Review.

1. Minimum mean square error estimator of random variable $Y$, after observing random variable $X$.
   - Constant estimator: $\hat{Y} = E[Y]$, error $= \sigma^2_Y$.
   - Linear estimator: $L(X) = E[Y|X] = \mu_Y + \sigma_Y \lambda X$, error $= \sigma^2 (1 - P_{XY})$.
   - Unconstrained estimator: $\hat{Y}(X) = E[Y|X]$, error $= \sigma^2_Y E[(E[Y|X])^2]$.

2. Law of large numbers
   Suppose that $X_1, X_2, \ldots, X_n$ are uncorrelated, each one has the same mean $\mu$, and their variance is bounded by $C$. Let $S_n = \frac{1}{n} \sum X_i$. Then, for any $s \geq 0$, we have
   $$P\left( \left\{ \frac{S_n - \mu}{\sqrt{\text{Var}(S_n)}} \geq s \right\}\right) = \frac{C}{\sqrt{\text{Var}(S_n)}}$$

3. Central limit theorem
   Suppose that $X_1, X_2, \ldots, X_n$ are independent, identically distributed, each with mean $\mu$ and variance $\sigma^2$. Let $S_n = \frac{1}{n} \sum X_i$. Then, for any $c$, we have
   $$\lim_{n \to \infty} P\left( \left\{ \frac{S_n - \mu}{\sqrt{\text{Var}(S_n)}} \leq c \right\}\right) = \Phi(c)$$
   where $\tilde{S}_n = \frac{S_n - \mu}{\sqrt{\text{Var}(S_n)}}$ is the standardized version of $S_n$.

4. Jointly Gaussian random variables
   We say random variables $X$ and $Y$ are jointly Gaussian random variables if any linear combination of $X$ and $Y$ is a Gaussian random variable.

Proposition: Suppose that $X$ and $Y$ have bivariate normal pdf with parameters $\mu_X, \mu_Y, \sigma_X, \sigma_Y$ and $\rho$.

$$f_{X,Y}(x,y) = \frac{1}{2\pi \sigma_X \sigma_Y \sqrt{1 - \rho^2}} \exp \left( \frac{-\left( \frac{x - \mu_X}{\sigma_X} \right)^2 + \left( \frac{y - \mu_Y}{\sigma_Y} \right)^2 - 2\rho \left( \frac{x - \mu_X}{\sigma_X} \right) \left( \frac{y - \mu_Y}{\sigma_Y} \right)}{2(1 - \rho^2)} \right)$$

(a) $X$ has $N(\mu_X, \sigma_X^2)$ distribution, $Y$ has $N(\mu_Y, \sigma_Y^2)$ distribution.
(b) For any $a, b \in \mathbb{R}$, $ax + by$ is a Gaussian random variable, i.e., $X$ and $Y$ are jointly Gaussian.
(a) \(X\) has \(N(\mu_X, \sigma_X^2)\) distribution, \(Y\) has \(N(\mu_Y, \sigma_Y^2)\) distribution.

(b) for any \(a, b \in \mathbb{R}\), \(ax + by\) is a Gaussian random variable, i.e., \(X\) and \(Y\) are jointly Gaussian.

(c) \(p_{XY} = \rho\)

(d) if \(\rho = 0\), then they are independent.

(e) For estimation of \(Y\) from \(X\), \(\hat{X}^*(X) = g^*(X)\). Equivalently, \(E(Y|X) = \hat{E}(Y|X)\), i.e., the best linear estimator of \(Y\) using \(X\), is the best unconstrained estimator.

(f) The conditional distribution of \(Y\) given \(X = u\) is \(N(\hat{E}(Y|X = u), \sigma_Y^2)\) where \(\sigma_Y^2\) is MSE for \(\hat{E}(Y|X)\).

Proof: We will prove it for \(\mu_X, \mu_Y = 0\) and \(\sigma_X, \sigma_Y = 1\). In the general case, follows by a linear transformation.

\[
\begin{align*}
\hat{f}_{Y|X}(u|v) &= \int_{-\infty}^{\infty} f_{X,Y}(u,v) \, dv \\
&= \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{u^2 + v^2}{2}\right) \cdot \frac{1}{\sqrt{2\pi (1-\rho^2)}} \exp\left(\frac{(v - \rho u)^2}{2(1-\rho^2)}\right) \,
\end{align*}
\]

For fixed \(u\), this is pdf of Gaussian random variable with distribution \(N(\rho u, 1-\rho^2)\).

\[
\begin{align*}
f_X(u) &= \int_{-\infty}^{\infty} f_{X,Y}(u,v) \, dv \\
&= \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{u^2}{2}\right)
\end{align*}
\]

Similarly, \(f_Y(v) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{v^2}{2}\right)\).

Let \(A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}\) for \(c, d\) s.t. \(ad - bc \neq 0\). Define \(\begin{pmatrix} W \\ Z \end{pmatrix} = A \begin{pmatrix} X \\ Y \end{pmatrix}\). Recall that class of bivariate normal pdfs are preserved via linear transformation. Hence \(\begin{pmatrix} W \\ Z \end{pmatrix}\) has bivariate normal pdf and by part (a), \(W = aX + bY\) has normal distribution.

(c) & (d) Notice that \(f_{X,Y}(u,v) = f_X(u) f_Y(v)\). Hence, by part (a),

\[
\begin{align*}
f_{Y|X}(v|u) &= \frac{1}{\sqrt{2\pi (1-\rho^2)}} \exp\left(\frac{(v - \rho u)^2}{2(1-\rho^2)}\right)
\end{align*}
\]

In particular, \(E(Y|X = u) = \rho u\). Hence, \(g^*(X) = \rho X\). Moreover,

\[
\begin{align*}
p_{XY} = E[XY] &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} uv \, f_{X,Y}(u,v) \, du \, dv \\
&= \int_{-\infty}^{\infty} u \, f_X(u) \int_{-\infty}^{\infty} v \, f_{Y|X}(v|u) \, dv \, du \\
&= \int_{-\infty}^{\infty} u \, f_X(u) \cdot \rho u \, du = \rho
\end{align*}
\]
4.37. [Transforming joint Gaussians to independent random variables]
Suppose $X$ and $Y$ are jointly Gaussian such that $X$ is $N(0, 9)$, $Y$ is $N(0, 4)$, and the correlation coefficient is denoted by $\rho$. The solutions to the questions below may depend on $\rho$ and may fail to exist for some values of $\rho$.

(a) For what value(s) of $a$ is $X$ independent of $X + aY$?
(b) For what value(s) of $b$ is $X + Y$ independent of $X - bY$?
(c) For what value(s) of $c$ is $X + cY$ independent of $X - cY$?
(d) For what value(s) of $d$ is $X + dY$ independent of $(X - dY)^3$?

4.30. [Law of Large Numbers and Central Limit Theorem]
A fair die is rolled $n$ times. Let $S_n = X_1 + X_2 + \ldots + X_n$, where $X_i$ is the number showing on the $i^{th}$ roll. Determine a condition on $n$ so the probability the sample average $\frac{S_n}{n}$ is within 1% of the mean $\mu_X$, is greater than 0.95. (Note: This problem is related to Example 4.10.6.)

(a) Solve the problem using the form of the law of large numbers based on the Chebychev inequality (i.e. Proposition 4.10.1 in the notes).

(b) Solve the problem using the Gaussian approximation for $S_n$, which is suggested by the CLT. (Do not use the continuity correction, because, unless $3.5n \pm (0.01)n\mu_X$ are integers, inserting the term 0.5 is not applicable).