Slides for the course

Statistics and econometrics

Part 3: Properties of estimators

European University Institute

Andrea Ichino

September 18, 2014

Outline

Finite sample properties

- Unbiasedness
- Efficiency
- Sufficiency

Asymptotic properties

- Asymptotic unbiasedness
- Consistency
- Asymptotic Distribution of Estimators
- Asymptotic efficiency

Invariance of ML estimators

Why are properties of estimators interesting?

We choose between estimators comparing their properties.

It is useful to distinguish between:

- Finite sample properties that hold for a given sample size n.
  - Unbiasedness
  - Efficiency
  - Sufficiency
- Asymptotic properties that hold when sample size goes to ∞:
  - Consistency
  - Asymptotic unbiasedness
  - Asymptotic efficiency
  - Asymptotic normality
- Other properties (e.g. Invariance)

Section 1

Finite sample properties

Subsection 1

Unbiasedness

Definition of Unbiasedness

An estimator $\hat{\theta}$ is unbiased for the parameter θ if

$$E(\hat{\theta}) = \theta \tag{1}$$

Example: for any distribution

$$X \sim f_X(X) \qquad \text{such that} \qquad E(X) = \mu \tag{2}$$

the sample mean is unbiased for the population mean

$$E(\hat{\mu}) = E\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n}\sum_{i=1}^{n} E(X_i) = \frac{1}{n} n\mu = \mu \tag{3}$$

Example of unbiased MM and biased ML estimators

$$f_X(x|\theta) = \frac{2X}{\theta^2} \qquad \text{for } 0 \leq X \leq \theta \tag{4}$$

The MM estimator for θ is unbiased:

$$E(X|\theta) = \int_0^\theta X \frac{2X}{\theta^2} dX = \frac{2}{3}\theta = \frac{1}{n}\sum_{i=1}^{n} X_i = \bar{X} \tag{5}$$

$$\hat{\theta}_{MM} = \frac{3}{2}\bar{X} \tag{6}$$

$$E(\hat{\theta}_{MM}) = E\left(\frac{3}{2}\bar{X}\right) = \frac{3}{2}E\left(\bar{X}\right) = \frac{3}{2}\frac{2}{3}\theta = \theta \tag{7}$$

The ML estimator is $\hat{\theta}_{ML} = X_{max}$ and it is obviously biased, because θ is the superior limit of the support.

But before choosing the MM estimator, in this case as in others, we should also evaluate other desirable properties.

Example: biased estimator of the variance of a normal

$$\hat{\sigma}^2_{ML} = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2 = \bar{\sigma}^2 \quad (8)$$

The ML estimator (sample variance) is biased for the true variance

$$\begin{aligned} E\left(\hat{\sigma}^2_{ML}\right) &= E\left(\frac{1}{n}\sum_{i=1}^{n} X_i^2 - \left(\frac{1}{n}\sum_{i=1}^{n}\bar{X}\right)^2\right) = \frac{1}{n}\left(\sum_{i=1}^{n} E(X_i^2) - E(n\bar{X}^2)\right) \\ &= \frac{1}{n}\left(\sum_{i=1}^{n}(\sigma^2 + \mu^2) - n\left(\frac{\sigma^2}{n} + \mu^2\right)\right) = \frac{n-1}{n}\sigma^2 \end{aligned} \quad (9)$$

An unbiased estimator is:

$$\hat{\sigma}^2 = \frac{n}{n-1}\hat{\sigma}^2_{ML} = \frac{n}{n-1}\frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2 = \frac{n}{n-1}\bar{\sigma}^2 \quad (10)$$

Why unbiasedness may not be a desirable property

Suppose that S is an unbiased estimator for θ

$$E(S) = \theta \tag{11}$$

Let's assume that the cost of mis-estimation is quadratic and equal to:

$$E(S - \theta)^2 = Var(S) \tag{12}$$

Now consider a generic biased estimator R that can always be written as

$$R = \alpha S + (1 - \alpha)K \tag{13}$$

where K is a constant.

Possible undesirability of unbiasedness (cont.)

