Mathematical Statistics

Anna Janicka

Lecture VII, 12.04.2021

ESTIMATOR PROPERTIES, PART III

(ASYMPTOTIC PROPERTIES)

Plan for Today

1. Asymptotic properties of estimators
   - *asymptotic unbiasedness*
   - consistency
   - asymptotic normality
   - asymptotic efficiency

2. Consistency, asymptotic normality and asymptotic efficiency of MLE estimators

Asymptotic properties of estimators

- Limit theorems describing estimator properties when $n \to \infty$
- In practice: information on how the estimators behave for large samples, *approximately*
- Problem: usually, there is no answer to the question what sample is large enough (for the approximation to be valid)

Consistency

Let $X_1, X_2, \ldots, X_n, \ldots$ be an IID sample (of independent random variables from the same distribution) . Let $\hat{g}(X_1, X_2, \ldots, X_n)$ be a sequence of estimators of the value $g(\theta)$.

$\hat{g}$ is a **consistent** estimator, if for all $\theta \in \Theta$, for any $\varepsilon > 0$:

$$\lim_{n\to\infty} P_\theta(|\hat{g}(X_1, X_2, \ldots, X_n) - g(\theta)| \leq \varepsilon) = 1$$

(i.e. $\hat{g}$ converges to $g(\theta)$ in probability)

Strong consistency

Let $X_1, X_2, \ldots, X_n, \ldots$ be an IID sample (of independent random variables from the same distribution). Let $\hat{g}(X_1, X_2, \ldots, X_n)$ be a sequence of estimators of the value $g(\theta)$.

$\hat{g}$ is **strong consistent**, if for any $\theta \in \Theta$:

$$P_\theta\left(\lim_{n\to\infty} \hat{g}(X_1, X_2, \ldots, X_n) = g(\theta)\right) = 1$$

(i.e. $\hat{g}$ converges to $g(\theta)$ almost surely)

Consistency – note

From the Glivenko-Cantelli theorem it follows that empirical CDFs converge almost surely to the theoretical CDF. Therefore, we should expect (strong) consistency from all sensible estimators.

Consistency = minimal requirement for a sensible estimator.

Consistency – how to verify?

- From the definition: for example with the use of a version of the Chebyshev inequality:

$$P(|\widehat{g}(X)-g(\theta)|\geq \varepsilon)\leq \frac{E(\widehat{g}(X)-g(\theta))^2}{\varepsilon^2}$$

Given that the MSE of an estimator is

$$MSE(\theta,\hat{g})=E_\theta(\hat{g}(X)-g(\theta))^2$$

we get a sufficient condition for consistency:

$$\lim_{n\to\infty} MSE(\theta,\hat{g})=0$$

- From the LLN

Consistency – examples

- For any family of distributions with an expected value: the sample mean $\overline{X}_n$ is a consistent estimator of the expected value $\mu(\theta)=E_\theta(X_1)$. Convergence from the SLLN.
- For distributions having a variance:
  $S_n^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2$ and $\hat{S}_n^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2$
  are consistent estimators of the variance
  $\sigma^2(\theta)=\mathrm{Var}_\theta(X_1)$. Convergence from the SLLN.

Consistency – examples/properties

- An estimator may be unbiased but inconsistent; e.g. $T_n(X_1, X_2, \ldots, X_n)=X_1$ as an estimator of $\mu(\theta)=E_\theta(X_1)$.

- An estimator may be biased but consistent; e.g. the biased estimator of the variance or any unbiased consistent estimator + 1/n.

