Econ-240A (1st Half - Fall 2009) Section 5

Owen Ozier

Friday, October 2nd

Contents

1 Example: Method of Moments, Maximum Likelihood, and Bias

When we left off last time, we had just calculated the method of moments estimator for a distribution. Re-deriving that quickly here:

1.1 Method of Moments

Consider a random sample, $X_1, ..., X_n$, from the distribution $f_X(x;\theta) = \theta x^{\theta-1} \cdot \mathbb{I}\{0 < x < 1\}$; $\theta > 0$.

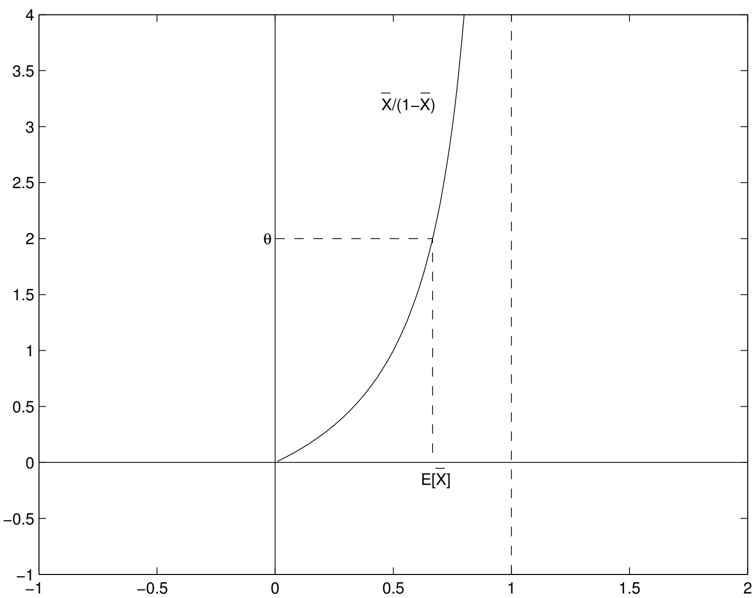
To find the method of moments estimator, $\hat{\theta}_{MM}$, we follow the procedure discussed in CB 7.2. First, we find the population moment(s) in terms of the parameter(s) of the distribution; then we substitute in the sample moment(s) and solve the system of equations that results. With one parameter, we just need the first moment:

$$\mathbb{E}[X] = \int_0^1 x \cdot \theta x^{\theta-1} dx = \frac{\theta}{\theta+1} x^{\theta+1}\Big|_0^1 = \frac{\theta}{\theta+1}$$

Substituting the sample moment, and solving:

$$\bar{X} = \frac{\hat{\theta}_{MM}}{\hat{\theta}_{MM} + 1}$$
$$\bar{X}(\hat{\theta}_{MM} + 1) = \hat{\theta}_{MM}$$
$$\bar{X} = \hat{\theta}_{MM}(1 - \bar{X})$$
$$\frac{\bar{X}}{1 - \bar{X}} = \hat{\theta}_{MM}$$

The next question is whether this is a *biased* estimator (see CB Definition 7.3.2).



In the figure above, we see that when the sample mean takes on its expectation exactly, the method of moments estimator exactly recovers θ. However, because $\bar{X}$ is random, and thus has some variation around its expected value, the upward curvature of the $\hat{\theta}_{MM}$ function (it is convex) suggests an application of Jensen's Inequality (CB Theorem 4.7.7):

$$\mathbb{E}\left[\frac{\bar{X}}{1 - \bar{X}}\right] \geq \frac{\mathbb{E}\left[\bar{X}\right]}{1 - \mathbb{E}\left[\bar{X}\right]}$$

and thus:

$$\mathbb{E}\left[\hat{\theta}_{MM}\right] \geq \theta$$

So the method of moments estimator is biased upwards, and when the sample is finite and thus the variance of $\bar{X}$ is nonzero, the inequality above is strict.

1.2 Maximum Likelihood

To find a maximum likelihood estimator for θ, To find a maximum likelihood estimator for θ, $\hat{\theta}_{MLE}$, we follow the procedure discussed after CB Definition 7.2.4. First, find the likelihood (or log-likelihood) function, use the FOC (first order condition) do optimize for θ, and confirm that we have maximized via the SOC (second order condition).

