

MIT OpenCourseWare
<http://ocw.mit.edu>

14.30 Introduction to Statistical Methods in Economics
Spring 2009

For information about citing these materials or our Terms of Use, visit: <http://ocw.mit.edu/terms>.

Problem Set #8 - Solutions

14.30 - Intro. to Statistical Methods in Economics

Instructor: Konrad Menzel

Due: Tuesday, April 28, 2009

Question One: Law of Large Numbers and Central Limit Theorem

Probably the two most important concepts that you will take away from this course are the Law of Large Numbers and the Central Limit Theorem and how they allow us to use averages to learn about the world around us.

1. State the Law of Large Numbers (please, just copy it down from the lecture notes).
 - Solution to 1: Suppose X_1, \dots, X_n is a sequence of i.i.d. draws with $\mathbb{E}[X_i] = \mu$ and $Var(X_i) = \sigma^2 < \infty$ for all i . Then for any $\epsilon > 0$ (typically a small value), the sample mean satisfies

$$\lim_n P(|\bar{X}_n - \mu| > \epsilon) = 0$$

We say that \bar{X}_n converges in probability to μ .

2. Explain what the Law of Large Numbers tells us about the average of an i.i.d. (independent, identically distributed) sample of the random variable X with mean μ and variance σ^2 .
 - Solution to 2: The law of large numbers tells us that the density of the average of an i.i.d. sample of a random variable X will be concentrated in an “epsilon ball” of radius ϵ . Or, more rigorously, for any $\epsilon > 0$, if we take an infinite sample of a random variable, the density of its mean will be concentrated at μ .

Suppose you wanted to know the unemployment rate for residents of Cambridge during April 2009. The “unemployed” are defined as “Persons 16 years and over who had no employment during the reference week, were available for work, except for temporary illness, and had made specific efforts to find employment sometime during the 4-week period ending with the reference week. Persons who were waiting to be recalled to a job from which they had been laid off need not have been looking for work to be classified as unemployed.” (Source: <http://www.econmodel.com/classic/terms/ur.htm> .)

Suppose you utilize a phone survey to sample the random variable $X = 1(Employed)$ where $1(\cdot)$ is the indicator function for whether someone is employed.

1. Write down an estimator, $\hat{\alpha}$, of the unemployment rate, α , which is the fraction of the labor force that is unemployed. Is your estimator a Method of Moments estimator for a Bernoulli random variable?

- Solution to 1: The estimator of choice would be $\hat{\alpha} \equiv \frac{1}{N} \sum_{i=1}^N X_i$. This estimator is a Method of Moments estimator which uses the mean of the distribution of X_i , which is a Bernoulli random variable since it only takes on the values 0 and 1.

2. Describe how the Law of Large Numbers applies to the estimator by stating what (at least three) conditions are required to hold about X in order for your estimator to be consistent (by the Law of Large Numbers you copied down from the lecture notes above).

- Solution to 2: The law of large numbers applies to this estimator since it is an average. In other words, as we survey more and more people, our estimator, $\hat{\alpha} \rightarrow^p \alpha$, where α is the true unemployment rate. The three conditions that we need are that the sample is (1) a collection of i.i.d. random variables (2) with finite mean and (3) finite variance.

3. For each condition you wrote down, comment on the plausibility of the assumption holding with respect to the unemployment rate.

- Solution to 3: Condition (1): The i.i.d.-ness of our sample can be implemented via a random phone survey. However, if we just select a “convenience sample” of people that happen to walk by 77 Mass. Ave. at noon, we’ll probably not have a representative sample of the population (i.e. our sample wouldn’t be independent or identically distributed).

Condition (2): The finite mean condition is guaranteed by the random variable being bounded between 0 and 1.

Condition (3): The finite variance condition is similarly guaranteed since a Bernoulli random variable has variance of $p(1-p)$ where p is the probability of drawing a 1 and is bounded between 0 and 1.

Now we’re going to take a closer look at the Central Limit Theorem.

