

Sampling Distributions



Sampling Distribution of \overline{X}

- The probability of a statistic is called the sampling distribution of the statistic. Let us consider our interest is \overline{X} .
- If $X_1, X_2, ..., X_n \sim N(\mu, \sigma^2)$, then $\overline{X} \sim N(\mu, \sigma^2/n)$.
- **Central Limit Theorem:** If random samples of size n observations are drawn from a population with mean μ and Std. σ , then when n large $(n \ge 30)$, the sampling distribution of \overline{X} will be approximately a normal distribution with mean μ and $\operatorname{Std} = \sigma/\sqrt{n}$.
- Example: The time required by workers to complete an assembly job has a mean of 50 minutes and Std. of 8 minutes. The supervisor intends to record 60 workers to complete one assembly job. What is the probability that the sample mean will be more than 52 minutes?

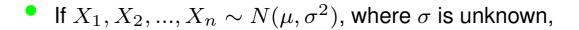
$$P(\overline{X} > 52) = P(Z > \frac{52 - 50}{8/\sqrt{60}})$$

$$= P(Z > 1.94)$$

$$= 1 - 0.9738 = 0.0262.$$



σ unknown



$$\frac{\overline{X} - \mu}{\sigma / \sqrt{n}} \sim t_{n-1}$$

t-distribution is symmetric distribution similar to normal but not normal. Its peak is lower and his tails are higher. As $n \to \infty$, the t-distribution converges to normal distribution.

Example:

Suppose that the weights of new born babies are normally distribution with mean 3 kgs. A random sample of size 10 is taken and showed that its standard deviation is 2.

- (a) Find the probability that the sample average is below 4.16 kgs.
- (b) What is the 90th percentile of the distribution of \overline{X} ?



Cont./Example



$$P(\overline{X} < 4.16) = P(t < \frac{4.16 - 3}{2/\sqrt{10}})$$

$$= P(t < 1.834)$$

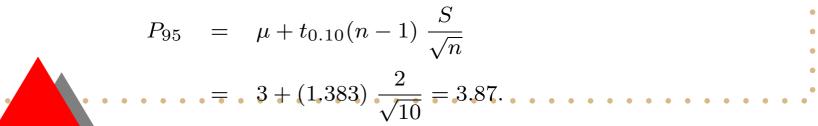
$$= 0.95.$$

(b) What is the 90th percentile of the distribution of X? Since

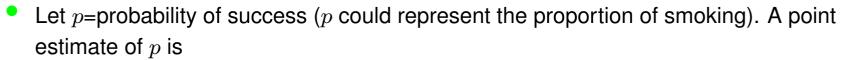
$$\frac{\overline{X} - \mu}{\sigma / \sqrt{n}} \sim t_9,$$

the 90th percentile of t-distribution with 9 degrees of freedom is 1.383. Therefore the 90th percentile of \overline{X} is





Distribution of \hat{p}



$$\hat{p} = \frac{X}{n} = \frac{\text{\# of successes}}{n}$$

Note X=number of successes $\sim Bin(p, npq)$. Then

$$E(\frac{X}{n}) = \mu = \frac{np}{n} = p,$$

and

$$\sigma_{\hat{p}}^2 = Var(\hat{p}) = \frac{npq}{n^2} = \frac{pq}{n}.$$

• For large n, $\hat{p} \sim N(p, \frac{pq}{n})$. Further,

$$\frac{\hat{p} - p}{\sqrt{pq/n}} \sim N(0, 1)$$



Example on \hat{p}

Example: 40% of the students at University of Jordan are smoking. If a sample of size 100 students is taken. Compute

- (a) The probability at least 50% of the student are smoking.
- (b) The probability that the sample proportion will be between 0.45 and 0.55.

Solution:

(a) The probability at least 50% of the student are smoking.

$$P(\hat{p} > 0.5) = P(Z > \frac{0.5 - 0.4}{\sqrt{\frac{(0.4)(0.6)}{100}}})$$
$$= P(Z > 2.04) = 1 - 0.9793 = 0.0207.$$

(b) The probability that the sample proportion will be between 0.45 and 0.55.

$$P(0.45 < \hat{p} < 0.55) = P(\frac{0.55 - 0.40}{\sqrt{\frac{(0.4)(0.6)}{100}}} < Z < \frac{0.55 - 0.4}{\sqrt{\frac{(0.4)(0.6)}{100}}})$$
$$= P(1.02 < Z < 3.06) = 0.9989 - 0.8461 = 0.1528.$$

Distribution of Sample Variance

- Chi-square distribution is not symmetric distribution and it is skewed to the right.
- If $Z \sim N(0,1)$, then $Z \sim \chi_1^2$ -chi-square with 1 degrees of freedom.
- If $Z_1, Z_2, ..., Z_n \sim N(0, 1)$, then $Z_1^2 + Z_2^2 + ... + Z_n^2 \sim \chi_n^2$.
- Note that

$$\frac{(n-1) S^2}{\sigma^2} = \sum_{i=1}^n \left(\frac{X_i - \overline{X}}{\sigma}\right)^2 \sim \chi_{n-1}^2$$

Example: In a sample of size n=6 is drawn from a population with $\sigma^2=10$. Find $P(S^2 > 18.4727).$

$$P(S^{2} > 18.4727) = P(\frac{(n-1)S^{2}}{\sigma^{2}} > \frac{5(18.4727)}{10})$$

$$= P(\frac{(n-1)S^{2}}{\sigma^{2}} > 9.2364)$$



Example:

Let $X_1, X_2, ..., X_{10} \sim N(\mu, \sigma^2 = 25)$. If S^2 is the sample variance, find the 90th percentile of S^2 .

Solution: We need to find the constant c such that $P(S^2>c)=0.10$. Using

$$\frac{(n-1)S^2}{\sigma^2} = \frac{9S^2}{25} \sim \chi_9^2$$

we have

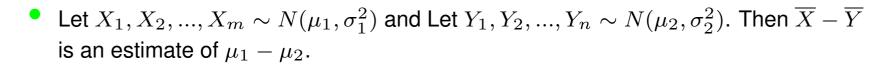
$$P(\frac{9S^2}{25} > \frac{9}{25}c) = 0.10$$

$$P(\chi_9^2 > 0.36c) = 0.10$$

That is, 0.36c = 14.6837 or c = 40.7881.



Distribution of $\overline{X} - \overline{Y}$



•
$$E(\overline{X} - \overline{Y}) = \mu_1 - \mu_2$$

$$Var(\overline{X} - \overline{Y}) = Var(\overline{X}) + (\overline{Y}) = (\sigma_1^2/m) + (\sigma_2^2/n).$$

•
$$\overline{X} - \overline{Y} \sim N(\mu_1 - \mu_2, \frac{\sigma_1^2}{m} + \frac{\sigma_2^2}{n})$$
. Therefore,

$$\frac{(\overline{X} - \overline{Y}) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{m} + \frac{\sigma_2^2}{n}}} \sim N(0, 1)$$

Example: The grades of section 1 (m=50) are normally distributed with mean 60 and std. 5 while the grades of section 2 (n=60) are normally distributed with mean 64 and std. 8. What is the probability that the sample average of section 1 is less than the sample average of section 2?



Cont./Example

Solution:

$$P(\overline{X} < \overline{Y}) = P(\overline{X} - \overline{Y} < 0)$$

$$= P(Z < \frac{0+4}{1.25})$$

$$= P(Z < 3.2)$$

$$= 0.9993$$