Here, the log likelihood function is:

$$l(\theta|x_1, ..., x_n) = n \log \theta + (\theta - 1) \sum_{i=1}^{n} \log x_i$$

Then solving from the FOC:

$$\frac{\partial l}{\partial \theta}(\theta|x_1, ..., x_n) = \frac{n}{\theta} + \sum_{i=1}^{n} \log x_i = 0$$

$$\frac{-n}{\sum_{i=1}^{n} \log x_i} = \hat{\theta}_{MLE}$$

Confirming with the SOC:

$$\frac{\partial^2 l}{\partial \theta^2}(\theta|x_1, ..., x_n) = -\frac{n}{\theta^2} < 0 \; \forall \theta > 0$$

So $\hat{\theta}_{MLE}$ maximizes the likelihood function.

In determining whether $\hat{\theta}_{MLE}$ is biased (and it usually is), the present functional form is not easy to evaluate. Consider a transformation that could simplify our situation: Let $Y_i = -\log X_i$. Then $y = g(x)$, and $g^{-1}(y) = x = e^{-y}$, and $\frac{d}{dy} g^{-1}(y) = -e^{-y}$, so we can use CB Theorem 2.1.5 to find the distribution of the new random variables Y_i:

$$f_Y(y) = f_x(g^{-1}(y)) \left| \frac{d}{dy} g^{-1}(y) \right| = \theta(e^{-y})^{\theta-1} \cdot e^{-y} = \theta e^{-\theta y}$$

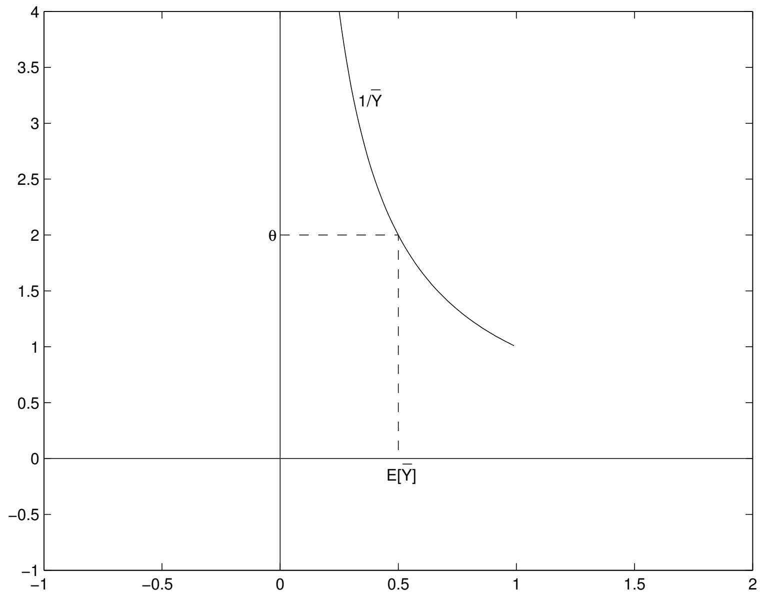
The support of Y is now $(0, \infty)$, and we recognize Y as having exponential distribution with expectation $1/\theta$. In these terms:

$$\frac{\partial l}{\partial \theta}(\theta|x_1, ..., x_n) = \frac{n}{\theta} + \sum_{i=1}^{n} \log x_i = 0$$

$$\hat{\theta}_{MLE} = \frac{n}{\sum_{i=1}^{n} y_i}$$

$$= \frac{1}{\overline{Y}}$$

Now, it is clear that if $\overline{Y} = \mathbb{E}[\overline{Y}]$, $\hat{\theta}_{MLE}$ will take on the true value of θ. But as in the case before, we graph the function, shown below, and see that Jensen's inequality will show the maximum likelihood estimator to be biased upwards:



2 Sufficiency and Completeness

As given in CB Definitions 6.2.1 and 6.2.21, a *sufficient* statistic is one for which the conditional distribution of the sample given a value of the statistic no longer depends on the parameter(s) of the original distribution; a *complete*[1] statistic is one for which:

$$\mathbb{E}[g(T)] = 0 \ \forall\theta \Rightarrow \text{Prob}\{g(T) = 0\} = 1 \ \forall\theta$$

2.1 Examples

While we have discussed the intuition for sufficiency–that we haven't thrown away anything from the sample that could have told us about θ–we have not really given as much intuition for completeness; it has to do with reducing the data so much that only (essentially) one unbiased estimator is possible from a complete statistic. Consider a random sample of 3 trials of a Bernoulli random variable (X_1, X_2, X_3)' with unknown probability of success $p \in (0, 1)$.

