Another CLT Example:

Alice & Bob Number Guessing Problem

Alice picks a 100 numbers randomly & independently from $f_X(x)$

$$S = X_1 + \cdots + X_{100} = \sum_{i=1}^{100} X_i$$

Bob must guess the sum $S_{100}$. Bob wins a prize if his guess $\hat{g}$ is within $\pm 2$ of $S_{100}$. 
Bob's guess \( g \) must fall here

\[
3_{100} - 2 \xrightarrow{S_{100}} S_{100} \xrightarrow{g \leq S_{100} + 2} g - 2 \leq S_{100}
\]

\[
S_{100} \leq g \leq S_{100} + 2 \implies S_{100} - 2 \leq g \implies S_{100} \leq g + 2
\]

\[
\implies g - 2 \leq S_{100} \leq g + 2 \text{ to win a prize}
\]

Bob guesses \( g = 55 \).

What's the probability that Bob wins a prize?

\[
\Pr(55 - 2 \leq S_{100} \leq 55 + 2) = ?
\]

---

**CLT Restatement**

\( X_1, \ldots, X_n \) IID

\[
E(X_i) = \mu \\
\sigma_{X_i}^2 = \sigma^2
\]

\[
M_n = \frac{X_1 + \cdots + X_n}{n} = \frac{S_n}{n}
\]

\[
S_n = X_1 + \cdots + X_n \\
E(M_n) = \mu , \quad \sigma_{M_n}^2 = \frac{\sigma^2}{n} , \quad \sigma_{M_n} = \frac{\sigma}{\sqrt{n}}
\]
$$Z_n = \frac{M_n - M}{\sigma_{M_n}} = \frac{\frac{X_1 + \cdots + X_n}{n} - M}{\frac{\sigma}{\sqrt{n}}} = \frac{S_n - M}{\sigma \sqrt{n}}$$

$$Z_n = \frac{S_n - nM}{\sigma \sqrt{n}}$$

This denominator is $\sigma_{S_n}$, the standard deviation of $S_n$

**CLT:** \[ \lim_{n \to \infty} P(Z_n \leq z) = P(Z \leq z) \]

*Standard Gaussian RV*

Back to Alice & Bob

![Diagram showing the central limit theorem and related statistics](image-url)
\[ f_X(x) \]

\[ \mu = \text{E}(X) = \frac{a+b}{2} \]
\[ \sigma_X^2 = \frac{(b-a)^2}{12} \]

\[ E(S_{100}) = 100 \mu = 500 \]

\[ \sigma_{S_{100}}^2 = \text{Var}(S_{100}) = \frac{100}{12} = \frac{25}{3} \]

\[ \sigma_{S_{100}} = \frac{10}{\sqrt{3}} = \frac{5}{\sqrt{3}} \]

\[ P(53 \leq S_{100} \leq 57) = ? \]
Standardize:

\[
\Pr\left(53 \leq S_{100} \leq 57\right) = 
\Pr\left(\frac{53-50}{\frac{5}{\sqrt{13}}} \leq \frac{S_{100} - 50}{\frac{5}{\sqrt{13}}} \leq \frac{57-50}{\frac{5}{\sqrt{13}}}\right) \sim Z_{100}
\]

\[
\Pr\left(\frac{3\sqrt{3}}{5} \leq Z \leq \frac{7\sqrt{3}}{5}\right)
\]
\[ \Pr \left( 1.039 \leq Z \leq 2.425 \right) = \Phi(2.425) - \Phi(1.039) \]
\[ = 0.9924 - 0.8508 = 0.142 \]

From the Gaussian Table, this corresponds to a 14.2% chance.

