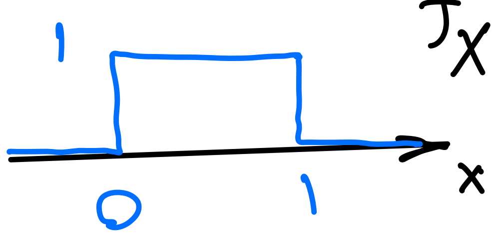
CS70-F21
L13
2 Dec.

Another CLT Example:

Alice & Bob Number Guessing Problem

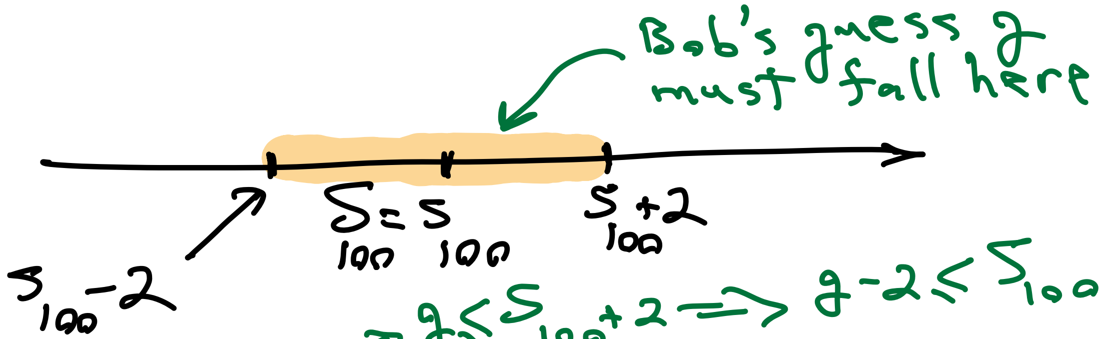
Alice picks a 100 numbers randomly & independently from $f_X(x)$



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

Bob must guess the sum S_{100}.

Bob wins a prize if his guess g is within ± 2 of S_{100}.



$$S_{100}-2 \le g \le S_{100}+2 \begin{cases} \nearrow g \le S_{100}+2 \Longrightarrow g-2 \le S_{100} \\ \searrow S_{100}-2 \le g \Longrightarrow S_{100} \le g+2 \end{cases}$$

$\Longrightarrow g-2 \le S_{100} \le g+2$ to win a prize

Bob guesses $g=55$.

What's the probability that Bob wins a prize?

$$\Pr(55-2 \le S_{100} \le 55+2) = ?$$

CLT Restatement

$X_1, \dots, X_n$ IID $\quad E(X_i) = \mu$

$\sigma_{X_i}^2 = \sigma^2$

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

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

$$Z_n = \frac{M_n - \mu}{\sigma_{M_n}} = \frac{\frac{X_1+\cdots+X_n}{n} - \mu}{\frac{\sigma}{\sqrt{n}}} = \frac{\frac{S_n}{n} - \mu}{\sigma/\sqrt{n}}$$