An example of a **sufficient** statistic for p is the entire sample, $T = (X_1, X_2, X_3)'$, a vector of three values. We haven't thrown anything away. This is not a **complete** statistic, however, because we can make more than one unbiased estimator of p from it: let $W_1 = X_1$, and let $W_2 = X_2$. While it is true that

[1] For those curious about the origin of the terminology, I can trace the use of the term "complete" in this context to the discussion of a "complete kernel" in Lehmann, E. L., and Scheffé, H. (1947): "On the Problem of Similar Regions," *Proceedings of the National Academy of Sciences*, 33 (12), 382-386. http://www.jstor.org/stable/87534. This is re-framed in terms of complete families of measures in Lehmann, E. L., and Scheffé, H. (1950): "Completeness, Similar Regions, and Unbiased Estimation: Part I," *Sankhyā: The Indian Journal of Statistics*, 10 (4), 305-340. http://www.jstor.org/stable/25048038, and is then discussed in terms of complete statistics in Basu, D. (1955): "On Statistics Independent of a Complete Sufficient Statistic," *Sankhyā: The Indian Journal of Statistics*, 15 (4), 377-380. http://www.jstor.org/stable/25048259.

$\mathbb{E}[W_1(T) - W_2(T)] = 0 \ \forall p \in (0,1)$, since $\mathbb{E}[X_1] = \mathbb{E}[X_2] = p$, it is not true that $\text{Prob}\{W_1(T) - W_2(T) = 0\} = 1 \ \forall p \in (0,1)$. In fact, $\text{Prob}\{X_1 = X_2\} = p^2 + (1-p)^2$.

An example of a **complete** statistic is the first observation, $T = X_1$. It is complete because T can only take on two values (since the X_i are Bernoulli): 0 and 1. Thus any function of T can also only take on at most two values. Let $g(T)$ be defined so that $g(0) = a$ and $g(1) = b$. In order to get $\mathbb{E}[g(T)] = 0 \ \forall p \in (0,1)$, we need $pb + (1-p)a = 0 \ \forall p \in (0,1)$. Thus, $a - b = 0$ and $a = 0$, so $b = 0$ as well. This implies $\text{Prob}\{g(T) = 0\} = 1 \ \forall p \in (0,1)$. This is not a **sufficient** statistic, however, because the conditional distribution of the random vector $(X_1, X_2, X_3)'$ given a value for $T = X_1$ still depends on the parameter, p. In particular, the random vector is $(T, 0, 0)'$ with probability $(1-p)^2$; $(T, 0, 1)'$ or $(T, 1, 0)'$ with probability $p(1-p)$ each; and is $(T, 0, 0)'$ with probability $(1-p)^2$.

An example (which we saw in lecture) of a statistic that is **both sufficient and complete** is the sum, $T = X_1 + X_2 + X_3$. Conditional on the sum being zero or three, there is only one possible arrangement; conditional on the sum being one, the three arrangements (0,0,1)', (0,1,0)', and (1,0,0)' are equally likely. Likewise when the sum is two. Since these probabilities no longer depend on p, T is sufficient. Since T only takes on four different values, we can characterize functions of T by $g(0) = a, g(1) = b, g(2) = c, g(3) = d$, and can show that if $\mathbb{E}[g(T)] = 0 \ \forall p \in (0,1)$, then $a = b = c = d = 0$, and thus $\text{Prob}\{g(T) = 0\} = 1 \ \forall p \in (0,1)$, and so T is also complete. More simply, we might also appeal to CB Theorem 6.2.6/6.2.10 and 6.2.25 for sufficiency and completeness here.

An example of a statistic that is **neither sufficient nor complete** could be the first two values: $T = (X_1, X_2)'$. It is not complete for the same reason that $T = (X_1, X_2, X_3)'$ is not complete, given above. It is not sufficient, though, because conditional on observing a value of T, the distribution of the random sample is still dependent on p: X_3 is still 1 with probability p and 0 with probability $(1-p)$.

2.2 Discussion

Generally (for the purposes of this class), sufficiency is easiest to show via either the factorization theorem (CB 6.2.6) or the exponential family result (CB 6.2.10). Arguing that a statistic is not sufficient can be done either directly, by finding the conditional distribution, or by making a convincing argument that no factorization along the lines of CB 6.2.6 is possible. Discrete random variables often yield conditional distributions more easily than do continuous ones, though.