1. State the Central Limit Theorem (please, just copy it down from the lecture notes).

- Solution to 1: Suppose X_1, \dots, X_n is a random sample of size n from a given distribution with mean μ and variance $\sigma^2 < \infty$. Then for any fixed number x ,

$$\lim_{n \rightarrow \infty} P \left(\sqrt{n} \frac{\bar{X}_n - \mu}{\sigma} \leq x \right) = \Phi(x)$$

We say that $\sqrt{n}(\bar{X}_n - \mu)$ converges in distribution (some people also say “converges in law”) to a normal with mean μ and variance σ^2 , or in symbols:

$$\sqrt{n}(\bar{X}_n - \mu) \rightarrow^d N(0, \sigma^2).$$

2. Explain what the Central Limit Theorem tells us about the average of an i.i.d. (independent, identically distributed) sample of the random variable X with mean μ and variance σ^2 .

- Solution to 2: The Central Limit Theorem (CLT) tells us that the average of an i.i.d. sample (random sample) of a random variable X will converge in distribution to a Normal. This means that no matter what distribution X was, we will be able to compare its average to the Normal distribution.

3. Describe how the Central Limit Theorem applies to the estimator you wrote down for the unemployment rate by stating what (at least three) conditions are required to hold about X in order for your estimator to be asymptotically normally distributed. Are these conditions different from those required for the Law of Large Numbers to apply?

- Solution to 3: Since the unemployment rate estimator is just the average of a random sample of the Bernoulli random variable of whether someone is unemployed or not, we know that its distribution will be Normal. The three conditions that are necessary to hold are the same as for the Law of Large Numbers (LLN) to hold: (1) i.i.d. random sample, (2) finite mean, and (3) finite variance.

4. Write down the distribution that your estimator converges to as $N \rightarrow \infty$, where N is the number of people you surveyed.

- Solution to 4: We write:

$$\sqrt{N}(\hat{\alpha} - \alpha) \rightarrow^d N(0, \alpha(1 - \alpha)).$$

5. Write down an estimator for the variance of X . Briefly comment on the assumptions required of the random variable $Y = X^2$ in order for your estimator to be consistent.

- Solution to 5: One estimator of the variance of X would be to make use of the variance identity $Var(X) = \mathbb{E}[X^2] - \mathbb{E}[X]^2$ to construct an estimator using the two Method of Moments estimators for the first and second noncentral moments:

$$\hat{\sigma}^2 = \left(\frac{1}{N} \sum_{i=1}^N X_i^2 \right) - \left(\frac{1}{N} \sum_{i=1}^N X_i \right)^2.$$

Equivalently, we could use the sample analog to the variance using our estimator of the unemployment rate and write:

$$\hat{\sigma}^2 = \frac{1}{N} \sum_{i=1}^N (X_i - \hat{\alpha})^2.$$

For either of these estimators, we need $\mathbb{E}[X^2] < \infty$ which is entirely reasonable for this example, since X is bounded. However, we also need the variance of the variance to be bounded: $Var((X - \alpha)^2) = \mathbb{E}[(X - \alpha)^4] - \mathbb{E}[(X - \alpha)^2]^2 < \infty$. A sufficient condition for this to hold is $\mathbb{E}[X^4] < \infty$. The nice thing about Bernoulli random variables is that $0 \leq \mathbb{E}[X^{k+1}] = \mathbb{E}[X^k] \leq 1$ for $k \geq 1$. So, in general, we can easily bound all of the moments of the Bernoulli random variable.

6. Now, use the fact that X is a Bernoulli random variable to write down a different estimator of the variance of X as a method of moments estimator (i.e. a function of your consistent estimator of the unemployment rate). Although the formula looks different, are these two estimators numerically identical? Do they need the same assumptions to hold for the Law of Large Numbers to apply?