Also for R we can define the cost of mis-estimation as:

$$
\begin{aligned}
E\quad & (R-\theta)^2 = E\left((R-E(R)) + (E(R)-\theta)\right)^2 \qquad (14)\\
= \quad & E\left((R-E(R))^2\right) + E\left((E(R)-\theta)^2\right) + E\left(2(R-E(R)(E(R)-\theta)\right)\\
= \quad & Var(R) + (E(R)-\theta)^2\\
= \quad & \alpha^2 Var(S) + (\alpha E(S) - (1-\alpha)K - \theta)^2\\
= \quad & \alpha^2 Var(S) + (1-\alpha)^2 (K-\theta)^2
\end{aligned}
$$

And we can always find a value of α such that

$$E(R-\theta)^2 < E(S-\theta)^2 = Var(S) \qquad (15)$$

Starting from an unbiased estimator I can construct a biased estimator with smaller variance and mis-estimation error.

Possible undesirability of unbiasedness (cont.)

It is easy to see graphically and intuitively that unbiasedness may not be desirable if it comes at the cost of a higher estimation error.

Biased but more precise estimators may be preferable to unbiased estimators

Moreover, within the class of unbiased estimators we need to define other criteria to choose which estimator we prefer

We now turn to the property of Efficiency, which allows us to rank the desirability of a set of unbiased estimators.

Subsection 2

Efficiency

Definition of Efficiency

Let $\hat{\theta}_1$ and $\hat{\theta}_2$ be two unbiased estimators of θ. If

$$Var(\hat{\theta}_1) < Var(\hat{\theta}_2) \tag{16}$$

then $\hat{\theta}_1$ is more efficient than $\hat{\theta}_2$.

The relative efficiency or relative precision of $\hat{\theta}_1$ with respect to $\hat{\theta}_2$ is

$$\frac{Var(\hat{\theta}_1)}{Var(\hat{\theta}_2)} \tag{17}$$

Within the set of unbiased estimators we clearly prefer the most efficient ones.

Example: Efficiency of the sample mean

The variance of the sample mean using the entire sample of size n:

$$Var(\hat{\mu}) = Var\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{\sigma^2}{n} \tag{18}$$

The variance of the sample mean using $k < n$ observations is :

$$Var(\hat{\omega}) = Var\left(\frac{1}{k}\sum_{i=1}^{k} X_i\right) = \frac{\sigma^2}{k} \tag{19}$$

The sample mean using all observations is relatively more efficient:

$$\frac{Var(\hat{\mu})}{Var(\hat{\omega})} = \frac{k}{n} \tag{20}$$

In general, it is not a good idea to throw away sample observations (but there are exceptions as we will see later).

The Cramer-Rao lower bound

If $\hat{\theta}$ is unbiased for θ given a random sample of size n, then:

$$\begin{aligned} Var(\hat{\theta}) &\geq \frac{1}{\mathcal{I}_n(\theta)} \\ &= \frac{1}{E_X\left(S(\theta, X)\right)^2} = \frac{1}{E_X\left(\frac{\partial l(X|\theta)}{\partial\theta}\right)^2} = \frac{-1}{E_X\left(\frac{\partial^2 l(X|\theta)}{\partial\theta^2}\right)} \end{aligned} \tag{21}$$

where $l(X|\theta) = ln(L(X|\theta))$ is the log likelihood and the RHS of the inequality is the Cramer-Rao lower bound

Using this theorem we can tell whether an estimator is a Minimum Variance Unbiased Estimator .

Note the relationship between the Cramer-Rao lower bound and the Fisher Information for a sample of size n.

An unbiased ML estimator reaches the Cramer-Rao lower bound also in small sample.

Regularity conditions for the Cramer-Rao lower bound

Some regularity conditions are needed for the Cramer-Rao lower bound to exist

- $f(X|\theta)$ must be
  - continuous
  - with continuous first order and second order derivatives
- the set of values X for which $f(X|\theta) \neq 0$ must not depend on θ, i.e. the support of the underlying distribution cannot depend on the parameter to be estimated.

