3RD WED OCT 11

Reminder: Submit only python notebook for HW

<u>Review</u>/Recap $\quad \mathbb{E}X, \ \mathrm{Var}(X), \ SD(X) \leftarrow$ HOW TO CALCULATE APPROX VIA SIM.

- Poisson($\lambda$) random variable
- Exponential waiting times in a Poisson process
- Over dispersion If $X \sim \mathrm{Poisson}(\lambda)$, $\mathbb{E}X = \lambda$, $\mathrm{Var}(X) = \lambda$

  $R \sim \exp(\lambda)$

  Given $R$, (write $X \mid R$) $X$ is Poisson($R$)

  Estimate Var or SD of $X$

$$\boxed{\mathrm{Var}(X) = \mathbb{E}[\mathrm{Var}(X \mid Y)] + \mathrm{Var}(\mathbb{E}[X \mid Y])}$$

- ESTIMATING EXPECTATIONS VIA SIMULATION

<u>TODAY</u> Central limit theorem

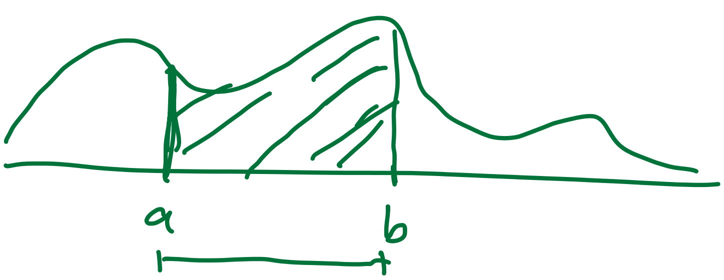
---

$$\mathbb{E}X = \begin{cases} \sum_k k \, \underbrace{\mathbb{P}(X=k)}_{\text{mass function}} & \text{discrete} \\ \int_{-\infty}^{\infty} x f_X(x)\,dx & \text{continuous random variable w/ density } f_X(x) \end{cases}$$

$\mathrm{Var}(X) = \mathbb{E}[(X - \mathbb{E}X)^2]$ measures dispersion of random variable

$= \mathbb{E}X^2 - 2\mathbb{E}[X\,\mathbb{E}[X]] + (\mathbb{E}X)^2$

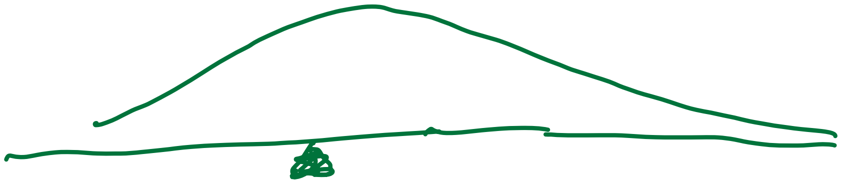
$= \mathbb{E}X^2 - (\mathbb{E}X)^2$




density $f(x) \geq 0$

$\int_a^b f(x)\,dx = \mathbb{P}(a \leq X \leq b)$

$$\int_{-\infty}^{\infty} f_X(x)\,dx = 1$$



---

Way to simulate $\bar{X}$

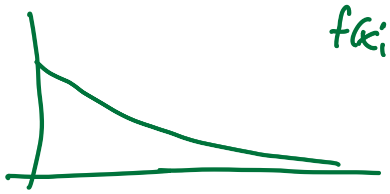
approx $\mathbb{E}(X)$

$X_1, X_2, \ldots, X_N$

$\frac{1}{N}\sum_{i=1}^{N} X_i$

$X \sim \exp(\mu) \longleftrightarrow$ density function $f(x;\mu) = \mu e^{-\mu x} \quad x \geq 0$

$f(x;r) = \frac{1}{r} e^{-\frac{1}{r}x} \quad x \geq 0$



$R_1, R_2, \ldots R_{100000}$ each comes from this prob distn

$X_1 \sim \text{Poisson}(R_1)$

$$\mathbb{P}(X_1 = k) = \frac{e^{-R_1} R_1^k}{k!}$$

$X_2 \sim \text{Poisson}(R_2)$

---

np.mean($x$) $\quad x = x_1, \ldots, x_n$

$$= \frac{x_1 + \cdots + x_n}{n}$$

$x_i = \begin{cases} 1 \\ 0 \end{cases} \quad$ np.mean($x$) $= \frac{\#1\text{'s}}{n}$