Compare with his likelihood of winning if he had guessed so! 

\[ \Pr \left( 48 \leq S_{100} \leq 52 \right) = ? \]

\[ = \Pr \left( \frac{48-50}{\frac{5}{\sqrt{3}}} \leq \frac{S_{100} - 50}{\frac{5}{\sqrt{3}}} \leq \frac{52-50}{\frac{5}{\sqrt{3}}} \right) = \Pr \left( \frac{-2\sqrt{3}}{5} \leq Z_{100} \leq \frac{2\sqrt{3}}{5} \right) \]

\[ \approx \Pr \left( 0.628 \leq Z \leq 0.6928 \right) = \Pr \left( |Z| \leq 0.6928 \right) = 2\Phi(0.6928) - 1 \]

CLT Approximation
Recall: \( f(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} \)

\[
\Pr(|z| \leq \alpha) = \Phi(\alpha) - \Phi(-\alpha) \\
= \Phi(\alpha) - [1 - \Phi(\alpha)] \\
= 2\Phi(\alpha) - 1
\]

From the table we know
\[
\Phi(0.69) = 0.7549 \\
\Phi(0.70) = 0.7580
\]

We can approximate down
\[
\Phi(0.6928) \approx \Phi(0.69) = 0.7549
\]

So, \( \Pr(\text{Bob wins if he guesses 50}) \)
\[
\approx \Pr(|Z| \leq 0.69) = 2\Phi(0.69) - 1 = 0.51
\] (51%)
Estimation:

RV $Y$ that I want to estimate. No observation
Estimate $Y$ using a fixed number $\hat{j}$

Error: $E = Y - \hat{j}$

Criterion: Minimize the mean of the squared error (MMSE)
Minimize $E(E^2) = E[(Y - \hat{j})^2]$ \[ E(E^2) \]

Let $Z = Y - \hat{j}$

$E(Z) = E(Y) - \hat{j}$

$\sigma_Z^2 = \text{var}(Z) = \sigma_Y^2$
\[ \sigma_Z^2 = E(Z^2) - E^2(Z) \]
\[ E(Z^2) = \sigma_Z^2 + E^2(Z) \]
\[ E[(Y-\hat{Y})^2] = \sigma_Y^2 + [E(Y-\hat{Y})]^2 \]
\[ \text{MSE} \]

Can I make \( E(Y-\hat{Y}) = 0 \)?
\[ E(Y-\hat{Y}) = E(Y) - \hat{Y} = 0 \Rightarrow \hat{Y} = E(Y) \]

Optimal Mean Squared Error:
\[ \text{MSE} = E(\varepsilon^2) = E[(Y - E(Y))^2] = \sigma_Y^2 \]
Mind Blown!

The mean is the optimal estimator, and it results in an $\text{MSE} = \sigma^2$. 

$$\text{MSE} = \sigma^2 + [\text{E}(Y) - \hat{\alpha}]^2$$

$$E[(Y - \text{E}(Y))^2] \leq E[(Y - \hat{\alpha})^2]$$

for all $\hat{\alpha}$. 
What if we have an observation $X$?

Point estimate of $Y$ in the MMSE sense

$b/c \ X=x \ is \ a \ sample \ value \ (point \ value) \ of \ X.$

$\hat{Y} = \mathbb{E}(Y \mid X=x) = g(x)$

$\mathbb{E}[\left( Y - \mathbb{E}(Y \mid X) \right)^2] \leq \mathbb{E}[\left( Y - g(x) \right)^2]$

A fn $g(x)$.
Linear MSE Estimation:

\[ \hat{Y}(X) = aX + b \]

\[ MSE = \frac{1}{n} \sum_{i=1}^{n} (\hat{Y}_i - aX_i - b)^2 \]

Let \( Z = Y - aX \)

\[ MSE = \mathbb{E} [(Z - b)^2] \]

What's the optimal \( b \)?

\[ b = \mathbb{E}(Z) = \mathbb{E}(Y - aX) = \mathbb{E}(Y) - a\mathbb{E}(X) \]

\[ MSE = \mathbb{E} \left[ (Y - aX - \mathbb{E}(Y) + a\mathbb{E}(X))^2 \right] \]

Grouped terms by color:

\[ \sigma^2 \]

\[ \sigma^2 \]
\[
E[(y - E(y))^2] + \alpha^2 E[(x - E(x))^2]
\]

\[-2\alpha E[(x - E(x))(y - E(y))] \]

