

Sampling Distributions & Control

Limit Theorem

Statistics concerns with conclusions & predictions resulting from chance outcomes

↳ these chance outcomes constitute a subset → samples

by using random samples we draw generalizations about population from which they came from

by using random variables we built certain functions of them these functions are used to estimate or make decisions about unknown population parameters

For example suppose that we want to estimate a population mean

If we obtain random sample of n observations $y_1, y_2, y_3, \dots, y_n$ then estimator of μ from sample mean is

$$\bar{y} = \frac{1}{n} \sum_{i=1}^n y_i$$

Goodness of this estimate depends on behavior of random var $y_1, y_2, \dots, y_n$ & effect of this behavior has on $\bar{y}$

Random variable $\bar{y}$ is a function of random variables $y_1, y_2, \dots, y_n$ & constant sample size, n .

Random var. $\bar{y}$ is an example of a statistic

A statistic is a function of observable random variables in a sample & unknown constants

Statistics are used to make inferences about unknown population parameters

Statistic itself is a random variable since it is a function of random variables

Probability distr. of statistic is called as its sampling distr.

Sampling distr. of a statistic gives a theoretical model of ~~positive~~ frequency histogram of possible values of statistic that we would observe through repeated sampling

Most of relative frequency distributions can be modelled adequately by a normal prob. distr.

Observable random variables in a random sample $y_1, y_2, \dots, y_n$ are independent with a common normal density function. If $y_1, y_2, \dots, y_n$ be a random sample of size n from a normal distr. with mean μ and variance σ^2 then the sample mean

$$\bar{y} = \frac{1}{n} \sum_{i=1}^n y_i \text{ is}$$

normally distributed with mean

$$E(\bar{y}) = \mu \quad V(\bar{y}) = \frac{\sigma^2}{n}$$

Under conditions of this theorem $\bar{y}$ is normally distributed with mean $\mu_{\bar{y}} = \mu$ & variance $\sigma_{\bar{y}}^2 = \sigma^2/n$

$$Z = \frac{\bar{y} - \mu_{\bar{y}}}{\sigma_{\bar{y}}} = \frac{\bar{y} - \mu}{\sigma/\sqrt{n}} = \sqrt{n} \left(\frac{\bar{y} - \mu}{\sigma} \right) \sim N(0,1)$$

Ex:

Show that if $y_{11}, y_{12}, \dots, y_{1n_1}$ and $y_{21}, y_{22}, \dots, y_{2n_2}$ are independent random variables with first n_1 constitutes a random sample from an infinite population with mean μ_1 & σ_1^2 and other n_2 constitutes a random sample from an infinite population with mean μ_2 & variance σ_2^2 . Then

$$a) E(\bar{y}_1 - \bar{y}_2) = \mu_1 - \mu_2$$

$$b) \text{Var}(\bar{y}_1 - \bar{y}_2) = \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}$$

$$\begin{aligned} a) E(\bar{y}_1 - \bar{y}_2) &= \frac{1}{n_1} \sum_{i=1}^{n_1} E(y_{1i}) - \frac{1}{n_2} \sum_{i=1}^{n_2} E(y_{2i}) \\ &= \frac{n_1}{n_1} \cdot \mu_1 - \frac{n_2}{n_2} \mu_2 = \mu_1 - \mu_2 \end{aligned}$$

$$\begin{aligned}
 b) \text{Var}(\bar{y}_1 - \bar{y}_2) &= \sum \frac{1}{n_1^2} \underbrace{\text{Var}(y_{1i})}_{\Sigma = n_1 \sigma_1^2} + \sum \frac{1}{n_2^2} \text{Var}(y_{2i}) \\
 &= \frac{1}{n_1^2} n_1 \sigma_1^2 + \frac{1}{n_2^2} n_2 \sigma_2^2 \\
 &= \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}
 \end{aligned}$$

Ex:

If $y_1, y_2, \dots, y_n$ are independent random variables having identical Bernoulli distributions with parameter θ then sample mean is proportion of successes in n trials detected by θ verify that

$$a) E(\bar{\theta}) = \theta \quad b) \text{Var}(\bar{\theta}) = \frac{\theta(1-\theta)}{n}$$

$M_Y = [1 + \theta(e^t - 1)] \rightarrow$ moment generating func. of bernoulli

$$M_{\bar{Y}} = [1 + \theta(e^{t/n} - 1)]^n$$

$$M' = n [1 + \theta(e^{t/n} - 1)]^{n-1} \cdot \frac{\theta}{n} e^{t/n} \quad \left. \begin{array}{l} \text{derivative} \\ \text{bernoulli} \\ p(y) = p^y (1-p)^{1-y} \\ \mu = n \cdot p, \sigma^2 = npq \end{array} \right\}$$

$$M'(0) = \theta \Rightarrow E(\bar{\theta}) = \theta$$

$$\begin{aligned}
 M'' &= \theta [1 + \theta(e^{t/n} - 1)]^{n-1} \cdot \frac{1}{n} e^{t/n} \\
 &\quad + \theta e^{t/n} (n-1) [1 + \theta(e^{t/n} - 1)]^{n-2} \cdot \frac{\theta}{n} e^{t/n}
 \end{aligned}$$

$$M''(0) = \frac{\theta}{n} + \frac{\theta^2(n-1)}{n}$$

$$\text{Var}(\theta) = \sigma^2 = \frac{\theta}{n} + \frac{\theta^2(n-1)}{n} - \theta^2$$

$$= \frac{\theta + n\theta^2 - \theta^2 - n\theta^2}{n} = \frac{\theta(1-\theta)}{n}$$