A different expression for the Cramer-Rao lower bound

It is easy to show that

$$E_X\left(\frac{\partial l(X|\theta)}{\partial\theta}\right)^2 = \sum_{i=1}^{n} E_X\left(\frac{\partial \ln f(X|\theta)}{\partial\theta}\right)^2 = nE_X\left(\frac{\partial \ln f(X|\theta)}{\partial\theta}\right)^2 \quad (22)$$

which explains the expression of the Cramer-Rao lower bound in Larsen and Marx book.

See Casella and Berger or other equivalent texts for a proof of the Cramer-Rao inequality.

Example: the sample mean reaches the lower bound

$$X \sim f_X(x|\mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \tag{23}$$

$$l(X|\mu) = -\frac{n}{2} ln(2\pi) - \frac{n}{2} ln(2\sigma^2) - \sum_{i=1}^{n} \frac{(X_i - \mu)^2}{2\sigma^2} \tag{24}$$

$$\frac{\partial l(X|\mu)}{\partial \mu} = \frac{1}{\sigma^2} \sum_{i=1}^{n} (X_i - \mu) \tag{25}$$

$$E\left(\frac{\partial l(X|\mu)}{\partial \mu}\right)^2 = E\left(\frac{1}{\sigma^4} \sum_{i=1}^{n} (X_i - \mu) \sum_{j=1}^{n} (X_j - \mu)\right) \tag{26}$$

which, because of the independence of sample observations is:

$$E\left(\frac{\partial l(X|\mu)}{\partial \mu}\right)^2 = \frac{n}{\sigma^4}\sigma^2 = \frac{n}{\sigma^2} = \frac{1}{Var(\hat{\mu})} = \frac{1}{Var\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right)} \tag{27}$$

Subsection 3

Sufficiency

Definitions of Sufficient Statistic

Given a random sample $\{X_1, ..., X_n\}$ drawn from a distribution $f_X(X|\theta)$, a statistic $\hat{S} = h(X_1, ..., X_n)$ is sufficient for θ if the likelihood function can be factorized as:

$$L(X|\theta) = \prod_{i=1}^{n} f_X(x_i|\theta) = g(\hat{S}, \theta)b(X_1, ..., X_n) = \tag{28}$$

This means that:

- to maximize the likelihood we just need to maximize $g(\hat{S}, \theta)$;
- the ML estimator $\hat{\theta}_{ML}$ is just a function of the sufficient statistic $\hat{S}$;
- $\hat{S} = h(X_1, ..., X_n)$ summarizes all the useful information that the sample can provide to estimate θ.

The sample mean is sufficient for the Normal mean

$$
\begin{aligned}
L(X|\mu,\sigma^2) &= \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x_i-\mu)^2}{2\sigma^2}} \\
&= \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{\left((x_i-\bar{x})+(\bar{x}-\mu)\right)^2}{2\sigma^2}} \qquad (29) \\
&= \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x_i-\bar{x})^2+(\bar{x}-\mu)^2+2(x_i-\bar{x})(\bar{x}-\mu)}{2\sigma^2}} \qquad (30) \\
&= \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x_i-\bar{x})^2}{2\sigma^2}} \prod_{i=1}^{n} e^{-\frac{(\bar{x}-\mu)^2}{2\sigma^2}} \qquad (31) \\
&= b(x_1,\ldots,x_n|\sigma^2)g(\bar{x},\mu|\sigma^2) \qquad (32)
\end{aligned}
$$

The sample mean $\bar{x}$ contains all the information the sample can provide to estimate the mean of the normal, for given variance.

A MM estimator that is not sufficient

Recall the distribution

$$f_X(x|\theta) = \frac{2X}{\theta^2} \qquad \text{for } 0 \leq X \leq \theta \tag{33}$$

for which the MM estimator for θ

$$\theta^{MM} = \frac{3}{2}\bar{X} \tag{34}$$

is unbiased while the ML estimator

$$\hat{\theta}^{ML} = X_{max} \tag{35}$$

is biased.

