Mean and variance of sum of random variables

Let $X$ be a random vector and $Y$ be the sum of $X_1, X_2, \ldots, X_n$, that is:

$$Y = \mathbf{1}^T X$$

Then by linearity of expectation:

$$E(Y) = E(\mathbf{1}^T X) = \mathbf{1}^T E(X) = \sum_{i=1}^{n} E(X_i)$$

Example

Let $X_1, X_2, \ldots, X_n$ be i.i.d. Bern $(p)$.

and: $Y = \sum_{i=1}^{n} X_i \sim \text{Binom}(n, p)$

Then: $E(Y) = \sum_{i=1}^{n} E(X_i) = np$

Since: $Y = \mathbf{1}^T X$

$$\text{var}(Y) = E \left( (Y - E(Y))^2 \right) = E \left( (\mathbf{1}^T(X - E(X)))^2 \right)$$

$$= E \left( \mathbf{1}^T(X - E(X))(X - E(X))^T \mathbf{1} \right)$$

$$= \mathbf{1}^T E \left( (X - E(X))(X - E(X))^T \right) \mathbf{1}$$

$$= \mathbf{1}^T K_{XX} \mathbf{1}$$
In particular, if: \( X_1, X_2, \ldots, X_n \) are uncorrelated, that is: \( K_X \) is diagonal, then:
\[
\text{var}(Y) = \sum_{i=1}^{n} \text{var}(X_i)
\]

Also:
\[
\text{var}(Y) = \sum_{i=1}^{n} \text{var}(X_i) = np(1-p).
\]

**Gaussian random vectors**

An \( n \)-dimensional random vector \( X \) is called Gaussian if its joint pdf is of the form:
\[
f_X(x) = \frac{1}{(2\pi)^{n/2} |K|^{1/2}} e^{-\frac{(x-\mu)^T K^{-1} (x-\mu)}{2}}
\]

where \( \mu \) is the mean and \( K \) is the invertible covariance matrix, i.e. \( K \) is symmetric and positive-definite.

\( K \) is positive semi-definite if for every \( a \):
\[
a^T K a \geq 0, \quad \underline{K \geq 0}
\]

\( K \) is positive definite if for every \( a \neq 0 \):
\[
a^T K a > 0, \quad \underline{K > 0}
\]

**Alternative definition**

We say \( X \) is Gaussian if: for every vector \( a \), \( a^T X \) is Gaussian.

**Properties**

(1) If \( X \) is Gaussian, then: uncorrelation implies independence.

**Proof** If \( K \) is diagonal, then \( f_X(x) \) factorizes into \( \prod_{i=1}^{n} f_{X_i}(x_i) \). (Check!)

(2) Any linear transformation of \( X \) is also Gaussian.

For example: if \( Y = AX \) then \( Y \sim N(AY, AKX A^T) \).

**Proof** Note that for any \( X \) with mean \( \mu_x \) and covariance matrix \( K_X \):
\[
Y = AX \text{ has mean } A\mu_x \text{ and covariance matrix } AKX A^T.
\]

To show that \( Y \) is Gaussian, we can use the characteristic function (full proof later).

(3) Marginals of \( X \) are Gaussian. For example, if \( X = (X_1, X_2, X_3)^T \) is Gaussian, then \( (X_1, X_3)^T \) is Gaussian and so is \( X_2 \).
Proof: Consider: \[
\begin{pmatrix}
X_1 \\
X_2 \\
X_3
\end{pmatrix} = \begin{pmatrix}
1 & 0 & 0 \\
0 & 0 & 1
\end{pmatrix} \begin{pmatrix}
X'_1 \\
X'_2 \\
X'_3
\end{pmatrix}
\]
and: \[
X_2 = \begin{pmatrix}
0 & 1 & 0
\end{pmatrix} \begin{pmatrix}
X'_1 \\
X'_2 \\
X'_3
\end{pmatrix}.\] Then use property 2.

(4) Conditionals of \(X\) are Gaussian.
For example, if \[
\begin{pmatrix}
X_1 \\
X_2
\end{pmatrix} \sim N \left( \begin{pmatrix}
\mu_1 \\
\mu_2
\end{pmatrix}, \begin{pmatrix}
K_{11} & K_{12} \\
K_{21} & K_{22}
\end{pmatrix} \right)
\]
Then: \[
X_2 \mid X_1 = X'_1 \sim N \left( \frac{K_{21}K_{11}^{-1}(X_1 - \mu_1) + \mu_2}{K_2 - K_{21}K_{11}^{-1}K_{12}}, \frac{K_{22} - K_{21}K_{11}^{-1}K_{12}}{K_2} \right)
\]
In particular, the MMSE estimate of \(X_2\) given \(X_1\) is:
\[
\frac{K_{21}K_{11}^{-1}(X_1 - \mu_1) + \mu_2}{K_2 - K_{21}K_{11}^{-1}K_{12}}
\]
and the corresponding MSE is \(K_{22} - K_{21}K_{11}^{-1}K_{12}\).

