Gaussian if every finite linear combination of variables in the set is Gaussian.

\(X \sim N(\mu, K)\) mean vector \(\mu\), covariance \(K\)

**Theorem:**
1) \((X_i : i \in I)\) jointly Gaussian, then each \(X_i\) is Gaussian
2) If \((X_i : i \in I)\) are Gaussian, and independent (for any finite collection), then \((X_i : i \in I)\) are jointly Gaussian.
3) \((X_i : i \in I)\) jointly Gaussian,

\(Y_i\)'s finite linear combinations of \(X_i\)'s; \(Z_k\)'s limits of sequence \(m(y_{i,j})\)

\((Y_i : i \in I)\), \((Z_k : k \in K)\) are each jointly Gaussian RVs.
Proposition 2.8 from the lecture notes

\( X_n \in \mathcal{N}, \forall n \)

\( X_n \Rightarrow X \sim (\text{a.s., w.s., p, d}) \Rightarrow X \sim \text{Gaussian} \)

4) Characteristic function of \( X \in \mathcal{N}_v(\mu, K) \), \( \text{dom}(X) = \mathbb{R} \)

\( \phi_X(u) = e^{i \mu^T u - \frac{1}{2} u^T K u} \)

5) If \( K \) is diagonal, then components of \( X \) are independent.

\( X, Y \) are jointly Gaussian vectors; they are independent if \( \text{Cov}(X,Y) = 0 \)

6) \( f_X(x) = \frac{1}{(2\pi)^{n/2} |K|^{1/2}} \exp \left( -\frac{(x - \mu)^T K^{-1} (x - \mu)}{2} \right) \)

Problem: Suppose that \( X, Y \) are jointly Gaussian, with parameters \( E[X] = \mu_X, E[Y] = \mu_Y, \text{Var}(X) = \sigma_X^2, \text{Var}(Y) = \sigma_Y^2, E[XY] = \sigma_{XY} \). Are the following claims true or false?

1. \( X, Y \) are Gaussian.

2. \( X \in \mathcal{N}(\mu_X, \sigma_X^2), Y \in \mathcal{N}(\mu_Y, \sigma_Y^2) \)

3. \( Z = ax + by, a, b \in \mathbb{R} \), is Gaussian with mean \( \alpha \mu_X + b \mu_Y \).

4. \( Z \sim X, Y \) independent.

5. \( Z = ax + by, a, b \in \mathbb{R} \), is Gaussian with variance \( \alpha^2 \sigma_X^2 + b^2 \sigma_Y^2 \).

\( \text{Cov}(Z, Z') = \text{Cov}(ax + by, ax + by) = a^2 \text{Var}(X) + 2ab \text{Cov}(X, Y) + b^2 \text{Var}(Y) \)

6. The linear MMSE of \( X \) given \( Y \) is

\( X_L = \mu_X + \frac{\sigma_{XY}}{\sigma_Y^2} (Y - \mu_Y) \)

Linear MMSE = unrestricted MMSE

\( E[X|Y] = E[X] + \frac{\text{Cov}(X,Y)}{\text{Var}(Y)} (Y - E[Y]) \)

\( = \mu_X + \frac{\text{Cov}(X,Y)}{\sigma_Y^2} (Y - \mu_Y) = \mu_X + \frac{\sigma_{XY}}{\sigma_Y^2} (Y - \mu_Y) \)

Recall that \( \frac{\sigma_{XY}}{\sigma_X \sigma_Y} = \frac{\text{Cov}(X,Y)}{\sigma_X \sigma_Y} \)

7. The conditional distribution of \( Y|X \) is Gaussian.
Example (Extra)

Let $y$ be an unobservable RV with $E[y] = 0$, $\text{var}(x) = 4$. We observe

\[
\begin{align*}
Y_1 &= X + W_1, & E[W_1] = E[W_2] = 0 \\
Y_2 &= X + W_2, & \text{var}(W_1) = 4, \text{var}(W_2) = 4
\end{align*}
\]

$W_1, W_2, X$ are independent.