$1^{st}$ point if can simulate $Z_1, Z_2, \ldots, Z_N$ random variables, all w/ same same distribution

$$\mathbb{E}[Z_i] \cong \frac{Z_1 + \cdots + Z_n}{n}$$

$$Z_i = \begin{cases} 1 & \text{if } X_i > 2 \\ 0 & \text{otherwise} \end{cases}$$

$$\mathbb{E} Z_i = \mathbb{P}(X_i > 2)$$
$$= 1 \cdot \mathbb{P}(Z_i = 1) + 0 \cdot \mathbb{P}(Z_i = 0)$$

$$\mathbb{E} g(X) \cong \frac{g(X_1) + \cdots + g(X_n)}{n}$$

e.g. $g(x) = \begin{cases} 1 & \text{if } x \geq 2 \\ 0 & \text{o/w} \end{cases}$

$\mathrm{Var}(X) = \mathbb{E}[(X - \mathbb{E}X)^2]$ measure of dispersion

$SD(X) = \sqrt{\mathrm{Var}(X)}$

$SD(aX) = |a| \, SD(X)$

$\mathrm{Var}(aX) = a^2 \mathrm{Var}(X)$

$\bar{X} = \frac{1}{n} \sum_{i=1}^{n} X_i$

$X_1, X_2, X_3, \ldots, X_n$

$\mathrm{Var}(X) \simeq \frac{1}{n} \sum_{i=1}^{n} (X_i - \bar{X})^2 \leftarrow$ sample variance

MATHEMATICAL

$$\mathbb{E}(X) = \int_{-\infty}^{\infty} x f(x)\,dx$$

EMPIRICAL (DATA, SIMULATION)

$$\frac{X_1 + X_2 + \cdots + X_n}{n} = \bar{X}$$

$$\mathrm{Var}(X) = \mathbb{E}[(X - \mathbb{E}X)^2] = \int_{-\infty}^{\infty} (x - \mathbb{E}X)^2 f(x)\,dx \quad \longleftrightarrow \quad \frac{1}{n}\sum_i (X_i - \bar{X})^2 = \hat{\sigma}^2$$

$$\hat{\sigma} = \sqrt{\hat{\sigma}^2}$$

Challenge: Show that $\mathrm{Var}(X) = 2$!

$R \sim \exp(1)$

$X \mid R \sim \text{Poisson}(R)$

$\mathrm{Var}(X \mid R) = R$

$\mathbb{E}[\mathrm{Var}(X \mid R)] = \mathbb{E}[R] = 1$

$\mathbb{E}[X \mid R] = R$

$\mathrm{Var}(\mathbb{E}(X \mid R)) = \mathrm{Var}(R) = 1$

Poisson $(\lambda)$ r.v. $X$

$\mathbb{E}X = \lambda$

$\mathrm{Var}(X) = \lambda$

$$\boxed{\mathrm{Var}(X) = \mathbb{E}[\mathrm{Var}(X \mid Y)] + \mathrm{Var}(\mathbb{E}[X \mid Y])}$$

[illegible]

$\mathrm{Var}(X) = 1 + 1 = 2$

$SD(X) = \sqrt{2}$

---

$X \sim \text{Poisson}(5)$

$\mathbb{E}[X(X-1)]$

$g(x) = x(x-1)$

$\mathbb{E}g(x) \overset{\sim}{=} ?$

$x_1, \ldots, x_{1000}$ Poisson $(5)$

$\overline{X} \simeq \mathbb{E}X$

$x_1 - 1, x_2 - 1, \ldots, x_{1000} - 1$

$\underbrace{x_1(x_1-1)}_{y_1}, \underbrace{x_2(x_2-1)}_{y_2}, \ldots, \underbrace{x_{1000}(x_{1000}-1)}_{y_{1000}}$

$\overline{Y} \simeq \mathbb{E}Y_i = \mathbb{E}X(X-1)$

$\mathbb{E}[\exp(X)]$

$g(x) = \exp(x)$

$x_1, \ldots, x_{1000}$

$e^{x_1}, \ldots, e^{x_{1000}}$

$\text{np.mean}(\quad) = \dfrac{e^{x_1} + \ldots + e^{x_{1000}}}{n}$

$\approx \mathbb{E}[e^X]$

# Central Limit Theorem

identically distributed

Random variables $X_1, \ldots, X_n$ IID

independent

$\sigma^2 = Var(X_i)$

Suppose $\mu = \mathbb{E}X_i = \mathbb{E}X_1 = \mathbb{E}X_n$ is constant!