**Linear estimation**

\[
\begin{array}{c}
\xrightarrow{X} \\
f_X(x)
\end{array} \quad \begin{array}{c}
\xrightarrow{Y}
\end{array} \quad \begin{array}{c}
g(y)
\end{array} \quad \begin{array}{c}
\xrightarrow{X}
\end{array}
\]

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

**Answer:** \(g^*(y) = E[X | Y = y]\)

Find \(g^*(y)\) of the form: \(g^*(y) = h^T y + h_0\) that minimizes the:

\[
MSE = E\left[\frac{(X - h^T y - h_0)^2}{Z}\right]
\]

First, note that the best \(h_0\) should be:

\(h_0 = E[X - h^T Y] = E[X] - h^T E[Y]\)

Hence, the problem is equivalent to minimizing:

\[
E\left[\frac{(X - h^T Y - h_0)^2}{Z}\right]
\]

**Note:** \(E[X'] = 0\) and \(E[Y'] = 0\).
Therefore, it suffices to consider the case both \(X\) and \(Y\) have zero mean.
Hence, the problem (under the zero-mean condition) is to find \(h\) that minimizes:
\[ E[(X-b^TY)^2] = E[X^2 - 2b^TYX + b^TYX^Tb] = E[X^2] - 2b^TE[XY] + b^T\text{Var}(Y)b = J(b) \]

\[
\frac{dJ(h)}{dh} = -2KyX + 2Kyh = 0 \Rightarrow KyX = Kyh
\]

(If Ky is invertible, then \( h = K_y^{-1}KyX \))

**Theorem** The best linear estimator of \( X \) given \( Y \) is: (for zero-mean case)

\[ \hat{X} = K_y^{-1}KyX \]

and the corresponding MSE is: \( \text{var}(X) - Ky_X^TK_y^{-1}Ky_X \).

More generally, if \( \mu_X \) and \( \mu_Y \) are nonzero, then:

\[ \hat{X} = K_y^{-1}KyX(Y - \mu_Y) + \mu_X \]

**Examples:**

1. If \( Y \) is scalar, then:

\[ \hat{X} = \frac{\text{cov}(X,Y)}{\text{var}(Y)}(Y - E(Y)) + E[X] \]

and \( \text{MSE} = \text{var}(X) - \left( \frac{\text{cov}(X,Y)}{\text{var}(Y)} \right)^2 \)

2. Let \( X \) be a random variable with mean \( \mu \) and variance \( \sigma^2 \). Let \( Y_i = X + Z_i, \ i = 1, 2, ..., n \), where \( Z_i \) has zero mean and variance \( \sigma^2 \), and \( X, Z_1, Z_2, ..., Z_n \) are uncorrelated.

Find the best linear estimate of \( X \) given \( Y_1, ..., Y_n \) and the corresponding MSE.

**Note:**

\[ E(Y_i) = E(X) + E(Z_i) = \mu \]

and:

\[ \text{cov}(Y_i, Y_j) = E[(Y_i - \mu)(Y_j - \mu)] = E[(X - \mu + Z_i)(X - \mu + Z_j)] \]

\[ = \sigma^2 + \text{cov}(Z_i, Z_j) \]

\[ = \begin{cases} 
\sigma^2 & \text{if } i = j \\
0 & \text{if } i \neq j 
\end{cases} \]

\[ K_y = \begin{pmatrix}
p + N & p & p & \ldots & p \\
p & p + N & p & \ldots & p \\
p & p & p + N & \ldots & p \\
p & \ldots & \ldots & \ldots & p \\
p & \ldots & \ldots & p & p + N 
\end{pmatrix} \]
Similarly: \( K_{yx} = \begin{pmatrix} p \\ \vdots \\ p \end{pmatrix} \), hence: \( h^* = K_y^{-1} K_{yx} \)

To solve this, recall that \( K_y h^* = K_{yx} \) and by symmetry, we can guess \( h^* = \begin{pmatrix} a \\ \vdots \\ a \end{pmatrix} \).

\[
\begin{pmatrix} p+np \ 
\vdots 
\vdots 
\vdots 
p \end{pmatrix} \begin{pmatrix} a \\ \vdots \\ a \end{pmatrix} = \begin{pmatrix} p \\ \vdots \\ p \end{pmatrix}
\]

\[(np+n)a = p \Rightarrow a = \frac{p}{np+n} \]

and \( h^* = a \cdot 1 \)

Therefore, the best linear estimate is:

\[ \hat{\mu} = \frac{p}{np+n} \mathbf{1}^T (Y - \mu) + \mu \]

\[
\text{MSE} = p - K_{yx}^T K_y^{-1} K_{yx} = p - \frac{np^2}{np+n} = \frac{np}{np+n}
\]