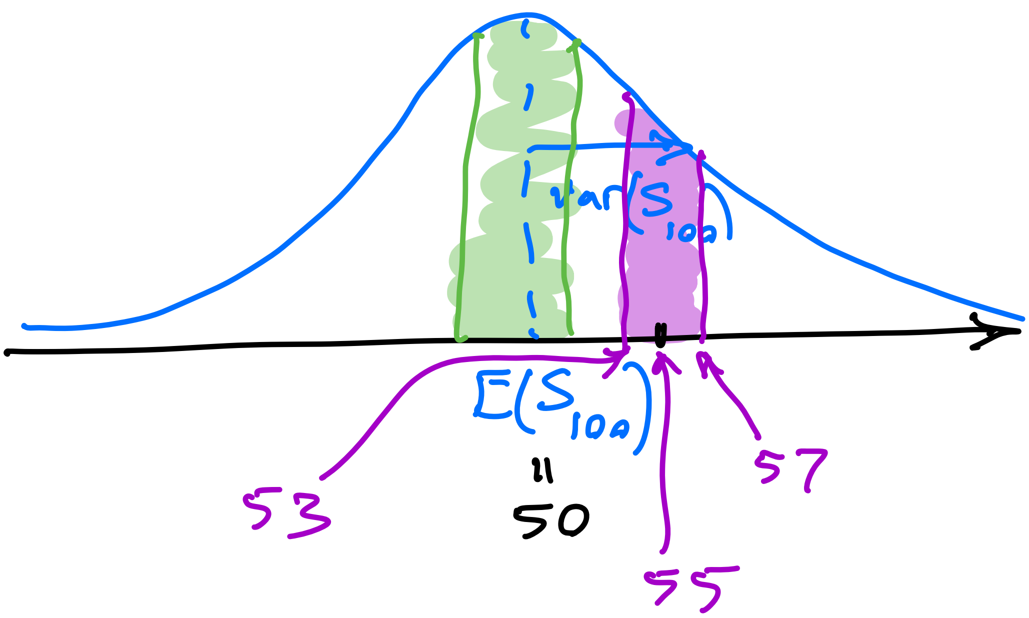
$$Z_n = \frac{S_n - n\mu}{\sigma\sqrt{n}}$$

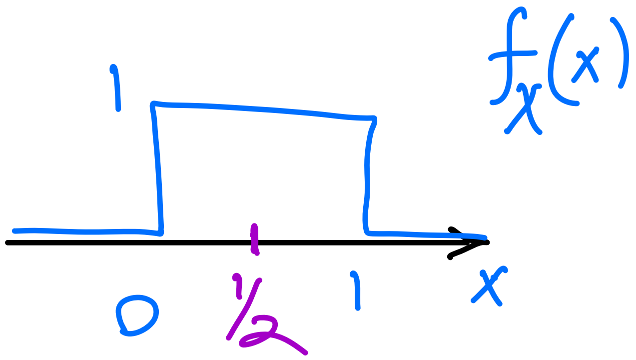
This denominator is σ_{S_n}, the standard deviation of S_n

CLT: $\lim_{n\to\infty} P(Z_n \le z) = P(Z \le z)$

Standard Gaussian RV.

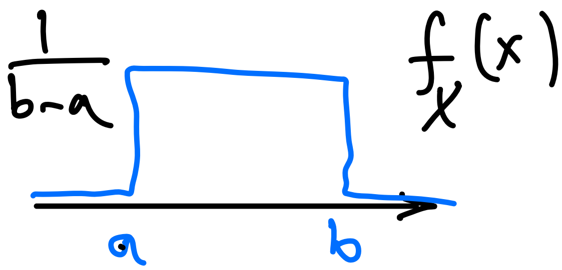
Back to Alice & Bob





$\mu = E(X_i) = 1/2$

$\sigma^2 = var(X_i) = \frac{1}{12}$



$E(X) = \frac{a+b}{2}$

$\sigma_X^2 = \frac{(b-a)^2}{12}$

Home Exercise. (*)

