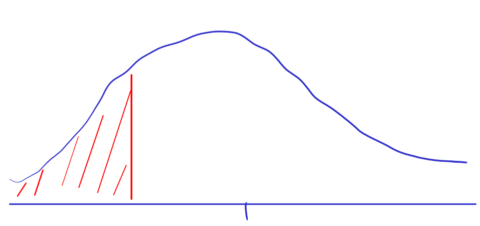
The Central Limit Theorem

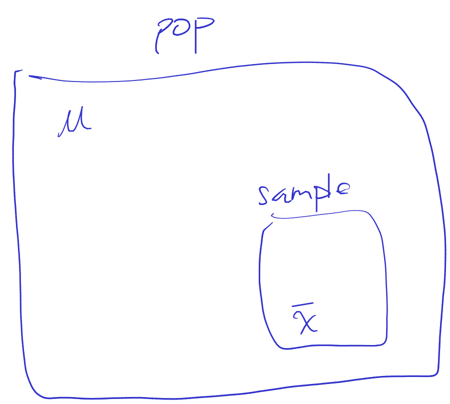
David Armstrong

UCI

$X \sim \text{Normal}(\mu, \theta)$



$E(X)$



Central Limit Theorem $\quad X \sim Dist(\mu, \sigma)$

Suppose we are interested in the sample mean $\bar{X}$, and we know that X has some distribution with $E[X] = \mu$ and $SD[X] = \sigma$.

- The MEAN of the SAMPLE AVERAGE is: $\quad \bar{x} = \frac{\Sigma x_i}{n}$

  $E\left[\bar{X}\right] = \mu_{\bar{X}}$

  As our sample size increases the mean of our sample will stay the same.

  PROOF: $E(\bar{x}) = E\left(\frac{\sum_{i=1}^{n} x_i}{n}\right) = \frac{1}{n}\left(E\left(\sum_{i=1}^{n} x_i\right)\right) = \frac{1}{n}\left(\sum_{i=1}^{n} E(x_i)\right)$

$$\frac{1}{n}\sum_{i=1}^{n} \mu = \frac{1}{n}(n\mu) = \mu$$

$$E(\bar{x}) = \mu$$

- The STANDARD DEVIATION of the SAMPLE AVERAGE is:

  $SD\left[\bar{X}\right] = \sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}}$

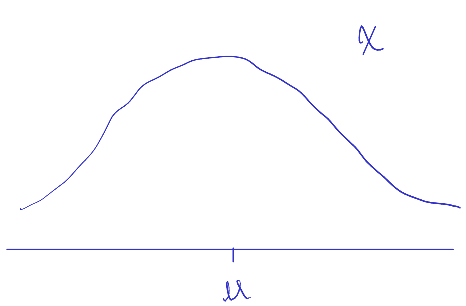
  In general, taking the average of larger sample sizes gives a more precise estimate of the true mean. Thus, the spread around the center gets smaller.

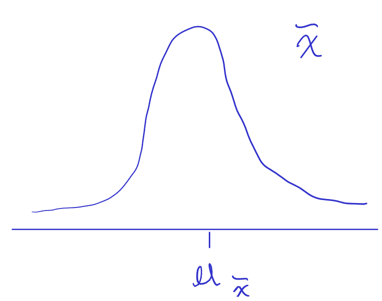
  PROOF:

$$Var(x) = Var\left(\frac{\sum_{i=1}^{n} x_i}{n}\right) = \frac{1}{n^2} Var\left(\sum_{i=1}^{n} x_i\right) = \frac{1}{n}\sum_{i=1}^{n} Var(x_i)$$

$$= \frac{1}{n^2}\sum_{i=1}^{n} \sigma^2 = \frac{1}{n^2} n\sigma^2 = \frac{\sigma^2}{n}$$

$$SD(\bar{x}) = \sqrt{Var(x)} = \sqrt{\frac{\sigma^2}{n}} = \frac{\sigma}{\sqrt{n}}$$





