

ECON 186 Class Notes: Probability and Statistics Part 3

Jijian Fan

Bias and Efficiency

- An estimator of a parameter, θ , is unbiased if the mean of its sampling distribution is equal to θ . That is, if

$$E(\hat{\theta}) = \theta$$

- where $\hat{\theta}$ is the parameter estimated from the sample whereas θ is the true population parameter.
- This is a desirable property for an estimator, but many estimators are unbiased. We also need another criteria to determine the best estimator.
- An unbiased estimator $\hat{\theta}_1$ is more efficient than another unbiased estimator $\hat{\theta}_2$, if the sampling variance of $\hat{\theta}_1$ is less than that of $\hat{\theta}_2$. That is,

$$\text{Var}(\hat{\theta}_1) < \text{Var}(\hat{\theta}_2)$$

Proof of Sample Mean Unbiasedness

- Consider independent and identically distributed random variables X_1, X_2, \dots, X_n where $E(X_1) = E(X_2) = \dots = E(X_n) = \mu$.

$$E\left(\frac{\sum X_i}{n}\right) = \frac{1}{n}E\left(\sum X_i\right) = \frac{1}{n}\sum E(X_i) = \frac{1}{n}n\mu = \mu$$

Proof of Sample Variance Unbiasedness

- Consider iid random variables that form a random sample, X_1, \dots, X_n . Also, $E(X_i) = \mu = E(\bar{X})$ and the population variance is $Var(X_i) = \sigma^2$.

$$\begin{aligned} E(s_x^2) &= E\left[\frac{\sum (X_i - \bar{X})^2}{n-1}\right] = \frac{1}{n-1} E\left[\sum (X_i^2 - 2\bar{X}X_i + \bar{X}^2)\right] \\ &= \frac{1}{n-1} E\left[\sum X_i^2 - 2\bar{X}\sum X_i + \bar{X}^2\sum 1\right] = \frac{1}{n-1} E\left[\sum X_i^2 - 2n\bar{X}^2 + n\bar{X}^2\right] \\ &= \frac{1}{n-1} \left[nE(X_i^2) - nE(\bar{X}^2)\right] = \frac{n}{n-1} E(X_i^2 - E(\bar{X}^2)) \end{aligned}$$

- So, we must find what $E(X_i^2)$ and $E(\bar{X}^2)$ are.

$$E(X_i^2) = Var(X_i) + (E(X_i))^2 = \sigma^2 + \mu^2$$

Proof of Sample Variance Unbiasedness

$$\begin{aligned} E(\bar{X}^2) &= \text{Var}(\bar{X}) + (E(\bar{X}))^2 = \text{Var}\left(\frac{1}{n}\sum X_i\right) + \mu^2 = \frac{1}{n^2}\text{Var}\left(\sum X_i\right) + \mu^2 \\ &= \frac{1}{n^2}\sum \text{Var}(X_i) + \mu^2 = \frac{1}{n^2}n\sigma^2 + \mu^2 = \frac{1}{n}\sigma^2 + \mu^2 \end{aligned}$$

- Then, plugging back in:

$$E(s_x^2) = \frac{n}{n-1} \left[\sigma^2 + \mu^2 - \left(\frac{1}{n}\sigma^2 + \mu^2 \right) \right] = \frac{n}{n-1} \left[\frac{n-1}{n}\sigma^2 \right] = \sigma^2$$

Bias and Efficiency

- Theorem: If X_1, \dots, X_n are a random sample from a population with mean μ and variance σ^2 , then \bar{X} is a random variable with mean μ and variance $\frac{\sigma^2}{n}$.
- Example: Suppose that μ is the mean value of parental income among UCSC students and σ^2 is the variance. Suppose that we take one random sample of 10 students and one random sample of 100 students where \bar{X}_1 is the mean value for the first 10 students sampled and \bar{X}_2 is the mean for all 100 students sampled.
- We want to see which sample is more efficient and check the bias of each. By the above theorem,

$$E(\bar{X}_1) = \mu = E(\bar{X}_2)$$

$$\text{Var}(\bar{X}_1) = \frac{\sigma^2}{10} \quad \text{and} \quad \text{Var}(\bar{X}_2) = \frac{\sigma^2}{100}$$

- So, both samples are unbiased, but $\text{Var}(\bar{X}_2) < \text{Var}(\bar{X}_1)$, so the larger sample is more efficient. This is one reason why having a larger sample size is beneficial.

