Mean and variance of sum of random variables

Let $X$ be a random vector and $Y$ be the sum of $X_1, \ldots, X_n$, that is $Y = \mathbf{1}^T X$.
Then by the 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. Bernoulli($p$) and $Y = \sum_{i=1}^n X_i \sim \text{Binom}(n, p)$.
Then $E(Y) = \sum_{i=1}^n E(X_i) = np$. Also $\text{Var}(Y) = \sum_{i=1}^n \text{Var}(X_i) = np(1-p)$.

Since $Y = \mathbf{1}^T X$,
$$\text{Var}(Y) = E[(Y - E(Y))^2] = E[(\mathbf{1}^T X - \mathbf{1}^T E(X))^2]$$
$$= E[\mathbf{1}^T (X - E(X))(X - E(X))^T \mathbf{1}]$$
$$= \mathbf{1}^T E[(X - E(X))(X - E(X))^T] \mathbf{1}$$
$$= \mathbf{1}^T K \mathbf{1}$$.

In particular, if $X_1, \ldots, X_n$ are uncorrelated, that is, $K_X$ is diagonal, then $\text{Var}(Y) = \sum_{i=1}^n \text{Var}(X_i)$.

Gaussian random vectors

A random vector $X$ is called Gaussian if its joint pdf is of the form
$$f_X(x) = \frac{1}{(2\pi)^{n/2} |K|^{1/2}} e^{-\frac{1}{2} (x - \mu)^T K^{-1} (x - \mu)}$$
where $\mu$ is the mean and $K$ is the invertible covariance matrix, i.e. $K$ is symmetric and positive definite.

Def $K$ is positive semidefinite if for any $a$, $a^T K a \geq 0$. (denoted as $K \succeq 0$)

$K$ is positive definite if for any $a \neq 0$, $a^T K a > 0$. (denoted as $K \succ 0$).

Alternative definition
We say $X$ is Gaussian if for any vector $a$, $a^T X$ is Gaussian.
(This definition covers degenerate cases e.g. $Z = (X, Y)$, $K = \left( \begin{array}{cc} 1 & \mu \\ \mu & K_X \end{array} \right)$ if $X \sim N(0, 1)$).
Properties
(1) If $X$ is Gaussian, then uncorrelation implies independence.
Proof: If $K$ is diagonal, then $f_X(x) = \prod_{i=1}^{n} f_X(x_i)$. (Check!)
(2) Any linear transformation of $X$ is also Gaussian. For example, if $Y = AX$, then $Y \sim N(A\mu_x, AK\mu A^T)$
Proof: Note that for any $X$ with mean $\mu_x$ and covariance matrix $K_x$
$Y = AX$ has mean $A\mu_x$ and covariance matrix $AKxA^T$.
To show that $Y$ is Gaussian, we can use the characteristic function.
(3) Marginal of $X$ are Gaussian. For example,
If $X = (X_1, X_2, X_3)$ is Gaussian, then $(X_1, X_3)^T$ is Gaussian and so is $X_2$.
Proof: Any projection is linear, so Property (2) implies the result.
For example, $(y_1) = (1, 0, 0) \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}$ and $x_2 = (0, 1, 0) \begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix}
(4) Conditionals of $X$ are Gaussian.
For example, if $x = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$ $\sim N \left( \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \begin{pmatrix} K_{11} & K_{12} \\ K_{21} & K_{22} \end{pmatrix} \right)$
then $X_1 | X = x_1 \sim N(\mu_1 + K_{12}(x_1 - \mu_2), K_{22} - K_{21}K_{12}^{-1}K_{12})$
In particular, the MMSE estimate of $X_2$ given $X_1$ is $K_{21}K_{12}^{-1}(x_1 - \mu_2)$
and the corresponding MSE is $K_{22} - K_{21}K_{12}^{-1}K_{12}$.

Linear estimation
\[ X \xrightarrow{\text{noisy channel}} Y \xrightarrow{g(Y)} X \]

- Find $g(x)$ that minimizes $\text{MSE} = \mathbb{E}[(X-g(Y))^2]$.
  Ans: $g(x) = \mathbb{E}[X|Y=y]$; hard to compute!

- Find $g(y)$ of the form $g(y) = h^T Y + h_0$ that minimizes $\mathbb{E}[(X-h^T Y + h_0)^2] = \mathbb{E}[(X-h^T Y)^2]$
First, note that the best $h_0$ should be $h_0 = \mathbb{E}[X - h^T Y] = \mathbb{E}[X] - h^T \mathbb{E}[Y]$.
Hence, the problem is equivalent to minimizing $\mathbb{E}[(X-EX) - h^T(Y-EY)]^2$.
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

\[
\]

\[
= J(h)
\]

Only the 2nd moments are required to find a linear estimate.

\[
\frac{dJ(h)}{dh} = 2k_{yx} + 2k_{yx} = 0 \quad \text{(heuristically)}
\]

\[\Rightarrow k_{yx} = k_{yx} h, \text{ if } k_{yx} \text{ is invertible, then } h = k_{yx}^{-1} k_{yx}\]

**Theorem:** The best linear estimator of \( X \) given \( Y \) is

\[
\hat{X} = k_{yx} k_{xy}^{-1} Y \quad \text{(for zero-mean case)}
\]

and the corresponding MSE is

\[
\text{Var}(\hat{X}) = k_{yx}^{-1} k_{yx}.
\]

More generally, if \( \mu_x \) and \( \mu_y \) are nonzero, then

\[
\hat{X} = \mu_x k_{yx}^{-1} (Y - EY).
\]

**Examples**

1. If \( Y \) is scalar, then \( \hat{X} = \frac{\text{Cov}(X,Y)}{\text{Var}(Y)} (Y - EY) + E(X) \)

   and MSE = \( \text{Var}(\hat{X}) = \frac{\text{Cov}(X,Y)^2}{\text{Var}(Y)} \)

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

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

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

   \[
   \text{Cov}(Y_i, Y_j) = E[(Y_i - \mu)(Y_j - \mu)] = E[(X_i - \mu + Z_i)(X_j - \mu + Z_j)] = P + \sigma_i \sigma_j = \sigma_i^2
   \]

   So \( k_{yx} = \begin{pmatrix}
   (1, 1, \ldots, 1)
   
   \end{pmatrix}
   \]

   Similarly, \( k_{xy} = \begin{pmatrix}
   (1, 1, \ldots, 1)
   
   \end{pmatrix}
   \]

   Hence, \( \hat{h} = k_{yx \sim} k_{xy} \). To solve this, recall that \( k_{yx} = k_{xy} \) and by symmetry we can guess \( \hat{h} = (a^T, \ldots, a^T) \), so \( nP + N \text{, } a = \frac{p}{np + N} \), and \( \hat{h} = a \). Therefore, the best linear estimate is

   \[
   \hat{X} = \frac{p}{np + N} (Y - H\mu) + \mu \quad \text{and MSE} = P - k_{yx} k_{xy} = \frac{np}{n} \]

   \[
   \]