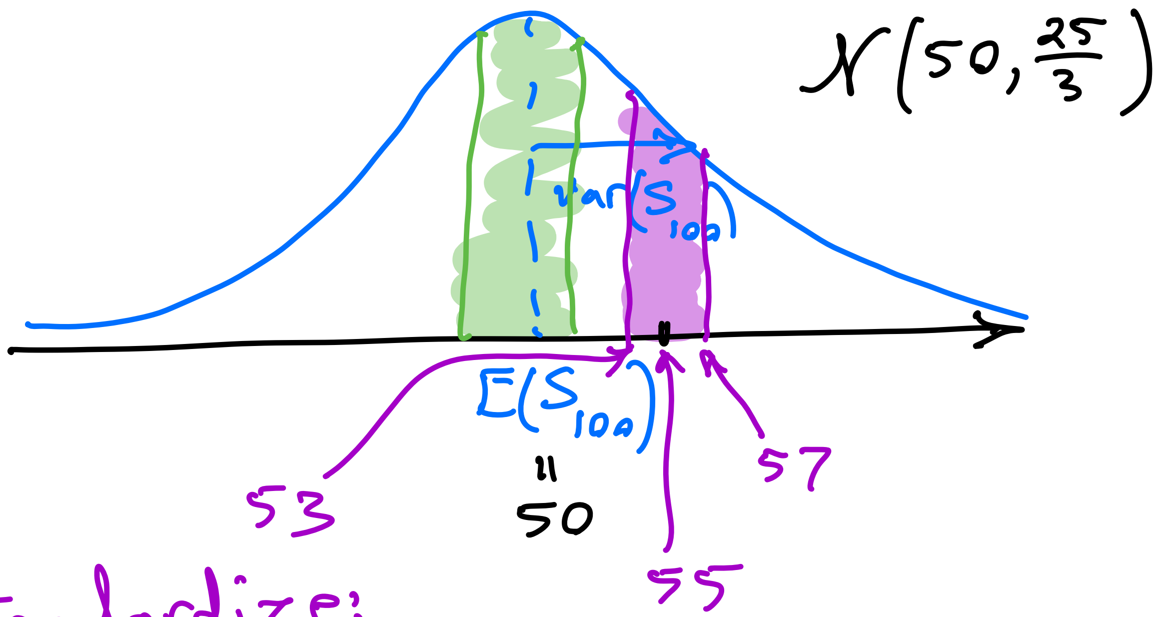
(*) After you attempt this yourself, see the end of this document for a derivation.

$$\boxed{E(S_{100}) = 100\mu = 50}$$

$$\sigma_{S_{100}}^2 = var(S_{100}) = \frac{100}{12} = \frac{25}{3}$$

$$\sigma_{S_{100}} = \frac{10}{2\sqrt{3}} = \frac{5}{\sqrt{3}}$$

$$P(53 \leq S_{100} \leq 57) = ?$$



Standardize:

$$\Pr(53 \le S_{100} \le 57) =$$

$$\Pr\left(\frac{53-50}{\frac{5}{\sqrt{3}}} \le \underbrace{\frac{S_{100}-50}{\frac{5}{\sqrt{3}}}}_{Z_{100}} \le \frac{57-50}{\frac{5}{\sqrt{3}}}\right) \approx$$

CLT Approximation

$$\Pr\left(\frac{3\sqrt{3}}{5} \le Z \le \frac{7\sqrt{3}}{5}\right)$$

$$= \Pr(1.039 \le Z \le 2.425)$$

$$= \Phi(2.425) - \Phi(1.039)$$

$$= 0.9924 - 0.8508$$ ← From the Gaussian Table

$$= 0.142$$ $\boxed{14.2\% \text{ Chance}}$

Compare w/ his likelihood of winning if he had guessed 50!

$$\Pr(48 \le S_{100} \le 52) = ?$$

$$= \Pr\left(\frac{48-50}{5/\sqrt{3}} \le \frac{S_{100}-50}{5/\sqrt{3}} \le \frac{52-50}{5/\sqrt{3}}\right) = \Pr\left(\frac{-2\sqrt{3}}{5} \le Z_{100} \le \frac{2\sqrt{3}}{5}\right)$$

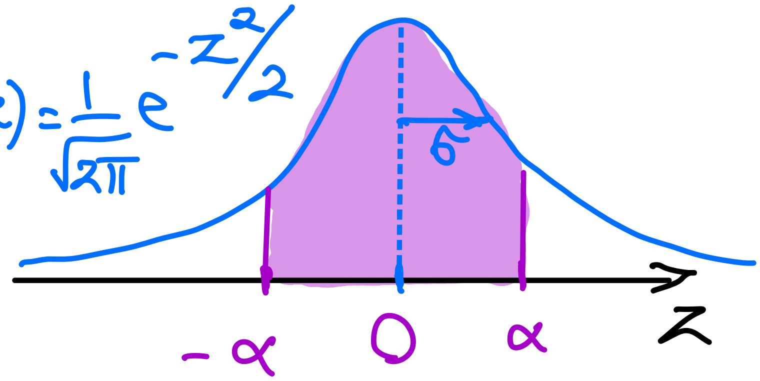
$$\approx \Pr(-0.6928 \le Z \le 0.6928) = \Pr(|Z| \le 0.6928)$$