Student's t-Distribution

- If we know the population mean and variance, we can fully characterize a normally distributed random variable. However, we rarely know this.
- In order to find statistical significance and perform hypothesis testing, we use the student's t-distribution, which describes samples drawn from a full population. The distribution varies based on sample size, and the larger the sample, the more the distribution resembles a normal distribution.
- We want to use the t-distribution rather than the normal either if our sample is very small or we do not know the population standard deviation.
- If $Z \sim N(0,1)$ and $X \sim \chi^2(n)$, and Z and X are independent, then the t-distribution with n degrees of freedom is

$$t(n) = \frac{Z}{\sqrt{X/n}} = Z\sqrt{\frac{n}{X}}$$

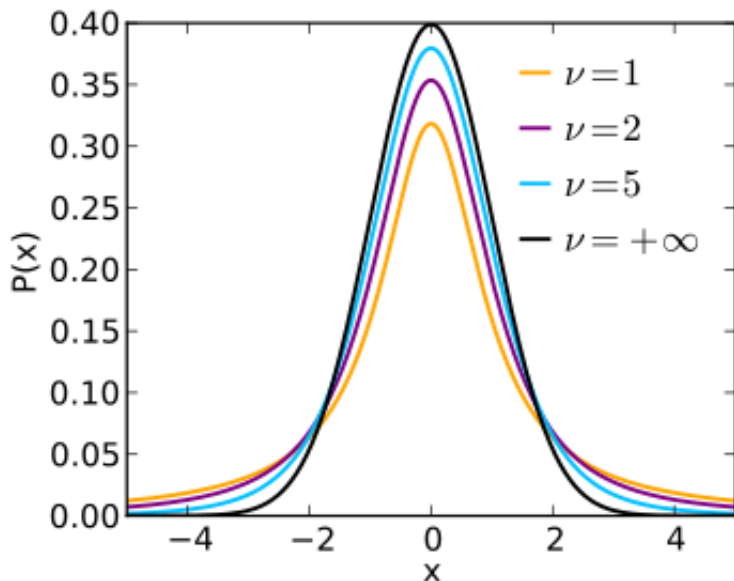
Student's t-Distribution

- Alternatively, let X_1, \dots, X_n be the numbers observed in a sample from a normally distributed population with mean μ . Then,

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

- The sampling distribution of this t-statistic or t-value is the t-distribution with $n - 1$ degrees of freedom.

Student's t-Distribution



Convergence in Probability

- Convergence in probability: A sequence Z_1, Z_2, \dots of random variables converges to b in probability if for every number $\varepsilon > 0$,

$$\lim_{n \rightarrow \infty} Pr(|Z_n - b| < \varepsilon) = 1$$

- This property is denoted by

$$Z_n \xrightarrow{P} b$$

- which is stated as Z_n converges to b in probability where the symbol \xrightarrow{P} is referred to as a probability limit.
- Intuitively, the above definition says that Z_n converges to b in probability if the probability that Z_n lies in each given interval around b , no matter how small this interval may be, approaches 1 as $n \rightarrow \infty$.

