Chapter 7 Overview

Introduction:

As we discussed in Chapter 6, it is important to remember the chain of information:

{unknown probability density function *f(x)*}

↓ (experimentation)

{sample data set}

↓ (statistical inference)

{estimate properties of *f(x)*}

That is, a certain random variable has some probability distribution that is unknown to us. We assemble a data set out of values "generated" by this random variable. We then study the data set to do our best to figure out what that unknown probability distribution looks like. We define different terms to distinguish between information about the data set and information about the unknown probability distribution.

A **statistic** is a characteristic or measure obtained by using the data values from a sample.

A **parameter** is a quantity that is a property of the original probability distribution.

So while the sample mean, sample median, sample variance, and sample quantiles from Chapter 6 are statistics, the unknown probability distribution has an actual mean, median, variance, and quantiles as defined in Chapter 2. The actual mean, median, variance and quantiles are all parameters.

In practice, since parameters are generally unknown, we must use statistical inference to estimate parameters. So essentially, statistics are what we know based on observations we make, while parameters are what we try to figure out as close as we can. Our first approach to estimating parameters is using point estimates.

7.1 Point Estimates

A **point estimate** of an unknown parameter θ is a statistic $\hat{\theta}$ that represents a "best guess" at the value of θ. There may be more than one sensible point estimate of a parameter. (For example, either the trimmed sample mean or the sample mean could be used to estimate the population mean, but one of these point estimates may be "better" than the other.)

Examples:

- The unknown probability distribution has a mean, μ. Then our sample mean $\bar{x}$ is a point estimate of μ.
- The unknown probability distribution has a standard deviation, σ. Then our sample standard deviation s is a point estimate of σ.

One extra very important thing to remember is that since our statistics are built on data that is random, this means that statistics are also random variables themselves! We've already seen this a bit in Chapter 2, when we studied $\bar{X}$, which we defined as the arithmetic average of a collection of i.i.d. random variables. Similarly, we have that $\bar{x}$ is the arithmetic average of a collection of multiple observations of the same random variable. In principle, these things are different, but mathematically, we focus on the similarities, since two observations of the same random variable functions the same way as one observation each from two different (but identical) random variables.

7.2 Properties of Point Estimates

Unbiased and Biased Point Estimates

A point estimate $\hat{\theta}$ for a parameter θ is said to be **unbiased** if $E(\hat{\theta}) = \theta$. Unbiasedness is a good property for a point estimate to possess.

If a point estimate is not unbiased, then its **bias** can be defined to be bias = $E(\hat{\theta}) - \theta$.

All other things being equal, the smaller the absolute value of the bias of a point estimate, the better.

Note that we can (often) figure out the bias of a point estimate without knowing the exact value of the parameter! But we do usually need a little information, such as the general type of random variable we're dealing with (i.e.: binomial, Poisson, normal, exponential, etc.) even if we don't know the actual values of the parameters.

The book illustrates this well with a Bernouilli random variable. In each trial, we either get success or failure. The parameter needed to completely define the random variable is *p*, the probability of success in each trial. We can take *n* observations from this random variable. Let *X* be the number of successes we observe. In this case, the sample mean is just $\frac{X}{n}$, so this should give us a point estimate of the unknown parameter *p*. Therefore, we'll call this point estimate $\hat{p}$. But we note that *X* is itself a binomial random variable, since it is the number of successes in a fixed number of Bernoulli trials! Since $E(X) = np$, this tells us that $E(\hat{p}) = E(\frac{X}{n}) = \frac{E(X)}{n} = \frac{np}{n} = p$. Therefore, the point estimate is unbiased!

Point Estimate of a Success Probability

Suppose that $X \sim \text{Bin}(n, p)$. Then $\hat{p} = \frac{X}{n}$ is an unbiased point estimate of the success probability *p*.

This same principle will work for situations other than Bernoulli trials to estimate the unknown mean μ.

Point Estimate of a Population Mean

If $X_1, \ldots, X_n$ is a sample of observations from a probability distribution with a mean μ, then the sample mean $\hat{\mu} = \bar{X}$ is an unbiased point estimate of μ.

In particular, when the underlying probability distribution is not symmetric:

1) a trimmed sample mean is in general not an unbiased point estimate of the population mean
2) a sample median is in general not an unbiased point estimate of the population median.

This does not mean these estimates should never be used, as they may still have a relatively small bias.

Point Estimate of a Population Variance

If $X_1, \ldots, X_n$ is a sample of observations from a probability distribution with a variance σ^2, then the sample variance $\hat{\mu}^2 = S^2 = \dfrac{\sum_{i=1}^{n} (X_i - \bar{X})^2}{n-1}$.

is an unbiased point estimate of the population variance σ^2.