How to estimate $\mu$? $\quad \bar{X}$ should estimate $\mu$

$$\mathbb{E}(\bar{X}) = \mathbb{E}\left[\frac{1}{n}\sum_{i=1}^n X_i\right] = \frac{1}{n}\sum_{i=1}^n \mathbb{E}X_i = \frac{1}{n}\sum_{i=1}^n \mu = \mu$$

$$Var(\bar{X}) = Var\left(\frac{1}{n}\sum_{i=1}^n X_i\right) = \frac{1}{n^2}\sum_{i=1}^n Var(X_i) = \frac{1}{n^2}\sum_{i=1}^n \sigma^2 = \sigma^2/n$$

$$Var(X_1) = Var(X_2) = \ldots = Var(X_n)$$

$$Var(aX) = a^2 Var(X)$$

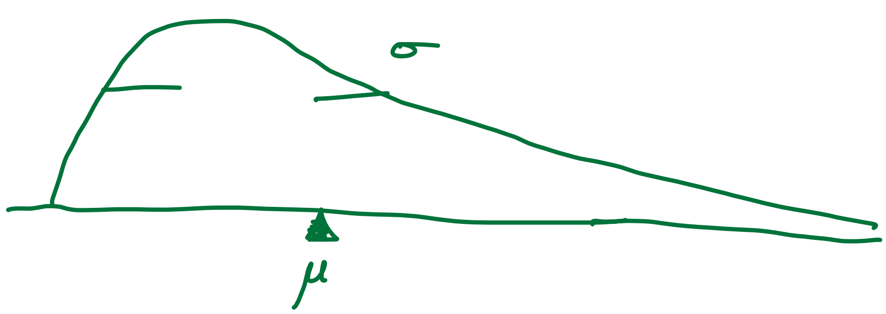
$$SD(\bar{X}) = \sigma/\sqrt{n}$$

Fundamental Theorem of Statistics

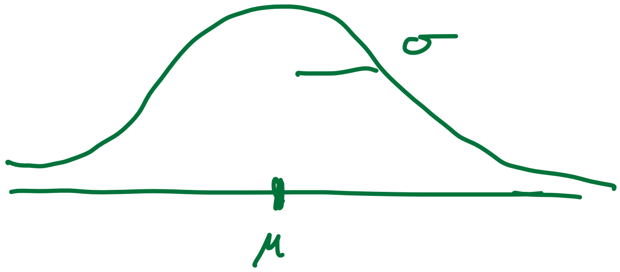
$X_1, \ldots, X_n$ IID $\quad \mu = \mathbb{E}X_i \quad \sigma^2 = Var(X_i)$

$\bar{X}$ has $\quad \mathbb{E}(\bar{X}) = \mu$

$Var(\bar{X}) = \sigma^2/n \qquad SD(\bar{X}) = \sigma/\sqrt{n}$

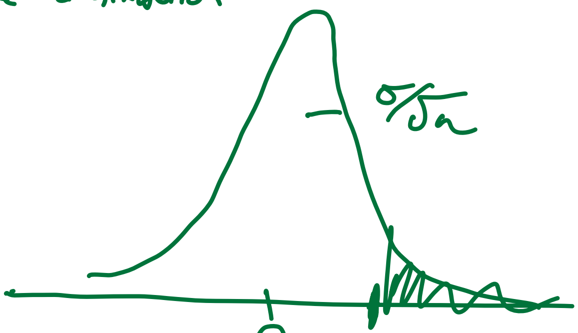


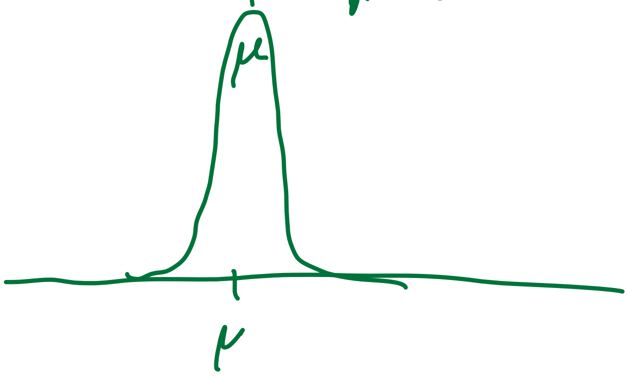

Guess $\mu$ is my goal!