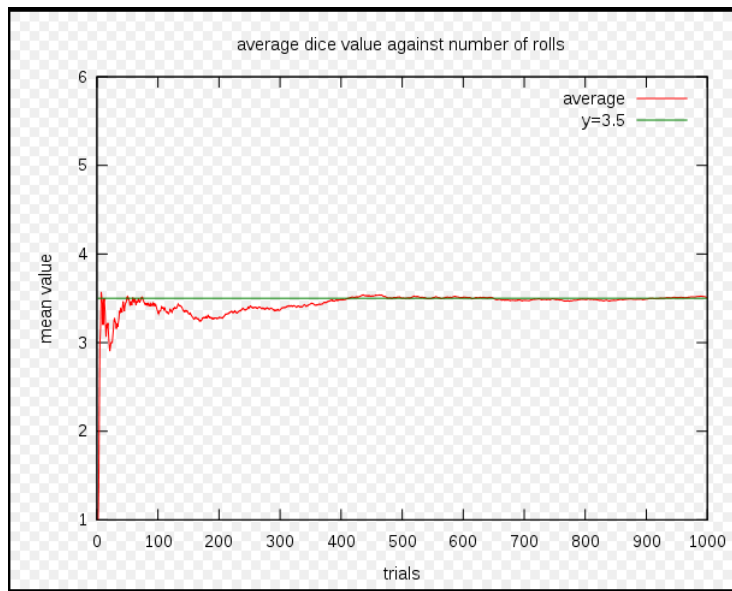
Law of Large Numbers

- Suppose that X_1, \dots, X_n form a random sample from a distribution for which the mean is μ and for which the variance is finite. Let \bar{X}_n denote the sample mean, then

$$\bar{X}_n \xrightarrow{P} \mu$$

- The above result is called the law of large numbers and it says there is a high probability that \bar{X}_n will be close to μ if the sample size n is large.
- So, if a large random sample is taken from a distribution with an unknown mean, the sample mean should come close to the true population mean.

Law of Large Numbers



Central Limit Theorem

- Central Limit Theorem (CLT): If the independent random variables X_1, \dots, X_n form a random sample of size n from a given distribution with mean μ and variance σ^2 , then for each fixed number x ,

$$\lim_{n \rightarrow \infty} Pr \left[\frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \leq x \right] = \Phi(x)$$

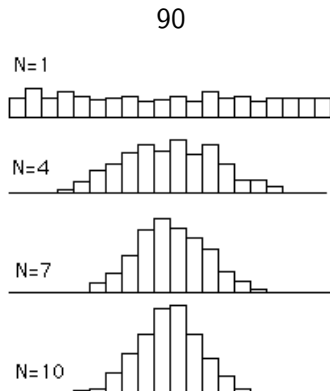
- where Φ denotes the cdf of the standard normal distribution.
- So, this very powerful theorem says that if we take a large random sample from any distribution with mean μ and variance σ^2 , the distribution of the random variable $\frac{\bar{X} - \mu}{\sigma/\sqrt{n}}$ will be approximately the standard normal distribution.

Central Limit Theorem

- So, the distribution of \bar{X} will be approximately normal with mean μ and variance $\frac{\sigma^2}{n}$ or the distribution of $\sum_{i=1}^n X_i$ will be approximately the normal distribution with mean $n\mu$ and variance $n\sigma^2$.
- Amazingly, no matter what the distribution of our data is, the CLT means that we can model a sample from any distribution as a student's t-distribution and thus use the t-test for hypothesis testing as long as the sample size is large enough.

Central Limit Theorem Example

- Below are shown the resulting frequency distributions each based on 500 means. For $N = 4$, 4 scores were sampled from a uniform distribution 500 times and the mean computed each time. The same method was followed with means of 7 scores for $N = 7$ and 10 scores for $N = 10$.



Central Limit Theorem Example

- Suppose that a fair coin is tossed 900 times. Approximate the probability of obtaining more than 495 heads using the Central Limit Theorem.
- Let $X_i = 1$ if a head is obtained on the i th toss and 0 otherwise. Then, $E(X_i) = \frac{1}{2}$ and $Var(X_i) = \frac{1}{4}$ (recall that the variance of the Bernoulli distribution is $p(1-p)$).
- So, the values X_1, \dots, X_{900} form a random sample of size $n = 900$ from a distribution with mean $\frac{1}{2}$ and variance $\frac{1}{4}$.

