GMM, Indirect Inference and Bootstrap

Estimators and their properties

Willi Mutschler

TU Dortmund

Winter 2015/2016

- Let X be a random variable (or random vector) representing a random experiment we are interested in
- We would like to say something about the distribution of X
- Usually, the distribution of X is unknown
- We have to collect information about the distribution by observing the random outcome n times
- Before the outcomes are actually observed, we may regard the n observations as random variables $X_1, \dots, X_n$

- The random variables $X_1, \dots, X_n$ are called a (simple) **random sample** from X, if
  1. each $X_i, i = 1, \dots, n$, is distributed in the same way as X,
  2. $X_1, \dots, X_n$ are stochastically independent.
- The sample elements are i.i.d.
- n is the sample size

- The joint density of the sample elements $X_1, \ldots, X_n$ is

$$f_{X_1,\ldots,X_n}(x_1,\ldots,x_n) = \prod_{i=1}^{n} f_{X_i}(x_i) = \prod_{i=1}^{n} f_X(x_i)$$

- Let $g : \mathbb{R}^n \longrightarrow \mathbb{R}$ be a real-valued function with n arguments, not containing any unknown parameters, then

$$T = g(X_1, \ldots, X_n)$$

  is called a **statistic** (or sample function)

Examples

- Sample mean:

$$\bar{X} = g(X_1, \dots, X_n) = \frac{1}{n} \cdot \sum_{i=1}^{n} X_i$$

- Sample variance:

$$\begin{aligned} S^2 &= g(X_1, \dots, X_n) = \frac{1}{n} \cdot \sum_{i=1}^{n} \left(X_i - \bar{X}\right)^2 \\ S^{*2} &= g(X_1, \dots, X_n) = \frac{1}{n-1} \cdot \sum_{i=1}^{n} \left(X_i - \bar{X}\right)^2 \end{aligned}$$

Examples

- Empirical distribution function

$$\hat{F}(x) = \frac{1}{n}\sum_{i=1}^{n} 1(X_i < x)$$

  where $1(A) = 1$ if A is true and $1(A) = 0$ else
- Empirical p-quantile

$$\hat{x}_p = inf\{x \in \mathbb{R} : \hat{F}(x) \geq p\}$$

Remarks:

- All concepts are easily generalized to the multivariate case
- The statistic $T = g(X_1, \ldots, X_n)$ is a function of random variables and hence also a random variable
- A statistic has a distribution (and thus an expectation and variance)
- Statistics are basic tools for estimation of parameters and hypothesis tests about parameters

- Let θ be a vector of unknown parameters we are interested in
- A statistic $\hat{\theta}(X_1, \ldots, X_n)$ is called **estimator (Schätzer)** of θ
- The realization $\hat{\theta}(x_1, \ldots, x_n)$ is called **estimate (Schätzwert)**
- The estimator $\hat{\theta}(X_1, \ldots, X_n)$ is a random vector
- The estimate $\hat{\theta}(x_1, \ldots, x_n)$ is a vector of real numbers
- Notation: Usually we simply write $\hat{\theta}$ for both, but $\hat{\theta}$ and $\hat{\theta}$ are not the same thing!

Example:

- Let $X \sim N(\mu, \sigma^2)$ with unknown parameters μ and σ^2
- We would like to estimate the parameter vector

$$\theta = \begin{bmatrix} \mu \\ \sigma^2 \end{bmatrix} = \begin{bmatrix} E(X) \\ Var(X) \end{bmatrix}$$

- A possible estimator of μ is

$$\hat{\mu} = \frac{1}{n} \sum_{i=1}^{n} X_i$$

- A possible estimator of σ^2 is

$$\hat{\sigma}^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \hat{\mu})^2$$

- The estimator of θ and the estimate are

$$\begin{aligned}
\hat{\theta} &= \begin{bmatrix} \hat{\mu} \\ \hat{\sigma^2} \end{bmatrix} = \begin{bmatrix} \frac{1}{n}\sum_{i=1}^{n} X_i \\ \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \hat{\mu})^2 \end{bmatrix} \\
\hat{\theta} &= \begin{bmatrix} \hat{\mu} \\ \hat{\sigma^2} \end{bmatrix} = \begin{bmatrix} \frac{1}{n}\sum_{i=1}^{n} x_i \\ \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \hat{\mu})^2 \end{bmatrix}
\end{aligned}$$