$X_1$

$\bar{X}$ has some distribution




$\mathbb{P}(|\bar{X} - \mu| > \varepsilon)$ is small as $n \longrightarrow \infty$

<u>Central Limit Theorem</u>

Suppose $X_1, X_2, \ldots$ are IID from some common distribution with $\mathbb{E}X_i = \mu$, $\mathrm{Var}(X_i) = \sigma^2$

$$\mathbb{P}\left( \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \le t \right) \longrightarrow$$


$$\int_{-\infty}^{t} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}\,dx$$

---

$\mathrm{Var}(\bar{X}) = (\sigma/\sqrt{n})^2 \qquad \mathbb{E}\bar{X} = \mu$

$$Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \qquad \mathbb{E}[Z] = \mathbb{E}\left[\frac{\sqrt{n}}{\sigma}(\bar{X} - \mu)\right]$$

$$= \frac{\sqrt{n}}{\sigma}\,\mathbb{E}(\bar{X} - \mu)$$

$$= \frac{\sqrt{n}}{\sigma}\left[ \mathbb{E}\bar{x} - \mu \right]$$

$$= 0$$

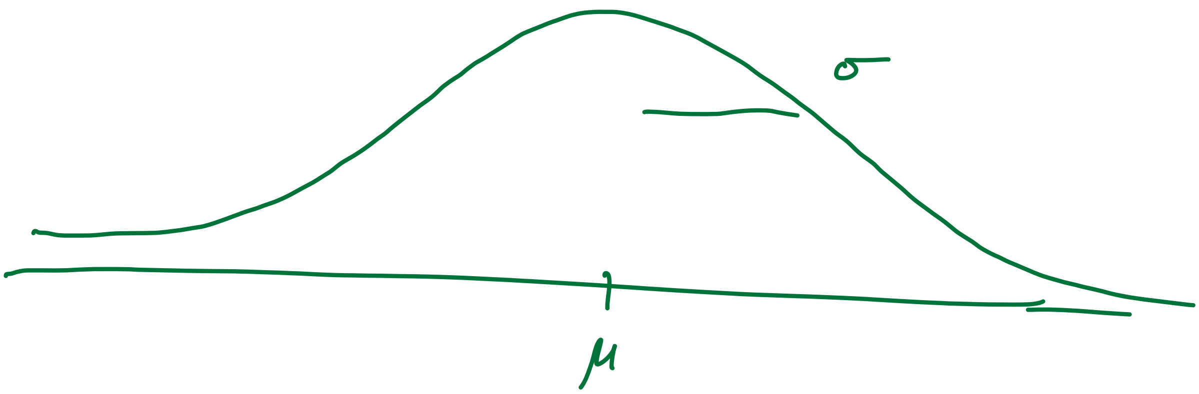
$$\mathrm{Var}(Z) = \frac{n}{\sigma^2}\left[\mathrm{Var}(\bar{x})\right] = \frac{n}{\sigma^2}\,\frac{\sigma^2}{n} = 1$$

Whatever the common distribution of $X_1, \ldots$

$$\frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \approx N(0,1)$$

$$\hookrightarrow f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}$$

$$\bar{X} \approx N(\mu, \sigma/\sqrt{n})$$




$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(x-\mu)^2}$$

$N(\mu, \sigma)$ density

$$\mathbb{E}N = \mu$$
$$\mathrm{Var}(N) = \sigma^2$$
$$\mathrm{SD}(N) = \sigma$$

$X_1, X_2, \ldots, X_{100}$ want to estimate $\mu = \mathbb{E}X$

suppose $\sigma^2 = \mathrm{Var}(X_i) = 3$

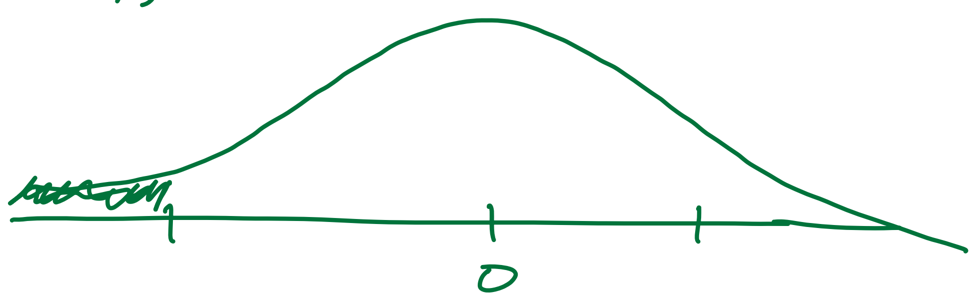
$$\mathbb{P}\left( |\bar{X} - \mu| > 0.05 \right)$$