We have seen that a good point estimate should have an expectation that is the same as the population parameter (unbiased). We also want one with as small a variance as possible. That is we want Var($\hat{\theta}$) to be as small as possible in order to be more confident that we have an estimate closer to the true value θ. An unbiased point estimate that has a smaller variance than any other unbiased point estimate is called a **minimum variance unbiased estimate** (MVUE).

It is important to note that if $X_1, \ldots, X_n$ is a sample of observations that are independently normally distributed with mean μ and variance σ^2, then the sample mean $\bar{X}$ is a minimum variance unbiased estimate of the mean μ.

To save time, we will omit relative efficiency and mean square error from this class, though you are free to read about them in the book, even though you won't be tested on them.

7.3 Sampling distributions

As we've mentioned, statistics are random variables themselves, so they have their own probability distributions. In this section, we investigate what the distributions are for some of the basic sample statistics we defined in Chapter 6. We'll call these **sampling distributions**.

To start, let's revisit the Bernoulli trial/Binomial random variable situation from the last section. We saw that if $X \sim \text{Bin}(n, p)$, then $\hat{p} = \dfrac{X}{n}$ is an unbiased point estimate of the success probability p. We also saw in Chapter 2 that Var(X) = $np(1\text{-}p)$, which means that

$Var\left(\hat{p}\right) = \dfrac{1}{n^2}Var(X) = \dfrac{np(1-p)}{n^2} = \dfrac{p(1-p)}{n}$. So we know the mean and variance of the sample proportion $\hat{p}$, but we still wish to know more about what kind of distribution $\hat{p}$ has.

But from the Central Limit Theorem from Chapter 5, we get the following.

Sample proportion

If $X \sim B(n, p)$, then the sample proportion $\hat{p} = \frac{X}{n}$ has the approximate distribution

$$\hat{p} \sim N\left(p, \ \frac{p(1-p)}{n}\right)$$

Furthermore, the standard deviation of $\hat{p}$ is called its <u>standard error</u> and is denoted by s.e.($\hat{p}$). That is, $\text{s.e.}(\hat{p}) = \sqrt{\frac{p(1-p)}{n}}$.

Since we don't actually know the real value of p, we don't actually know exactly what s.e.($\hat{p}$) is. Like other statistics, it is also a random variable. However, we can approximate it by replacing p with $\hat{p} = \frac{x}{n}$, where x is the observed number of successes. This gives us the estimation $\text{s.e.}(\hat{p}) \approx \frac{1}{n}\sqrt{\frac{x(n-x)}{n}}$.

The standard error will be used in the future to indicate the accuracy of a point estimation $\hat{p}$. Smaller values of the standard error indicate less variability and thus more accurate estimates. Increasing the sample size generally reduces the standard error.

Sample Mean:

We generalize beyond Bernoulli trials to suppose that $X_1, \ldots, X_n$ are independent, identically distributed RVs with a mean μ. (We can think of these as observations from a single RV.) We saw last section that the sample mean $\hat{\mu} = \bar{X}$ is an unbiased point estimate of the population mean μ. Once again, the Central Limit Theorem tells us how we can approximate $\hat{\mu}$ as a random variable.

Sample Mean

If $X_1, \ldots, X_n$ are observations from a population with mean μ and variance σ^2, then the sample mean $\hat{\mu} = \bar{X}$ has the approximate distribution

$$\hat{\mu} = \bar{X} \sim N\left(\mu, \ \frac{\sigma^2}{n}\right)$$

The standard error of the sample mean is $\text{s.e.}(\bar{X}) = \frac{\sigma}{\sqrt{n}}$.

Note again that the larger the sample size, the smaller the standard error becomes.

We end just by mentioning two more sampling distributions given in this section of the book, which will be used more in Chapter 8.

Sample Variance

If $X_1, \ldots, X_n$ are normally distributed with a mean μ and a variance σ^2, then the sample variance S^2 has the distribution (note below χ is used for the chi-square distribution).

$$S^2 \sim \sigma^2 \frac{\chi^2_{n-1}}{(n-1)}$$

Then since $\bar{X} \sim N\left(\mu,\ \sigma^2\right)$, we can convert $\bar{X}$ to a standard normal by subtracting the mean and dividing by the standard deviation. Thus, $\dfrac{\sqrt{n}(\bar{X} - \mu)}{\sigma} \sim N(0,\ 1)$. Combining this with the above to replace σ with S, we have

If $X_1, \ldots, X_n$ are normally distributed with a mean μ, then

$$\frac{\sqrt{n}(\bar{X} - \mu)}{S} \sim t_{n-1}.$$

The quantity $\dfrac{\sqrt{n}(\bar{X} - \mu)}{S}$ is called a *t*-statistic.

We'll skip Section 7.4 in this class, so this is the end of Chapter 7.