LECTURE 7: POINT ESTIMATION

MECO 7312.
INSTRUCTOR: DR. KHAI CHIONG
OCTOBER 21, 2020

A (point) estimator is any function $W(X_1, \ldots, X_n)$ of a random sample. An estimator is both a statistics and a random variable. An *estimate* is a realized value of this random variable.

In some cases, there are natural candidates to estimate a population parameter (such as estimating the population mean with the sample mean), but in other cases, it is more difficult. We will study various ways of coming up with sensible estimators, and evaluate these estimators.

1. Method of Moments

In the Method of Moments (MOM) approach, estimators are found by solving a system of simultaneous equations. These equations arise from equating the first k sample moments to the corresponding k population moments.

The k-th moment is:

$$m_k = \frac{1}{n}\sum_{i=1}^{n} X_i^k$$

The k-th population moment is:

$$\mathbb{E}[X^k] \equiv \mu_k(\theta_1, \ldots, \theta_l) \tag{1}$$

Implicit in Equation 1 above, the population moment depends on population parameters $\theta_1, \ldots, \theta_l$. For example, θ could be the population variance of the Normal distribution, or correlation of the multivariate Normal, or the rate parameter of the exponential distribution.

The Method of Moments estimator is justified through the WLLN and the SLLN:

$$\frac{1}{n}\sum_{i=1}^{n} X_i^k \to \mathbb{E}[X^k], \quad \text{almost surely and in probability as } n \to \infty$$

1.1. Example: parameters of the Normal distribution

Suppose $X_1, \ldots, X_n$ are iid $\mathcal{N}(\mu, \sigma^2)$. We would like to come up with estimators for μ and σ^2.

Equating the first sample moment with the first population moment:

$$\frac{1}{n}\sum_{i=1}^{n} X_i = \mathbb{E}[X] = \mu \tag{2}$$

Equating the second sample moment with the second population moment:

$$\frac{1}{n}\sum_{i=1}^{n} X_i^2 = \mathbb{E}[X^2]$$

$$\frac{1}{n}\sum_{i=1}^{n} X_i^2 = \mu^2 + \sigma^2 \tag{3}$$

Solving the system of equations 2 and 3, we obtain $\frac{1}{n}\sum_{i=1}^{n} X_i$ as an estimator for μ and $\frac{1}{n}\sum_{i=1}^{n} X_i^2 - (\frac{1}{n}\sum_{i=1}^{n} X_i)^2$ as an estimator for σ^2. The application of Method of Moments here results in some familiar estimators, but it does not recover the sample variance.

1.2. Example: parameters of the Uniform distribution

$X_1, \ldots, X_n \sim_{i.i.d} U[0, \theta]$, where θ is the parameter.

Equating the first sample moment and the first population moment:

$\frac{1}{n}\sum_{i=1}^{n} X_i = \mathbb{E}[X] = \frac{\theta}{2}$. Therefore, $\hat{\theta}^{MOM} = \frac{2}{n}\sum_{i=1}^{n} X_i = 2\bar{X}$.

Is this a reasonable estimator? Suppose the realized data is $x_1 = 1, x_2 = 2$, then $\hat{\theta}^{MOM} = 3$. What if we have $x_1 = 0.1, x_2 = 0.1, x_3 = 1$, then $\hat{\theta}^{MOM} = 0.8$.

What is the sampling distribution of $\hat{\theta}^{MOM}$? Since $\bar{X}$ is asymptotically Normal:

$$\sqrt{n}\left(\bar{X} - \frac{\theta}{2}\right) \to_d \mathcal{N}\left(0, \frac{\theta^2}{12}\right)$$

By either the Delta method or the Continuous Mapping Theorem,

$$\sqrt{n}\left(2\bar{X}-\theta\right) \to_d \mathcal{N}\left(0, \frac{\theta^2}{3}\right)$$

That is, $\hat{\theta}^{MOM} \approx \mathcal{N}(\theta, \frac{\theta^2}{3n})$. Consistency is implied here.

In general, the Method of Moments estimator is a function of sample moments, and sample moments are asymptotically Normal by CLT.

$$\sqrt{n}\left(\frac{1}{n}\sum_{i=1}^{n} X_i^k - \mathbb{E}[X^k]\right) \to_d \mathcal{N}(0, \mathrm{Var}(X^k))$$

We can therefore use Delta Method to derive the asymptotic distribution of the Method of Moments estimator. Other alternative methods are bootstrapping and simulation.

1.3. Example: parameters of the Binomial distribution

Let $X_1, X_2, \ldots, X_n$ be iid from binomial(k,p), that is $P(X_i = x|k,p) = \binom{k}{x}p^x(1-p)^{k-x}$ for $x = 0, 1, \ldots, k$.