\(\hat{b}_x y = \text{cov}(x, y)\)

\[
\text{MSE} = \hat{b}_y^2 + \alpha^2 \hat{b}_x^2 - 2\alpha \hat{b}_x y
\]

Must minimize \(\text{MSE}\) with respect to \(\alpha\):

\[
\frac{d\text{MSE}}{d\alpha} = 2\alpha \hat{b}_x^2 - 2 \hat{b}_x y = 0
\]

\[
\frac{d^2\text{MSE}}{d\alpha^2} = 2\hat{b}_x^2 > 0
\]

\(\alpha \hat{b}_x^2 - \hat{b}_x y = 0 \rightarrow \alpha = \frac{\hat{b}_x y}{\hat{b}_x^2}\)

\(b = E(y) - \alpha E(x)\)
\( \hat{\gamma}_L(X) = ax + b = \frac{\hat{\mu}_{xy}}{\hat{\sigma}_x^2} X + E(Y) - \frac{\hat{\mu}_{xy}}{\hat{\sigma}_x^2} E(X) \)

\( \hat{\gamma}_L(X) = E(Y) + \frac{\hat{\mu}_{xy}}{\hat{\sigma}_x^2} (X - E(X)) \)

\( \text{correction term} \)

If \( X, Y \) uncorrelated?

\( \hat{\mu}_{xy} = 0 \quad \hat{\sigma}_{xy} = \text{cov}(X,Y) \)

\( \hat{\gamma}_L(X) = E(Y) \)

Home: Write \( \hat{\gamma}_L(X) \) in terms of \( \rho = \frac{\hat{\mu}_{xy}}{\hat{\sigma}_x \hat{\sigma}_y} \). (*) After you attempt this, see the next page.
LMSE Estimator in Terms of the Correlation Coefficient:

\[ \hat{Y}_L(X) = \text{E}(Y) + \frac{\hat{\sigma}_{XY}}{\hat{\sigma}_X^2} [X - \text{E}(X)] \]

\( \rho = \frac{\hat{\sigma}_{XY}}{\hat{\sigma}_X \hat{\sigma}_Y} \Rightarrow \rho \frac{\hat{\sigma}_Y}{\hat{\sigma}_X} = \frac{\hat{\sigma}_{XY}}{\hat{\sigma}_X^2} \)

Plug into the expression for \( \hat{Y}_L(X) \):

\[ \hat{Y}_L(X) = \text{E}(Y) + \rho \frac{\hat{\sigma}_Y}{\hat{\sigma}_X} [X - \text{E}(X)] \]

Only uses means, variances, and covariances!
What's the Mean Squared-Error (MSE) of the Linear Estimator?

\[ E = Y - \hat{Y}_L(X) = Y - E(Y) - P \frac{\partial E_y}{\partial x} [X - E(X)] \]

Claim: \( E(E) = E[Y - \hat{Y}(X)] = 0 \)

That is, \( \hat{Y}_L(X) \) is an unbiased estimator.

Proof: \( E(E) = E\left\{ Y - E(Y) - P \frac{\partial E_y}{\partial x} [X - E(X)] \right\} \)

\( E(E) = E[Y - E(Y)] - P \frac{\partial E_y}{\partial x} E[X - E(X)] = 0. \)
Claim: \( \text{MSE} = E(E^2) = \frac{b^2}{\gamma} - \frac{\hat{b}_x^2}{\hat{b}_y^2} = (1 - \rho^2) \frac{b^2}{\gamma} \)

Proof:

\[ E(E^2) = \text{var}(E) \quad \text{Since } E(E) = 0 \]

\[ = \text{var}(y - aX - b) = \text{var}(y - aX) \]

\[ = \beta_y^2 + a^2 \beta_x^2 - 2a \beta_y \beta_x \]

Recall \( a = \frac{\hat{b}_y \hat{b}_x}{\hat{b}_y^2} \quad \Rightarrow \)

\[ \text{MSE} = E(E^2) = \frac{b^2}{\gamma} + \frac{\hat{b}_x^2}{\hat{b}_y^2} \frac{b^2}{\gamma} - 2 \frac{\hat{b}_x^2}{\hat{b}_y^2} \frac{b^2}{\gamma} \]

