Econ 210 - Statistics Review

Sidharth Sah[1]

September 8, 2023

[1] Thanks to Azeem Shaikh and Max Tabord-Meehan for useful material

Samples

- Statistics involves attempting to learn characteristics of probability distributions using finite samples of data
- If $X_1, X_2, ..., X_n$ are independent rv's with the same distribution, they are called Independent and Identically Distributed - **i.i.d.**
- We're generally going to assume that our samples are drawn i.i.d. from our distribution of interest

Estimator

- Say we wish to estimate a parameter, θ
- An Estimator is a function that goes from a sample to a guess of our paramter:

$$\hat{\theta}_n = \hat{\theta}_n(X_1, X_2, ..., X_n)$$

- Note that while θ is a number, $\hat{\theta}_n$ is a function/rv until it is actually calculated for a specific sample (similar to a conditional expectation)

Estimator Example - Sample Mean

- Say we want to estimate the mean of rv X - i.e. $\theta = E[X]$
- Let $X_1, X_2, ..., X_n$ be i.i.d. $\sim X$
- Natural estimator is the Sample Mean, often denoted $\overline{X}_n$

$$\overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$$

Properties of Estimators

- We will be interested in various properties of estimators in order to know when/if they are useful and how to interpret them
- Finite-sample properties are true for the estimator at any n
- Large-sample properties hold approximately as $n \to \infty$

Bias

- Bias is one oft-discussed finite-sample property - says if the estimator is "correct" in expectation:

$$Bias[\hat{\theta}_n] = E[\hat{\theta}_n] - \theta$$

- If $Bias[\hat{\theta}_n] = 0$, then $\hat{\theta}_n$ is called Unbiased

Bias Example - Sample Mean

- For $X_1, X_2, ..., X_n$ i.i.d. $\sim X$, the sample mean is an *unbiased* estimator for the population mean, $E[X]$:

$$
\begin{aligned}
Bias[\overline{X}_n] &= E[\overline{X}_n] - E[X] \\
&= E[\frac{1}{n}\sum_{i=1}^{n} X_i] - E[X] \\
&= \frac{1}{n}\sum_{i=1}^{n} E[X_i] - E[X] \\
&= \frac{1}{n} nE[X] - E[X] \\
&= 0
\end{aligned}
$$

Bias Example - Sample Variance

- $\hat{\sigma}_n^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X}_n)^2$ is sometimes called the *sample variance*
- This estimator as is, is downward biased:

$$Bias[\hat{\sigma}_n^2] = E[\hat{\sigma}_n^2] - \sigma_X^2 < 0$$

- Can "fix" this with a degrees of freedom adjustment:

$$\frac{n}{n-1}\hat{\sigma}_n^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X}_n)^2$$

is an unbiased estimator of variance

Variance of Estimators

- The *Variance* of an estimator, $Var(\hat{\theta}_n)$, is another important finite-sample property
- If we had two unbiased estimators, for instance, we'd generally prefer the one with a smaller variance - we'll be less likely to draw an estimate far away from the true value

Variance of Estimators Example - Sample Mean

- Can calculate the variance of $\overline{X}_n$ for $X_1, ..., X_n$ i.i.d. $\sim X$

$$
\begin{aligned}
Var(\overline{X}_n) &= Var(\frac{1}{n}\sum_{i=1}^{n} X_i) \\
&= \frac{1}{n^2}\sum_{i=1}^{n} Var(X_i) && \text{(Ind. of } X_i\text{)} \\
&= \frac{1}{n^2}\sum_{i=1}^{n} Var(X) && (X_i \sim X) \\
&= \frac{1}{n} Var(X)
\end{aligned}
$$

Sampling Distributions

- As $\hat{\theta}_n$ is itself a rv, it also has a probability distribution, which is called the Sampling Distribution
- This distribution is often hard to fully characterize, but there are exceptions...

Sampling Distribution Example - Normal Dist

- Say $X_1, ..., X_n$ are i.i.d. $\sim X$, where $X \sim N(\mu_X, \sigma_X^2)$. Consider the sample mean
- $\overline{X}_n$ is a linear combination of independent normals, so it itself is normal
- We also already derived the mean and variance of $\overline{X}_n$, so we can say:

$$\overline{X}_n \sim N(\mu_X, \frac{\sigma_X^2}{n})$$

Consistency

- Consistency is a *large-sample* property saying that the estimator will "get close" to the parameter as the size of the sample, n, grows large
- Formally, an estimator is consistent if it converges in probability to the parameter

Convergence in Probability

- A sequence of random variables, X_n, Converges in Probability to another rv or scalar, X, if, for any $\varepsilon > 0$, as $n \to \infty$

$$P(|X_n - X| > \varepsilon) \to 0$$

- This is notated $X_n \xrightarrow{p} X$
- So, $\hat{\theta}_n$ converges in probability to θ, and is a consistent estimator for θ, if

$$\hat{\theta}_n \xrightarrow{p} \theta$$

Weak Law of Large Numbers

- The Weak Law of Large Numbers says that the sample mean is a consistent estimator for the expectation, aka

$$\overline{X}_n \xrightarrow{p} E[X]$$

- This property requires that $X_1, ..., X_n$ be i.i.d. $\sim X$ and that $E[X^2] < \infty$