$$= 2\Phi(0.6928) - 1$$

CLT Approximation

Standard Gaussian

Recall: $f_Z(z) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2}$



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

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(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 $\hat{y}$

Error: $\varepsilon = Y - \hat{y}$

Criterion: Minimize the mean of the squared error (MMSE)

Minimize

$$E(\varepsilon^2) = E[(Y - \hat{y})^2]$$

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

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

$$\sigma_Z^2 = \mathrm{var}(Z) = \sigma_Y^2$$

$$\sigma_Z^2 = E(Z^2) - E^2(Z)$$

$$E(Z^2) = \sigma_Z^2 + E^2(Z)$$

$$\underbrace{E\left[(Y-\hat{y})^2\right]}_{\text{MSE}} = \sigma_Y^2 + \underbrace{\left[E(Y-\hat{y})\right]^2}_{\text{can influence this.}}$$

can't do anything with this

Can I make

$$E(Y-\hat{y}) = 0 \;?$$

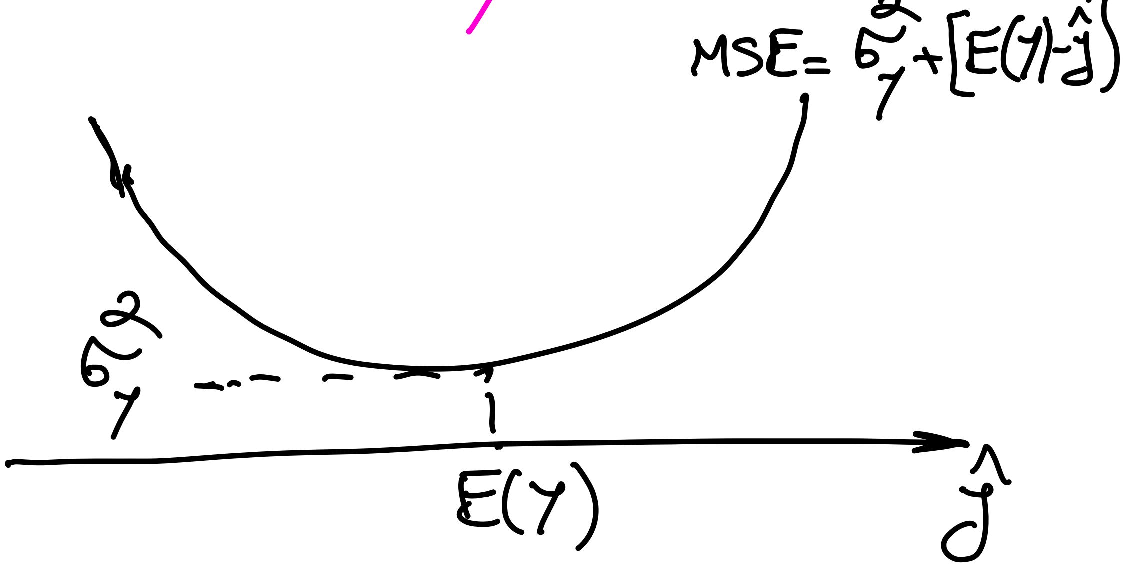
$$E(Y-\hat{y}) = E(Y) - \hat{y} = 0 \Rightarrow \boxed{\hat{y} = E(Y)}$$

Optimal Mean Squared-Error:

$$\text{MSE} = E(\varepsilon^2) = E\left[(Y - E(Y))^2\right] = \sigma_Y^2$$

Mind Blown!