Likewise, completeness is most easily shown by either appealing to the exponential family result (CB 6.2.25), or by showing a one-to-one mapping to any other case for which we know a complete statistic (the uniform distribution). Completeness can be argued directly via the definition, which is again typically easier for discrete random variables, but this is often much harder than appealing to an earlier completeness result.

3 Example: 2002 Midterm 2a-2e, Sufficiency, Completeness, UMVU

The midterm problem actually had nine parts, going on to testing and confidence intervals. Here, I present only the first five parts. I am essentially duplicating the original problem and solution as written by Prof.

Michael Jansson, adding explanation in a few places.

Let $X_1, ..., X_n$ be a random sample from a continuous distribution with pdf

$$f(x|\theta) = \frac{1}{2\sqrt{\theta}} x^{-1/2} \mathbf{1}(0 \leq x \leq \theta),$$

where $\theta \in \Theta = (0, \infty)$ is an unknown parameter and $\mathbf{1}(\cdot)$ is the indicator function.

3.1 (a) Method of Moments

Question: *Find a method of moments estimator $\hat{\theta}_{MM}$ of θ.*

As usual, we find the expectation in terms of θ, and solve for $\hat{\theta}$ substituting in the sample moment for the population moment.

$$\begin{aligned}
\mathbb{E}[X_i] &= \int_{-\infty}^{\infty} x f(x|\theta) dx = \int_0^\theta \frac{1}{2\sqrt{\theta}} x^{1/2} dx \\
&= \frac{2}{3} \frac{1}{2\sqrt{\theta}} x^{3/2} \Big|_{x=0}^{x=\theta} \\
&= \frac{1}{3\sqrt{\theta}} \theta^{3/2} \\
&= \frac{\theta}{3}
\end{aligned}$$

Substituting:

$$\frac{\hat{\theta}_{MM}}{3} = \bar{X} \qquad \text{and thus:} \qquad \hat{\theta}_{MM} = 3\bar{X}$$

(Where $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$.)

3.2 (b) Sufficiency

Question: *Find the likelihood function and show that $X_{(n)} = \max_{1 \leq i \leq n} X_i$ is a sufficient statistic for θ.*

Since this family of distributions has parameter-dependent support, it cannot be an exponential family (see discussion after CB Definition 3.4.5), so we are likely to appeal to the factorization criterion (CB Theorem 6.2.6) in arguing that a statistic is sufficient.

First, though, the likelihood function:

$$\begin{aligned}
L(\theta|X_1, ..., X_n) &= \prod_{i=1}^n f(X_i|\theta) = \prod_{i=1}^n \left[\frac{1}{2\sqrt{\theta}} X_i^{-1/2} \mathbf{1}(0 \leq X_i \leq \theta) \right] \\
&= \frac{1}{2^n \theta^{n/2}} \left(\prod_{i=1}^n X_i \right)^{-1/2} \cdot \mathbf{1}(X_{(1)} \geq 0) \cdot \mathbf{1}(X_{(n)} \leq \theta)
\end{aligned}$$

Where $X_{(1)} = \min_{1 \leq i \leq n} X_i$ and $X_{(n)} = \max_{1 \leq i \leq n} X_i$. Re-writing:

$$L(\theta|X_1, ..., X_n) = \left[\theta^{-n/2} \cdot \mathbf{1}(X_{(n)} \leq \theta) \right] \cdot \left[2^{-n} \left(\prod_{i=1}^n X_i \right)^{-1/2} \cdot \mathbf{1}(X_{(1)} \geq 0) \right]$$

Based on the factorization above, we can see that $X_{(n)}$ is the only function of the sample needed for the part of the likelihood function that depends on θ, so $X_{(n)}$ is sufficient for θ.

