stats6257asymp 625.7 Asymptotics

Gunnar Stefansson

December 1, 2015

Copyright This work is licensed under the Creative Commons Attribution-ShareAlike License. To view a copy of this license, visit http://creativecommons.org/licenses/by-sa/1.0/ or send a letter to Creative Commons, 559 Nathan Abbott Way, Stanford, California 94305, USA.

Contents

1 asymptotics

1.1 asymptotics

1.1.1 Handout

Asymptotics Suppose $X_1, X_2, \ldots$ are a sequence of random variables (usually i.i.d.) with density $f_\theta, \theta \in \Theta$ and $W_n = W_n(X_1, \ldots, X_n)$ is a sequence of estimators.
W_n is **consistent** if

$$W_n \xrightarrow{P_\theta} \theta, \theta \in \Theta$$

i.e. $P_\theta[|W_n - \theta| > \varepsilon] \xrightarrow{n \to \infty} 0, \forall \theta \in \Theta$. (Usually, $\Theta \subset \mathbb{R}$ or $\Theta \subset \mathbb{R}^k$ and $\tau : \Theta \to \mathbb{R}$ continuous.)

Example $X_1, X_2, \ldots \sim U(0, \theta)$ i.i.d. then with $W_n := max\{X_1, \ldots, X_n\}$ we know that $W_n \xrightarrow{P_\theta} \theta$ and we can easily show this by looking at the cdf $F_{(n)}$ for $X_{(n)} = W_n$ and note that

$$F_{(n)}(x) \to \begin{cases} 0, & \text{if } x < \theta \\ 1, & \text{else} \end{cases}$$

So $W_n \xrightarrow{D} \theta$ and since this is convergence to a constant, we also have $W_n \xrightarrow{P_\theta} \theta$.

From Chebyshev's theorem we know that if $EW_n = \theta$ and $VW_n \to 0$then $W_n \xrightarrow{P_\theta} \theta$.

Summary

Theorem If W_n is consistent for θ and $\{a_n\}, \{b_n\}$ are sequences such that $a_n \to 1, b_n \to 0$, then $a_n W_n + b_n$ is also consistent for θ.

Theorem MLEs are consistent under very general conditions (see below).
More formally: If $X_1, X_2, \ldots \sim f_\theta$ i.i.d., $\theta \in \Theta \subset \mathbb{R}$

$$L_{\mathbf{x}}(\theta) := \prod_{i=1}^{n} f_\theta(x_i), \hat{\theta} := argmax_{\theta \in \Theta} L_{\mathbf{x}}(\theta)$$

and $\tau : \Theta \to \mathbb{R}$ is continuous, then $\tau(\hat{\theta}) \xrightarrow{P_\theta} \tau(\theta)$.

Conditions

- A1 $X_1, X_2, \ldots$ are i.i.d., $X_i \sim f_\theta$
- A2 $f_\theta \neq f_{\theta'}$ if $\theta \neq \theta'$
- A3 $\theta \mapsto f_\theta$ is differentiable and $\{f_\theta\}, \theta \in \Theta$ have common support
- A4 Θ is an open set (all θ are interior points)
- A5 $x \mapsto f_\theta(x)$ is three-times differentiable with respect to θ and $\theta \mapsto \int f_\theta(x)dx$ can be differentiated under the integral sign
- A6 For $\theta_0 \in \Theta \exists c > 0$ and a function $M : \Omega \mapsto \mathbb{R}$ such that $|\frac{\delta^3}{\delta\theta^3} ln f_\theta(x)| \leq M(x)$ for $x \in \Omega$ and $\theta_0 - c \leq \theta \leq \theta_0 + c$ with $E[M(X_1)] < \infty$

Note These do not hold e.g. for $U(0,\theta)$ etc., but do hold for distributions such as the normal, gamma, Poisson, binomial, etc.

Efficiency Efficiency of an estimator is closely related to consistency. Where consistency has to do with the question: Does the estimator converge to the parameter it is estimating? Efficiency is concerned with the asymptotic variance of an estimator. (Note: In the following we define $VX = V(X) := \mathrm{Var}(X)$).