The mean is the optimal estimator, and it results in an MSE = σ_Y^2.



$$E\left[(Y - E(Y))^2\right] \le E\left[(Y - \hat{y})^2\right]$$

for all $\hat{y}$.

What if we have an observation X?

$$\underbrace{\hat{y} = E(Y|X=x)}_{\text{Point estimate}} = g(x)$$

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

lower case b/c $X=x$

Point estimate of Y in the MMSE sense

$$\hat{Y}_{MMSE} = E(Y|X) = g(X)$$

Random Variable

$$E\left[\left(Y - E(Y|X)\right)^2\right] \leq E\left[\left(Y - g(X)\right)^2\right]$$

$\forall$ fctns $g(X)$.

Linear MSE Estimation:

$$\hat{Y}(X) = aX + b \qquad \mathcal{E} = Y - \hat{Y}$$

$$= Y - (aX + b)$$

$$= Y - aX - b$$

$$MSE = \overset{\text{def}}{E}\Big[\underbrace{(\overbrace{Y - aX}^{Z} - b)^2}_{(Y-\hat{Y})^2}\Big]$$

Let $Z = Y - aX$

$$MSE = E\left[(Z-b)^2\right]$$

What's the optimal b?

$$b = E(Z) = E(Y - aX) = E(Y) - aE(X)$$

$$MSE = E\left[\left(Y - aX - E(Y) + aE(X)\right)^2\right]$$

$$= E\Big[\Big(\underbrace{(Y - E(Y))}_{\sigma_Y^2} - a\underbrace{(X - E(X))}_{\sigma_X^2}\Big)^2\Big]$$

Grouped terms by color

$$= \overbrace{E\left[\left(Y - E(Y)\right)^2\right]} + a^2 E\left[\left(X - E(X)\right)^2\right]$$

$$- 2a \underbrace{E\left[\left(X - E(X)\right)\left(Y - E(Y)\right)\right]}_{\sigma_{XY} = \mathrm{cov}(X, Y)}$$

$$MSE = \sigma_Y^2 + a^2 \sigma_X^2 - 2a\,\sigma_{XY}$$

Must minimize wrt a:

$$\frac{dMSE}{da} = 2a\sigma_X^2 - 2\,\sigma_{XY} = 0$$

$$\frac{d^2MSE}{da^2} = 2\sigma_X^2 > 0$$

$$a\sigma_X^2 - \sigma_{XY} = 0 \longrightarrow \boxed{a = \frac{\sigma_{XY}}{\sigma_X^2}}$$

$$\boxed{b = E(Y) - aE(X)}$$

$$\hat{Y}_L(X) = aX + b = \frac{\sigma_{XY}}{\sigma_X^2} X + E(Y) - \frac{\sigma_{XY}}{\sigma_X^2} E(X)$$

$$\hat{Y}_L(X) = E(Y) + \underbrace{\frac{\sigma_{XY}}{\sigma_X^2}(X - E(X))}_{\text{correction term}}$$

If X, Y uncorrelated?

$$\sigma_{XY} = 0 \qquad \sigma_{XY} = \text{cov}(X,Y)$$

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

Home: write $\hat{Y}_L(X)$ in terms of $\rho = \frac{\sigma_{XY}}{\sigma_X \sigma_Y}$.

(*) After you attempt this, see the next page.

LMSE Estimator in Terms of the Correlation Coefficient ρ:

$$\hat{Y}_L(X) = E(Y) + \frac{\sigma_{XY}}{\sigma_X^2}\left[X - E(X)\right]$$

$$\rho = \frac{\sigma_{XY}}{\sigma_X \sigma_Y} \Longrightarrow \rho \frac{\sigma_Y}{\sigma_X} = \frac{\sigma_{XY}}{\sigma_X^2}$$

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

$$\hat{Y}_L(X) = E(Y) + \rho \frac{\sigma_Y}{\sigma_X}\left[X - E(X)\right]$$

$$\boxed{\begin{aligned} \hat{Y}_L(X) &= E(Y) + \frac{\sigma_{XY}}{\sigma_X^2}\left[X - E(X)\right] \\ &= E(Y) + \rho \frac{\sigma_Y}{\sigma_X}\left[X - E(X)\right] \end{aligned}}$$

Only uses means, variances, and covariances!

