Lecture 25

What is sampling distribution of $\bar{X}$ if $X_1, \ldots, X_n$ are iid Normal(μ, σ)?

From last time

$$E\left[\frac{\sum_{i=1}^{n} X_i}{n}\right] = \frac{1}{n}\sum_{i=1}^{n} E[X_i]$$

$$= \frac{1}{n}(\overbrace{\mu + \mu + \cdots + \mu}^{n \text{ times repeated}})$$

$$= \frac{n\mu}{n}$$

$$= \mu$$

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

$$= \frac{1}{n^2}(\overbrace{\sigma^2 + \sigma^2 + \cdots + \sigma^2}^{n \text{ times repeated}})$$

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

which means that

$$SD(\bar{X}) = SD\left(\frac{\sum X_i}{n}\right) = \sqrt{\frac{\sigma^2}{n}} = \frac{\sigma}{\sqrt{n}}$$

Also from last time, when the X_i are all normally distributed so is any linear combination

So we conclude for this case that

$$\bar{X} \sim N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$$

ie the sampling distribution for $\bar{X}$ when original data is normal is also normal.

what is the sampling distribution of T_n if $X_1, \ldots, X_n$ are iid $N(\mu, \sigma)$

$$E\left[\sum_{i=1}^{n} X_i\right] = \sum_{i=1}^{n} E[X_i]$$

$$= \underbrace{\mu + \mu + \cdots + \mu}_{n \text{ of them}} = n\mu$$

$$Var(T_n) = Var\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} Var(X_i)$$

$$= \sigma^2 + \cdots + \sigma^2$$

$$= n\sigma^2$$

so

$$SD(T_n) = \sqrt{Var(T_n)} = \sqrt{n\sigma^2} = \sqrt{n}\,\sigma$$

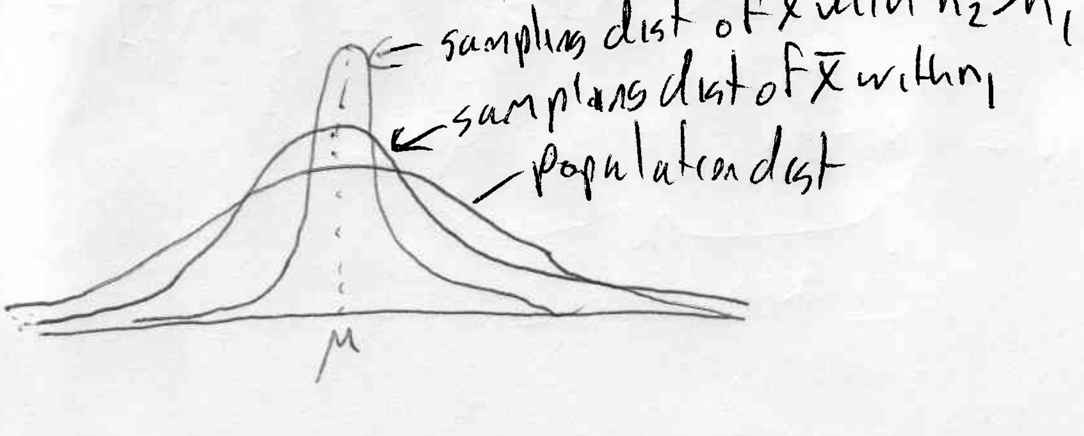
Again T_n is a linear combination of X_i so T_n is also normally distributed. We conclude from this that

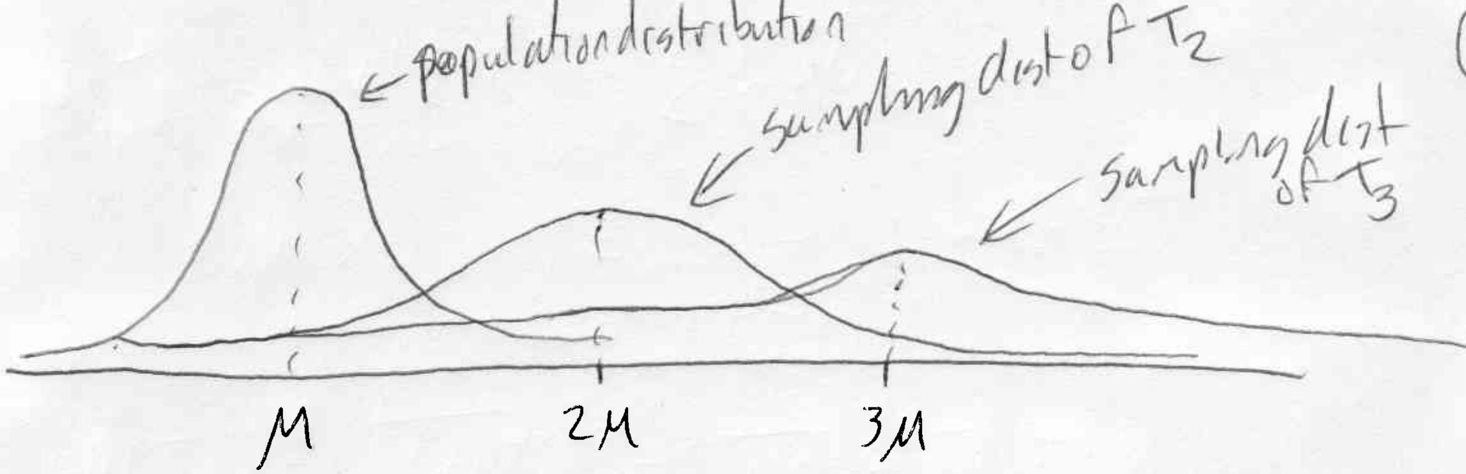
$$T_n \sim N(n\mu, \sqrt{n}\,\sigma)$$

From these results what do we learn?