Definition If $nVT_n \mapsto \sigma^2$ then σ^2 is the **limiting variance** of T_n.

Example $V\bar{X}_n = \frac{\sigma^2}{n}$ and $nV\bar{X}_n = \sigma^2$ and we are e.g. interested in $\sqrt{n}\bar{X}$.

Example Consider $X_1, X_2, .. \sim n(\mu, \sigma^2)$ and define $Y_n := \bar{X}_n$ for any given n. Then $(Y_n)_{n\geq 1}$ is a sequence of estimators. We have seen that $EY_n = \mu$ and $VY_n = \frac{\sigma^2}{n}$. So the limiting variance of Y_n is $\lim_{n\to\infty} nV\bar{X}_n = \sigma^2$. We also note that $\sqrt{n}(Y_n - \mu) \xrightarrow{D} n(0,\sigma^2)$ $(*)$.

We are interested in estimating $\frac{1}{\mu}$ by using $\frac{1}{X_n}$. We let $g(t) = \frac{1}{t}$ and then $g(Y_n) = \frac{1}{Y_n} = \frac{1}{X_n}$. By carrying out straightforward calculations we arrive at the following conclusion. For any given n, $Eg(Y_n) = \infty$ and $Vg(Y_n) = \infty$ and thus the limiting variance of $g(Y_n)$ is ∞.

Let's assume that $g'(\mu)$ exists and is not zero then we can use the delta method to estimate the variance as $n \to \infty$. Since $(*)$ it follows from the delta method that $\sqrt{n}(g(Y_n) - g(\mu)) \xrightarrow{D} n(0, \sigma^2 (g'(\mu))^2)$. Here we have an estimate for the (limit) variance of $g(Y_n)$ which is finite i.e. $Vg(Y_n) = \sigma^2 (g'(\mu))^2 < \infty$. We refer to $\sigma^2 (g'(\mu))^2$ as the **asymptotic variance**.

Definition If $W_n = W_n(X_1, ...X_n)$ where $\{X_i\}_{i=1}^{\infty}$ are i.i.d. f_θ and $\sqrt{n}(W_n - \tau(\theta)) \xrightarrow{D} n(0,\sigma^2)$, then σ^2 is the **asymptotic variance** of W_n. (Note: σ^2 may be a function of θ).

Definition With W_n as above, W_n is **asymptotically efficient** if $\sqrt{n}(W_n - \tau(\theta)) \xrightarrow{D} n(0, v(\theta))$ with $v(\theta) = \frac{(\tau'(\theta))^2}{E_\theta[(\frac{\delta}{\delta\theta} ln f_\theta(X))^2]}$ (Cramer-Rao lower bound).

1.2 are

1.2.1 Handout

Notes 4-6

4 *20131108_115613.jpg*
MLEs are asymptotically efficient:
<u>Theorem:</u>
Under regularity conditions $A1 - A6$, with $X_1, X_2, ... \sim f_\theta$ iid, $\hat{\theta} := \arg\max_{\theta\in\Theta} L_{\mathbf{x}_n}(\theta)$, where $L_{\mathbf{x}_n} = \prod_{i=1}^{n} f_\theta(x_i)$ and $\tau : \Theta \to \mathbf{R}$ continuous, $\sqrt{n}(\tau(\hat{\theta} - \tau(\theta)) \xrightarrow{D} n(0, r(\theta))$ with $r(\theta)$ given as CRLB.
"<u>Proof</u>":
Write the log-likelihood as

$$l_{\mathbf{x}_n}(\theta) := \sum_{i=1}^{n} ln f_\theta(x_i)$$

and write the Taylor expansion of $l'_{\mathbf{x}_n}$ as

$$l'_{\mathbf{x}_n}(\theta) = l'_{\mathbf{x}_n}(\theta_0) + (\theta - \theta_0) l''_{\mathbf{x}_n}(\theta_0) + R.$$