- Solution to 6: Since X is a Bernoulli random variable, we can remember that the variance is $p(1 - p)$, where p is the probability of drawing a success or 1 for the random variable. A different estimator would be:

$$\begin{aligned}
 \tilde{\sigma}^2 &= \hat{\alpha}(1 - \hat{\alpha}) \\
 &= \left(\frac{1}{N} \sum_{i=1}^N X_i \right) \left(1 - \frac{1}{N} \sum_{i=1}^N X_i \right) \\
 &= \frac{1}{N} \sum_{i=1}^N X_i - \left(\frac{1}{N} \sum_{i=1}^N X_i \right) \left(\frac{1}{N} \sum_{i=1}^N X_i \right) \\
 &= \frac{1}{N} \sum_{i=1}^N X_i^2 - \left(\frac{1}{N} \sum_{i=1}^N X_i \right)^2 \\
 \tilde{\sigma}^2 &= \hat{\sigma}^2
 \end{aligned}$$

For the Bernoulli random variable, the two estimators are, in fact, numerically identical. However, as we used the fact that $X^2 = X$ for a Bernoulli random variable (i.e. $1^2 = 1$ and $0^2 = 0$) in the fourth equation, this result will not generally hold for other random variables, although we may be able to similarly construct Maximum Likelihood Estimators from Method of Moments estimators. As should be expected, since the two estimators are numerically identical, they certainly will need the same assumptions to hold for the Law of Large Numbers to apply.

7. Use your estimators for the average unemployment rate and the variance of X : How many people do you need to call if you want your estimator of α to be within 0.002 (i.e. 0.2% unemployment) with 95% probability? (Assume that since unemployment rose from 8.1 to 8.5 from February to March that your expectations are that it will rise to 8.7 in April.)

- Solution to 7: This is just a similar power calculation as we did for the last problem set where we more rigorously note the conditions necessary for the CLT to apply (recognize that it is just an approximation):

$$\begin{aligned}
 \lim_{n \rightarrow \infty} P \left(\sqrt{n} \frac{\bar{X}_n - \mu}{\sigma} \leq x \right) &= \Phi(x) \\
 P \left(\sqrt{n} \frac{|\bar{X}_n - \mu|}{\sigma} \leq x \right) &\approx \Phi(x) - \Phi(-x)
 \end{aligned}$$

Recognizing that $\hat{\alpha}$ is just an average, we just plug in the relevant pieces to the above CLT approximation and the estimators:

$$\begin{aligned}
 P(|\hat{\alpha} - \alpha| < 0.002) &= 0.95 \\
 P\left(\sqrt{N}\frac{|\hat{\alpha} - \alpha|}{\sqrt{\tilde{\sigma}^2}} < \sqrt{N}\frac{0.002}{\sqrt{\tilde{\sigma}^2}}\right) &= 0.95 \\
 P\left(\sqrt{N}\frac{\hat{\alpha} - \alpha}{\sqrt{\hat{\alpha}(1 - \hat{\alpha})}} < -\sqrt{N}\frac{0.002}{\sqrt{\hat{\alpha}(1 - \hat{\alpha})}}\right) &= 0.025 \\
 \Phi\left(-\sqrt{N}\frac{0.002}{\sqrt{\hat{\alpha}(1 - \hat{\alpha})}}\right) &= 0.025 \\
 -\sqrt{N}\frac{0.002}{\sqrt{\hat{\alpha}(1 - \hat{\alpha})}} &= \Phi^{-1}(0.025) = -1.96 \\
 N &= \left\lceil \left(\frac{1.96}{0.002}\sqrt{\hat{\alpha}(1 - \hat{\alpha})}\right)^2 \right\rceil
 \end{aligned}$$

where we need to use our expectations of the unemployment rate this month of $\mathbb{E}[\hat{\alpha}] = 0.087$ to get an estimate of the variance. Alternatively, we could use March's unemployment rate (8.5%) to get an estimate of the variance. We'll use both and see how the sample size differs:

$$\begin{aligned}
 N &= \left\lceil \left(\frac{1.96}{0.002}\sqrt{0.087(1 - 0.087)}\right)^2 \right\rceil = 76286 \\
 N &= \left\lceil \left(\frac{1.96}{0.002}\sqrt{0.085(1 - 0.085)}\right)^2 \right\rceil = 74696
 \end{aligned}$$