A MM estimator that is not sufficient (cont.)

Consider two random samples of size $n = 3$:

$$S_1 = \{3, 4, 5\} \qquad\qquad S_2 = \{1, 3, 8\}$$

For both samples the MM estimate is

$$\hat{\theta}_e^{MM} = \frac{3}{2}\bar{x} = \frac{3}{2}\frac{12}{3} = 6 \tag{36}$$

but the estimator, as well as the sample mean, are not sufficient: the two samples convey different information on what θ might be:

- S_1 is compatible with the possibility that $\theta = 5$;
- this possibility is incompatible in S_2.

Example 5.6.2 in Larsen and Marx shows that $\hat{\theta}^{ML} = X_{max}$ is sufficient as intuition would suggest.

Sufficient statistics and MVUE (Blackwell theorem)

Given a random sample $\{X_1, ..., X_n\}$ drawn from $f_X(X|\theta)$, if

- $\hat{\theta}$ is a MVUE and
- $\hat{S} = h(X_1, ..., X_n)$ is a sufficient statistic for θ,

then $\hat{\theta}$ is function of $\hat{S}$ only and not directly of the sample

The converse is not true: not all functions of sufficient statistics are MVUE.

But if we want an MVUE we can restrict our search to functions of sufficient statistics

Rao-Blackwell criterium for MVUE

An estimator $\hat{\theta}$ is a MVUE for θ if and only if *f*or any other estimator $\tilde{\theta}$ that is unbiased for θ the following equality holds

$$Cov(\hat{\theta}; \hat{\theta} - \tilde{\theta}) = 0 \tag{37}$$

Section 2

Asymptotic properties

Why are asymptotic properties important

We obviously always work with finite samples, but:

1. we would feel unconfortable in using an estimator that had undesirable properties in the hypothetical case in which the sample size could go to ∞.
2. A finite sample may be sufficiently “large” for asymptotic results to hold with a very good approximation, even if its actual size is effectively not so large.
3. Small sample properties are often difficult to characterize and less attractive than asymptotic properties.
4. Asymptotic hypothesis testing is easy to define and perform, while it may be more problematic in small sample.

Notation for the asymptotic analysis of estimators

Asymptotics studies how the sequence of estimators that we obtain for each sample size behaves when the sample size increases towards ∞.

Given an estimator $\hat{\theta}$ for the parameter θ, we denote its sequence, when sample size n increases, as $\hat{\theta}_n$.

Note that for each element $\hat{\theta}_n$ in the sequence, the estimator (the recipe) is the same except that it is applied to a sample of different (and actually larger) size.

As a companion to the pages that follow see also the Appendix (Part 11 of the slides): *Some basic asymptotic results*

Subsection 1

Asymptotic unbiasedness

Definition of Asymptotic Unbiasedness

$\hat{\theta}_n$ is asymptotically unbiased for θ if

$$\lim_{n \to +\infty} E(\hat{\theta}_n) = \theta \tag{38}$$

For example, we know that the sample variance (ML estimator) is biased for the population variance

$$E\left(\frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2\right) = \frac{n-1}{n}\sigma^2 \tag{39}$$

But it is asymptotically unbiased because:

$$\lim_{n \to +\infty} \frac{n-1}{n}\sigma^2 = \sigma^2 \tag{40}$$

Subsection 2

Consistency

Definition of consistency

Let $\hat{\theta}_n$ denote a sequence of estimators for each sample size n.