Central Limit Theorem Example

- The Central Limit Theorem tells us that $\sum_{i=1}^{900} X_i$ will be approximately normal with mean $900 \left(\frac{1}{2}\right) = 450$ and variance $900 \left(\frac{1}{4}\right) = 225$, with standard deviation $\sqrt{225} = 15$.
- So, the variable $Z = \frac{H-450}{15}$ will have approximately the standard normal distribution. Thus,

$$\begin{aligned} Pr(H > 495) &= Pr\left(\frac{H - 450}{15} > \frac{495 - 450}{15}\right) \\ &= Pr(Z > 3) \approx 1 - \phi(3) = 1 - .9987 = 0.0013 \end{aligned}$$

Hypothesis Testing

- Suppose that you see an advertisement by a student who is offering to tutor economics classes for \$10 an hour. You want to know how close this rate is to the average. What if the average is \$20 an hour? How about \$15? \$10.5? Is this sufficiently “close?”
- To assess this question, we must perform four steps.
 - ▶ 1) Survey a random sample of tutors
 - ▶ 2) Calculate sample mean \bar{X} .
 - ▶ 3) Calculate standard error of \bar{X} .
 - ▶ 4) Compare \bar{X} to \$10.

Hypothesis Testing

- We want to formally test whether \$10 is actually statistically similar to the average. If we want to test that the average tutoring rate is \$10 an hour, then we call this the null hypothesis. That is, the null hypothesis is what we assume is true unless our test proves otherwise. The alternative hypothesis is any alternative.
- Notation:
 - ▶ H_0 : *null hypothesis*
 - ▶ H_1 or H_a : *alternative hypothesis*
- In our example,
 - ▶ H_0 : $\mu = 10$
 - ▶ H_1 : $\mu \neq 10$ (or $\mu > 10$ or $\mu < 10$)
- $\mu \neq 10$ is a two-tailed test and $\mu > 10$ and $\mu < 10$ are one-tailed tests.

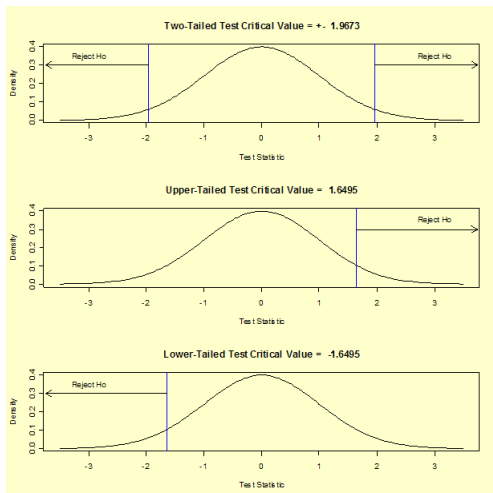
Hypothesis Testing

- Generally we conduct a test by using an estimator from our sample and checking to see how different it is from some specified value such as \bar{X} .
- Usually our estimator will have a t-distribution, F-distribution, or χ^2 -distribution.
- Intuitively, we conduct the test by creating some critical region where we reject the null hypothesis if the estimator falls in that critical region.
- If the estimator does not fall in the critical region then we “accept” (technically fail to reject) the null hypothesis.

Types of Error

- Type I error: The procedure may lead to a rejection of the null hypothesis when the null hypothesis is true.
- Type II error: The procedure may fail to reject the null hypothesis when it is false.
- The probability of a type I error is the size of the test. It is commonly referred to as the significance level (α).
- The power of a test is the probability that it will correctly lead to rejection of a false null hypothesis.

Hypothesis Testing



Hypothesis Testing

- Let us return to our tutoring example.
- $H_0 : \mu = 10$
- $H_1 : \mu \neq 10$
- Assume that the distribution of tutoring rates is normal with mean μ and variance σ^2 .
- Suppose we have a sample of 16 tutors with $\bar{X} = 11.80$ and $s^2 = 9$
- Recall that $\frac{\sqrt{n}(\bar{X}-\mu)}{s} \sim t(n-1)$. We have all of this information except for μ which is what we are trying to test. So, we can use the t-test.
- That is, we model our sample as a t-distribution and assess whether the t-statistic is greater than some critical value, where the critical value is based on the significance level we are trying to test. If it is, we reject the null.