To see whether the Current Population Survey (CPS), which measures unemployment, is using a large enough sample each month, I investigated their website: "Each month, 2,200 highly trained and experienced Census Bureau employees interview persons in the 60,000 sample households for information on the labor force activities (jobholding and jobseeking) or non-labor force status of the members of these households during the survey reference week (usually the week that includes the 12th of the month)." (Source: http://www.bls.gov/cps/cps_htgm.htm.) Thus, with 60,000 households, the CPS likely obtains the working status of more than 75,000 people that are in the labor force in order to get a tight margin of error on their monthly unemployment statistics. However, since they oversample certain demographics in order to say things about subpopulations, they may need even more households to obtain a precise estimate of overall unemployment.

It should be noted, however, that under usual economic circumstances, where unemployment rates hover around 5.0%, a 0.2% margin would only require a sample of 45,619. Thus, higher unemployment rates actually make precision harder for the CPS, since the same level of precision in unemployment rate estimates requires much larger sample sizes (up until 50% unemployment where the variance

of a Bernoulli random variable is maximized).

Thus, any time you ever use the unemployment rate for statistical analyses, you should remember that these are all measured with error! We don't know the exact unemployment rate, but only know it within a reasonable ($\sim 0.2\%$) margin of error! This means that even if we estimate the unemployment rate to be 8.7% in April with our sample of 75,000 people, we actually can't be absolutely certain that unemployment has gone up from our 8.5% (with a margin of error of about 0.2%) from our estimate of if 8.7% (also with a margin of error of about 0.2%). So, understand the statistics! :)

Question Two: Unbiasedness v. Consistency

First, what is the difference between unbiasedness and consistency? Second, prove that the sample average, $\frac{1}{N} \sum_{i=1}^N X_i$, is an unbiased estimator of a sample of N i.i.d. random variables, X_1, \dots, X_N , where $\mathbb{E}[X_i] = \mu$. Third, show that it is a consistent estimator of μ under one additional assumption and give the assumption that you need to make.

- Solution: Per the notes, “An estimator $\hat{\theta} = \hat{\theta}(X_1, \dots, X_n)$ is unbiased for θ if $\mathbb{E}_{\theta_0}[\hat{\theta}] = \theta_0$ for all values of θ_0 .” The definition of consistency follows: “For a sample X_1, \dots, X_n , we say that $\hat{\theta}$ is a consistent estimator for θ if as we increase n , the estimator converges in probability to θ_0 , i.e. for all $\epsilon > 0$,

$$\lim_{n \rightarrow \infty} P_{\theta_0} \left(\left| \hat{\theta}(X_1, \dots, X_n) - \theta_0 \right| < \epsilon \right) = 1$$

for all values of θ_0 .”

Thus, unbiasedness is a finite-sample argument which says that the average (expected) result of an estimator is equal to the true parameter. Consistency, on the other hand, says that an estimator will converge to the true parameter as $n \rightarrow \infty$. Consistency is sometimes referred to as “asymptotic unbiasedness,” although unbiasedness in finite samples is not required for consistency. In other words, any bias goes to zero as $n \rightarrow \infty$. Proving that the sample average is unbiased is the same thing we've proved time and time again:

$$\mathbb{E}[\bar{X}_n] = \frac{1}{n} \mathbb{E}[X_1 + \dots + X_n] = \frac{1}{n} (\mathbb{E}[X_1] + \dots + \mathbb{E}[X_n]) = \frac{1}{n} n \mathbb{E}[X_i] = \mathbb{E}[X_i] = \mu.$$

Showing that $\bar{X}_n \xrightarrow{p} \mu$ just requires a LLN to apply. The only additional assumption we need is $Var(X) = \sigma^2 < \infty$. With this, we can invoke the LLN and state, for $\epsilon > 0$:

$$\begin{aligned} \lim_{n \rightarrow \infty} P_{\mu} (|\bar{X}_n - \mu| > \epsilon) &= 0 \\ \lim_{n \rightarrow \infty} P_{\mu} (|\bar{X}_n - \mu| < \epsilon) &= 1 \end{aligned}$$

which gives us consistency of the sample mean, \bar{X}_n .