Since the MLE also maximizes $l_{\mathbf{x}_n}$, it satisfies $l'_{\mathbf{x}_n}(\hat{\theta}) = 0$ and we obtain

$$0 = l'_{\mathbf{x}_n}(\hat{\theta}) = l'_{\mathbf{x}_n}(\theta_0) + (\hat{\theta} - \theta_0) l''_{\mathbf{x}}(\theta_0) + R$$

or

$$\hat{\theta} - \theta_0 = \frac{-l'_{\mathbf{x}_n}(\theta_0)}{l''_{\mathbf{x}_n}(\theta_0)} + \tilde{R}$$

$$\Rightarrow \sqrt{n}(\hat{\theta} - \theta_0) = \frac{\frac{1}{\sqrt{n}} l'_{\mathbf{x}_n}(\theta_0)}{-\frac{1}{n} l''_{\mathbf{x}_n}(\theta_0)} + \tilde{R}$$

and we note that

5 $$-\frac{1}{n} l''_{\mathbf{X}_n}(\theta_0) = -\frac{1}{n} \sum_{i=1}^{n} \frac{\partial^2}{\partial \theta^2} ln f_\theta(X_i) \xrightarrow{P_\theta} -E\left[\frac{\partial^2}{\partial \theta^2} ln f_\theta(X_i)\right] = I(\theta)$$

(a) $$\frac{1}{\sqrt{n}} l'_{\mathbf{X}_n}(\theta_0) = \sqrt{n}\left(\frac{1}{n}\sum_{1}^{n} \frac{d}{d\theta} ln f_{\theta_0}(X_i)\right) \xrightarrow{D} n(0, I(\theta)).$$

and hence

$$\sqrt{n}(\hat{\theta} - \theta) \xrightarrow{D} n(0, \tfrac{1}{I(\theta)}).$$

<u>Example</u>:The above theorem shows that it is typically the case that MLE's are efficient and consistent. We want to note that this phrase is somewhat redundant, as efficiency is defined only when the estimator is asymptotically normal and, as we will illustrate, asymptotic normality implies consistency. Suppose that

$$\sqrt{n}\frac{W_n - \mu}{\sigma} \sim Z \text{ in distribution,}$$

where $Z \sim n(0,1)$. By applying Slutsky's Theorem (Theorem 5.5.17) we conclude

$$W_n - \mu = \left(\frac{\sigma}{\sqrt{n}}\right)\left(\sqrt{n}\frac{W_n - \mu}{\sigma}\right) \to lim_{n\to\infty}\left(\frac{\sigma}{\sqrt{n}}\right) Z = 0$$

so $W_n - \mu \to 0$ in distribution. From Theorem 5.5.13 we know that convergence in distribution to a point is equivalent to convergence in probability, so W_n is consistent estimator of μ

1 *20131115_093137.jpg* :

<u>Estimating variances</u>

Recall that if $\sqrt{n}(Y_n - \mu) \xrightarrow{D} n(0, \sigma^2)$ then $\sqrt{n}(g(Y_n) - g(\mu)) \xrightarrow{D} n(0, \sigma^2(g'(\theta))^2)$ [Delta method], and if $\hat{\theta}$ is the MLE for θ, so $\tau(\hat{\theta})$ is the MLE for $\tau(\theta)$, we have:

$$\sqrt{n}(\tau(\hat{\theta}) - \tau(\theta)) \xrightarrow{D} n\left(0, \frac{(\tau'(\theta))^2}{E\left[-\frac{\partial^2}{\partial\theta^2} lnL_{\mathbf{X}}(\theta)\right]}\right)$$

since $I_n(\theta) := E\left[\left(\frac{\partial}{\partial\theta} lnL_{\mathbf{X}}(\theta)\right)^2\right] = E\left[-\frac{\partial^2}{\partial\theta^2} lnL_{\mathbf{X}}(\theta)\right]$ [the information number of sample].

It follows that

$$V[\tau(\hat{\theta})] \simeq \frac{[\tau'(\theta)]^2}{I_n(\theta)} \simeq \frac{[\tau'(\hat{\theta})]^2}{-\frac{\partial^2}{\partial\theta^2} lnL_{\mathbf{x}}(\theta)}$$

(here "$\simeq$" means asymptotic).

Note that we can write $I_n(\hat{\theta}) = -\frac{\partial^2}{\partial\theta^2} lnL_{\mathbf{x}}(\hat{\theta})$ for the observed information number.