Asymptotic normality

$\hat{g}(X_1, X_2, ..., X_n)$ is an **asymptotically normal** estimator of $g(\theta)$, if for any $\theta \in \Theta$ there exists $\sigma^2(\theta)$ such that, when $n \to \infty$

$$\sqrt{n}\left(\hat{g}(X_1, X_2, \ldots, X_n) - g(\theta)\right) \xrightarrow{D} N(0, \sigma^2(\theta))$$

Convergence in distribution, i.e. for any a

$$\lim_{n\to\infty} P_\theta\left(\frac{\sqrt{n}}{\sigma(\theta)}\left(\hat{g}(X_1, X_2, ..., X_n) - g(\theta)\right) \le a\right) = \Phi(a)$$

in other words, the distribution of $\hat{g}(X_1, X_2, ..., X_n)$ is for large n similar to $N(g(\theta), \frac{\sigma^2}{n})$

Asymptotic normality – properties

- An asymptotically normal estimator is consistent (not necessarily strongly).
- A *similar* condition to unbiasedness – the expected value of the asymptotic distribution equals $g(\theta)$ (but the estimator *does not need to* be unbiased).
- **Asymptotic variance** defined as $\sigma^2(\theta)$ or $\sigma^2(\theta)/n$ – the variance of the asymptotic distribution

Asymptotic normality – what it is not

- For an asymptotically normal estimator we <u>usually</u> have:

$$E_\theta \hat{g}(X_1, X_2, \ldots, X_n) \xrightarrow{n\to\infty} g(\theta)$$

$$n \operatorname{var} \hat{g}(X_1, X_2, \ldots, X_n) \xrightarrow{n\to\infty} \sigma^2(\theta)$$

but <u>these properties needn’t hold</u>, because convergence in distribution does not imply convergence of moments.

Asymptotic normality – example

- Let $X_1, X_2, \ldots, X_n, \ldots$ be an IID sample from a distribution with mean μ and variance σ^2. On the base of the CLT, for the sample mean we have

$$\sqrt{n}(\bar{X} - \mu) \xrightarrow{D} N(0, \sigma^2)$$

In this case the asymptotic variance, σ^2/n, is equal to the estimator variance.

Asymptotic normality – how to prove it

In many cases, the following is useful:

Delta Method. Let T_n be a sequence of random variables such that for $n\to\infty$ we have

$$\sqrt{n}(T_n - \mu) \xrightarrow{D} N(0,\sigma^2)$$

and let $h:\mathbb{R}\to\mathbb{R}$ be a function differentiable at point μ such that $h'(\mu)\neq 0$. Then

$$\sqrt{n}\left(h(T_n) - h(\mu)\right) \xrightarrow{D} N(0,\sigma^2(h'(\mu))^2)$$

μ, σ^2 are functions of θ

usually used when estimators are functions of statistics T_n, which can be easily shown co converge on the base of CLT

Asymptotic normality – examples cont.

In an exponential model: $MLE(\lambda) = \frac{1}{\overline{X}}$

From CLT, we get

$$\sqrt{n}\left(\overline{X} - \tfrac{1}{\lambda}\right) \xrightarrow{D} N\left(0, \tfrac{1}{\lambda^2}\right)$$

so from the Delta Method for $h(t)=1/t$:

$$\sqrt{n}\left(\tfrac{1}{\overline{X}} - \lambda\right) \xrightarrow{D} N\left(0, \tfrac{1}{\lambda^2} \cdot \left(-\tfrac{1}{(1/\lambda)^2}\right)^2\right)$$

so $\frac{1}{\overline{X}}$ is an asymptotically normal (and consistent) estimator of λ.

Asymptotic efficiency

For an asymptotically normal estimator $\hat{g}(X_1, X_2, \ldots, X_n)$ of $g(\theta)$ we define **asymptotic efficiency** as

$$\text{as.ef}(\hat{g}) = \frac{\left(g'(\theta)\right)^2 n}{\sigma^2(\theta) \cdot I_n(\theta)},$$

where $\sigma^2(\theta)/n$ is the asymptotic variance, i.e. for $n \to \infty$

$$\sqrt{n}\left(\hat{g}(X_1, X_2, \ldots, X_n) - g(\theta)\right) \xrightarrow{D} N(0, \sigma^2(\theta))$$

modification of the definition of efficiency to the limit case, with the asymptotic variance in place of the normal variance

$$\text{as.ef}(\hat{g}) = \frac{\left(g'(\theta)\right)^2}{\sigma^2(\theta) \cdot I_1(\theta)}$$

Relative asymptotic efficiency

Relative asymptotic efficiency for asymptotically normal estimators $\hat{g}_1(X)$ and $\hat{g}_2(X)$

$$\text{as.ef}\,(\hat{g}_1,\hat{g}_2) = \frac{\sigma_2^2(\theta)}{\sigma_1^2(\theta)} = \frac{\text{as.ef}\,(\hat{g}_1)}{\text{as.ef}\,(\hat{g}_2)}$$

Note. A less (asymptotically) efficient estimator may have other properties, which will make it preferable to a more efficient one.