Find the linear MISE estimator of $x$ given $Y_1, Y_2$

$$\hat{x}_L = aY_1 + bY_2 + c$$

Orthogonality principle

1) $E[(X - aY_1 - bY_2 - c)] = 0 \quad \Rightarrow \quad c = 0$

2) $E[(X - aY_1 - bY_2)Y_1] = 0, \ E[(X - aY_1 - bY_2)Y_2] = 0$

\[
\begin{align*}
E[Y_1] - aE[Y_1^2] - bE[Y_2]E[Y_1] = 0
\end{align*}
\]

\[
\begin{align*}
E[X(X + w_1)] - aE[(X + w_1)^2] - bE[(X + w_1)(X + w_2)] = 0
\end{align*}
\]

\[
\begin{align*}
E[X^2] + E[Y_1Y_2] - a(E[X^2] + E[w_1^2]) - b(E[X^2] + E[w_2^2]) = 0
\end{align*}
\]

$$4 - 5a - 4b = 0 \quad \quad 5a + 4b = 4$$

Similarly, ...

$$4a + 8b = 4$$

\[
\begin{align*}
\Rightarrow \quad a = \frac{2}{3}, \quad b = \frac{1}{6}
\end{align*}
\]

$$\hat{x}_L = \frac{2}{3}Y_1 + \frac{1}{6}Y_2$$

Example Suppose $x, y$ are jointly Gaussian, zero mean, $X = \begin{pmatrix} 4 \\ 3 \end{pmatrix}$

$E[x^2 | y] = ?$

Observe the following: $X, Y$ jointly Gaussian vectors, $y \in \mathbb{R}$

$X | Y = y \sim N(X_L(y); \text{Cov}(e))$

and $X_L(y) = E[X | Y = y]$

If $\text{Cov}(y)$ is non-singular

$$E[X | Y = y] = X_L(y) = E[X] + \text{Cov}(X, Y) \text{Cov}(Y)^{-1} (y - E[Y])$$

$\text{Cov}(e) = \text{Cov}(X) - \text{Cov}(X, Y) \text{Cov}(Y)^{-1} \text{Cov}(Y, X)$

and if $\text{Cov}(e)$ is non-singular

$$f_{X | Y = y}(x | y) = \frac{1}{(2\pi)^{w/2} |\text{Cov}(e)|^{1/2}} \exp\left(-\frac{1}{2} (x - X_L(y))^T \text{Cov}^{-1}(e) (x - X_L(y)) \right)$$
Proof Consider \( X_e(Y) = \text{linear function of } Y \)

\[ e = X - X_e(Y) = \text{linear function of } X_1, Y \]

Orthogonality \( E[e] = 0 \)

\( \text{principle } \text{Cov}(e, Y) = 0 \implies Y, e \text{ are jointly Gaussian, independent} \)

Next, rewrite \( X = e + X_e(Y) \)

given \( Y = y \), \( e | Y = y \sim N(0, \text{Cov}(e)) \)

Hence \( X | Y = y \sim N(X_e(y), \text{Cov}(e)) \)

Also, \( E[X | Y] = X_e(Y) \)

Example solution \( X, Y \) are jointly Gaussian, zero mean, \( K = \begin{pmatrix} 4 & 3 \\ 3 & 3 \end{pmatrix} \)

Given \( Y = y \), the conditional distribution of \( X \) is

\[ N \left( \frac{\text{Cov}(X, Y)}{\text{Var}(Y)} y, \text{Cov}(X) - \frac{\text{Cov}^2(X, Y)}{\text{Var}(Y)} \right) \]

\[ N \left( \frac{3y}{2}, 3 \right) \]

\[ E[X^2 | Y = y] = \left( \frac{3y}{2} \right)^2 + 3 \]
Suppose $X$ is a RV with $E[X^4] = 30$.
- Derive an upper bound on $P(|X| > 10)$. 
- Find a distribution for $X$ s.t. that the upper bound holds with equality.

\[
P(|X| > 10) \leq \frac{E[f(X)]}{f(10)}
\]

Take $f(X) = X^4$.

\[
P(|X| > 10) \leq \frac{E[X^4]}{10^4} = \frac{30}{10^4} = 0.003
\]

\[
P(X = 10) = 0.003
\]

\[
P(X = 0) = 1 - 0.003
\]

Let $U_1, U_2, \ldots$ be a sequence of independent random variables, each uniformly distributed in $[0,1]$.

For what values of $c > 0$ does there exist a $b > 0$ (depending on $c$) s.t.

\[
P(U_1 + \ldots + U_n > c n < e^{-bn} \quad \forall n \geq 1
\]

\[
E[U_1] = \frac{1}{2} \quad E[U_1 + \ldots + U_n] = \frac{n}{2}
\]

Want $c > 1/2$.

If $c = 1/2$, prob is exactly $0.5$.

Define $X_i = U_i - c U_{n+1}$.

$X_i$'s are iid, $E[X_i] = \frac{1}{2} - c \frac{1}{2} = \frac{1-c}{2}$.

\[
P(X_1 + \ldots + X_n \geq 0) \leq e^{-bn}
\]

\[
\Rightarrow \quad 0 > \frac{c-1}{2} \quad \Rightarrow \quad c < 1
\]