Statistics ST2334 Topic 6: Sampling Distributions (part b)

t-distribution and χ^2-distribution

2011/2012 Semester 2

Sampling Distribution of the $\bar{X}$

Let $X_1, X_2, \ldots, X_n$ be independent identically distributed with $E(X) = \mu$ and $Var(X) = \sigma^2$.

- Then $E(\bar{X}) = \mu$ and $Var(\bar{X}) = \sigma^2/n$.
- If X_i is Normal,

$$\frac{\bar{X} - \mu}{\sigma/\sqrt{n}} = Z \sim N(0, 1)$$

- If n is large, regardless of X_i's distribution, by CLT,

$$\frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \approx Z \sim N(0, 1)$$

- But what if we don't know σ?

Unknown σ

- Not knowing σ does not invalidate the results. However, we need a practical way to get around it.
- Recall that we estimate σ by the sample standard deviation

$$s = \sqrt{\frac{1}{n-1}\sum_{i}^{n}(X_i - \bar{X})^2}$$

- We showed that this was the sensible thing to do in the sense that $E(s^2) = \sigma^2$.
- Note however that σ is a constant, while s is random. Substituting s for σ will change the distribution of our statistic.

t-statistic

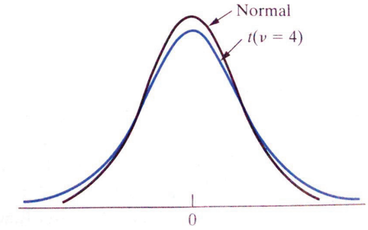
- If the X_i is normal, then

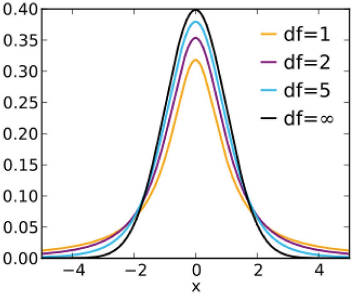
$$t = \frac{\bar{X} - \mu}{s/\sqrt{n}}$$

  is a random variable having the t distribution with parameter $\nu = n - 1$.
- The above is true for all n. However, it is mainly used in the case where n is small. This is because for n large, the t-distribution is close to the normal distribution.

t-distribution

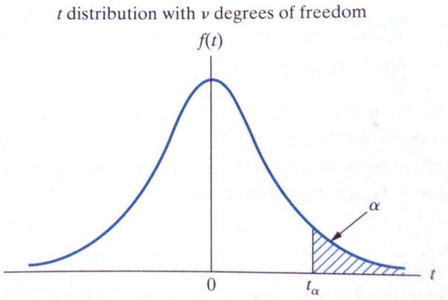
- The t-distribution (some times called the student-t) is denoted by $t(\nu)$.
- The shape of the pdf is similar to that of the normal distribution.
- ν is the degrees of freedom, and as $\nu \to \infty$, $t(\nu) \to Z$.





Properties of t-distribution

- The density function of t distribution is bell shaped, centered and being symmetric at 0.
- t distribution approaches N(0,1) as the parameter $\nu \to \infty$. when $n \geq 30$, we can replace it by N(0, 1).
- Table 4 contains selected values of t_α defined below for various values of ν, such that $P(t > t_\alpha) = \alpha$



Example: Mean score?

The lecturer of a class announced that the mean score of the midterm is 16 out of 30. A student doubts it, so he randomly chose 5 classmates and asked them for their scores: 20, 19, 24, 22, 25.

- Should the student believe that the mean score is 16? Assume the scores is approximately normally distributed.
- The student has n=5 sampled data

$$X_1 = 20, X_2 = 19, X_3 = 24, X_4 = 22, X_5 = 25.$$

- If $\mu = 16$,

$$t = \frac{\bar{X} - \mu}{s/\sqrt{n}} = \frac{\bar{X} - 16}{s/\sqrt{5}}$$

  should follow a t distribution with $\nu = 5 - 1 = 4$.

Example: Mean score?

- Now that with the observed data $\bar{X} = 22, s = 2.55$.

$$t = \frac{22 - 16}{2.55/\sqrt{5}} = 5.26$$

- Check Table 4, $t_{0.005} = 4.604$. This is saying that there is only a 0.005 chance that t is larger than 4.604, provided the lecturer is telling the truth that $\mu = 16$.
- Should the student believe him based on his findings?

What if σ is unknown AND X_i is NOT normal?

- If n is large, CLT tells us that

$$\frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \approx Z \sim N(0,1)$$

- Since n is large, s is a good estimate of σ and thus

$$\frac{\bar{X}-\mu}{s/\sqrt{n}} \approx \frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \approx Z \sim N(0,1)$$

  as well.
- If n is not large?

The Sampling Distribution of the Variance σ^2

Let $X_1, X_2, \ldots, X_n$ be independent identically distributed with $E(X) = \mu$ and $Var(X) = \sigma^2$.

- Recall the sample variance is defined as

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

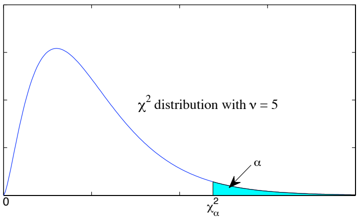
  and that $E(s^2) = \sigma^2$.

- If X_i is normal, then,

$$\chi^2 = \frac{(n-1)s^2}{\sigma^2} = \frac{\sum_{i=1}^{n}(X_i - \bar{X})^2}{\sigma^2}$$

  is a random variable with chi-squared distribution with parameter $\nu = n - 1$.

Chi-squared Distribution



- All chi-squares values are nonnegative.
- The chi-square distribution is a family of curves, each is determined by the degrees of freedom ν.
- All the density functions have a long right tail.
- Values can be found in Table 5.

Example: χ^2

- Suppose 6 random samples are drawn from N(μ, 4), define the sample variance

$$s^2 = \frac{1}{5}\sum_{i=1}^{n}(X_i - \bar{X})^2$$

Find c such that $P(s^2 > c) = 0.05$

- $\frac{5s^2}{4} \sim \chi^2$ with $\nu = 5$. Hence,

$$\begin{aligned} & P(s^2 > c) = P(5s^2/4 > 5c/4) \\ \Rightarrow \quad & 5c/4 = \chi^2_{0.05} = 11.07 \\ \Rightarrow \quad & c = 8.86 \end{aligned}$$

Some general identities

- Let $Z_1, Z_2, \ldots, Z_\nu$ be independent $N(0,1)$, then

$$\chi^2_\nu = Z_1^2 + Z_2^2 + \cdots + Z_n u^2$$

follows a chi-squared distribution with ν degrees of freedom.

- If $X \sim \chi^2_{\nu_1}$ and $Y \sim \chi^2_{\nu_2}$ are independent, then

$$X + Y \sim \chi^2_{\nu_1+\nu_2}$$

- If $Z \sim N(0,1)$ and $X \sim \chi^2_\nu$ are independent, then

$$t = \frac{Z}{\sqrt{X/\nu}}$$

follows the t-distribution with ν degrees of freedom.