- Why do we need the complex theoretical concept of estimators as random variables?
- Note that **the** estimator of θ does not exist, there are always many possible estimators
- Example: Let $\theta = Var(X)$; two possible estimators of θ are

$$
\begin{aligned}
\hat{\theta}_1(X_1, \ldots, X_n) &= \frac{1}{n} \sum_{i=1}^{n} \left(X_i - \bar{X}\right)^2 \\
\hat{\theta}_2(X_1, \ldots, X_n) &= \frac{1}{n-1} \sum_{i=1}^{n} \left(X_i - \bar{X}\right)^2
\end{aligned}
$$

Important questions:

- How can we compare different estimators?
- What is a good estimator?
- Which criteria should a good estimator satisfy?
- Is there an optimal estimator?
- How can we find good estimators?

- We distinguish two groups of properties:
    - small (finite) sample properties
    - asymptotic properties
- We consider finite sample properties first
- For simplicity, we only consider univariate estimators
- Thought experiment: repeated samples

- An estimator $\hat{\theta}(X_1, \ldots, X_n)$ is called **unbiased** for θ if

$$E\left(\hat{\theta}\right) = \theta$$

- The bias is defined as

$$bias(\hat{\theta}) = E(\hat{\theta}) - \theta$$

- Generalization to multivariate case is obvious

- How can two unbiased estimators of the unknown parameter θ be compared to each other?
- Let $\hat{\theta}_1$ and $\hat{\theta}_2$ be two unbiased estimators for θ. The estimator $\hat{\theta}_1$ is **relatively more efficient** than $\hat{\theta}_2$, if

$$Var(\hat{\theta}_1) \leq Var(\hat{\theta}_2)$$

for all possible θ and $Var(\hat{\theta}_1) < Var(\hat{\theta}_2)$ for at least one possible θ

Estimators and their properties

Mean squared error

- How can two biased estimators be compared?
- Let $\hat{\theta}$ be an arbitrary estimator for θ. Then

$$\begin{aligned} MSE(\hat{\theta}) &= E\left[\left(\hat{\theta}-\theta\right)^2\right] \\ &= Var\left(\hat{\theta}\right)+\left[bias(\hat{\theta})\right]^2 \end{aligned}$$

  is called the **mean-squared error** of the estimator
- If the estimator is unbiased, its MSE is equal to its variance
- If $MSE(\hat{\theta}_1) < MSE(\hat{\theta}_2)$, then $\hat{\theta}_1$ is more MSE-efficient

- What happens if the sample size goes to infinity?
- Practical relevance: How do estimators behave in large samples?
- We consider a sequence of estimators $\hat{\theta}_n(X_1, \dots, X_n)$ for $n = 1, 2, \dots$
- Consistency
- Asymptotic normality
- Asymptotic efficiency

- An estimator $\hat{\theta}_n(X_1, \ldots, X_n)$ is called **consistent** for θ, if

$$plim\ \hat{\theta}_n(X_1, \ldots, X_n) = \theta$$

- Sufficient (but not necessary) condition for consistency:

$$\begin{aligned} \lim_{n\to\infty} E(\hat{\theta}_n) &= \theta \\ \lim_{n\to\infty} Var(\hat{\theta}_n) &= 0 \end{aligned}$$

- Consistency is a basic and very important property of estimators

- **Attention:**
  Consistency and (asymptotic) unbiasedness are not the same thing
- An estimator can be
  - consistent and unbiased
  - inconsistent and unbiased
  - consistent and biased
  - inconsistent and biased
  - consistent and asymptotically unbiased
  - inconsistent and asymptotically unbiased
  - consistent and asymptotically biased
  - inconsistent and asymptotically biased

- An estimator $\hat{\theta}_n(X_1, \ldots, X_n)$ for θ is called **asymptotically normal**, if there is a sequence of real numbers $\theta_1, \theta_2, \ldots$ and a function $V(\theta)$ such that
$$\sqrt{n} \cdot \left(\hat{\theta}_n - \theta_n\right) \xrightarrow{d} U \sim N(0, V(\theta))$$
- Alternative notation:
$$\hat{\theta}_n \overset{appr}{\sim} N(\theta_n, V(\theta)/n)$$
- Generalization to the multivariate case

- The estimator

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

  for the expectation $E(X)$ is consistent and asymptotically normal under some mild regularity conditions