$\hat{\theta}_n$ is consistent for θ if it converges in probability to θ, i.e. if:

$$Pr(|\hat{\theta}_n - \theta| < \varepsilon) > 1 - \delta \qquad \text{for } n \to +\infty \text{ and } \forall \varepsilon > \delta > 0 \tag{41}$$

or equivalently if

$$\lim_{n \to +\infty} = Pr(|\hat{\theta}_n - \theta| > \varepsilon) = 0 \qquad \forall \varepsilon \tag{42}$$

Sufficient conditions for consistency

Using Chebyshev's inequality we can write that for every estimator $\hat{\theta}_n$,

$$Pr(|\hat{\theta}_n - E(\hat{\theta}_n)| > \epsilon) < \frac{Var(\hat{\theta}_n)}{\epsilon^2} \tag{43}$$

A set of sufficient conditions for the Consistency of $\hat{\theta}$ is, therefore that:

$$\begin{aligned} E(\hat{\theta}) &= \theta \rightarrow \text{unbiasedness} && (44) \\ \lim_{n \to +\infty} Var(\hat{\theta}_n) &= 0 \rightarrow \text{the variance goes to zero when } n \text{ increases.} \end{aligned}$$

Consistency of the sample mean

We know already that the sample mean is unbiased for $\hat{\mu}$:

$$E(\hat{\mu}_n) = E\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n}\sum_{i=1}^{n} E(X_i) = \frac{1}{n} n\mu = \mu \tag{45}$$

We also know that

$$Var(\hat{\mu}_n) = E(\hat{\mu} - \mu)^2 = E\left(\frac{1}{n}\sum_{i=1}^{n} X_i - \mu\right)^2 = \frac{n}{n^2}\sigma^2 = \frac{\sigma^2}{n} \tag{46}$$

and since

$$\lim_{n\to+\infty} Var(\hat{\mu}_n) = \lim_{n\to+\infty} \frac{\sigma^2}{n} = 0 \tag{47}$$

we can conclude, using Chebyshev inequality, that the sample mean is consistent for the population mean.

Subsection 3

Asymptotic Distribution of Estimators

When small sample distributions are unknown ...

Estimators are transformations of the sample random variables .

Using the CLT and the Delta method, we can derive the parameters of the asymptotic (normal) distribution of an estimator .

This can be done without the need of assumptions concerning the underlying distribution of the random variables in the sample for which the estimator is used.

We summarize here the most useful results without proofs (for which see the reading list).

These results are crucial for hypothesis testing when small sample distributions are unknown.

Asymptotic Distribution of the ML estimator

Let $\hat{\theta}_n^{ML}$ be the ML estimator for a sample size n.

The asymptotic distribution of the ML estimator is

$$\sqrt{n}(\hat{\theta}_n^{ML} - \theta) \quad \xrightarrow{d} \quad \text{Normal}(0, \frac{n}{\mathcal{I}_n(\theta)}) \tag{48}$$

or equivalently

$$\sqrt{n}(\hat{\theta}_n^{ML} - \theta) \quad \xrightarrow{d} \text{Normal}(0, \frac{1}{\mathcal{I}_1(\theta)}) \tag{49}$$

where

$$\mathcal{I}_n(\theta) = E_X\left(\frac{\partial \ln(L(x|\theta))}{\partial\theta}\right)^2 = -E_X\left(\frac{\partial^2 \ln(L(x|\theta))}{\partial\theta^2}\right)$$

and $\mathcal{I}_n(\theta) = n\mathcal{I}_1(\theta)$

The ML estimator is always asymptotically efficient

An example: exponential distribution

Consider a random sample x_1, x_2,..., x_n from the pdf

$$x \sim f(x, \theta) = \theta e^{-x\theta} \quad \text{with} \quad \theta \geq 0 \quad \text{and} \quad x > 0. \tag{50}$$

where, using rules of integration, we can derive

$$E(X) = \frac{1}{\theta} \quad \text{and} \quad Var(X) = \frac{1}{\theta^2} \tag{51}$$

The likelihood function is:

$$L(x|\theta) = \theta^n e^{-\theta \sum_{i=1}^{i=n} x_i} = \theta^n e^{-\theta n \bar{x}} \tag{52}$$

and the log likelihood is

$$l(x|\theta) = n log \theta - \theta \sum_{i=1}^{n} x_i \tag{53}$$

An example: exponential distribution (cont.)

