Anna Janicka

Mathematical Statistics 2020/2021
Lecture 6

1. Asymptotic Properties of Estimators – cont.

During the previous lecture, we introduced the distinction between small and large sample behavior of estimators, and studied two asymptotic properties: unbiasedness and consistency. Here, we will continue the topic with two more concepts: asymptotic normality and asymptotic efficiency.

1.1. Asymptotic normality.

Definition 1. *We will say that $\hat{g}$ is an* **asymptotically normal** *estimator of $g(\theta)$, if for any $\theta \in \Theta$ there exists $\sigma^2(\theta)$ such that, when $n \to \infty$*

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

The notation $\xrightarrow{D}$ signifies convergence in distribution: for any $a \in \mathbb{R}$ we have

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

where $\Phi(\cdot)$ denotes the CDF of the standard normal distribution.

In other words, an estimator is asymptotically normal if the distribution of $\hat{g}(X_1, X_2, \ldots, X_n)$ is for large n similar to $\mathcal{N}\left(g(\theta), \frac{\sigma^2(\theta)}{n}\right)$.

The notion of asymptotic normality is stronger than consistency, i.e. an estimator which is asymptotically normal is consistent (although not necessarily strongly consistent). Note that although the definition of asymptotic normality includes a concept which is similar to unbiasedness – i.e., the expected value of the asymptotic normal distribution which fits the distribution of the estimator is equal to $g(\theta)$ – this does not necessarily imply that the estimator itself is unbiased (even asymptotically, although in the latter case counterexamples may be thought of as "pathological" and are not frequently encountered in practice). A similar case holds for the variance: although for large samples, the estimator distribution may be approximated with the use of a normal distribution with a variance equal to $\frac{\sigma^2(\theta)}{n}$, this does not imply that the variance of the estimator is equal to this value (even in the limit). However, the value of $\frac{\sigma^2(\theta)}{n}$ is called the **asymptotic variance**.[1]

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, we see that $\bar{X}$ is an asymptotically normal estimator of the mean:

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

In this case, the asymptotic variance $\frac{\sigma^2}{n}$ is exactly equal to the estimator variance.

In many cases, the proof that an estimator is asymptotically normal may be conducted with the use of the following handy theorem:

Theorem 1. Delta Method. *Let T_n be a sequence of random variables such that for $n \to \infty$ we have*

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

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

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

[1] Some authors define the asymptotic variance as $\sigma^2(\theta)$.

In the above formulation, μ and σ are in fact functions of the parameter θ, governing the probability distribution. This theorem, which allows us to say when a function of an estimator which is asymptotically normal is also asymptotically normal, proves extremely useful especially in cases when studying estimators which are functions of statistics T_n, which can be easily shown co converge on the base of CLT.

Example: Let $X_1, X_2, \ldots, X_n, \ldots$ be an IID sample from an exponential distribution with parameter $\lambda > 0$. We have seen that the $MLE(\lambda) = 1/\bar{X}$. Finding the distribution of $1/\bar{X}$ in order to study the properties of the MLE estimator is possible, but can be avoided. From the CLT, we have that the average is an asymptotically normal estimator of the inverse of λ:

$$\sqrt{n}(\bar{X} - \frac{1}{\lambda}) \xrightarrow{D} \mathcal{N}(0, \frac{1}{\lambda^2}).$$

Therefore, using the Delta Method for $h(x) = \frac{1}{x}$ we get

$$\sqrt{n}(\frac{1}{\bar{X}} - \lambda) \xrightarrow{D} \mathcal{N}(0, \frac{1}{\lambda^2}(-\frac{1}{\frac{1}{\lambda^2}})^2),$$

that is: the MLE estimator is asymptotically normal, and the asymptotic variance is equal to $\frac{\lambda^2}{n}$.

Asymptotic normality is a welcome property of estimators. If an estimator is asymptotically normal, then for large samples the distribution of this estimator is approximately normal, meaning that for further calculations which we would like to perform (for example: hypothesis testing) we may use the normal distribution instead of the exact distribution. Asymptotic normality of an estimator also makes another asymptotic property – efficiency – well defined.