- Consistency $\longrightarrow$ laws of large number
- Asymptotic normality $\longrightarrow$ central limit theorems

- **Weak law of large numbers**: Let $X_1, X_2, \ldots$ be a sequence of i.i.d. random variables with $E(X_i) = \mu$ and $Var(X_i) = \sigma^2 < \infty$
- Consider the sequence of random variables

$$\bar{X}_n = \frac{1}{n} \sum_{i=1}^{n} X_i$$

- Then $plim\ \bar{X}_n = \mu$

Remarks:

- The law of large number states that $\bar{X}_n$ is consistent for $E(X) = \mu$
- For every (arbitrarily small) $\epsilon > 0$, the probability that the sample mean $\bar{X}_n$ deviates around μ by less than $\pm\epsilon$ converges to zero as the sample size goes to infinity
- Generalization to multivariate case is obvious
- Both the assumption of independence and the assumption of identical distributions may be weakened

- **Central limit theorem**: Let X be a random variable with $E(X) = \mu$ and $Var(X) = \sigma^2 < \infty$, and let $X_1, \dots, X_n$ be a random sample of X
- Consider the sequence of random variables

$$Z_n = \sqrt{n}\frac{\bar{X}_n - \mu}{\sigma}$$

- Then

$$Z_n \xrightarrow{d} U \sim N(0,1)$$

- Common notations:

$$\sqrt{n}\frac{\bar{X}_n - \mu}{\sigma} \overset{d}{\to} U \quad \sim \quad N(0,1)$$
$$\sqrt{n}\left(\bar{X}_n - \mu\right) \overset{d}{\to} U \quad \sim \quad N(0,\sigma^2)$$
$$\bar{X}_n \overset{appr}{\sim} N\left(\mu, \frac{\sigma^2}{n}\right)$$

- Convenient (but wrong) notation: $\sqrt{n}\left(\bar{X}_n - \mu\right) \overset{d}{\to} N(0,\sigma^2)$

- **Multivariate central limit theorem**: Let $X = (X_1, \dots, X_m)'$ be a random vector with $E(X) = \mu$ and $Cov(X) = \Sigma$
- Let $X_1, \dots, X_n$ be a (multivariate) random sample of X and

$$\bar{X}_n = \frac{1}{n} \sum_{i=1}^{n} X_i$$

- Then

$$\sqrt{n} \left(\bar{X}_n - \mu \right) \xrightarrow{d} U \sim N(0, \Sigma)$$

Estimators of moments

- Let $X_1, \dots, X_n$ be a random sample of X, then

$$\hat{\mu}_p = \frac{1}{n} \sum_{i=1}^{n} X_i^p$$

is an estimator for the p-th raw moment μ_p of X and

$$\hat{\mu}'_p = \frac{1}{n} \sum_{i=1}^{n} (X_i - \hat{\mu}_1)^p$$

is an estimator for the p-th central moment μ'_p of X

Weak law of large numbers for moments

- Let $X_1, X_2, \dots$ be a sequence of iid random variables with

$$\begin{aligned} E(X_i^p) &= \mu_p \\ E(X_i^{2p}) &= \mu_{2p} < \infty \end{aligned}$$

- Then *plim* $\hat{\mu}_p = \mu_p$
- Attention: The assumption $\mu_{2p} < \infty$ is *not* innocuous!

Central limit theorem for moments

- Let $X_1, X_2, \ldots$ be a sequence of iid random variables with

$$
\begin{aligned}
E(X_i^p) &= \mu_p \\
E(X_i^{2p}) &= \mu_{2p} < \infty
\end{aligned}
$$

- Then

$$\sqrt{n}\left(\hat{\mu}_p - \mu_p\right) \xrightarrow{d} U \sim N(0, Var\left(\hat{\mu}_p\right))$$

  where

$$Var\left(\hat{\mu}_p\right) = \frac{\mu_{2p} - \mu_p^2}{n}$$

Fundamental theorem of mathematical statistics

- Define

$$\Delta_n = sup_{x \in \mathbb{R}} \left| \hat{F}(x) - F(x) \right|.$$

- Let $X_1, X_2, \ldots$ be a sequence of iid random variables with distribution function $F(x)$. Then

$$P\left(lim_{n \to \infty} \Delta_n = 0\right) = 1.$$

- The empirical distribution function $\hat{F}$ converges uniformly to the cumulative distribution function F.