NB: For any finite sample we have

$$V[\tau(\hat{\theta})] \geq \frac{[\tau'(\theta)]^2}{I_n(\theta)}$$

so this underestimates the actual variance!

6 *20131115_093148.jpg* :

Definition: Let W_n, V_n be sequences of random variables such that

$$\sqrt{n}(W_n - \tau(\theta)) \xrightarrow{D} n(0, \sigma_w^2)$$
$$\sqrt{n}(V_n - \tau(\theta)) \xrightarrow{D} n(0, \sigma_v^2)$$

then the asymptotic relative efficiency of V_n with respect to W_n is

$$ARE(V_n, W_n) = \frac{\sigma_w^2}{\sigma_v^2}$$

Interpretation: If you need a sample size n to satisfy some "large sample" quantity criteria with W_n, then you need a sample size m s.t. $\frac{\sigma_v^2}{m} = \frac{\sigma_w^2}{n}$ for the same result with V_n, i.e. you need $m = n\frac{\sigma_v^2}{\sigma_w^2}$. Equivalently, a "large sample" confidence interval becomes longer/shorter in proportion to $\sqrt{ARE}$.

Example: Let $X_1, X_2, \ldots \sim P(\lambda)$, i.i.d. We want to estimate $P[X_1 = 0] = e^{-\lambda} =: \tau(\lambda)$. Consider two estimators:

$$\hat{\tau}_1 := \frac{1}{n}\sum_{i=1}^{n} I_{[X_i=0]} \sim b(e^{-\lambda}, 1)$$
$$\hat{\tau}_2 := e^{-\hat{\lambda}} = \tau(\hat{\lambda})$$

where $\hat{\lambda} = \frac{1}{n}\sum_{i=1}^{n} X_i$ is the MLE.

Note that $\hat{\tau}_1$ is unbiased but though $\hat{\tau}_2$ is biased, it is consistent and asymptotically efficient. We know that

$$E[\hat{\tau}_1] = e^{-\lambda}$$
$$V[\hat{\tau}_1] = \frac{1}{n} e^{-\lambda}(1 - e^{-\lambda})$$

and we know that

$$\sqrt{n}(\hat{\tau}_2 - \tau(\lambda)) \xrightarrow{D} n(0, \lambda(\tau'(\lambda))^2) = n(0, \lambda e^{-2\lambda})$$

and

$$\sqrt{n}(\hat{\tau}_1 - \tau(\lambda)) \xrightarrow{D} n(0, e^{-\lambda}(1 - e^{-\lambda}))$$

so

$$ARE(\hat{\tau}_1, \tau(\hat{\lambda})) = \frac{\lambda e^{-2\lambda}}{e^{-\lambda}(1 - e^{-\lambda})} = \frac{\lambda}{e^{\lambda} - 1}$$

i.e. $\hat{\tau}_2$ beats $\hat{\tau}_1$ for any $\lambda > 0$.

1.3 are etc

1.3.1 Handout

A note on robustness - the median

Suppose we have a sample, or sequence $X_1,\ldots,X_n \sim f$, where f is a continuous pdf with corresponding cdf F and population median μ, i.e. $F(\mu)=1/2$ and $F'=f$.

Suppose n is odd and consider the first n ordered values of the sample median, i.e.

$$M_n := \tilde{X}_n := median\{X_i\}_{i=1,\ldots,n} = X_{(n+1)/2:n}$$

where $X_{1:n} \leq \ldots \leq X_{n:n}$.

Consider the task of evaluating $\lim_{n\to\infty} P[\sqrt{n}(M_n-\mu)\leq a]$, i.e. finding the limiting distribution of M_n. First note that $\sqrt{n}(M_n-\mu)\leq a \Leftrightarrow M_n \leq \mu+a/\sqrt{n} \Leftrightarrow$ at least half of the $X's$ are $\leq \mu+a/\sqrt{n}$. So let

$$Y_i = \begin{cases} 1 & \text{for } X_i \leq \mu+a/\sqrt{n} \\ 0 & \text{else} \end{cases}$$

to obtain $Y_i \sim b(F(\mu+a/\sqrt{n}),1)$ and $\sum Y_i \sim b(p_n := F(\mu+a/\sqrt{n}),n)$ and $\sqrt{n}(M_n-\mu)\leq a \Leftrightarrow \sum Y_i \geq \frac{n+1}{2}$. So Y_i is a Binomial (or Bernoulli) r.v. with success probability $p_n = F\left(\mu+\frac{a}{\sqrt{n}}\right)$.

