STA 111: Probability & Statistical Inference

Lecture Ten – Statistical Inference and Estimators
D.S. Sections 7.1 & 7.9

Instructor: Olanrewaju Michael Akande

Department of Statistical Science, Duke University

Outline

- Questions from Last Lecture
- Point Estimation
- Unbiasedness and Minimum Variance Unbiased Estimators
- Robust Estimators
- Recap

Introduction

– Today we will begin our introduction to statistics/statistical inference.

– Usually in practice, we observe data first without knowing the true distribution. Then the question of interest becomes: given the observed data, what can we say about the underlying population?

– To do that we will first learn about estimators and desirable properties of estimators.

– We will also learn about unbiased estimators, robust estimators and minimum variance unbiased estimators.

Point Estimates

Statisticians often provide two things:

- a point estimate of some quantity of interest, and
- a statement of the uncertainty in that estimate.

Usually, other disciplines only provide the point estimate.

A parameter is some property of a distribution (or density function), such as the mean, median, standard deviation, and so forth.

A point estimate for a parameter is some statistic $h(X_1, \ldots, X_n)$ which, when evaluated for a random sample, gives a sensible approximation to the parameter.

Desirable Properties of Point Estimates

The Central Limit Theorem indicates one of many approaches. If the parameter of interest is the population mean μ, then the statistic $h(X_1, \ldots, X_n) = \bar{X}$ provides a sensible estimate of μ.

In particular, we know the uncertainty in that estimate: $\sigma/\sqrt{n}$.

In fact, the law of large numbers suggests looking at averages as estimators of expectations since averages (and sums) converge to the respective expectations.

There are several desirable properties one can want from a point estimate:

- unbiasedness
- minimum variance (i.e, minimum uncertainty)
- minimum mean squared error.

We discuss these in the context of several estimation strategies.

Common Examples

Besides the mean, other point estimates for common parameters are:

- the sample proportion X/n for the population (or binomial) proportion p.
- the sample variance,

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

for the population variance.

Common Examples

- the average squared deviation,

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

  for the population variance.
- the 10% trimmed sample mean for the population mean; this is the average of the sample after removing the largest 5% of the values and the smallest 5% of the values.

Notice that a point estimator has to be random since it is a function of a random sample from some distribution, but the true parameter itself is constant.

Unbiased Estimates

A point estimate $\hat{\theta} = h(X_1, \ldots, X_n)$ is said to be an unbiased estimator θ for a population parameter θ if $\mathbb{E}[\hat{\theta}] = \theta$.

The bias in a point estimate is $\mathbb{E}[\hat{\theta}] - \theta = \text{bias}(\hat{\theta})$. For unbiased estimates, this is clearly zero.

The mean squared error (MSE) of a point estimate is

$$\mathbb{E}[\hat{\theta} - \theta]^2 = \mathbb{V}[\hat{\theta}] + [\text{bias}(\hat{\theta})]^2.$$

The MSE has many attractive features. In particular, it is sometimes possible to trade-off a small bias for a large reduction in variance, and this leads to better accuracy.

Examples

If X has the $\text{Bin}(n, p)$ distribution, then the sample proportion X/n is an unbiased estimate for the parameter p. To see that, notice that

$$\mathbb{E}[X/n] = \frac{1}{n}\mathbb{E}[X] = \frac{1}{n}np = p.$$

The sample mean $\bar{X}$ of a random sample is unbiased for the population mean μ. We know this from the properties of linear combinations:

$$\mathbb{E}[\bar{X}] = \mathbb{E}[\frac{1}{n}\sum_{i=1}^{n} X_i] = \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}[X_i] = \frac{1}{n}(n\mu) = \mu.$$

In this case we also know the variance of the estimator. Recall that

$$\mathbb{V}[\bar{X}] = \left(\frac{1}{n}\right)^2 \sum_{i=1}^{n}\sigma^2 = \left(\frac{1}{n}\right)^2 (n\sigma^2) = \frac{\sigma^2}{n}.$$

Examples

Example 1: Let X be uniformly distributed on the interval $[0, \theta]$ where $f(x) = 1/\theta$ so that θ is the unknown parameter. You have a random sample $X_1, \ldots X_n$ and use the statistic $\hat{\theta}_1 = Z = \max\{X_1, \ldots, X_n\}$ as an estimate of θ. Then

$$F(x) = \mathbb{P}[X \leq x] = \int_0^x 1/\theta \, dt = x/\theta.$$

and from the lecture 8, $G(z) = \mathbb{P}[Z \leq z] = \mathbb{P}[\max\{X_1, \ldots, X_n\} \leq z]$, and

$$\begin{aligned} \mathbb{P}[\max\{X_1, \ldots, X_n\} \leq z] &= \mathbb{P}[X_1 \leq z \text{ and } \cdots \text{ and } X_n \leq z] \\ &= \prod_{i=1}^{n} \mathbb{P}[X_i \leq z] \\ &= \prod_{i=1}^{n} \frac{z}{\theta} = \left(\frac{z}{\theta}\right)^n. \end{aligned}$$

Examples

So the distribution of the sample maximum is $G(z) = (z/\theta)^n$ for $0 \leq z \leq \theta$ and thus the probability density function of the maximum is $g(z) = n(1/\theta)^n z^{n-1}$ on $0 \leq z \leq \theta$.