What's the Mean Squared-Error (MSE) of the Linear Estimator?

$$\mathcal{E} = Y - \hat{Y}_L(X) = Y - E(Y) - \rho \frac{\sigma_Y}{\sigma_X}[X - E(X)]$$

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

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

Proof: $E(\mathcal{E}) = E\left\{Y - E(Y) - \rho \frac{\sigma_Y}{\sigma_X}[X - E(X)]\right\}$

$$E(\mathcal{E}) = \underbrace{E[Y - E(Y)]}_{0} - \rho \frac{\sigma_Y}{\sigma_X} \underbrace{E[X - E(X)]}_{0} = 0.$$

Claim: $MSE \triangleq E(\varepsilon^2) = \sigma_Y^2 - \dfrac{\sigma_{XY}^2}{\sigma_X^2} = (1-\rho^2)\sigma_Y^2$

Proof:

$E(\varepsilon^2) = var(\varepsilon)$ since $E(\varepsilon) = 0$

b doesn't affect spread

$= var(Y - aX - b) = var(Y - aX)$

$= \sigma_Y^2 + a^2\sigma_X^2 - 2a\sigma_{XY}$

Recall $a = \dfrac{\sigma_{XY}}{\sigma_X^2} \Longrightarrow$

$$MSE = E(\varepsilon^2) = \sigma_Y^2 + \frac{\sigma_{XY}^2}{\sigma_X^{\cancel{4}\,2}}\cancel{\sigma_X^2} - 2\frac{\sigma_{XY}^2}{\sigma_X^2}$$

$$\boxed{MSE = \sigma_Y^2 - \frac{\sigma_{XY}^2}{\sigma_X^2} = (1-\rho^2)\sigma_Y^2 \leq \sigma_Y^2}$$

Recall $\rho = \dfrac{\sigma_{XY}}{\sigma_X \sigma_Y}$

As long as X, Y are correlated, observation is helpful (MSE reduced from no-observ. case)

Interpretation

$\rho = 0$ (i.e. X, Y uncorrelated)

$\Longrightarrow \hat{Y}_L(X)$ provides no useful information; in particular,

$$\hat{Y}(X) = E(Y)$$

as though no observation was made.

$\rho = \pm 1 \Longrightarrow MSE = 0 \Longrightarrow$

$Y = \hat{Y}(X)$ with probability 1

$\Longleftrightarrow$ Y is a linear function of X

$\Longleftrightarrow$ Y & X are linearly dependent.

Afterthought on M_n, S_n, CLT:

$X_1, \ldots, X_n$ IID w/

$E(X_i) = \mu$ $\quad$ $var(X_i) = \sigma^2$

$$S_n = X_1 + \cdots + X_n$$

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

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

- Standardize S_n:

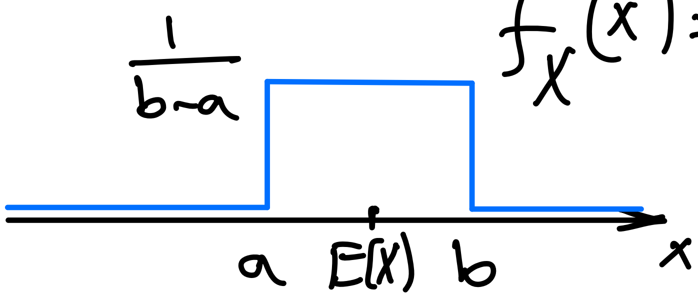
$$Z_n = \frac{S_n - E(S_n)}{\sigma_{S_n}} = \frac{S_n - n\mu}{\sigma\sqrt{n}}$$

Now divide the numerator & denominator by n:

$$Z_n = \frac{\frac{S_n}{n} - \frac{n\mu}{n}}{\frac{\sigma\sqrt{n}}{n}} = \frac{M_n - \underbrace{\mu}_{E(M_n)}}{\underbrace{\frac{\sigma}{\sqrt{n}}}_{\sigma_{M_n}}}$$

So, Z_n is also the standardized version of M_n.