Relative asymptotic efficiency – examples.
Is the mean better than the median?

Relative asymptotic efficiency – examples.
Is the mean better than the median?

Depends on the distribution!

a) normal model N(μ, σ^2): σ^2 known

$$\sqrt{n}\left(\overline{X}-\mu\right)\xrightarrow{D}N(0,\sigma^2)$$

$$\sqrt{n}\left(\text{mêd}-\mu\right)\xrightarrow{D}N(0,\tfrac{\pi\sigma^2}{2})$$

$$\text{as.ef}(\text{mêd},\overline{X})=\tfrac{2}{\pi}<1$$

b) Laplace model Lapl(μ, λ) λ known

$$\sqrt{n}\left(\overline{X}-\mu\right)\xrightarrow{D}N(0,\tfrac{2}{\lambda^2})$$

$$\sqrt{n}\left(\text{mêd}-\mu\right)\xrightarrow{D}N(0,\tfrac{1}{\lambda^2})$$

$$\text{as.ef}(\text{mêd},\overline{X})=2>1$$

c) some distributions do not have a mean...

Theorem: For a sample from a continuous distribution with density $f(x)$, the sample median is an asymptotically normal estimator for the median m (provided the density is continuous and $\neq 0$ at point m):

$$\sqrt{n}\left(\text{mêd}-m\right)\xrightarrow{D}N(0,\tfrac{1}{4(f(m))^2})$$

Consistency of ML estimators

Let $X_1, X_2, \ldots, X_n, \ldots$ be a sample from a distribution with density $f_\theta(x)$. If $\Theta \subseteq \mathbb{R}$ is an open set, and:

- all densities f_θ have the same support;
- the equation $\frac{d}{d\theta}\ln L(\theta) = 0$ has exactly one solution, $\hat{\theta}$.

Then $\hat{\theta}$ is the $MLE(\theta)$ and it is consistent

Note. MLE estimators do not have to be unbiased!

Asymptotic normality of ML estimators

Let $X_1, X_2, \ldots, X_n, \ldots$ be a sample with density $f_\theta(x)$, such that $\Theta \subseteq \mathbb{R}$ is open, and $\hat{\theta}$ is a consistent m.l.e. (for example, fulfills the assumptions of the previous theorem), and

- $\frac{d^2}{d\theta^2} \ln L(\theta)$ exists
- Fisher Information may be calculated, $0 < I_1(\theta) < \infty$
- the order of integration with respect to x and derivation with respect to θ may be changed

then $\hat{\theta}$ is asymptotically normal and

$$\sqrt{n}\left(\hat{\theta} - \theta\right) \xrightarrow{D} N\left(0, \frac{1}{I_1(\theta)}\right)$$

Asymptotic normality of ML estimators

Additionally, if $g:\mathbb{R}\to\mathbb{R}$ is a function differentiable at point θ, such that $g'(\theta) \neq 0$, and $\hat{g}(X_1, X_2,..., X_n)$ is $MLE(g(\theta))$, then

$$\sqrt{n}\left(\hat{g}(X_1, X_2,\ldots,X_n) - g(\theta)\right) \xrightarrow{D} N(0, \frac{(g'(\theta))^2}{I_1(\theta)})$$

Asymptotic efficiency of ML estimators

If the assumptions of the previous theorems are fulfilled, then the ML estimator (of θ or $g(\theta)$) is asymptotically efficient.

Asymptotic normality and efficiency of ML estimators – examples

- In the normal model: the mean is an asymptotically efficient estimator of μ
- In the Laplace model: the median is an asymptotically efficient estimator of μ

Summary: basic (point) estimator properties

- bias
- variance
- MSE
- efficiency

- asymptotic unbiasedness
- consistency
- asymptotic normality
- asymptotic efficiency