The first order condition is

$$\frac{dl(X|\theta)}{d\theta} = S(\theta, X) = \frac{n}{\theta} - \sum_{i=1}^{n} X_i = 0 \tag{54}$$

which can be solved to obtain the ML estimator is

$$\hat{\theta}_{ML} = \frac{n}{\sum_{i=1}^{n} X_i} \tag{55}$$

The S.O.C is also satisfied since:

$$\frac{d^2 l(X|\theta)}{(d\theta)^2} = \frac{-n}{\theta^2} < 0 \tag{56}$$

An example: exponential distribution (cont.)

The final ingredient to derive the asymptotic distribution is the Fisher Information:

$$\mathcal{I}_n(\theta) = E_X\left(\frac{\partial \ln(L(X|\theta))}{\partial\theta}\right)^2 = -E_X\left(\frac{\partial^2 \ln(L(X|\theta))}{\partial\theta^2}\right) = n\mathcal{I}_1(\theta)$$

The computation is easy using the negative of the expectation of the second derivative, which is a constant:

$$\mathcal{I}_n(\theta) = \frac{n}{\theta^2} \quad (57)$$

$$\mathcal{I}_1(\theta) = \frac{1}{\theta^2} \quad (58)$$

(59)

An example: exponential distribution (cont.)

We now have all the elements to state what is the asymptotic distribution of the ML estimator

$$\sqrt{n}(\hat{\theta}_n^{ML} - \theta) \quad \xrightarrow{d} \quad \text{Normal}(0, \frac{1}{\mathcal{I}_1(\theta)}) \tag{60}$$

$$\sqrt{n}(\hat{\theta}_n^{ML} - \theta) \quad \xrightarrow{d} \quad \text{Normal}(0, \theta^2) \tag{61}$$

Asymptotic Distribution of the MM estimator

Let $\hat{\theta}_n^{MM}$ be the MM estimator for a sample size n, that is obtained by solving equations of the form

$$\sum_{i=1}^{n} g(x_i, \theta_n^{MM}) = 0 \tag{62}$$

with

- $E(g(X, \theta)) = 0$
- g is twice continuously differentiable

then

$$\sqrt{n}(\hat{\theta}_n^{MM} - \theta) \quad \xrightarrow{d} \text{Normal}(0, nV) \tag{63}$$

where

$$V = [E(g_\theta(X, \theta))]^{-1}[E(g(X, \theta)^2)][E(g_\theta(X, \theta))]^{-1} \tag{64}$$

An example: exponential distribution (cont.)

Continuing the previous example, the MM estimator for θ solves:

$$(g(X,\theta)) = (E(X) - \frac{\sum X_i}{n}) = (E(X) - \bar{X}) = 0 \quad (65)$$

To compute V we first need to derive

$$\begin{aligned} E(g(X,\theta)^2) &= E((E(X) - \bar{X})^2) && (66)\\ &= E(E(X)^2 + \bar{X}^2 - 2E(X)\bar{X}) && (67)\\ &= E(X)^2 + E(\bar{X}^2) - 2E(X)^2 && (68)\\ &= E(\bar{X}^2) - E(X)^2 && (69)\\ &= \frac{Var(X)}{n} = \frac{1}{n\theta^2} && (70) \end{aligned}$$

where in the last step we have used

$$E(\bar{X}^2) = Var(\bar{X}) + E(X)^2 = \frac{Var(X)}{n} + E(X)^2 \quad (71)$$

An example: exponential distribution (cont.)

To compute V we also need $E(g_\theta(X,\theta))$. Since:

$$g(X,\theta) = E(X) - \bar{X} = \frac{1}{\theta} - \bar{X} \tag{72}$$

then

$$E(g_\theta(X,\theta)) = -\frac{1}{\theta^2} \tag{73}$$

and

$$\begin{aligned} V &= [E(g_\theta(X,\theta))]^{-1}[E(g(X,\theta)^2)][E(g_\theta(X,\theta))]^{-1} && (74) \\ &= \left(-\frac{1}{\theta^2}\right)^{-1}\left(\frac{1}{n\theta^2}\right)\left(-\frac{1}{\theta^2}\right)^{-1} = \frac{\theta^2}{n} && (75) \end{aligned}$$

An example: exponential distribution (cont.)