Ex: Show that formula of sample variance can be written as

$$s^2 = \frac{n \left(\sum_{i=1}^n y_i^2 \right) - \left(\sum_{i=1}^n y_i \right)^2}{n(n-1)}$$

$$s^2 = \sum_{i=1}^n \frac{(y_i - \bar{y})^2}{n-1} = \frac{1}{n-1} \left(\sum_{i=1}^n y_i^2 - 2\bar{y} \sum_{i=1}^n y_i + n\bar{y}^2 \right)$$

$$= \frac{1}{n-1} \left(\sum y_i^2 - 2n\bar{y}^2 + n\bar{y}^2 \right)$$

$$= \frac{\sum y_i^2}{n-1} - \frac{n\bar{y}^2}{n-1} = s^2$$

Let $y_1, y_2, \dots, y_n$ be a random sample of size n from a normal distr with mean μ & variance σ^2 and $z_i = \frac{(y_i - \mu)}{\sigma}$ are independent identically distributed

$N(0,1) \quad i = 1, 2, \dots, n$

$$\sum_{i=1}^n z_i^2 = \sum_{i=1}^n \left(\frac{y_i - \mu}{\sigma} \right)^2 \sim \chi_n^2$$

chi-square with n degrees of freedom

If $y \sim N(0,1)$, $y^2 \sim \chi_{v=1}^2$

If y has a standard normal distr, its square has a chi-square distr with 1 degree of freedom

If $y_1, y_2, \dots, y_n$ are independent random variables having standard normal distributions, then

$$X = \sum_{i=1}^n y_i^2 \sim \chi_{v=n}^2$$

If $y_1, y_2, \dots, y_n$ are independent random variables having χ^2 distr with $v_1, v_2, \dots, v_n$ degrees of freedom then

$$X = \sum_{i=1}^n y_i \sim \chi_{v_1+v_2+\dots+v_n}^2 \text{ degrees of freedom}$$

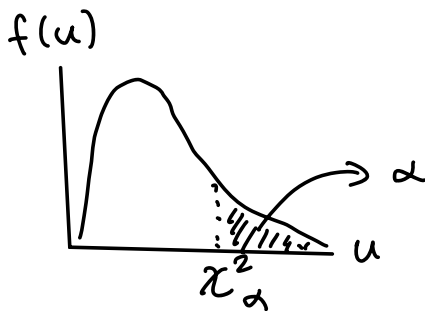
If u_1 & u_2 are independent variables and $u_1 \sim \chi^2_{v_1}$
 & $u_2 \sim \chi^2_{v_2} \Rightarrow u_1 + u_2 \sim \chi^2_{v_1+v_2}$

If $\bar{y}$ & s^2 are mean & variance of a random sample of size n from a normal population with mean μ & variance σ^2 , then

a) $\bar{y}$ & s^2 are independent of each other

$$b) W = \frac{(n-1) s^2}{\sigma^2} = \frac{1}{\sigma^2} \sum_{i=1}^n (u_i - \bar{y})^2 \sim \chi^2_{n-1}$$

$$s^2 = \frac{1}{n-1} \sum_{i=1}^n (u_i - \bar{y})^2$$



ratio of
sample variance to
population variance
= W

Ex: Find median value of χ^2 distr. corresponding to

a) 3 b) 28 c) 40 degrees of freedom (df)

$$a) \chi^2_{0.50, 3} = 8.34$$

$$b) \chi^2_{0.50, 28} = 27.3$$

$$c) \chi^2_{0.50, 40} = 39.3$$

Solve using the table!

$$\sqrt{n} (\bar{y} - \mu) / \sigma \sim N(0, 1)$$

difficulty is that σ is unknown in most applications
 when σ is unknown, it should be replaced by an estimate, s ,

$$\sqrt{n} \left(\frac{\bar{y} - \mu}{s} \right) \rightarrow \text{inference}$$

$\sqrt{n} \left(\frac{\bar{y} - \mu}{s} \right) \sim \text{Student's } t\text{-distribution}$
 $n-1$ degrees of freedom