1.2. **Asymptotic efficiency.** For an asymptotically normal estimator $\hat{g}$ of the value $g(\theta)$ (i.e. an estimator for which we can calculate the asymptotic variance), we can introduce the following concept:

Definition 2. *We define* **asymptotic efficiency** *as*

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

where $\frac{\sigma^2(\theta)}{n}$ *is the asymptotic variance.*

Note that this is a modification of the definition of efficiency to the limit case, where the variance of the estimator is substituted with the asymptotic variance. Note also that if the sample is IID, we have

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

Just as in the finite sample case, also relative efficiency of estimators may be defined:

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

allowing us to compare asymptotic efficiency of two estimators of the same value.

In general, asymptotic efficiency is a welcome property of an estimator. There are cases, however, where a less asymptotically efficient estimator may be preferred to a more efficient one – for example, if small sample properties of the estimator are better, or if the more efficient estimator uses such properties of the distribution, which the researcher is unsure of (cf. Problem 1 from Problem set 7&8).

Examples: For symmetric distributions, the mean coincides with the median. Assume that we are interested in estimating the center of the distribution. Should we use the sample average or the sample median to do it? In some cases, the response is straightforward: if the distribution does not have a mean (as is the case for the Cauchy distribution for example), the sample average will not be a consistent estimator of the center of the distribution (it will not converge). In other cases, some insight may be gained with an analysis of asymptotic efficiency.

Before we look at examples of two distributions, let us formulate a theorem describing the asymptotic properties of the sample median:

Theorem 2. *Let $X_1, X_2, \ldots, X_n, \ldots$ be an IID sample from a continuous distribution with density $f(x)$, such that the density is continuous and different from 0 for the median m. Then, the sample median is an asymptotically normal estimator of the median m, and*

$$\sqrt{n}(\hat{med} - m) \xrightarrow{D} \mathcal{N}(0, \frac{1}{4(f(m))^2}).$$

In consequence:

(1) For a normal distribution $\mathcal{N}(\mu, \sigma^2)$, we have that:

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

and

$$\sqrt{n}(\hat{med} - \mu) \xrightarrow{D} \mathcal{N}(0, \frac{\sigma^2 \pi}{2}),$$

so the relative asymptotic efficiency for $\bar{X}$ and $\hat{med}$ estimators of μ is equal to

$$\text{as.ef}(\bar{X}, \hat{med}) = \frac{\frac{\sigma^2 \pi}{2}}{\sigma^2} = \frac{\pi}{2} > 1,$$

which means that the average is more asymptotically efficient. Meanwhile,

(2) For a Laplace distribution (with density $f(x) = \frac{\lambda}{2} e^{-\lambda|x-\mu|}$), we have that

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

and

$$\sqrt{n}(\hat{med} - \mu) \xrightarrow{D} \mathcal{N}(0, \frac{1}{\lambda^2}),$$

so the relative asymptotic efficiency for $\bar{X}$ and $\hat{med}$ estimators of μ is equal to

$$\text{as.ef}(\bar{X}, \hat{med}) = \frac{\frac{1}{\lambda^2}}{\frac{2}{\lambda^2}} = \frac{1}{2} < 1,$$

which means that the median is more asymptotically efficient.

2. Asymptotic properties of ML estimators

We will conclude our considerations of asymptotic properties of estimators by formulating a couple of theorems describing the asymptotic properties of MLE estimators.

Theorem 3. *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*

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

then $\hat{\theta}$ is the m.l.e. of θ and it is consistent.

Theorem 4. *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 the m.l.e. $\hat{\theta}$ is consistent (for example, the distribution fulfills the assumptions of the previous theorem), and*

(1) *$\frac{d^2}{d\theta^2} \ln L(\theta)$ exists;*
(2) *Fisher information may be calculated, and $0 < I_1(\theta) < \infty$;*
(3) *the order of integration with respect to x and derivation with respect to θ may be reversed,*

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

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

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

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

As a consequence, comparing the asymptotic variances in the theorem above with the expression in the definition of asymptotic efficiency, we get that

Theorem 5. *Let $X_1, X_2, \ldots, X_n, \ldots$ be a sample from a distribution with density $f_\theta(x)$. If the regularity conditions from the previous theorems are fulfilled, then the m.l. estimators are asymptotically efficient.*

Therefore, if certain regularity conditions are fulfilled, we have that the m.l. estimators of θ or $g(\theta)$ are: consistent, asymptotically normal and asymptotically efficient. This is why the maximum likelihood estimation technique is, in most cases, the method of choice for parameter estimation. Even though, as we have seen, the m.l. estimators need not be unbiased.