We now have all the elements to state what is the asymptotic distribution of the MM estimator

$$\sqrt{n}(\hat{\theta}_n^{MM} - \theta) \quad \xrightarrow{d} \quad \text{Normal}(0, nV) \tag{76}$$

$$\sqrt{n}(\hat{\theta}_n^{MM} - \theta) \quad \xrightarrow{d} \quad \text{Normal}(0, \theta^2) \tag{77}$$

Note that in this case the asymptotic distributions of the MM and ML estimators coincide.

But this is not always the case as the next example shows.

Asymptotics of MM and ML for the Pareto distribution

Recall the example in Part II of the slides.

Let X denote individual income. The Pareto's Law claims that

$$P(X \geq x) = \left(\frac{\nu}{X}\right)^{\theta} \quad \Rightarrow \quad F_X(X|\theta,\nu) = 1 - \left(\frac{\nu}{X}\right)^{\theta} \qquad (78)$$

where ν is the (known) minimum income in the population and $\theta > 1$.

Thus by differentiation the pdf is:

$$X \sim f_X(X|\theta,\nu) = \theta\nu^{\theta}\left(\frac{1}{X}\right)^{\theta+1} \quad , \quad X > \nu; \theta > 1 \qquad (79)$$

The MM estimator for θ in the Pareto distribution

$$E(X) = \int_{\nu}^{\infty} X\theta\nu^{\theta}\left(\frac{1}{X}\right)^{\theta+1} dX = \theta\nu^{\theta}\int_{\nu}^{\infty} X^{-\theta} dX = \frac{\theta\nu}{\theta - 1} \tag{80}$$

The Method of Moments estimates θ solving for $\hat{\theta}$:

$$E(X) = \frac{\hat{\theta}\nu}{\hat{\theta} - 1} = \frac{1}{n}\sum_{i=1}^{n}(X_i) = \bar{X} \tag{81}$$

which gives the estimator

$$\hat{\theta}^{MM} = \frac{\bar{X}}{\bar{X} - \nu} \tag{82}$$

And given a random sample $x_1, ..., x_n$, an estimate

$$\theta_e^{MM} = \frac{\bar{x}}{\bar{x} - \nu} \tag{83}$$

Asymptotic variance of the MM estimator

$$(g(x,\theta))^2 = \left(\frac{\theta\nu}{\theta-1} - \bar{X}\right)^2 \tag{84}$$

$$\begin{aligned} E\left((g(x,\theta))^2\right) &= E(\bar{X}^2) - \left(\frac{\theta\nu}{\theta-1}\right)^2 \\ &= \frac{Var(X)}{n} = \frac{1}{n}\frac{\nu^2\theta}{(\theta-1)^2(\theta-2)} \end{aligned} \tag{85}$$

where in the last line we have used

$$E(\bar{X}^2) = Var(\bar{X}) + E(X)^2 = \frac{Var(X)}{n} + \left(\frac{\theta\nu}{\theta-1}\right)^2 \tag{86}$$

and the expression for the variance of a Pareto distribution which holds for $\theta > 2$.

Asymptotic variance of the MM estimator (cont.)

We also need

$$E(g_\theta(X,\theta)) = \frac{-\nu}{(\theta-1)^2} \tag{87}$$

in order to derive

$$\begin{aligned} V &= [E(g_\theta(X,\theta))]^{-1}[E(g(X,\theta)^2)][E(g_\theta(X,\theta))]^{-1} && (88)\\ &= \left(\frac{(\theta-1)^2}{-\nu}\right)\left(\frac{1}{n}\frac{\nu^2\theta}{(\theta-1)^2(\theta-2)}\right)\left(\frac{(\theta-1)^2}{-\nu}\right) = \frac{\theta^2}{n} && (89)\\ &= \frac{(\theta-1)^2\theta}{n(\theta-2)} && (90) \end{aligned}$$