Hypothesis Testing

- In our tutoring example, suppose that we want to examine if \bar{X} is statistically equal to μ at the 5% significance level. Since we are performing a two-tailed test, we use $\alpha = 0.025$ to find our significance. Then, with 15 degrees of freedom, we can see from the t-table that the t-critical value is $t_c = 2.131$.
- In the two-tailed case, if $\frac{\sqrt{n}(\bar{X}-\mu)}{s} < -t_c$ or $\frac{\sqrt{n}(\bar{X}-\mu)}{s} > t_c$, then reject H_0 . In this example:

$$\frac{\sqrt{16}(11.80 - 10)}{3} = 2.4 > 2.131 \rightarrow \text{Reject } H_0$$

Hypothesis Testing

- More generally, we can compare the means of two samples to see if they are equal, we do not only have to compare to a constant.
- Suppose we have two samples T and C with two sample means, \bar{X}_T and \bar{X}_C with variances var_T and var_C , and sample sizes n_T and n_C .
- Then, we can define the t-statistic as

$$t = \frac{\bar{X}_T - \bar{X}_C}{\sqrt{\frac{var_T}{n_T} + \frac{var_C}{n_C}}}$$

- The degrees of freedom for the test is

$$n_T + n_C - 2$$

Confidence Intervals

- We now know how to tell if a null hypothesis is true or not at a certain confidence level, but we would also like to be able to compute intervals containing the true parameter with some desired level of confidence.
- Say $\alpha = 0.05$, which we call a 5% significance level. Think of this as the proportion of random samples where the true parameter falls in our estimated range.
- For a population with unknown mean μ and unknown standard deviation, a confidence interval for the population mean is $\bar{X} \pm t_c \frac{s}{\sqrt{n}}$.

Confidence Interval Example

- So, for our tutoring example, the 95% confidence interval is

$$\begin{aligned} & \left(\bar{X} - t_c \frac{s}{\sqrt{n}}, \bar{X} + t_c \frac{s}{\sqrt{n}} \right) \\ & = \left(11.8 - 2.131 * \frac{3}{\sqrt{16}}, 11.8 + 2.131 * \frac{3}{\sqrt{16}} \right) = (10.2, 13.4) \end{aligned}$$

- So, since 10 does not lie in our 95% confidence interval, we can say that we are 95% confident that 10 is not the population mean. Are we 99% confident though?

$$\begin{aligned} & \left(\bar{X} - t_c \frac{s}{\sqrt{n}}, \bar{X} + t_c \frac{s}{\sqrt{n}} \right) \\ & = \left(11.8 - 2.947 * \frac{3}{\sqrt{16}}, 11.8 + 2.947 * \frac{3}{\sqrt{16}} \right) = (9.6, 14.0) \end{aligned}$$

- Since 10 does lie in this confidence interval, we cannot say that 10 is not the population mean with 99% confidence. Could also have used t-test to determine this.

P-values

- We have just explored the concept of hypothesis testing, which tells us whether we should reject some hypothesis with some confidence level? However, what if we want to know more generally what the probability is that our result is actually more extreme than some random event?
- The p-value is the probability, under the assumption of the null hypothesis, of obtaining a result equal to or more extreme than what was actually observed.
- Another interpretation is: The p-value is the smallest level α_0 such that we would reject the null-hypothesis at level α_0 with the observed data.

P-values

- To calculate p-values, we can use the following expressions

- ▶ For a right tail event:

$$Pr(X \geq x|H)$$

- ▶ For a left tail event:

$$Pr(X \leq x|H)$$

- ▶ For a double tailed event:

$$2 * \min\{Pr(X \leq x|H), Pr(X \geq x|H)\}$$

P-values

- So, we can calculate p-values directly from our calculated t-statistic.
- Consider our previous example where we calculated a t-statistic of 2.4 with 15 degrees of freedom and we were performing a two-tailed test.
- From the t-table, we can see that our p-value is roughly 0.03. The interpretation is that if the mean tutoring rate in the population actually was 10, the sample mean of 11.80 that we got from our one sample would happen randomly about about 3% of the time.
- So, we say that this result is significant at the 5% level, but it is not significant at the 1% level.