1 Asymptotic Properties of Estimators

Last time, we discussed some finite-sample properties of estimators. Here are some common asymptotic properties:

- consistency
- asymptotic normality
- efficiency (will talk about this much more later)

Definition 1.1. We say an estimator $\hat{\theta}_n$ for θ is consistent if $\hat{\theta}_n$ converges in probability to θ as $n \to \infty$.

Thus, the weak LLN is saying that for iid random variables, the sample mean is consistent for the population mean.

Caution: the property of consistency may seem similar to unbiasedness, but these are really two different properties. An unbiased estimator is not necessarily consistent, and a consistent estimator is not necessarily unbiased.

Definition 1.2. Asymptotic Normality. An estimator $\hat{\theta}$ is asymptotically normal if it converges in distribution to a normal RV.

So by the CLT, if we have an iid random sample, then the sample mean is asymptotically normal. We will see soon that there are other estimators that are asymptotically normal too.

It is also possible to extend finite-sample notions like bias, variance, MSE, etc. to asymptotic versions. For example, efficiency involves a lower asymptotic variance. These properties will be discussed more coming up.

1.1 Sample Mean and Variance

Let $X_1, X_2, \ldots, X_n \overset{\text{iid}}{\sim} (\mu, \sigma^2)$.

Definition 1.3. The sample mean and variance are defined as

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

and

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

respectively.

They have a number of desirable properties, including unbiasedness as well as consistency.

Note that the estimator $\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2$ is NOT unbiased! Why does this happen? Intuitively, this is because when we go to estimate a variance, we have to estimate the mean μ as part of the formula. That is, we are forced to use the data not only for estimating the desired estimand (the variance) but also for estimating the mean. And the $X_i's$ in the variance estimator formula have to sum to $n\bar{X}$, which means that they are not free to fully vary in that formula. This constraint means that we lose one degree of freedom, which is why $n-1$ is the correct factor for unbiasedness.

2 Confidence Intervals

So far in section, we have only spoken about estimation, which gives information about a parameter in the form of a single point. If instead we want a range of possible values for the parameter, then we can form a confidence interval.

Definition 2.1. Confidence Interval. A $1-\alpha(100)\%$ confidence interval for θ is an interval $C_n = C_n(X)$ computed from the data such that $\mathbb{P}_\theta(\theta \in C_n) \geq 1-\alpha$ for all θ.

Recall that a confidence interval is random, since it's a function of the data!

Please visit this website to explore confidence intervals in more depth: `https://istats.shinyapps.io/ExploreCoverage/`. The website has a number of other helpful simulations, such as on the CLT (`https://istats.shinyapps.io/SampDist_Prop/`).

2.1 Interpretation of Confidence Intervals

How do we interpret a confidence interval in frequentist statistics? By definition, a $(1-\alpha)100\%$ confidence interval is constructed to contain θ with probability $\geq 1-\alpha$. For example, let $\alpha = 0.05$. If we repeatedly take a random sample from a population and calculate a 95% confidence interval from it, then at least $(1-\alpha)100\%$ of those intervals are expected to contain θ.

Caution: We cannot interpret a *specific* confidence interval in terms of probability – only confidence. That is, once we have actually obtained a confidence interval from a dataset, we CANNOT say that θ is contained in it with probability $\geq 1 - \alpha$. For example, suppose we obtain a 95% confidence interval of (3.2, 7.6). An incorrect interpretation: θ is between 3.2 and 7.6 with probability 0.95. A correct interpretation: with 95% confidence, we estimate that θ is between 3.2 and 7.6.

3 Method of Moments Estimators

One of the oldest ways to estimate a parameter is though the method of moments (MoM). Recall the steps for computing a MoM estimator for a parameter:

(a) Write the moments in terms of the parameter.

(b) Rearrange the terms, so that the parameter is written in terms of the moments.

(c) Plug in the sample moments for the moments.

In general, MoM estimators do not have as good theoretical properties as some other estimators (such as maximum likelihood estimators). But one advantage is that they are typically easy to calculate.

4 Exercises

Problem 1.

Let $X_1, X_2, \ldots, X_n \overset{\text{iid}}{\sim} Bern(p)$. What is the MoM estimator of p?