Chebychev's Inequality

- The proof of the WLLN requires the use of Chebychev's Inequality, which states that for an rv X and $\varepsilon > 0$,

$$P(|X| > \varepsilon) \leq \frac{E[X^2]}{\varepsilon^2}$$

- Proof:

$$\mathbb{1}\{|X| > \varepsilon\} \leq \frac{X^2}{\varepsilon^2}$$

$$E[\mathbb{1}\{|X| > \varepsilon\}] \leq E[\frac{X^2}{\varepsilon^2}] \qquad \text{(Prop of Expecs)}$$

$$P(|X| > \varepsilon) \leq \frac{E[X^2]}{\varepsilon^2}$$

Proof of WLLN

- Fix an $\varepsilon > 0$

$$\begin{aligned} P(|\overline{X}_n - E[X]| > \varepsilon) &\leq \frac{E[|\overline{X}_n - E[X]|^2]}{\varepsilon^2} \qquad \text{(Chebychev)} \\ &\leq \frac{Var(\overline{X}_n)}{\varepsilon^2} \\ &\leq \frac{Var(X)}{n\varepsilon^2} \to 0 \text{ as } n \to \infty \\ &\Rightarrow P(|\overline{X}_n - E[X]|) \to 0 \text{ as } n \to \infty \end{aligned}$$

$$\Rightarrow \overline{X}_n \xrightarrow{p} E[X]$$

Continuous Mapping Theorem

- Suppose that, for sequences of rv's X_n and Y_n and scalars x and y, $X_n \xrightarrow{p} x$, $Y_n \xrightarrow{p} y$. For any function g that is continuous at (x, y):

$$g(X_n, Y_n) \xrightarrow{p} g(x, y)$$

- This is stated for two sequences of rv's, but is true for any finite number of rv's

CMT Example - Sample Variance

- The sample variance (w/out the degrees of freedom adjustment) is not unbiased, but *is* consistent, under assumptions that $X_1, ..., X_n$ iid $\sim X$ and $E[X^4] < 0$
- Proof:

$$\hat{\sigma}_n^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X}_n)^2$$
$$= \frac{1}{n}\sum_{i=1}^{n} X_i^2 - \overline{X}_n^2$$

- See Supplemental Note about sample variance for greater detail on this step

CMT Example - Sample Variance (cont.)

- WLLN implies that

$$\frac{1}{n}\sum_{i=1}^{n} X_i^2 \xrightarrow{p} E[X^2]$$

$$\overline{X}_n \xrightarrow{p} E[X]$$

- Then, we apply the CMT for $g(x, y) = x - y^2$ (cont. at any finite (x, y)) and say

$$\hat{\sigma}_n^2 \xrightarrow{p} Var(X)$$

Limiting Distributions

- Estimators have sampling distributions defined at any finite n. However, we said these are often difficult to characterize
- Sometimes, as $n \to \infty$, the sampling distributions of a sequence of θ_n (with ever increasing sample sizes) will start to resemble more familiar distributions. This is the Limiting Distribution of an estimator

Convergence in Distribution

- Say that X_n is a sequence of rv's and that X is a continuous rv. Then, we say that X_n Converges in Distribution to X if

$$P(A_n \leq t) \to P(A \leq t) \; \forall t$$

- This is denoted $A_n \xrightarrow{d} A$

Central Limit Theorem

- This definition leads us to an important result - the Central Limit Theorem. Let $X_1, ..., X_n$ be iid $\sim X$ and suppose that $E[X^2] < \infty$. Then,

$$\sqrt{n}(\overline{X_n} - E[X]) \xrightarrow{d} N(0, Var[X])$$

Slutsky's Lemma

- CLT is often used in conjunction with Slutsky's Lemma - for sequences of rv's X_n and Y_n, rv X, and scalar y, such that $X_n \xrightarrow{d} X$ and $Y_n \xrightarrow{p} y$,
  (i) $X_n Y_n \xrightarrow{d} Xy$
  (ii) $X_n + Y_n \xrightarrow{d} X + y$
  (iii) $X_n / Y_n \xrightarrow{d} X / y$ whenever $y \neq 0$

CLT/Slutsky Example - Convergence to Std Normal

- From CLT, we know that

$$\sqrt{n}(\overline{X_n} - E[X]) \xrightarrow{d} N(0, \sigma_X^2)$$

- We also know that

$$\hat{\sigma}_n \xrightarrow{p} \sigma_X$$

- Further assume that $\sigma_X > 0$. Then, by Slutskty

$$\frac{1}{\hat{\sigma}_n}\sqrt{n}(\overline{X_n} - E[X]) \xrightarrow{d} \frac{1}{\sigma_X}N(0, \sigma_X^2)$$

$$\xrightarrow{d} N(0, 1)$$

Hypothesis Testing

- The above is the basic argument used in Hypothesis Testing
- For hypothesis testing, we need a:
  - Null Hypothesis (H_0): A statement about a parameter we "want" to disprove
  - Alternative Hypothesis (H_1): What we "want" to prove
  - Test statistic (T_n): Function of the data such that "large" values of T_n suggest H_0 is likely to be false
  - Critical value (c): Defines what we mean by large
  - Decision rule: Says we reject H_0 if and only if $T_n > c$