If W & Z are independent random variables $W \sim \chi^2_v$,
 $Z \sim N(0,1)$

$$T = \frac{Z}{\sqrt{W/v}} \sim t_v$$

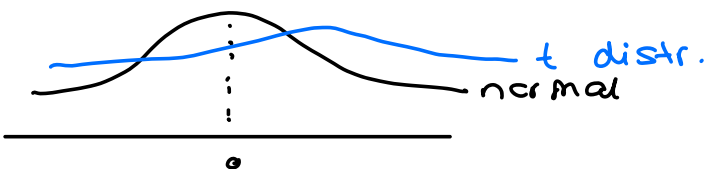
$$T = \frac{\bar{z}}{\sqrt{w/v}} = \frac{\sqrt{n} (\bar{y} - \mu) / s}{\sqrt{[(n-1) s^2 / (r^2)] / (n-1)}} = \sqrt{n} \left(\frac{\bar{y} - \mu}{s} \right) \sim t_{n-1}$$

$$f(t) = \frac{\Gamma\left(\frac{v+1}{2}\right)}{\sqrt{\pi v} \Gamma\left(\frac{v}{2}\right)} \cdot \left(1 + \frac{t^2}{v}\right)^{-\frac{v+1}{2}} \quad \text{for } -\infty < t < \infty$$

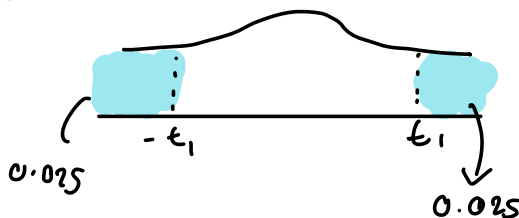
Like s.t. normal density func. t density func. is symmetric about zero

A t -distributed r.v. has same expected value as a s.t. normal r.v.

s.t. normal r.v. $\rightarrow$ has variance 1
 t distributed r.v. $\rightarrow$ has variance greater than 1



Ex:



find value of t_1 ^{with 9 degrees of freedom} for which
a) the shaded areas on right is 0.05

b) total shaded area is 0.05
c) total unshaded area is 0.95
d) shaded area on left is 0.01

e) the area to the left of $t_1 = 0.90$

a) If total shaded area on right is 0.05 then area to the left of t_1 is $1 - 0.05 = 0.95$. t_1 represents 95th percentile therefore we should look at $t_9^{0.95} = 1.83$

b) If total shaded area is 0.05 then shaded area on right is 0.025 by symmetry. The area left of t_1 is $1 - 0.025 = 0.975 \rightarrow t_1$ represents 97.5th percentile $t_9^{0.975} = 2.26$

c) $1 - 0.99 = 0.01 \Rightarrow 0.01/2 = 0.005$

$$t_9^{0.995} = 3.25$$

d) $t_9^{0.99} = 2.82$ for the left it is -2.82

e) $t_9^{0.90} = 1.38$

If S_1^2 & S_2^2 are variances of independent random samples of size n_1 & n_2 from normal populations with variances σ_1^2 & σ_2^2

$$F = \frac{S_1^2 / \sigma_1^2}{S_2^2 / \sigma_2^2} = \frac{\sigma_2^2 S_1^2}{\sigma_1^2 S_2^2} \sim F_{n_1-1, n_2-1}$$

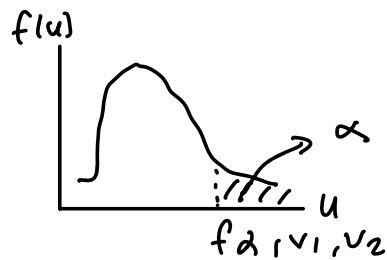
If W_1 & W_2 are independent random variables having χ^2 distributions with v_1 & v_2 degrees of freedom

$$F = \frac{W_1 / v_1}{W_2 / v_2} \sim f_{v_1, v_2}$$

prob density of f . distr is

$$f(f) = \frac{\Gamma\left(\frac{v_1 + v_2}{2}\right)}{\Gamma\left(\frac{v_1}{2}\right)\Gamma\left(\frac{v_2}{2}\right)} \cdot \left(\frac{v_1}{v_2}\right)^{v_1/2} \cdot f^{\frac{v_1}{2}-1} \left(1 + \frac{v_1}{v_2} f\right)^{-\frac{1}{2}(v_1 + v_2)}$$

for $f > 0$



Ex: If X has an F distr. with v_1 & v_2 d.f. show that $Y = \frac{1}{X}$ has F distr. with v_2 & v_1 d.f.

$$F = \frac{W_1 / v_1}{W_2 / v_2} \quad \begin{array}{l} W_1 \sim \chi^2_{v_1} \\ W_2 \sim \chi^2_{v_2} \end{array}$$

$$\frac{1}{F} = \frac{W_2 / v_2}{W_1 / v_1} \rightarrow \frac{1}{F} \sim F_{v_2, v_1}$$

Ex:

$$F_{5,9}^{0.105} = 3.48$$

$$F_{7,12}^{0.10} = 4.64$$

$$F_{2,14}^{0.005} = 3.74$$