Problem 2.

Let $X \sim Bin(n, p)$. What is the MoM estimator for p? How does it relate to the result in the previous question?

Problem 3.

Find a MoM estimator for $X_1, X_2, \ldots, X_n \overset{\text{iid}}{\sim} Uniform(a, b)$.

Problem 4.

If $\hat{v}$ is a consistent estimator of $\text{Var}(\hat{\theta})$, then is it also true that $\sqrt{\hat{v}}$ is a consistent estimator of $se(\hat{\theta})$? Show why or why not.

Problem 5.

If $\hat{v}$ is an unbiased estimator of $\text{Var}(\hat{\theta})$, then is it also true that $\sqrt{\hat{v}}$ is an unbiased estimator of $se(\hat{\theta})$? Show why or why not.

Problem 6.

Let $X_1, X_2, \dots, X_n \overset{\text{iid}}{\sim} Bern(p)$. Recall that the variance of X_i is $p(1-p)$, so by the Central Limit Theorem, $\hat{p} = \bar{X} \implies N(p, p(1-p)/n)$. Equivalently, we can write

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

But we more commonly use this result, which replaces the true standard error with an estimate of the standard error:

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

Why is this okay to do? That is, why does asymptotic normality still hold when using the plug-in estimate of the standard error?

Solution

Part (a)

The first moment is

$$\mu_1 = \mathbb{E}(X_i) = p.$$

In general, the first sample moment is $\frac{1}{m}\sum_{i=1}^{m} Y_i$, where m is the sample size. So the MoM estimator here is $\hat{p}_{MOM} = \frac{1}{n} X_i = \bar{X}$.

Part (b)

The first moment is

$$\mu_1 = \mathbb{E}(X) = np.$$

Rearranging, we have

$$p = \frac{\mu_1}{n}.$$

Here, μ_1 simply equals X, so the MoM estimator is $\hat{p}_{MOM} = \frac{X}{n}$.

A Binomial RV represents the number of successes in n independent Bernoulli trials, so X here is the same as $\sum_{i=1}^{n} X_i$ in the first problem. Thus, the MoM estimator for p is the same for either case, which makes sense since these are statistically equivalent experiments.

Part (c)

Suppose $\hat{v}$ is a consistent estimator for $\text{Var}(\hat{\theta})$. By definition then, $\hat{v} \xrightarrow{p} \text{Var}(\hat{\theta})$. Let $g(y) = \sqrt{y}$ be the square root function, which is continuous. By the Continuous Mapping Theorem, we have that $g(\hat{v}) \xrightarrow{p} g(\text{Var}(\hat{\theta}))$, i.e., $\sqrt{\hat{v}} \xrightarrow{p} se(\hat{\theta})$. So indeed, it is a consistent estimator.

Part (d)

Let $s = \sqrt{\hat{v}}$. Then by a property of the variance, we have

$$\text{Var}(s) = \mathbb{E}(s^2) - (\mathbb{E}(s))^2,$$

so

$$\begin{aligned}(\mathbb{E}(s))^2 &= \mathbb{E}(s^2) - \text{Var}(s)\\ &= \mathbb{E}(\hat{v}) - \text{Var}(s)\\ &= \text{Var}(\hat{\theta}) - \text{Var}(s) \text{ (by def. of } \hat{v} \text{ being unbiased for Var}\hat{\theta})\\ <&\text{Var}(\hat{\theta}).\end{aligned}$$

Thus, $\sqrt{\hat{v}}$ is biased for estimating $se(\hat{\theta})$ and actually underestimates it. This counterexample shows that in general unbiasedness is not preserved for functions of RVs, even if the function is continuous.

Part (e)

Firstly, $\hat{p}(1-\hat{p})/n$ is a consistent estimator for $p(1-p)/n$. This is a pretty well-known fact, but it can also be easily verified using the Weak Law of Large Numbers etc. By what we previously proved in section, we know too then that $\sqrt{\hat{p}(1-\hat{p})/n}$ is consistent for $\sqrt{p(1-p)/n}$. Finally, by Slutsky's Theorem, we have that $\frac{\hat{p}-p}{\sqrt{\hat{p}(1-\hat{p})/n}} \implies N(0,1)$.