$\bar{X}$ becomes less variable as n increases

T_n becomes more variable as n increases.





what do we do if $X_1, \ldots, X_n$ are iid but not normal? The answer is that we use a very important theorem in statistics known as the central limit theorm

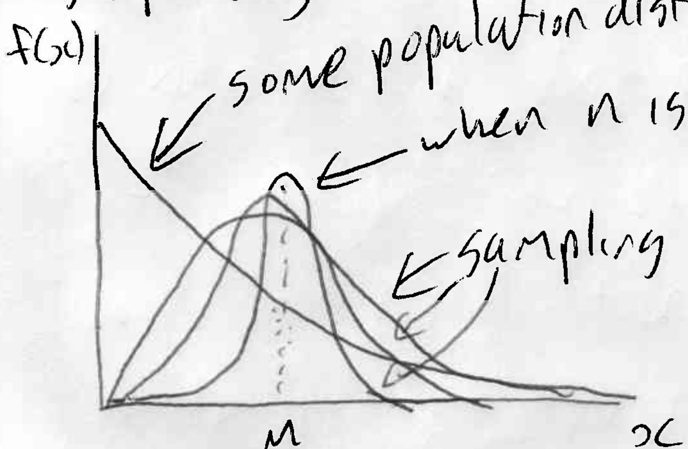
<u>Central Limit Theorem</u> (CLT)

Let $X_1, X_2, \ldots X_n$ be a random sample from some distribution with mean μ and variance σ^2. Then if the sample size, n, is large $\bar{X}$ has approximately normal distribution with mean $\mu_{\bar{X}} = \mu$ and ~~st~~variance $\sigma^2_{\bar{X}} = \frac{\sigma^2}{\sqrt{n}}$. Similarly T_n has approximately

Normal distribution with mean $\mu_{T_n} = n\mu$ and variance $\sigma^2_{T_n} = n\sigma^2$. The approximations improve as n increases.

<u>Proof</u> Omitted because it involves advanced mathematics

Graphically



<u>Note</u> that the binomial approximation using the normal distribution we discussed earlier falls out as a consequence of the CLT

Recall that a Binomial RV X with parameters n, p can be represented as the sum of n independent bernoulli r.v each with mean p.

ie $X = X_1 + \cdots X_n$ $\quad E[X] = E[X_1 + \cdots + X_n] = np$

$Var(X) = Var(X_1 + \cdots + X_n) = np(1-p)$

so when n is large $X = X_1 + \cdots + X_n$ is approximately Normal with mean np and sd $\sqrt{np(1-p)}$

and another* interesting quantity is

$$\hat{p} = \frac{X}{n} \quad \text{the sample proportion}$$

$$E[\hat{p}] = E\left[\frac{X}{n}\right] = \frac{1}{n}E[X] = \frac{1}{n}E[X_1 + \cdots + X_n]$$

$$= \frac{1}{n} np = p$$

$$Var(\hat{p}) = Var\left(\frac{X}{n}\right) = \frac{1}{n^2} Var(X) = \frac{1}{n^2} Var(X_1 + \cdots + X_n)$$

$$= \frac{1}{n^2} np(1-p)$$

$$= \frac{p(1-p)}{n}$$

Some examples

1. Suppose a bulldozer is removing rubble and placing it into a dump truck. Each load has mean weight 1.1 tons and standard deviation 0.3 tons. A dump truck can be loaded with no more than 50 tons of material before it is overloaded. What is the probability that the truck is not

overloaded after 47 loads of material? Assume independence

Solution Let T_n = total weight after n loads

so

$$T_{47} = X_1 + \dots X_{47}$$

$$E[T_{47}] = 47(1.1) = 51.7$$

$$Var(T_{47}] = 47(0.3)^2 = 4.23$$

$$SD(T_{47}) = 2.0567 \ (4dp)$$

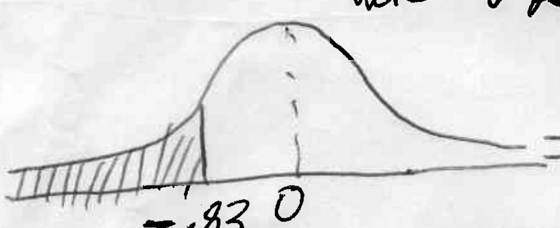
By CLT T_{47} is approx $N(51.7, 2.0567)$

so

$$P(\text{not overloaded}) = P(T_{47} \leq 50)$$

note approximation here

$$\approx P\left(\frac{T_{47} - 51.7}{2.0567} \leq \frac{50 - 51.7}{2.0567}\right)$$



$$= P(Z < -0.83)$$

$$= .2033$$

2. Suppose that a factory produces nuts and bolts. Quality control for the factory says that ideally the nuts should have mean diameter 12mm with sd .05 mm.

The inspection unit takes a SRS of 100 nuts each hour from the production line and uses it to judge whether the product is being produced within specifications. In particular, if it is very unlikely that the mean diameter of the 100 nuts is the value it takes (or more extreme) then it is likely that the machine is out of spec. Suppose that $\bar{X} = 12.014$. Would we conclude that the machine needs adjustment?

Since n is large $\bar{X}$ has approximately $N(12, \frac{.05}{\sqrt{100}})$ by CLT

$$P(\bar{X} \geq 12.014) \approx P\left(\frac{\bar{X}-12}{.005} \geq \frac{12.014-12}{.005}\right)$$

$$= P(Z \geq 2.8)$$

$$= 1 - \Phi(2.8)$$

$$= 1 - .9974$$

$$= .0026$$

Since this probability is small, it is likely that the machine is out of spec.