As long as \( X, Y \) are correlated, observation is helpful (MSE reduced from no-observance case)
Interpretation

\[ \rho = 0 \quad (\text{i.e. } X, Y \text{ uncorrelated}) \]
\[ \implies \hat{y}(X) \text{ provides no useful information; in particular, } \]
\[ Y(X) = E(Y) \]
\[ \text{as though no observation was made.} \]

\[ \rho = \pm 1 \implies \text{MSE} = 0 \implies \]
\[ Y = \hat{y}(X) \text{ with probability } 1 \]
\[ \iff Y \text{ is a linear function of } X \]
\[ \iff Y \& X \text{ are linearly dependent.} \]
Afterthought on $M_n, S_n, CLT$:

$X_1, \ldots, X_n$ IID w/ 
$E(X_i) = \mu$ \quad $\text{var}(X_i) = \sigma^2$

$S_n = X_1 + \cdots + X_n$

$M_n = \frac{S_n}{n} = \frac{X_1 + \cdots + X_n}{n}$ \quad Sample Mean

Standardizing $S_n$ & $M_n$ gets us to the same standardized random variable

• Standardize $S_n$:

$Z_n = \frac{S_n - E(S_n)}{\sqrt{S_n}} = \frac{S_n - n\mu}{\sigma \sqrt{n}}$
Now divide the numerator & denominator by \( n \):

\[
Z_n = \frac{S_n - \frac{n \mu}{\sqrt{n}}}{\frac{\sigma \sqrt{n}}{n}} = \frac{M_n - \mu}{\frac{\sigma \sqrt{n}}{n}}
\]

So, \( Z_n \) is also the standardized version of \( M_n \).
Mean & Variance of a Uniform PDF

\[ f_X(x) = \begin{cases} \frac{1}{b-a} & \text{if } a \leq x \leq b \\ 0 & \text{elsewhere} \end{cases} \]

**Mean:**

\[
E(X) = \int_{a}^{b} x f_X(x) \, dx = \frac{1}{b-a} \int_{a}^{b} x \, dx = \frac{1}{b-a} \left[ \frac{x^2}{2} \right]_{a}^{b}
\]

\[
= \frac{b^2 - a^2}{2(b-a)} = \frac{(b-a)(b+a)}{2(b-a)} = \frac{a+b}{2}
\]

\[
E(X) = \frac{a+b}{2} \quad \text{Not surprisingly, mid-point between } a \text{ and } b.
\]
Variance:

\[ \sigma_X^2 = E[(X - E(X))^2] \]

Let \( Y = X - E(X) \) \( \Rightarrow \) \( E(Y) = 0 \)

\[ \sigma_Y^2 = \sigma_X^2 \]

\[ f_Y(y) = f_X(y + E(X)) \]

Left-shifted copy of \( f_X(\cdot) \).

Since \( E(Y) = 0 \), we know \( \sigma_Y^2 = E(Y^2) \)

\( \Rightarrow \sigma_X^2 = E(Y^2) \)

\[ E(Y^2) = \int_{-\infty}^{\infty} y^2 f_Y(y) \, dy = \frac{1}{b-a} \int_{-\frac{b-a}{2}}^{\frac{b-a}{2}} y^2 \, dy \]

\[ = \frac{2}{b-a} \int_0^{\frac{b-a}{2}} y^2 \, dy = \frac{2}{b-a} \frac{y^3}{3} \bigg|_0^{\frac{b-a}{2}} \]

Since integrand is an even function,

\[ \int_0^{\frac{b-a}{2}} y^2 \, dy = \frac{2}{3} \left( \frac{b-a}{2} \right)^3 \]
Summary:

\[
\frac{1}{b-a} \begin{cases} 
\frac{1}{b-a} & \text{if } a \leq x \leq b \\
0 & \text{elsewhere.} 
\end{cases}
\]

Mean: \[ E(X) = \frac{a+b}{2} \quad \text{midpoint} \]

Variance: \[ \sigma^2 = \frac{(b-a)^2}{12} \]