Mean & Variance of a Uniform PDF



$$f_X(x) = \begin{cases} \frac{1}{b-a} & \text{if } a \le x \le b \\ 0 & \text{elsewhere.} \end{cases}$$

Mean:

$$E(X) = \int_{-\infty}^{\infty} x f_X(x)\,dx = \frac{1}{b-a}\int_a^b x\,dx = \frac{1}{b-a}\left.\frac{x^2}{2}\right|_a^b$$

$$= \frac{b^2 - a^2}{2(b-a)} = \frac{\cancel{(b-a)}(b+a)}{2\cancel{(b-a)}} = \frac{a+b}{2}$$

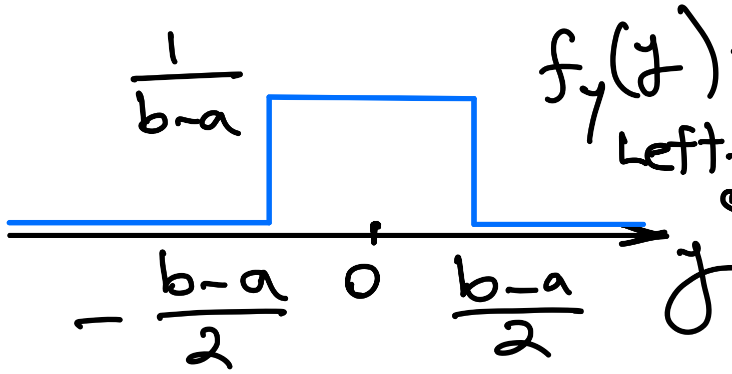
$$E(X) = \frac{a+b}{2}$$

Not surprisingly, midpoint between a and b.

Variance:

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

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 $f_X()$.

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

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

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

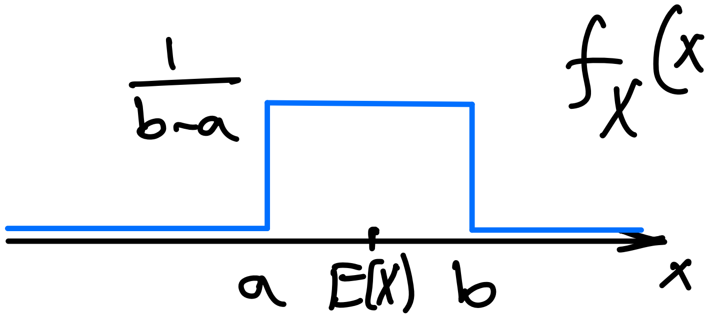
$$= \frac{2}{b-a}\int_0^{\frac{b-a}{2}} y^2\,dy = \frac{2}{b-a}\left.\frac{y^3}{3}\right|_0^{\frac{b-a}{2}}$$

since integrand is an even function

$$E(Y^2) = \frac{\cancel{2}}{\cancel{b-a}} \cdot \frac{(b-a)^{\cancel{3}\,2}}{3 \cdot 2^{\cancel{3}\,2}} = \frac{(b-a)^2}{12}.$$

$$\Longrightarrow \sigma_X^2 = \frac{(b-a)^2}{12}$$

Summary:



$$f_X(x) = \begin{cases} \frac{1}{b-a} & \text{if } a \le x \le b \\ 0 & \text{elsewhere.} \end{cases}$$

Mean: $E(X) = \frac{a+b}{2}$ midpoint

Variance: $\sigma_X^2 = \frac{(b-a)^2}{12}$