Since we know the density, we can find the expected value of Z, where Z is the sample maximum and check if it is unbiased:

$$
\begin{aligned}
\mathbb{E}[\hat{\theta}_1] = \mathbb{E}[Z] &= \int_0^\theta z * \frac{n}{\theta^n} z^{n-1}\, dz = \frac{n}{\theta^n} \frac{1}{n+1} z^{n+1} \Big|_0^\theta \\
&= \frac{n}{n+1}\theta.
\end{aligned}
$$

Examples

So the estimator $\hat{\theta}_1$ of θ has a small bias:

$$\frac{n}{n+1}\theta - \theta = -\frac{\theta}{n+1}$$

.

One can make $\hat{\theta}_1$ into an unbiased estimator by using the new estimator

$$\hat{\theta}_2 = \frac{(n+1)\hat{\theta}_1}{n} = \frac{(n+1)}{n}\max\{X_1, \ldots, X_n\}.$$

Minimum Variance Unbiased Estimators

Usually, a first requirement for a good estimator of a parameter is that it be unbiased. When there are several unbiased estimators, one should use the one that has smallest variance.

This is not the only way to frame the problem of selecting an estimator. For example, one might want the estimator which:

- minimizes the the mean squared error,
- has the largest probability of being within some fixed distance from the true value,
- is unbiased and minimizes something more practical than the variance.

Examples

Consider again the case of a random sample from the Unif$(0, \theta)$ distribution. Then for any $X \sim$ Unif$(0, \theta)$, $\mathbb{E}[X] = \frac{\theta}{2}$. Clearly, $\mathbb{E}[\bar{X}] = \frac{\theta}{2}$ so $\hat{\theta}_3 = 2\bar{X}$ is an unbiased estimator of θ.

We now have two candidate estimators:

$$\hat{\theta}_2 = \frac{(n+1)}{n} \max\{X_1, \ldots, X_n\} \text{ and } \hat{\theta}_3 = 2\bar{X}.$$

Which has the smaller variance?

Since $\hat{\theta}_3$ is a linear combination, its variance is $\frac{4\sigma^2}{n}$ where σ^2 is the variance of the Unif$(0, \theta)$ distribution. You can check that the variance of the Unif$(0, \theta)$ distribution is $\theta^2/12$. Thus

$$\mathbb{V}[\hat{\theta}_3] = \frac{\theta^2}{3n}.$$

Examples

To find the variance of $\hat{\theta}_2$ we first find

$$\mathbb{E}[Z^2] = \int_0^\theta z^2 g(z)\, dz = \int_0^\theta z^2 * \frac{n}{\theta^n} z^{n-1}\, dz = \frac{n}{n+2}\theta^2.$$

Since $\mathbb{V}[Z] = \mathbb{E}[Z^2] - (\mathbb{E}[Z])^2$, we have

$$\mathbb{V}[Z] = \frac{n}{n+2}\theta^2 - \left[\frac{n}{n+1}\theta\right]^2 = \left[\frac{n}{(n+2)(n+1)^2}\right]\theta^2.$$

Since $\hat{\theta}_2 = \frac{(n+1)}{n} Z$, then

$$\mathbb{V}[\hat{\theta}_2] = \left(\frac{n+1}{n}\right)^2 \left[\frac{n}{(n+2)(n+1)^2}\right]\theta^2 = \frac{1}{n(n+2)}\theta^2.$$

A little algebra shows that $n(n+2) > 3n$ for all $n > 1$, so $\hat{\theta}_2$ is better than $\hat{\theta}_3$.

Robust Estimators

Previously, we claimed to like estimators that are unbiased, have minimum variance, and/or have minimum mean squared error. Typically, one cannot achieve all of these properties with the same estimator.

An estimator may have good properties for one distribution, but not for another. We saw that $\frac{n}{n-1}Z$, for Z the sample maximum, was excellent in estimating θ for a $\text{Unif}(0, \theta)$ distribution. But it would not be excellent for estimating θ every pdf supported on $[0, \theta]$.

A robust estimator is one that works well across many families of distributions. In particular, it works well when there may be outliers in the data.

Robust Estimators

The 10% trimmed mean is a robust estimator of the population mean. It discards the 5% largest and 5% smallest observations, and averages the rest. Obviously, one could trim by some fraction other than 10%, but this is a commonly-used value.

Surveyors distinguish errors from blunders. Errors are measurement jitter attributable to chance effects, and are approximately Gaussian. Blunders occur when the guy with the theodolite is standing on the wrong hill.

A trimmed mean throws out the blunders and averages the good data. If all the data are good, one has lost some sample size. But in exchange, you are protected from the corrosive effect of outliers.

Recap

Today we covered:

- Point Estimation
- Unbiasedness and Minimum Variance Unbiased Estimators
- Robust Estimators