And thus

$$\sqrt{n}(\hat{\theta}_n^{MM} - \theta) \xrightarrow{d} \text{Normal}(0, nV) \tag{91}$$

$$\sqrt{n}(\hat{\theta}_n^{MM} - \theta) \xrightarrow{d} \text{Normal}(0, \frac{(\theta-1)^2\theta}{(\theta-2)}) \tag{92}$$

The ML estimator for θ in the Pareto distribution

The likelihood of the random sample $x_i, ..., x_n$ is

$$L(X|\theta,\nu) = \prod_{i=1}^{n} \theta\nu^{\theta} \left(\frac{1}{X_i}\right)^{\theta+1} \tag{93}$$

$$l(X|\theta,\nu) = \ln L(X|\theta,\nu) = n\ln\theta + n\theta\ln\nu - (\theta+1)\sum_{i=1}^{n} lnX_i \tag{94}$$

The first order condition is

$$\frac{n}{\theta} + n\ln\nu - \sum_{i=1}^{n} lnX_i = 0 \tag{95}$$

Hence the estimator and estimate are :

$$\hat{\theta}^{ML} = \frac{n}{-n\ln\nu + \sum_{i=1}^{n} lnX_i}; \qquad \theta_e^{ML} = \frac{n}{-n\ln\nu + \sum_{i=1}^{n} lnx_i} \tag{96}$$

Asymptotic variance of the ML estimator

The asymptotic variance is the inverse of the Fisher Information:

$$\mathcal{I}_n(\theta) = E_X\left(\frac{\partial \ln(L(X|\theta))}{\partial\theta}\right)^2 = -E_X\left(\frac{\partial^2 \ln(L(x|\theta))}{\partial\theta^2}\right) = n\mathcal{I}_1(\theta)$$

The computation is easy using the negative of the expectation of the second derivative, which is a constant:

$$\mathcal{I}_n(\theta) = \frac{n}{\theta^2} \quad (97)$$

$$\mathcal{I}_1(\theta) = \frac{1}{\theta^2} \quad (98)$$

and therefore

$$\sqrt{n}(\hat{\theta}_n^{ML} - \theta) \xrightarrow{d} \text{Normal}(0, \frac{1}{\mathcal{I}_1(\theta)}) \quad (99)$$

$$\sqrt{n}(\hat{\theta}_n^{ML} - \theta) \xrightarrow{d} \text{Normal}(0, \theta^2) \quad (100)$$

Comparison of ML and MM asymptotic variances

It is easy to check that

$$AsyVar(\hat{\theta}^{ML}) = \theta^2 < \frac{\theta(\theta-1)^2}{\theta-2} = AsyVar(\hat{\theta}^{MM}) \qquad (101)$$

The MM estimator of the parameter θ of the Pareto distribution is less efficient asymptotically than the ML estimator.

Subsection 4

Asymptotic efficiency

Definition of Asymptotic Efficiency

$\hat{\theta}_n$ is asymptotically efficient for θ in a given class of estimators if for any other estimator $\tilde{\theta}_n$ in the same class:

$$AsyVar(\hat{\theta}_n) \leq AsyVar(\tilde{\theta}_n) \tag{102}$$

where AsyVar denotes the variance of the asymptotic distribution

This is for example the case for the ML and MM estimators given that

$$\frac{1}{\mathcal{I}_1(\theta)} \leq nV \tag{103}$$

Indeed we know that the ML estimator is asymptotically the most efficient. Which is not always true for the MM estimator

Section 3

Invariance of ML estimators

Invariance: a useful property of the ML estimator

If

$$\hat{\theta}^{ML} \text{ is the ML estimator for } \theta,$$

then, for any continuous function $g(.)$,

$$g(\hat{\theta}^{ML}) \text{ is the ML estimator for } g(\theta).$$

Note that:

- in general invariance does not preserve unbiasedness: if $\hat{\theta}^{ML}$ is unbiased for θ, $g(\hat{\theta}^{ML})$ may be biased for $g(\theta)$;
- if $g(.)$ is linear then invariance preserves unbiasedness;
- invariance always preservers consistency, even for non linear $g(.)$ functions (see Continuous Mapping Theorem for P-Convergence).