LECTURE NOTE 1: PARAMETERS AND ESTIMATORS

The goal of this lecture is to give you a refresher on key concepts in statistics and to set notation for the rest of the quarter.

The properties of random variables (e.g., expectations, variances, covariances) are called *population parameters*. For example, a random variable X has mean (or expected value) $\mu_X = E[X]$, variance $\sigma_X^2 = V[X]$, and standard deviation $\sigma_X = \sqrt{\sigma_X^2} = \sqrt{V[X]}$. The mean measures the central tendency of X, while the variance and standard deviation meaure the dispersion of X.

We will be interested in the properties of *samples*, which are called *statistics*. Here, we consider the properties of a sample of N observations of the random variable X: $X_1, X_2, \cdots, X_N$. The observations are mutually independent draws from the distribution of X. We will say they are are "independently and identically distributed," or "i.i.d." Note the implication: *each observation has its own probability distribution*, and we assume that all of these individual distributions are identical. We will derive statistics of the sample to estimate population parameters for the distribution of X. We will call such statistics *estimators*.

Usually, we will use Greek letters to refer to parameters and Greek letters with hats to refer to their corresponding estimators. For example, suppose we are interested in a regression parameter β. We will typically use $\hat{\beta}$ to refer to an estimator for β. For generic discussions about parameters and estimators, we will use the parameter θ and the estimator $\hat{\theta}$. When I write θ and $\hat{\theta}$, they could be standing in for μ and $\hat{\mu}$, or for σ and $\hat{\sigma}$, or for β and $\hat{\beta}$, or for ρ and $\hat{\rho}$.

We often want $\hat{\theta}$ to satisfy some of the following properties:

1. Unbiasedness: $E[\hat{\theta}] = \theta$.
2. Consistency: As the sample size grows, $\hat{\theta}$ gets closer and closer to θ. Mathematically, we express this property in the following way: as $N \to \infty$, $Pr\left[|\hat{\theta} - \theta| > \varepsilon\right] \to 0$ for any constant ε. We say that $\hat{\theta}$ "converges in probability" to θ, and we write $\hat{\theta} \xrightarrow{p} \theta$.
3. Efficiency (or Precision): $\hat{\theta}$ has the smallest possible variance $V[\hat{\theta}] = E\left[\left(\hat{\theta} - E[\hat{\theta}]\right)^2\right]$.

Properties (1) and (2) are both about the accuracy of $\hat{\theta}$, so we will typically seek estimators that either satisfy (1) and (3) or satisfy (2) and (3). Note that all three properties, especially (1) and (3), are features of the distribution of $\hat{\theta}$. This fact bears repeating: $\hat{\theta}$ has its own distribution. Make sure you understand this abstract concept, for it underlies much of the rest of the course.

Property (3) is about the variance of $\hat{\theta}$, which captures how 'noisy' $\hat{\theta}$ is. If the variance of $\hat{\theta}$ is high, then $\hat{\theta}$ is not very informative, even if it is unbiased or consistent. Hence our desire to find an estimator with the smallest possible variance.

We use the same word, *variance*, to refer to:

1. the expected squared deviation of X from its expected value
2. the expected squared deviation of $\hat{\theta}$ from its expected value

However, when we take the square roots of these quantities, as is common in applied statistical research, we use different names. As noted above, the square root of the variance of X is called the *standard deviation* of X. In contrast, the square root of the variance of $\hat{\theta}$ is called the *standard error* of $\hat{\theta}$. Students mix up these terms a lot! Just remember: standard deviations are for variables, standard errors are for estimators.