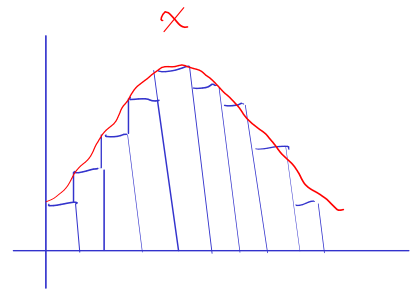
Central Limit Theorem

- Suppose $X \sim Normal\,(\mu, \sigma)$ and $n \geq 1$

  Then $\bar{X} \sim Normal\left(\mu_{\bar{X}} = \mu, \sigma_{\bar{X}} = \dfrac{\sigma}{\sqrt{n}}\right)$

  $$Z = \frac{\bar{X} - \mu_{\bar{X}}}{\sigma_{\bar{X}}} = \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}}$$

  $\overset{x}{N} \longrightarrow \overset{\bar{x}}{N}$

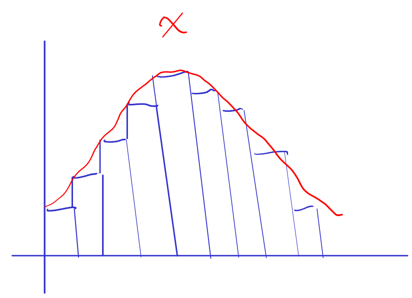


- Suppose $X \sim ApproximatelyNormal\,(\mu, \sigma)$ and $n \geq 1$

  Then $\bar{X} \sim ApproximatelyNormal\left(\mu_{\bar{X}} = \mu, \sigma_{\bar{X}} = \dfrac{\sigma}{\sqrt{n}}\right)$

  $$Z = \frac{\bar{X} - \mu_{\bar{X}}}{\sigma_{\bar{X}}} = \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}}$$

  $\overset{x}{AN} \longrightarrow \overset{\bar{x}}{AN}$



- The Central Limit Theorem (CLT):

  Draw a Simple Random Sample (SRS) of size $n \geq 30$ from any non-normal population with $E[X] = \mu$ and $SD[X] = \sigma$, then the sample mean has a sampling distribution that is approximately normal.

  Suppose $X \sim NonNormal\,(\mu, \sigma)$ and $n \geq 30$

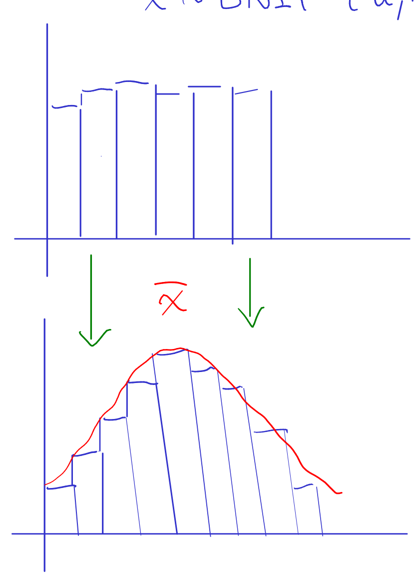
  Then $\bar{X} \sim ApproximatelyNormal\left(\mu_{\bar{X}} = \mu, \sigma_{\bar{X}} = \dfrac{\sigma}{\sqrt{n}}\right)$

  $$Z = \frac{\bar{X} - \mu_{\bar{X}}}{\sigma_{\bar{X}}} = \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}}$$

$x \sim UNIF\,(a, b)$

$E(x) = \mu = \dfrac{b+a}{2}$

$Var(x) = \dfrac{(b-a)^2}{12}$



Example: In engineering, weights of people are considered so that airplanes and elevators aren't overloaded, chairs won't break. Men's weights are normally distributed with a mean of 173 lbs., and a standard deviation of 30 lbs.

a. What is the distribution for a one randomly selected man's weight?

$X \sim \text{Normal}(\mu = 173, \sigma = 30)$

b. What is the probability a randomly selected man weighs more than 180 lbs.?

$$P(X > 180) = 1 - P(X \leq 180)$$
$$= 1 - \text{pnorm}(180, 173, 30)$$
$$= 0.4078$$

c. What is the distribution of the average men's weight if we are considering a SRS of 9 men?

