Mathematical Statistics

Anna Janicka

Lecture VII, 1.04.2019

ESTIMATOR PROPERTIES, PART III

CONFIDENCE INTERVALS – INTRO

Plan for Today

1. Asymptotic properties of estimators – cont.
   - asymptotic normality
   - asymptotic efficiency
2. Consistency, asymptotic normality and asymptotic efficiency of MLE estimators
3. Interval estimation – confidence intervals

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 $n \to \infty$

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

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 usually 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 these properties needn't hold, 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

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

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 $h: \mathrm{R} \to \mathrm{R}$ 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 $h(t)=1/t$:

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

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?

Depends on the distribution!

a) normal model N(μ, σ^2):

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

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

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

b) Laplace model Lapl(μ, λ)

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

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

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

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 ≠0 at point *m*):

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

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 $\Theta \subseteq \mathrm{R}$ 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*(θ) 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 $\Theta \subseteq \mathrm{R}$ 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

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

Asymptotic normality of ML estimators

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

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

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

Examples

Summary: basic (point) estimator properties

- bias
- variance
- MSE
- efficiency

- consistency
- asymptotic normality
- asymptotic efficiency

Interval estimation – confidence intervals

- We do not provide a single value estimate, but rather a lower and an upper bound for the estimate (the true value will fit into these bounds with given probability)
- We estimate with given precision

Confidence interval

Let $g(\theta)$ be a function of unknown parameter θ, and let $\overline{g} = \overline{g}(X_1, X_2, \ldots, X_n)$ and $\underline{g} = \underline{g}(X_1, X_2, \ldots, X_n)$ be statistics

Then, $[\underline{g}, \overline{g}]$ is a **confidence interval** for $g(\theta)$ with a confidence level 1-α, if for any θ

$$P_\theta\left(\underline{g}(X_1, X_2, \ldots, X_n) \leq g(\theta) \leq \overline{g}(X_1, X_2, \ldots, X_n)\right) \geq 1 - \alpha$$

Confidence intervals – use and interpretation

- Typically: α is a small number, for example 1-α = 0,95 or 1-α = 0,99
- The condition from the definition means: the random interval $[\underline{g},\overline{g}]$ includes the unknown value $g(\theta)$ with given (high) probability.
- If we calculate the *realization* of the confidence interval (e.g. $\underline{g}=1, \overline{g}=3$) then we CAN'T say that the unknown parameter is included in the range with probability 1-α anymore!

the parameter is either in the interval or not – the event is not random, it is just something we don't know.

Confidence intervals – construction

- The confidence interval depends on the underlying probability distribution
- Usually, normal samples are considered (the distribution most frequently observed in nature)

Confidence intervals – construction cont.

- Convenient method: we look for random variables which depend on sample data and parameter values, but whose *distributions* do not depend on unknown parameters (pivotal method)
- If $U = U(X_1, X_2, ..., X_n, \theta)$ is such a function, then we look for confidence intervals [*a*,*b*] such that

$$P_\theta\left(a \le U \le b\right) \ge 1 - \alpha$$

- Usually we look for „symmetric” CI

$$P_\theta\left(U < a\right) \le \frac{\alpha}{2}, \qquad P_\theta\left(U > b\right) \le \frac{\alpha}{2}$$