Lecture 20

Last time we discussed the sampling distribution of the sample proportion. Today we begin our discussion on sampling distribution of the mean.

Suppose that $X_1, X_2, \ldots X_n$ are measurements of a r.v. X on n individuals choosen using a SRS. Note that X_i should therefore by independent so sample mean

$$\bar{x} = \frac{\sum X_i}{n} = \frac{1}{n}\left[X_1 + X_2 + \cdots + X_n\right]$$

by the rules about expected (mean) values of r.vs

$$\mu_{\bar{x}} = \frac{1}{n}(\mu_X + \mu_X + \cdots \mu_X)$$

$$= \frac{n\mu_X}{n} = \mu_X$$

and the rules for variances

$$\sigma^2_{\bar{x}} = \left(\frac{1}{n}\right)^2 (\sigma_X^2 + \sigma_X^2 + \cdots + \sigma_X^2) = \left(\frac{1}{n}\right)^2 n\sigma_X^2 = \frac{\sigma_X^2}{n}$$

ie

$\mu_{\bar{x}} = \mu_x$ (mean value of $\bar{x}$ is population mean of original data)

$\sigma_{\bar{x}} = \dfrac{\sigma_x}{\sqrt{n}}$ (std deviation of $\bar{x}$ is population std deviation of original data divided by square root of sample size)

<u>Example</u>

Suppose a chemistry measurement has mean 160 mg and standard deviation 1.5 mg

What is the mean and standard deviation of a single measurement?

$\mu_x = 160 \quad \sigma_x = 1.5$

What is the mean and standard deviation of the mean of 3 measurements?

$\mu_{\bar{x}} = 160 \quad \sigma_{\bar{x}} = \dfrac{1.5}{\sqrt{3}}$

what is the mean and standard deviation of the mean of 10 measurements

$\mu_{\bar{x}} = 160 \quad \sigma_{\bar{x}} = \dfrac{1.5}{\sqrt{10}}$

What is mean and standard deviation of the sample mean of 25 measurements?

$\mu_{\bar{x}} = 160 \quad \sigma_{\bar{x}} = \frac{1.5}{\sqrt{25}} = .3$

Notice that the standard deviation of the sample mean goes down (ie gets smaller) as the sample size increases.

What is the distribution of $\bar{X}$?

If the original $X_1, \ldots X_n$ are all from a normal distribution with mean μ and standard deviation σ then $\bar{X}$ has exactly the normal distribution with mean μ and standard deviation $\frac{\sigma}{\sqrt{n}}$

Notation

$N(\mu, \sigma)$ is read as "Normal distribution with mean μ and standard deviation σ"

So For example

$N(\mu, \frac{\sigma}{\sqrt{n}})$ would be read as "Normal distribution with mean μ and standard deviation $\frac{\sigma}{\sqrt{n}}$"

<u>What if the distribution of $X_1, \ldots, X_n$ is not normal?</u>

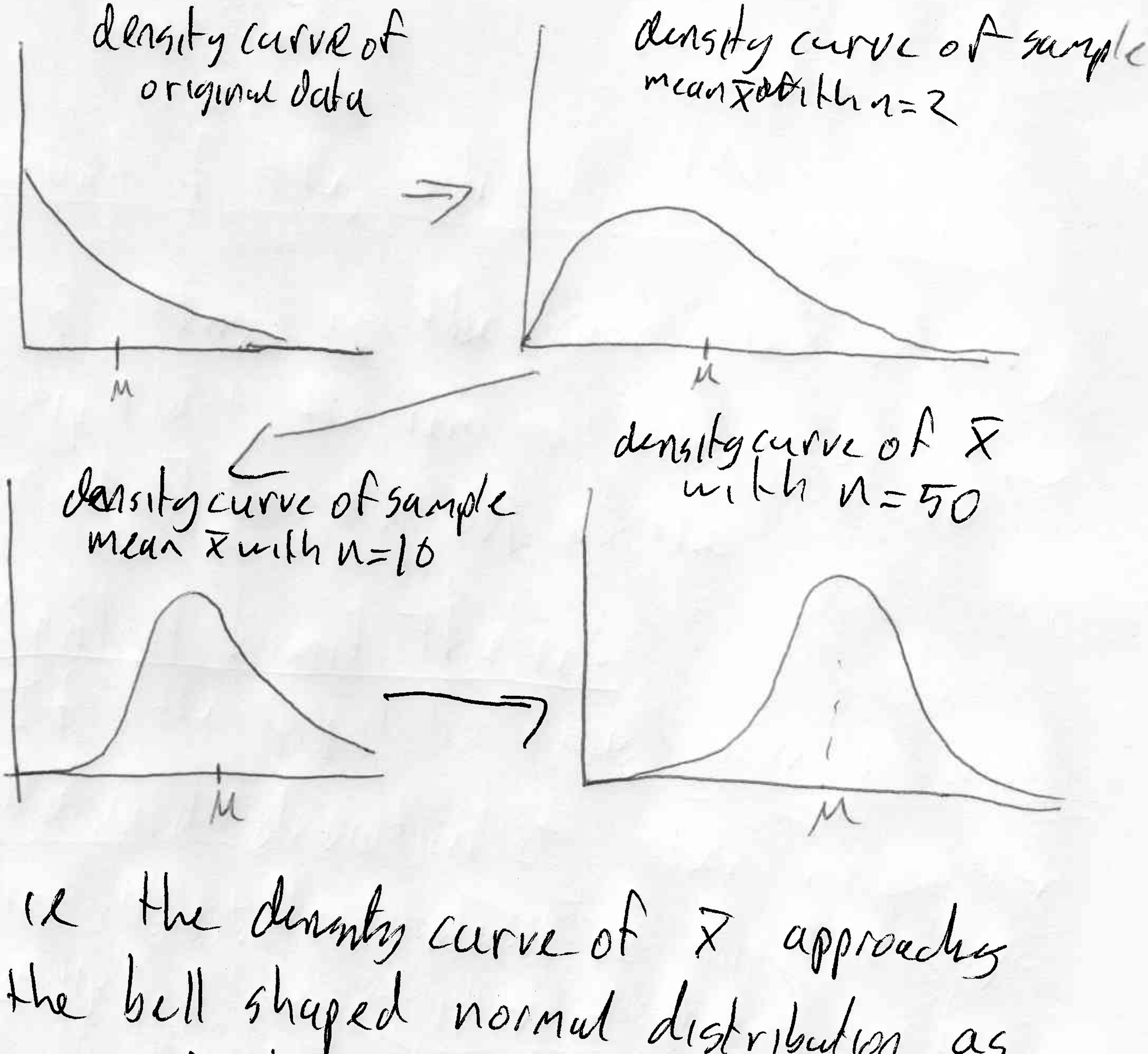
If this is the case then a very important theorem in statistics helps us out. It is known as the.......

