Chapter 6 - Point Estimation

When we assume a class of models (like Gamma, or Normal, ...) for the source of data, and we have a random sample of data collected, it is common to be interested in the value of parameters of the assumed model. The value of a model parameter must be estimated from the sample data. For example, we might use the sample mean to estimate the population mean – this seems so obvious you may wonder why we need a chapter to describe it! But some parameter estimators (functions of the sample data designed to estimate a parameter) are not so obvious. This chapter 6 tells how to figure out these estimators.

Usually when we estimate a parameter, we would like to know how accurate the estimate is. We deal with this in future chapters. For now, we are just concerned with the "point" estimate itself, and not how accurate the point estimate is.

What is a good estimator? There are a few criteria that have been proposed:

Unbiasedness – the estimator must have the property that it gives the true parameter value **on average.**

Minimum Variance – the estimator must have the smallest variance possible.

Minimum Variance Unbiased Estimator (MVUE) The estimator has the minimum variance among all unbiased estimators.

Minimim Mean Squared Error (MMSE) The estimator has the smallest average variability around the true parameter value (as measured by the squared deviation).

MVUE is traditional in statistics, though hard to justify – who wants to be correct on average, rather than close all the time (MMSE)?

Minimum Variance is unworkable – the estimator "0" has minimum variance always, and is totally useless!

The sample mean and sample SD are MVUE of the population mean and variance, respectively.

Measuring the quality of an estimator. The most common way is with the Standard Error (SE). The SE of an estimator is really just the SD of the sampling distribution of the estimator. For example, the SE of the sample mean (SEM) = SD of the population $/\sqrt{n}$ and this is usually estimated by s/ $\sqrt{n}$

There are some mathematical ways to generate estimators of parameters in specific situations – other than the common ones of estimating means or SDs.

The two methods we will introduce in this course are
1. The Method of Moments
2. The Maximum Likelihood Method.

The method of moments:

$E(X^k)$ is called the kth moment of the RV X (or the "kth population moment") and $(1/n)\sum_{i=1}^{n} x_i^{k}$ is called the kth sample moment.
If the probability model for X has a parameter θ, the equation
$E(X^k) = (1/n)\sum_{i=1}^{n} x_i^{k}$
will be an equation in θ, for each k.

Typically, we need as many equations (for k=1,2,3,...) as there are parameters, in order to be able to solve for the parameter as a function of the data $x_1, x_2, ..., x_n$.

For example, suppose X is exponential with mean λ^{-1}. Then we set

$\lambda^{-1} = (1/n)\sum_{i=1}^{n} x_i$ so the estimate of λ, which is usually denoted $\hat{\lambda}$,

would be $\hat{\lambda} = ((1/n)\sum_{i=1}^{n} x_i)^{-1}$
This $\hat{\lambda}$ is the Method-of-moments estimator of λ.

2. Maximum Likelihood Method

Suppose X has density from a class of densities $f_X(x)$. When we say "class" here we are thinking of the density having an unspecified parameter, and any value of the parameter would give, via this function, a particular density. The object is to use data values to tell you which parameter value you should use. In order to focus on the parameter, we will re-write the density as $f_X(x; \theta)$. To keep things simple, suppose we just have one data value X=x. If you had to guess θ on the basis of this one X=x, how would you do it? Remembering that $f_X(x; \theta)$ is telling you how likely X=x is, wouldn't you choose a value of θ that made this as large as possible? Such a θ would be, as a function of x, the maximum likelihood estimator of θ.

For example, suppose X is Poisson with mean m. $P(X=x)=(e^{-m}m^{x})/x!$
What value of m maximizes this function? If you set the derivative =0, you find $\hat{m}=x$.
This $\hat{m}$ is the maximum likelihood estimator (MLE) of m.

Sometimes neither method produces satisfactory estimators – there are some examples in Ch 6.

Of course, once a point estimator for a parameter is determined, it is of interest to know how precise the estimate is. It might seem that it is impossible for sample to both provide a parameter estimate and also to say how accurate it is likely to be. We saw one example last lecture: when we use the sample mean to estimate the population mean, the estimated standard error of the mean ($s/n^{1/2}$) does just that. But actually there is a general method that is always available and easy to understand – it is called the **bootstrap**.

Here is the general idea:

A random sample of size n is selected from an unknown population. A parameter of the population (such as the mean or SD) is estimated by a statistic $t(x_1,x_2,...,x_n)$. That is $t(x_1,x_2,...,x_n)$ estimates some parameter θ. Now we want to know how accurate this estimate of θ is. Here is how to do it with the bootstrap:
1. Take a random sample of size n, with replacement, from the original sample of size n. We call this a bootstrap sample.
2. Calculate the value of t() for the bootstrap sample, and save it for later use.
3. Repeat step 1. and 2. several hundred times.
4. Compute the standard deviation of the several hundred values of t().
The result is an estimate of the SD of the original $t(x_1,x_2,...,x_n)$.

I will demo this process in class to show how to compute the variability of the sample mean without assuming you know the square root law. Then we can compare the bootstrap method with the theoretical approach (square root law). It works!