Both k and p are unknown parameters to be estimated.

$$\frac{1}{n}\sum_{i}^{n} X_i = kp$$
$$\frac{1}{n}\sum_{i}^{n} X_i^2 = kp(1-p) + k^2p^2$$

Solving the equations above in terms of k and p:

$$\hat{p} = 1 - \frac{\frac{1}{n}\sum_i^n X_i^2 - \bar{X}^2}{\bar{X}}$$
$$\hat{k} = \frac{\bar{X}^2}{\bar{X} - \frac{1}{n}\sum_i^n X_i^2 + \bar{X}^2} = \frac{\bar{X}^2}{\bar{X} - \frac{1}{n}\sum_i^n (X_i - \bar{X})^2}$$

If the data are $(1, 1, 0, 0)$, then $\hat{p} = 0.5$ and $\hat{k} = 1$. If the data are $(3, 4, 5)$, then $\hat{p} = 0.83$ and $\hat{k} = 4.8$.

1.4. Example: mixture distribution

Apart from the special cases above, computing the population moments often involve difficult integrations that necessitate the use of computers. In Pearsons original paper, the density was a mixture of two normal density functions:

$$f(x|\theta) = \lambda \frac{1}{\sqrt{2\pi}} e^{-(x-\mu_1)^2/2} + (1-\lambda)\frac{1}{\sqrt{2\pi}} e^{-(x-\mu_2)^2/2}$$

where $\theta = (\lambda, \mu_1, \mu_2)$ are parameters to be estimated. Mixtures of normals are used to fit data that has multiple modes.

$$\begin{aligned} \frac{1}{n}\sum X_i &= \int x f(x|\theta)\, dx = \lambda\mu_1 + (1-\lambda)\mu_2 \\ \frac{1}{n}\sum X_i^2 &= \int x^2 f(x|\theta)\, dx = \lambda\mu_1^2 + (1-\lambda)\mu_2^2 + 1 \\ \frac{1}{n}\sum X_i^3 &= \int x^3 f(x|\theta)\, dx = \lambda\mu_1(\mu_1^2+3) + (1-\lambda)\mu_2\left(\mu_2^2+3\right) \end{aligned}$$

We can symbolically evaluate the integrals using Mathematica.

The above system of equations can be solved numerically. Suppose our data is $(-2,-1,1,2)$, then $\hat{m}_1 = 0$, $\hat{m}_2 = 2.5$, $\hat{m}_3 = 0$, where $\hat{m}_k = \frac{1}{n}\sum x_i^k$ denotes the realized k-sample moment. It follows that $\hat{\lambda} = 0.5$, $\hat{\mu}_1 = -1.22$, $\hat{\mu}_2 = 1.22$.

What if the data are $(-2,-1,3,4,5)$? Check that $\hat{m}_1 = 1.8$, $\hat{m}_2 = 11$, $\hat{m}_3 = 41.4$. Use Mathematica to see that $\hat{\lambda} = 0.411$, $\hat{\mu}_1 = -1.31$, $\hat{\mu}_2 = 3.97$.

1.5. Generalized Method of Moments (GMM)

In general, we can formulate a system of equations to solve for unknown θ by coming up with functions $g_1, \ldots, g_K$. Each equation below is called a moment condition.

$$\begin{aligned} \mathbb{E}_\theta[g_1(X)] &= \frac{1}{n}\sum_{i=1} g_1(X_i) \\ &\vdots \\ \mathbb{E}_\theta[g_K(X)] &= \frac{1}{n}\sum_{i=1} g_K(X_i) \end{aligned}$$

Implicit here is the assumption that $X_1, \ldots, X_n$ are i.i.d from a density $f(x;\theta)$, where θ is an unknown parameter that enters into the pdf. Therefore the population expectation on the left-hand side is taken with respect to $f(x;\theta)$.

The left-hand side of the above system of equations may be non-linear functions of the parameters. That is, $\mathbb{E}_\theta[g_k(X)]$ may be non-linear functions of θ. We numerically find θ that "best" satisfies the system of equations above. That is, the GMM estimator finds θ that minimizes:

$$\sum_{k=1}^{K} \left(\mathbb{E}[g_k(X,\theta)] - \frac{1}{n}\sum_{i=1}^{n} g_k(X_i) \right)^2$$

By letting $g(X) = e^X$ or $g(X) = \frac{1}{X}$, we can obtain moment conditions like $\mathbb{E}[e^X]$ or $\mathbb{E}[\frac{1}{X}]$.

Example:

Again let $X_1, \ldots, X_n \sim_{i.i.d} U[0,\theta]$, where θ is the parameter.