Hypothesis Testing Example - Two-Sided Test

- For a *two-sided test* about the expectation of a rv, X:
  - Null Hypothesis (H_0): $E[X] = \mu_0$
  - Alternative Hypothesis (H_1): $E[X] \neq \mu_0$
  - Test statistic (T_n):

$$T_n = \frac{\sqrt{n}}{\hat{\sigma}_n^2}(|\overline{X_n} - \mu_0|)$$

  - The critical value and decision rule will be determined by our sensitivity to different kinds of errors...

Type 1 and Type 2 Errors

- Type 1 Error - Rejecting H_0 when it is true
- Type 2 Error - Not rejecting H_0 when it is false
- We usually can't say with certainty whether H_0 is true or false, but we can say things about the probabilities of Type 1 and 2 errors

Critical Values

- We set the critical value so as to determine the probability of Type 1 Error - thus, c is such that, under the assumption that H_0 is true:

$$P(T_n > c) \approx \alpha$$

for $\alpha \in (0, 1)$ that is chosen by the researcher

- α, the likelihood of falsely rejecting the null, is called the <u>Significance Level</u>

Critical Values Example - Two-Sided Test

- $H_0 : E[X] = \mu_0$, $H_1 : E[X] \neq \mu_0$. Assume that $0 < \sigma_X^2 < \infty$. We want the significance level α - ie we want:

$$P(T_n > c) = P(\frac{\sqrt{n}}{\hat{\sigma}_n}(|\overline{X_n} - \mu_0|) > c) = \alpha$$

- Under the assumption that $E[X] = \mu_0$, this implies

$$P(\frac{\sqrt{n}}{\hat{\sigma}_n}(|\overline{X_n} - E[X]|) > c) = \alpha$$

Critical Values Example - Two-Sided Test (cont.)

- Earlier, we showed that the left-hand size converges in distribution to a standard normal. Using this:

$$P(\frac{\sqrt{n}}{\hat{\sigma}_n}(|\overline{X_n} - E[X]|) > c) \approx 1 - \Phi(c) + \Phi(-c)$$
$$= 2(1 - \Phi(c))$$
$$\text{(Symmetry of normal)}$$

- Thus, we can get the appropriate critical value in terms of our pre-chosen significance level:

$$2(1 - \Phi(c)) = \alpha$$
$$\Rightarrow c = \Phi^{-1}(1 - \frac{\alpha}{2})$$

P-Values

- The P-value ($\hat{p}_n$) is a continuous measure that tells us the smallest value of α under which we would reject the test
- Note that rejecting at a smaller significance level means that we would reject at any higher significance level - if we reject at α', we'll reject at any $\alpha'' > \alpha'$

P-Values Example - Two-Sided Test

- The two-sided hypothesis test we developed will reject at the significance level α if:

$$T_n > \Phi(1 - \frac{\alpha}{2})$$
$$\Rightarrow \alpha > 2(1 - \Phi(T_n))$$

- We reject if the above statement is true, so we reject at any significance level greater than $2(1 - \Phi(T_n))$. Thus,

$$\hat{p}_n = 2(1 - \Phi(T_n))$$

Confidence Sets

- A Confidence Set, $C_n = C_n(X_1, ..., X_n)$, is a set of values constructed such there is a pre-specified probability of our parameter falling within the set:

$$P(\theta \in C_n) \approx 1 - \alpha$$

where α is selected by the researcher

Confidence Sets Example - Two-Sided Test

- Imagine performing hypothesis test of $H_0 : E[X] = \mu_0$ at significance level α for *every possible* value μ_0 and putting every non-rejected value in C_n. When we test $H_0 : E[X] = E[X]$ there is an α probability of rejection. Thus, there is a $1 - \alpha$ chance that we *don't* reject the true value, and $E[X] \in C_n$
- We know we reject H_0 for a given μ_0 if:

$$\frac{\sqrt{n}}{\hat{\sigma}_n}(|\overline{X_n} - \mu_0|) > \Phi^{-1}(1 - \frac{\alpha}{2})$$

Confidence Sets Example - Two-Sided Test (cont.)

- Thus, we fail to reject when:

$$\frac{\sqrt{n}}{\hat{\sigma}_n}(|\overline{X_n} - \mu_0|) \leq \Phi^{-1}(1 - \frac{\alpha}{2})$$

$$|\overline{X_n} - \mu_0| \leq \frac{\hat{\sigma}_n}{\sqrt{n}}\Phi^{-1}(1 - \frac{\alpha}{2})$$

$$\Rightarrow C_n = [\overline{X_n} \pm \frac{\hat{\sigma}_n}{\sqrt{n}}\Phi^{-1}(1 - \frac{\alpha}{2})]$$

$$\Rightarrow P(E[X] \in C_n) \approx 1 - \alpha$$