$\bar{x} \sim \text{Normal}\left(\mu_{\bar{x}} = 173, \sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{30}{\sqrt{9}} = \frac{30}{3} = 10\right)$

d. If 9 men are randomly selected (say to be in an elevator), what is the probability that their average weight is more than 180 lbs.

$$P(\bar{x} > 180) = 1 - P(\bar{x} \leq 180)$$
$$= 1 - \text{pnorm}(180, 173, 10)$$
$$= 0.2420$$

$X \sim \text{Right Skewed}(\mu = 60, \sigma = 25)$

Example: A rental car company has noticed that the distribution of the number of miles customers put on rental cars per day is right skewed. The distribution has a mean of 60 miles and a standard deviation of 25 miles. A random sample of 120 rental cars is selected.

$n = 120$

a. Describe the sampling distribution of the average number of miles driven per day for the sample of 120 rental cars. Use the appropriate notation.

$\bar{x} \sim AN(\mu_{\bar{x}} = \mu = 60, \sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{25}{\sqrt{120}} \approx 2.2822)$

b. What is the probability that the mean number of miles driven per day for the sample of 120 cars is less than 54?

$P(\bar{x} < 54) = \text{pnorm}(54, 60, \frac{25}{\sqrt{120}})$

≈ 0.0043

c. What is the probability that the total number of miles driven per day in the sample of 120 cars exceeds 7400?

$P(x > 7400) = P(\bar{x} > \frac{7400}{120})$

$= 1 - P(\bar{x} < \frac{7400}{120})$

$= 1 - \text{pnorm}(\frac{7400}{120}, 60, \frac{25}{\sqrt{120}})$

$= 0.2326$

$\frac{x}{N} \longrightarrow \frac{\bar{x}}{N}$ $n \geq 1$ $\qquad$ $\frac{x}{OD} \longrightarrow \frac{\bar{x}}{AN}$ $n \geq 30$

$\frac{x}{AN} \longrightarrow \frac{\bar{x}}{AN}$ $n \geq 1$

Example: Which statement is correct regarding the Central Limit Theorem:

~~A~~. All variables have approximately normal shaped distributions if a random sample contains at least 30 observations. $N \rightarrow N$

~~B~~. Population distributions are normal whenever the population size is large. AN

✓ C. For non-normal populations, the sampling distribution of the sample mean is approximately normal with a sufficiently large random sample.

~~D~~. The sampling distribution of the sample mean looks identical to the population distribution with a large sample size. $OD \rightarrow AN$

Inverse Calculations: This Z-score calculation can also be rearranged to solve for a sample mean:

$$Z = \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \qquad \longrightarrow \qquad \bar{X} = Z\frac{\sigma}{\sqrt{n}} + \mu \qquad \text{qnorm}$$

Example: The amounts of telephone bills for all households in a large city have a distribution that is not normal with a mean of \$75 and a standard deviation of \$27. A random sample of 90 households is selected from this city.

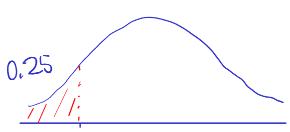
$x \sim OD(\mu = 75, \sigma = 27)$

a. What is the probability that the sample average telephone bill will be less than \$70?

$\bar{x} \sim (\mu_{\bar{x}} = 75, \sigma_{\bar{x}} = \frac{27}{\sqrt{90}})$

$P(\bar{x} < 70) = \text{pnorm}(70, 75, \frac{27}{\sqrt{90}}) = 0.0395$

b. What is the average telephone bill cost representing the 25th percentile?



$z_c = \text{qnorm}(0.25, 0, 1) = -0.6745$

$\bar{x}_c = \text{qnorm}(0.25, 75, \frac{27}{\sqrt{90}}) = \$73.08$

$\bar{x} = z_c \frac{\sigma}{\sqrt{n}} + \mu = -0.6745(\frac{27}{\sqrt{90}})(75) = \$73.08$