PSTAT 120B - Outline [Shravan Shenoy]

1) - A random sample is a collection of independent and identically distributed (iid) random variables
   - $Y_1, Y_2, Y_3 \ldots Y_n$ are a random sample if they have properties that allow them to be written as:

$$Y_1, \ldots, Y_n \overset{iid}{\sim} f(y)$$

2) - A statistic is a function of the observable random variables in a sample and known constants
   - any function $h(\mathbf{Y})$

3) - The term "sampling distribution" refers to the probability distributions of statistics.
   - The sampling distribution of a statistic provides a theoretical model for the relative frequency histogram of the possible values of the statistic that would observed through repeated sampling

4) Central Limit Theorum - The sample average of n random variables converges to a Gaussian distribution, which can be stated in several equivalent ways:

$$\cdot \sqrt{n}\,(\bar{Y}_n - \mu) \xrightarrow{d} N(0, \sigma^2) \qquad \cdot \frac{\sqrt{n}\,(\bar{Y}_n - \mu)}{\sigma} \xrightarrow{d} N(0,1)$$

$$\cdot \frac{\bar{Y}_n - \mu}{\sigma / \sqrt{n}} \xrightarrow{d} N(0,1)$$

5) When $Y_1, \ldots Y_n \overset{iid}{\sim} N(\mu, \sigma^2)$, the sampling distribution of $\bar{Y} = n^{-1} \sum_{i=1}^{n} Y_i$ is normal with mean $\mu_{\bar{Y}} = \mu$ and variance $\sigma^2_{\bar{Y}} = \sigma^2 / n$

6) When $Y_1, \ldots Y_n \overset{iid}{\sim} f(y)$ with common mean μ and var σ^2,

- The mean and variance of sample mean $\overline{Y}$ is:

  - mean of $\overline{Y} = \mu$ · variance $(Y_i) = \frac{\sigma^2}{n}$

- The approximate sampling distribution of $\overline{Y} = n^{-1} \sum_{i=1}^{n} Y_i$ is:

  - $\frac{\overline{Y} - \mu}{\sigma/\sqrt{n}} \sim N(0,1)$ or $\overline{Y} \sim N(\mu, \sigma^2/n)$

PSTAT 120B HW 2

Shravan Shenoy

1) $\mathbb{E}(\sum_{i=1}^{n} a_i Y_i)$

$= \mathbb{E}(a_1 Y_1 + a_2 Y_2 \dots a_n Y_n)$

$= \int\int (a_1 Y_1 + a_2 Y_2 \dots a_n Y_n) f(Y_1 \dots Y_n) dy_1 \dots dy_n$

$= \int\int a_1 Y_1 f(Y_1 \dots Y_n) dy_1 \dots dy_n +$
$\int\int a_2 Y_2 f(Y_1 \dots Y_n) dy_1 \dots dy_n +$
$\int\int a_n Y_n f(Y_1 \dots Y_n) dy_1 \dots dy_n$

$= a_1 \int Y_1 f_1(Y_1) dy_1 + a_2 \int Y_2 f_2(Y_2) dy_2 + a_n \int Y_n f_n(Y_n) dy_n$

$= \boxed{\sum_{i=1}^{n} a_i \mathbb{E} Y_i}$

2) $Var[\sum_{i=1}^{n} a_i Y_i]$

using $Var(X) = E[(X - E(X))^2]$

$\Rightarrow E[(\sum_{i=1}^{n} a_i Y_i - \sum_{i=1}^{n} a_i Y_i E[Y_i])^2]$

$\Rightarrow E[\sum_{i=1}^{n} a_i^2 (Y_i - E(Y_i))^2 + \sum_{i=1}^{n}\sum_{j=1}^{n} a_i a_j (Y_i - E[Y_i])(Y_j - E[Y_j])$

$\Rightarrow E(\sum_{i=1}^{n} a_i^2 (Y_i - E[Y_i])^2) + E(\sum_{i=1}^{n}\sum_{j=1}^{n} a_i a_j (Y_i - E[Y_i])$

$\Rightarrow \sum_{i=1}^{n} a_i^2 E[(Y_i - E[Y_i])^2] + \sum_{i=1}^{n}\sum_{j=1}^{n} a_i a_j E[(Y_i - E[Y_i])(Y_j - E[Y_j])]$

$= \boxed{\sum_{i=1}^{n} a_i^2 \, var(Y_i) + 2\sum_{i<j} a_i a_j \, cov(Y_i, Y_j)}$

3) $Y_1, Y_2, Y_3 \dots Y_n \overset{iid}{\sim} N(\mu, \sigma^2)$

$\bar{Y} = n^{-1} \sum_{i=1}^{n} Y_i$

$\mathbb{E}^{\wedge}(t(\frac{X-a}{b})) = e^{\frac{-at}{b}} \mathbb{E}^{\wedge}(\frac{t}{b} X)$

$\Rightarrow \boxed{\bar{Y} \sim N(\mu, \sigma^2/n)}$

- The MGF for the normal dist. with parameters μ and σ is $m(t) = e^{\mu t + \sigma^2 t^2/2}$
- Since the Y_i's are independent, the mgf of $Z = \frac{\bar{Y} - \mu}{\sigma/\sqrt{n}}$ is: $M_Z(t) = E(e^{\bar{Y} t})$

$\Rightarrow \prod_{i=1}^{n} E(e^{\frac{1}{n} t \bar{Y}}) = e^{\mu t \sum \frac{1}{n} + (t^2 \sigma^2/2) \sum \frac{1}{n^2}}$

$\boxed{\sim Dist(mean\ \mu \sum \frac{1}{n}, var\ \sigma^2 \sum \frac{1}{n^2})}$

4) $Y_1 \dots Y_n$ (random sample)

common mean μ, common var σ^2

$P(\bar{Y} \leq x) = P((Y_1 + Y_2 \dots + Y_n)/n \leq x)$
$= P(Y_1 \dots Y_n \leq nx)$

$E[Y_i] = \mu$; $\boxed{E[\bar{Y}] = n\mu}$

$Var[Y_i] = \sigma^2$; $\boxed{Var(\bar{Y}) = n\sigma^2}$

$P\left(\frac{(Y_1 \dots Y_n) - n\mu}{n\sigma^2} \leq \frac{nx - n\mu}{\sqrt{n\sigma^2}}\right)$

$\Rightarrow P\left(\frac{(Y_1 \dots Y_n) - n\mu}{n\sigma^2} \leq \frac{n(x - \mu)}{\sqrt{n}\,\sigma}\right)$

$\lim_{n \to \infty} \frac{(Y_1 + \dots Y_n) - n\mu}{n\sigma^2} \to N(0,1)$

$\Rightarrow P(\bar{Y} \leq x) = P(Z \leq \frac{\sqrt{n}(x - \mu)}{\sigma})$

$= \boxed{F_Z\left(\frac{\sqrt{n}(x - \mu)}{\sigma}\right)}$

5) $\mu = 7 \quad \sigma = .5$ 64 workers

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{.5}{\sqrt{64}}$$

$$z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}}$$

$$P(\bar{x} \le 6.90) = P\left(\frac{\bar{x}-\mu}{\sigma/\sqrt{n}} \le \frac{6.90-7}{.5/8}\right)$$

$$= P\left(\frac{\bar{x}-\mu}{\sigma/\sqrt{n}} \le -.10/.0625\right)$$

$$= P(z \le -1.6)$$

$$= \Phi(-1.6)$$

$$= 1 - \Phi(1.6)$$

$$= \boxed{.0548}$$