STK4011 and STK9011 Autumn 2016

Point estimation - asymptotics

Covers (most of) the material from sections 10.1.1, 10.1.2 and 10.1.3

Ørnulf Borgan
Department of Mathematics
University of Oslo

Consistency

Let $W_1, W_2, ...$ be sequence of estimators for θ

A typical situation is that we have a sequence $X_1, X_2, ...$ of iid random variables with pdf or pmf $f(x|\theta)$ and define $W_n = W_n(X_1, ..., X_n)$

For example we may have $W_n = \bar{X}_n = \sum_{i=1}^n X_i / n$

Definition 10.1.1

A sequence of estimators $W_1, W_2, ...$ is a consistent sequence of estimators of the parameter θ if, for every $\varepsilon > 0$ and every $\theta \in \Theta$, we have

$$\lim_{n\to\infty} P_\theta(|W_n - \theta| < \varepsilon) = 1$$

For short we say that W_n is consistent

We remember that the mean squared error (MSE) of an estimator W_n of θ is given as $\mathrm{E}(W_n - \theta)^2$

Further the bias of W_n is given by $\mathrm{Bias}_\theta W_n = \mathrm{E}_\theta W_n - \theta$

We have $\mathrm{E}_\theta(W_n - \theta)^2 = \mathrm{Var}_\theta W_n + \left(\mathrm{Bias}_\theta W_n\right)^2$

Theorem 10.1.3

If a sequence of estimators $W_1, W_2, ...$ of the parameter θ satisfy

i) $\lim_{n\to\infty} \mathrm{Var}_\theta W_n = 0$

ii) $\lim_{n\to\infty} \mathrm{Bias}_\theta W_n = 0$

for every $\theta \in \Theta$, then W_n is a consistent estimator for θ

Limiting and asymptotic variance

Consider a sequence of estimators $W_1, W_2, ...$ for θ

Consistency is a very weak property

We also need to look at how the variance (and the bias) behaves as n increases

For all sensible estimators $\mathrm{Var} W_n \to 0$ so we may instead consider $k_n \mathrm{Var} W_n$ for a normalizing sequence of constants k_n (often $k_n = n$)

Definition 10.1.7

For a sequence of estimators $W_1, W_2, ...$, if $\lim_{n\to\infty} k_n \mathrm{Var} W_n = \tau^2 < \infty$, then τ^2 is called the limiting variance

Example 10.1.8
Let $X_1, X_2, ...$ be iid and $n(\mu, \sigma^2)$ distributed and consider $\bar{X}_n$

$n\mathrm{Var}\bar{\mathrm{X}}_n = \sigma^2$ for all n, so the limiting variance is σ^2

Now consider $W_n = 1/\bar{X}_n$

Then $\mathrm{Var}W_n = \infty$ for all n, so the limiting variance does not exist

But by the approximation

$$W_n = 1/\bar{X}_n \approx 1/\mu - (1/\mu^2)(\bar{X}_n - \mu)$$

we obtain

$$\mathrm{E}W_n \approx \frac{1}{\mu} \quad \text{and} \quad \mathrm{Var}W_n \approx \frac{\sigma^2}{n\mu^4}$$

The example illustrates that we need another concept than the limiting variance

Definition 10.1.9
For a sequence of estimators $W_1, W_2, ...$, suppose that $k_n(W_n - \tau(\theta)) \to n(0, \sigma^2)$ in distribution, then σ^2 is called the asymptotic variance or the variance of the limiting distribution of W_n

Example 10.1.8 (continued)
For the situation of example 10.1.8, the delta method gives that (cf. example 5.5.25):

$$\sqrt{n}\left(\frac{1}{\bar{X}_n} - \frac{1}{\mu}\right) \to n\left(0, \frac{\sigma^2}{\mu^4}\right)$$

Thus $W_n = 1/\bar{X}_n$ has asymptotic variance σ^2/μ^4

Asymptotical efficient estimators

Let $X_1, X_2, ...$ be a sequence of iid random variables with pdf or pmf $f(x|\theta)$ and let W_n be an estimator for $\tau(\theta)$ based on $X_1, ..., X_n$

We assume that expressions of the form

$$\int_{-\infty}^{\infty} W_n(x) f(x|\theta) dx$$

may be differentiated with respect to θ by changing the order of differentiation and integration (summation in the case of pmf)

Remember that if W_n is an unbiased estimator for $\tau(\theta)$, we have by the Cramér-Rao inequality (Corollary 7.3.10) that

$$\mathrm{Var}_\theta W_n \geq \frac{[\tau'(\theta)]^2}{nI_1(\theta)}$$

where

$$I_1(\theta) = \mathrm{E}_\theta\left[\left(\frac{\partial}{\partial\theta}\log f(X|\theta)\right)^2\right]$$

is the information in one observation

Assume now that we have

$$\sqrt{n}\left(W_n - \tau(\theta)\right) \to n(0, v(\theta))$$

Then we may use the approximation

$$\mathrm{Var}\left(\sqrt{n}W_n\right) \approx v(\theta)$$

to obtain

$$\mathrm{Var}\left(W_n\right) \approx \frac{v(\theta)}{n}$$

Therefore if $\nu(\theta) = \left[\tau'(\theta)\right]^2 / I_1(\theta)$, we have

$$\mathrm{Var} W_n \approx \frac{\left[\tau'(\theta)\right]^2}{n I_1(\theta)}$$

and W_n will asymptotically achieve the Cramér-Rao lower bound

This motivates the following definition:

Definition 10.1.11

Let $X_1, X_2, \ldots$ be a sequence of iid random variables with pdf or pmf $f(x \mid \theta)$ and let W_n be an estimator for $\tau(\theta)$ based on $X_1, \ldots, X_n$

Then W_n is asymptotically efficient if

$$\sqrt{n}\left(W_n - \tau(\theta)\right) \to n(0, v(\theta))$$

in distribution, where

$$v(\theta) = \frac{\left[\tau'(\theta)\right]^2}{I_1(\theta)}$$

Theorem 10.1.13

Let $X_1, X_2, \ldots$ be a sequence of iid random variables with pdf or pmf $f(x \mid \theta)$, let $\hat{\theta} = \hat{\theta}_n$ be the maximum likelihood estimator based on $X_1, \ldots, X_n$, and let $\tau(\theta)$ be a differentiable function of θ. Then under «certain regularity conditions» (cf. page 516) we have that

$$\sqrt{n}\left(\tau(\hat{\theta}) - \tau(\theta)\right) \to n(0, v(\theta))$$

where

$$v(\theta) = \left[\tau'(\theta)\right]^2 \big/ I_1(\theta)$$

The most important regularity condition is that one may change the order of differentiation (with respect to θ) and integration/summation (as for the Cramér-Rao inequality)

For the maximum likelihood estimator $\tau(\hat{\theta})$ we have

$$\sqrt{n}\left(\tau(\hat{\theta}) - \tau(\theta)\right) \to n\left(0, \left[\tau'(\theta)\right]^2 \big/ I_1(\theta)\right)$$

Hence we may use the approximation

$$\mathrm{Var}\,\tau(\hat{\theta}) \approx \frac{\left[\tau'(\theta)\right]^2}{n I_1(\theta)}$$

This may be estimated using expected information

$$\widehat{\mathrm{Var}}\,\tau(\hat{\theta}) = \frac{\left[\tau'(\theta)\right]^2\Big|_{\theta=\hat{\theta}}}{n I_1(\theta)\Big|_{\theta=\hat{\theta}}}$$

or observed information

$$\widehat{\mathrm{Var}}\,\tau(\hat{\theta}) = \frac{\left[\tau'(\theta)\right]^2\Big|_{\theta=\hat{\theta}}}{-\sum_{i=1}^{n}\left(\partial^2 \big/ \partial\theta^2\right)\log f(X_i \mid \theta)\Big|_{\theta=\hat{\theta}}}$$

Example 10.1.14

$X_1, X_2, \ldots.$ are iid Bernoulli random variables with success probability p

ML estimator $\hat{p} = \sum_{i=1}^{n} X_i / n$

Consider estimation of the odds $\tau(p) = p/(1-p)$

We have $\tau'(p) = 1/(1-p)^2$ and $I_1(p) = 1/\left[p(1-p)\right]$

Therefore: $\sqrt{n}\left(\frac{\hat{p}}{1-\hat{p}} - \frac{p}{1-p}\right) \to n\left(0\ ,\ \frac{p}{(1-p)^3}\right)$

Both methods for estimating the variance give

$$\widehat{\mathrm{Var}}\left(\frac{\hat{p}}{1-\hat{p}}\right) = \frac{\hat{p}}{n(1-\hat{p})^3}$$

The following definition is useful for comparing the large sample properties of two estimators:

Definition 10.1.16

If two estimators W_n and V_n are satisfying

$$\sqrt{n}\left(W_n - \tau(\theta)\right) \to n(0, \sigma_W^2)$$

$$\sqrt{n}\left(V_n - \tau(\theta)\right) \to n(0, \sigma_V^2)$$

in distribution, then the asymptotic relative efficiency (ARE) of V_n with respect to W_n is

$$\mathrm{ARE}\left(V_n, W_n\right) = \frac{\sigma_W^2}{\sigma_V^2}$$

Example 10.1.17

$X_1, X_2, \ldots.$ are iid Poisson(λ) random variables

We want to estimate

$$\tau(\lambda) = P(X = 0) = e^{-\lambda}$$

We consider the estimators

$$e^{-\hat{\lambda}} \quad \text{where } \hat{\lambda} = \sum_{i=1}^{n} X_i / n$$

$$\hat{\tau} = \frac{1}{n}\sum_{i=1}^{n} I\{X_i = 0\}$$

We find

$$\mathrm{ARE}(\hat{\tau}, e^{-\hat{\lambda}}) = \frac{\lambda e^{-2\lambda}}{e^{-\lambda}\left(1-e^{-\lambda}\right)} = \frac{\lambda}{e^{\lambda} - 1}$$