Question Three: Avoiding Vocabulary Ambiguity

Avoid ambiguity in your understanding and use of similar terms.

1. Define the term “statistic.” Is a statistic a random variable?

- Solution to 1: A statistic is a function of a sample (data) X_1, \dots, X_n . And, a function of random variables is also a random variable.

2. Define the term “estimator.” Is an estimator a random variable? What’s the difference between an estimator and a statistic? Or is this just semantics?

- Solution to 2: An estimator $\hat{\theta}$ of θ is a statistic (i.e. a function of X_1, \dots, X_n), $\hat{\theta} = \hat{\theta}(X_1, \dots, X_n)$. An estimator is a random variable since it is a statistic which is also a random variable. Further, an estimator is a statistic which has a particular population parameter θ which it is intended to estimate. Thus, there is a difference, although generally the two will be interchanged outside of economics.

3. Define the term “realization” of an estimator.

- Solution to 3: A realization of an estimator is the evaluation of the estimator using a realization of a sample (or a collection of draws from the random variables X_1, \dots, X_n . This is also called an estimate.

4. Define the term “estimate.”

- Solution to 4: An estimate is the realization of an estimator (the function of the realizations of the sample).

5. Define the “standard deviation” of a random variable, X .

- Solution to 5: The standard deviation of a random variable is $\sqrt{Var(X)}$.

6. Define the “standard error” of an estimator, $\hat{\theta}(X)$.

- Solution to 6: The standard error of an estimator is the standard deviation of the estimator: $\sqrt{Var(\hat{\theta}(X))}$. So, we can think of the standard error as the standard deviation of the random variable (estimator) $\hat{\theta}(X)$.

Question Four: The Delta Method

Give the standard error of the estimator $\hat{\theta}_Z = \frac{1}{N} \sum_{i=1}^N Z_i^2$ for a standardized random variable Z with standardized kurtosis $\mathbb{E}[Z^4] = \delta^4$. Assume that N is “large.” (Hint: What is the asymptotic distribution of $\hat{\theta}_Z$?)

- Solution: The standard error of the estimator $\hat{\theta}_Z$ is just its standard deviation, but we just derive the variance first:

$$\begin{aligned}
\text{Var}(\hat{\theta}_Z) &= \text{Var}\left(\frac{1}{N} \sum_{i=1}^N Z_i^2\right) \\
&= \frac{1}{N^2} \sum_{i=1}^N \text{Var}(Z_i^2) \\
&= \frac{1}{N^2} \sum_{i=1}^N \mathbb{E}(Z_i^4) - \mathbb{E}(Z_i^2)^2 \\
&= \frac{1}{N^2} \sum_{i=1}^N \delta^4 - 1^2 \\
\text{Var}(\hat{\theta}_Z) &= \frac{\delta^4 - 1}{N}
\end{aligned}$$

This gives us a standard error of $\sqrt{\text{Var}(\hat{\theta}_Z)} = \sqrt{\frac{\delta^4 - 1}{N}}$. And, since $\hat{\theta}_Z$ is just an average, $\hat{\theta}_Z$ is Normally distributed.

Now, perform a change of variables to obtain the standard error of the estimator $\hat{\theta}_X = \frac{1}{N} \sum_{i=1}^N (X_i - \mu)^2$ for a random variable X with $\mathbb{E}[X] = \mu$, $\text{Var}(X) = \sigma^2$, and standardized kurtosis $\frac{\mathbb{E}[(X-\mu)^4]}{\sigma^4} = \delta^4$.