Consider the first 3 sample moment and the corresponding population moments:

$$\begin{aligned}
\frac{1}{n}\sum_{i=1}^{n} X_i &= \frac{\theta}{2} \\
\frac{1}{n}\sum_{i=1}^{n} X_i^2 &= \frac{\theta^2}{3} \qquad \text{, since } \mathbb{E}[X^2] = \frac{\theta^2}{3} \\
\frac{1}{n}\sum_{i=1}^{n} X_i^3 &= \frac{\theta^3}{4}
\end{aligned}$$

The Generalized Method of Moments estimator would solve the following minimization problem:

$$\hat{\theta} = \underset{\theta}{\operatorname{argmin}} \left(\frac{1}{n}\sum_{i=1}^{n} X_i - \frac{\theta}{2} \right)^2 + \left(\frac{1}{n}\sum_{i=1}^{n} X_i^2 - \frac{\theta^2}{3} \right)^2 + \left(\frac{1}{n}\sum_{i=1}^{n} X_i^3 - \frac{\theta^3}{4} \right)^2 \tag{4}$$

Suppose again that realized data is $x_1 = 0.1, x_2 = 0.1, x_3 = 1$. If we use only the first moment, then $\hat{\theta}^{MOM} = 0.8$. However, if we use all three moments, we can show numerically that $\hat{\theta} = 1.01502$ in the following minimization.

$$\hat{\theta} = \underset{\theta}{\text{argmin}} \left(1.2/3 - \frac{\theta}{2}\right)^2 + \left(1.02/3 - \frac{\theta^2}{3}\right)^2 + \left(1.002/3 - \frac{\theta^3}{4}\right)^2 \tag{5}$$

When there is only one unknown, but two equations, we say that the parameter is over-identified. More generally, when the number of moment conditions is greater than the number of parameters to be estimated, we say that the parameters are over-identified.

Over-identification allows us to check whether our assumed model of data-generating process is valid. This is known as the Sargan-Hansen J-test. The intuition is that if the objective function in 5 is almost minimized at zero, then we do not reject the hypothesis that the data-generating process is $U[0, \theta]$.

Hansen, Lars Peter (1982)[1] shows that the GMM estimator has good large sample properties, it is strongly consistent and asymptotically normal under some assumptions. That is,

$$\sqrt{n}(\hat{\theta}^{GMM} - \theta) \to_d \mathcal{N}(0, V)$$

Where V is the asymptotic variance $\hat{\theta}^{GMM}$. Hansen (1982) then shows how to calculate V as a function of the data (that is, he provides a consistent estimator for V). Later on, the asymptotic distribution of $\hat{\theta}^{GMM}$ can be used for *inference*: hypothesis testing and confidence interval regarding $\hat{\theta}^{GMM}$.

1.6. Simple linear regression

Consider a random variable Y is generated as: $Y = a + bX + \epsilon$, where a and b are some unknown parameters, and $\epsilon \sim \mathcal{N}(0, \sigma^2)$. Moreover, X is a random variable such that $\mathbb{E}[X\epsilon] = 0$. Now suppose we have a random sample $Y_1, \ldots, Y_n$ of Y, and a random sample $X_1, \ldots, X_n$ of X. We wish to estimate a and b via the moment conditions below:

$$0 = \mathbb{E}[\epsilon] = \mathbb{E}[Y - a - bX] = \frac{1}{n}\sum_{i=1}^{n} Y_i - a - b\frac{1}{n}\sum_{i=1}^{n} X_i \tag{6}$$

$$0 = \mathbb{E}[X\epsilon] = \mathbb{E}[X(Y - a - bX)] = \frac{1}{n}\sum_{i=1}^{n} X_i Y_i - a\frac{1}{n}\sum_{i=1}^{n} X_i - b\frac{1}{n}\sum_{i=1}^{n} X_i^2 \tag{7}$$

Multiplying Equation 6 by $\bar{X}$ and then subtracting it from Equation 7:

[1] "Large Sample Properties of Generalized Method of Moments Estimators". Econometrica (1982).

$$0 = \frac{1}{n}\sum_{i=1}^{n} X_i Y_i - \bar{X}\bar{Y} - b\frac{1}{n}\sum_{i=1}^{n} X_i^2 + b\bar{X}^2$$

$$b = \frac{\frac{1}{n}\sum_{i=1}^{n} X_i Y_i - \bar{X}\bar{Y}}{\frac{1}{n}\sum_{i=1}^{n} X_i^2 - \bar{X}^2}$$

Which is the sample covariance divided by the sample variance. Moreover,

$$a = \bar{Y} - b\bar{X}$$

Hence, the method of moments estimators give rise to the usual formulas for calculating regression coefficients! What about σ^2? We can just add an additional moment condition $\mathbb{E}[\epsilon^2] = \sigma^2$.

Note: we can even relax the assumption that ϵ is Normally distributed!