<u>Central Limit Theorem</u>

If the population has mean μ and standard deviation σ, then no matter the distribution when we take a large sample the sampling distribution of $\bar{X}$ is approximately $N(\mu, \frac{\sigma}{\sqrt{n}})$

In otherwords when the sample size is large enough we can use the normal distribution when dealing problems about $\bar{x}$.

Graphical representation of CLT



ie the density curve of $\bar{x}$ approaches the bell shaped normal distribution as n gets larger.

Example

A machine produces parts for a car. This part has mean weight 150g and, because the machine has variability, standard deviation 10g. Each hour a SRS of 100 of the parts is taken to make sure the machine is still ~~still~~ producing quality parts. Suppose that an entire hours output will have to be discarded if the mean weight of the sample is below 147.5g. what is the probability of this happening?

$$\mu_{\bar{x}} = 150g \quad \sigma_{\bar{x}} = \frac{10}{\sqrt{100}} = \frac{10}{10} = 1g$$

because $n = 100$ is large the CLT tells us that $\bar{x}$ is approximately Normally distributed with mean 150 and stddev 1g.

approximately equal due to CLT

$$P(X < 147.5) \approx P\left(\frac{X-150}{1} < \frac{147.5-150}{1}\right)$$

$= P(z < -2.5)$

$= .0062$ (From table)

Example

An experiment to compare the nutrional value of ~~two~~ two feeding regimins for chickens is carried out on 40 chicks. At random 20 are assigned to diet 1 and 20 to diet 2. At the end of the experiment the weight gains are measured. Call $\bar{x}$ the mean weight gain for diet 1 and $\bar{y}$ the mean weight gain for diet 2. At the end of experiment inference about which diet is better will be based on the difference $\bar{y}-\bar{x}$.

Suppose $\mu_x = 360$ and $\sigma_x = 55$ and $\mu_y = 385$ $\sigma_y = 50$.

What is the mean of $\bar{y}-\bar{x}$?

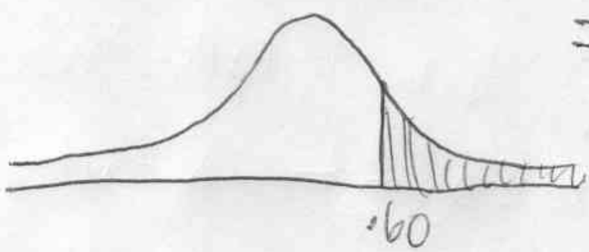
$$\mu_{\bar{y}-\bar{x}} = \mu_{\bar{x}} - \mu_{\bar{y}} = 385 - 360 = 25g$$

$$\sigma_{\bar{y}-\bar{x}} = \sqrt{\sigma_{\bar{y}}^2 + \sigma_{\bar{x}}^2} = \sqrt{\frac{55^2}{20} + \frac{50^2}{20}} = 16.6208$$

Assuming that $\bar{x}$ and $\bar{y}$ are normally distributed then $\bar{y} - \bar{x}$ is normally distributed with mean 25g and std dev 16.6208

what is the probability that diet 2 has average weight gain of more than 35g above that of diet 1?

$$P(\bar{y} - \bar{x} \geq 35) = P\left(\frac{(\bar{y} - \bar{x}) - 25}{16.6208} \geq \frac{35 - 25}{16.6208}\right)$$



$$= P(Z \geq .60)$$

$$= 1 - P(Z < .60)$$

$$= 1 - .7257$$

$$= .2743$$