3.3 (c) Maximum Likelihood Estimator

Question: *Show that $\hat{\theta}_{ML} = X_{(n)}$ is the maximum likelihood estimator of θ.*

The solution, in Prof. Jansson's words: Now, $L(\theta|X_1, .., X_n) = 0$ for $\theta < X_{(n)}$, while

$$L(\theta|X_1, ..., X_n) = \theta^{-n/2} \cdot \left[2^{-n} \left(\prod_{i=1}^{n} X_i \right)^{-1/2} \cdot \mathbf{1}(X_{(1)} \geq 0) \right]$$

is a decreasing function of θ for $\theta \geq X_{(n)}$. As a consequence, the maximum likelihood estimator of θ is $\hat{\theta}_{ML} = X_{(n)}$. (Note that if the original support had been $0 \leq x < \theta$ with strict rather than weak inequality at θ, the likelihood function would have used $\mathbf{1}(X_{(n)} < \theta)$ rather than $\mathbf{1}(X_{(n)} \leq \theta)$, and no MLE would exist; $\hat{\theta} = X_{(n)}$ would have likelihood zero.)

3.4 (d) Completeness

Question: *Show that the sufficient statistic $X_{(n)}$ is complete. (Hint: Use the fact that $\sqrt{X_i} \sim$ i.i.d. $U[0, \sqrt{\theta}]$.)*

Since this is not an exponential family, we will have to show the result either directly or via one-to-one transformation to a statistic whose completeness we already have. Since we know that the maximum value in a sample is complete for the uniform distribution, the hint seems like it will allow us to show completeness. First, note that $Y_i = \sqrt{X_i} \sim U[0, \sqrt{\theta}]$. Since $\sqrt{\cdot}$ is a monotonically increasing transformation, $Y_{(n)} = \sqrt{X_{(n)}}$, meaning that order is preserved. Also, since $Y_i = \sqrt{X_i}$, $X_i = Y_i^2$ is the transformation that recovers an X_i from a Y_i.

That said, suppose we have any function $g(\cdot)$ such that $\mathbb{E}[g(X_{(n)})] = 0$. Now, let us define a function $g'(y) = g(y^2)$ in terms of the original function, $g(\cdot)$, and a transformation that will turn y's back into x's. This means that $\mathbb{E}[g'(Y_{(n)})] = \mathbb{E}[g(X_{(n)})] = 0$ by assumption. But because $Y_{(n)}$ is complete–as discussed in both lecture and the textbook–we have that:

$$\mathbb{E}[g(X_{(n)})] = 0 \iff \mathbb{E}[g'(Y_{(n)})] = 0 \Rightarrow \text{Prob}\{g'(Y_{(n)}) = 0\} = 1 \, \forall \theta > 0 \iff \text{Prob}\{g(X_{(n)}) = 0\} = 1 \, \forall \theta > 0$$

So, in summary, we have shown that

$$\mathbb{E}[g(X_{(n)})] = 0 \;\Rightarrow\; \text{Prob}\{g(X_{(n)}) = 0\} = 1 \, \forall \theta > 0$$

and thus $X_{(n)}$ is complete.

3.5 (e) UMVU Estimator

Question: *Find a uniform minimum variance unbiased estimator of θ. (Hint: Use part (d) and the fact that Prob$\{X_{(n)} \leq t\} = (t/\theta)^{n/2}$ for $0 \leq t \leq \theta$.)*

First of all, "recall" CB Theorem 7.5.1, the Lehmann-Scheffé theorem: an unbiased estimator based on a complete sufficient statistic is UMVU.

We have shown that $X_{(n)}$ is complete and sufficient, so it remains to find an unbiased estimator of θ based on $X_{(n)}$. Having been given the CDF for $X_{(n)}$, we can find its PDF and then its expectation, as a starting point:

$$f_{X_{(n)}}(t|\theta) = \frac{d}{dt}(t/\theta)^{n/2} = (n/2)\theta^{-n/2} t^{(n-2)/2} \cdot \mathbf{1}(0 \leq t \leq \theta)$$

Thus:

$$\begin{aligned}
\mathbb{E}[X_{(n)}] &= (n/2)\theta^{-n/2}\int_0^\theta t\cdot t^{(n-2)/2}dt = (n/2)\theta^{-n/2}\int_0^\theta t^{n/2}dt \\
&= \frac{(n/2)}{((n+2)/2)}\theta^{-n/2}t^{(n+2)/2}\Big|_0^\theta \\
&= \frac{n}{n+2}\theta^{-n/2}\theta^{(n+2)/2} \\
&= \frac{n}{n+2}\theta
\end{aligned}$$

Therefore,

$$\hat{\theta}_{UMVU} = \frac{n+2}{n}X_{(n)}$$

is an unbiased estimator of θ, and because $X_{(n)}$ is complete and sufficient, it is also a UMVUE of θ.