- Solution: The standard error of $\hat{\theta}_X$ is very straightforward to obtain. Since X is just a location-scale transformation of Z , we just need to use the inverse transformation $\sigma Z + \mu = X$ which gives us the transformation $\sigma^2 \hat{\theta}_Z = \hat{\theta}_X$:

$$\begin{aligned}
\hat{\theta}_Z &\sim N\left(1, \frac{\delta^4 - 1}{N}\right) \\
\hat{\theta}_X &\sim N\left(\sigma^2, \sigma^4 \frac{\delta^4 - 1}{N}\right)
\end{aligned}$$

We want to verify this more rigorously:

$$\begin{aligned}
\text{Var}(\hat{\theta}_X) &= \text{Var}\left(\frac{1}{N} \sum_{i=1}^N (X_i - \mu)^2\right) \\
&= \text{Var}\left(\frac{\sigma^2}{N} \sum_{i=1}^N \left(\frac{X_i - \mu}{\sigma}\right)^2\right) \\
&= \sigma^4 \text{Var}(\hat{\theta}_Z) \\
\text{Var}(\hat{\theta}_X) &= \sigma^4 \frac{\delta^4 - 1}{N}
\end{aligned}$$

Thus, the standard error is $\sqrt{\text{Var}(\hat{\theta}_X)} = \sigma^2 \sqrt{\frac{\delta^4 - 1}{N}}$.

A more general version of obtaining the distribution of transformations of random variables that are normally distributed is called the Delta Method: Wikipedia: Delta Method. However, for this simple, univariate transformation, you should be able to just use the methods you've already learned about transformations of random variables.

Question Five: Maximum Likelihood Estimators

Maximum likelihood estimators are very commonly used in economics.

1. Give the likelihood function of a sample of N i.i.d. Poisson random variables, X_1, \dots, X_N .

- Solution to 1: We have the joint density $f(x_1, \dots, x_N) = \prod_{i=1}^N f(x_i)$ where $f(x_i) = \frac{e^{-\lambda} \lambda^{x_i}}{x_i!}$. The likelihood function is precisely this joint density:

$$L(\lambda|x_1, \dots, x_N) = \prod_{i=1}^N \frac{e^{-\lambda} \lambda^{x_i}}{x_i!}.$$

2. Give the log-likelihood function and simplify.

- Solution to 2: The log-likelihood function is simply the log of the density:

$$\begin{aligned} \mathcal{L}(\lambda|x_1, \dots, x_N) &= \log \left(\prod_{i=1}^N \frac{e^{-\lambda} \lambda^{x_i}}{x_i!} \right) \\ &= \sum_{i=1}^N \log \left(\frac{e^{-\lambda} \lambda^{x_i}}{x_i!} \right) \\ &= \sum_{i=1}^N -\lambda + x_i \log(\lambda) - \log(x_i!) \\ &= -N\lambda + \log(\lambda) \sum_{i=1}^N x_i - \sum_{i=1}^N \log(x_i!) \end{aligned}$$

3. Take the first order conditions and solve for the maximum likelihood estimator of λ .

- Solution to 3: We differentiate the log-likelihood function:

$$\begin{aligned} \frac{\partial \mathcal{L}(\lambda|x)}{\partial \lambda} &= -N\lambda + \log(\lambda) \sum_{i=1}^N x_i - \sum_{i=1}^N \log(x_i!) \\ 0 &= -N + \frac{1}{\lambda} \sum_{i=1}^N x_i - 0 \\ \hat{\lambda} &= \frac{1}{N} \sum_{i=1}^N x_i \end{aligned}$$

4. How does the MLE from (3) compare to the Method of Moments (MoM) estimator from lecture?

- Solution to 4: The Method of Moments estimator from lecture was exactly the same estimator. Thus, the MLE for λ is just the sample average. An alternative estimator using the second noncentral moment of the X_i 's could be $\tilde{\lambda} = \sqrt{\frac{1}{N} \sum_{i=1}^N x_i^2 - \frac{1}{N} \sum_{i=1}^N x_i}$. Do you think this estimator is unbiased? Consistent? However, since it isn't the MLE, it probably is not the most efficient estimator (i.e. $Var(\tilde{\lambda}) > Var(\hat{\lambda})$).