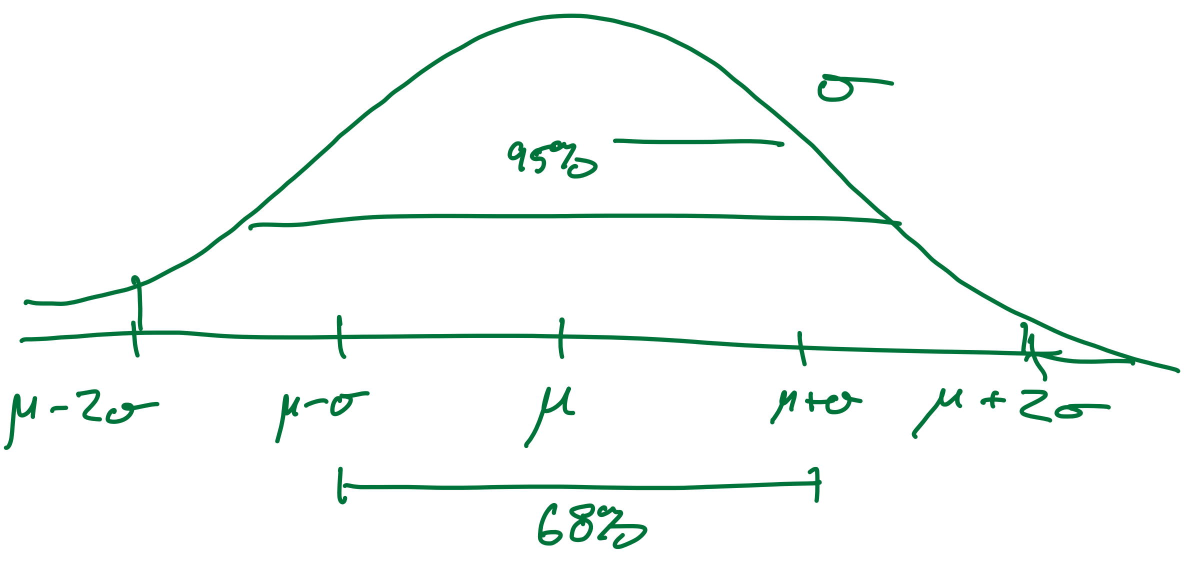
Central Limit Theorem: $\dfrac{\bar{X} - \mu}{\sqrt{3}/\sqrt{100}} \approx N(0,1)$

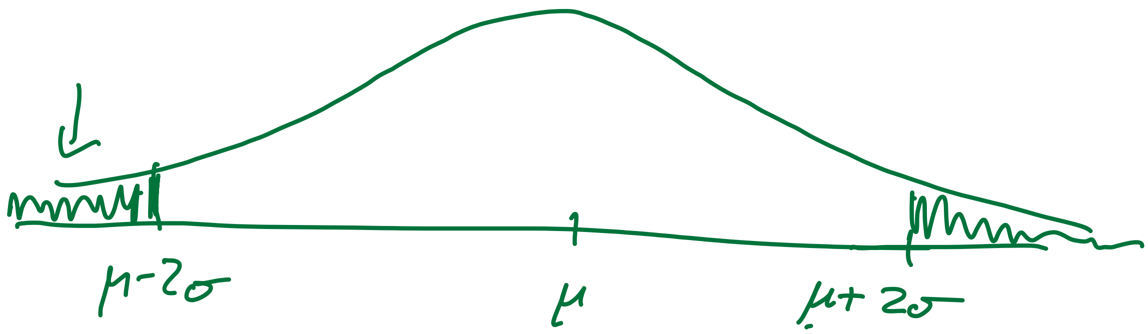
$$\mathbb{P}\left( \frac{|\bar{X} - \mu|}{\sqrt{3}/10} > \frac{0.05}{\sqrt{3}/10} \right)$$

$$\approx \mathbb{P}\left( |Z| > 10 \times 0.05/\sqrt{3} \right) = 2\,\mathbb{P}\left( Z < 10 \times 0.05/\sqrt{3} \right)$$

$Z \sim N(0,1)$




$\mu - 2\sigma$
$\mu$
$\mu + 2\sigma$