Doing some algebra we get

$$P[\sqrt{n}(M_n-\mu)\leq a] = P[\sum Y_i \geq \frac{n+1}{2}] = P\left[\frac{\sum Y_i - np_n}{\sqrt{np_n(1-p_n)}} \geq \frac{\frac{n+1}{2}-np_n}{\sqrt{np_n(1-p_n)}}\right]. \quad (1)$$

Since $\sum Y_i$ is Binomial its e.v. and variance are $EY_i = np_n$ and $VY_i = np_n(1-p_n)$. Looking at the limit of p_n we see that $\lim_{n\to\infty} p_n = \lim_{n\to\infty} F(\mu+\frac{a}{\sqrt{n}}) = F(\mu) = \frac{1}{2}$. From this we infer that $\frac{\sum Y_i - np_n}{\sqrt{np_n(1-p_n)}} \to Z$ (standard normal) by the CLT.

We would like to evaluate the right hand side in the last P in (1) so we carry out the calculations

$$\begin{aligned}
\lim_{n\to\infty}\frac{\frac{n+1}{2}-np_n}{\sqrt{np_n(1-p_n)}} &= \lim_{n\to\infty}\frac{n(\frac{1}{2}-p_n)+\frac{1}{2}}{\sqrt{np_n(1-p_n)}}\\
&= \lim_{n\to\infty}\frac{\sqrt{n}(\frac{1}{2}-p_n)}{\sqrt{p_n(1-p_n)}}+\underbrace{\lim_{n\to\infty}\frac{1}{\sqrt{np_n(1-p_n)}}}_{=0}\\
&= \lim_{n\to\infty}\frac{-\left(p_n-\frac{1}{2}\right)}{\sqrt{p_n(1-p_n)}/\sqrt{n}}\\
&= \lim_{n\to\infty}\frac{1}{\sqrt{p_n(1-p_n)}}\cdot\frac{-\left(F\left(\mu+\frac{a}{\sqrt{n}}\right)-F(\mu)\right)}{1/\sqrt{n}}\\
&= \frac{1}{1/2}\cdot\lim_{h_n\to 0}\frac{-\left(F\left(\mu+ah_n\right)-F(\mu)\right)}{h_n}\ ,\quad \left(h_n:=\frac{1}{\sqrt{n}}\right)\\
&= 2(-aF'(\mu))\\
&= -2af(\mu).
\end{aligned}$$

We conclude

$$P[\sqrt{n}(M_n-\mu)\le a]\to P[Z\ge -2af(\mu)]=P\left[\frac{-Z}{2f(\mu)}\le a\right]$$

and $\frac{-Z}{2f(\mu)}\sim n\left(0,\frac{1}{[2f(\mu)]^2}\right)$. We therefore have shown

$$\sqrt{n}(M_n-\mu)\ \xrightarrow{\mathcal{D}}\ n\left(0,\frac{1}{[2f(\mu)]^2}\right).$$

Recall that if $Var[X_i]=\sigma^2$ and $E[X_i]=\mu$, then $\sqrt{n}(\bar{X}_n-\mu)\ \xrightarrow{\mathcal{D}}\ n(0,\sigma^2)$.
For symmetric distributions $F(E[X_i])=1/2$ where we can compare $\bar{X}$ and $\tilde{X}$ for such distributions.

Case 1:
$X_i\sim n(\mu,\sigma^2)$. The limiting variance for $\bar{X}_n$ is σ^2, but for $\tilde{X}_n$ it is $\frac{1}{4f(\mu)^2}=\frac{\pi\sigma^2}{2}$ and

$$ARE(\tilde{X}_n,\bar{X}_n)=\frac{\sigma^2}{\frac{\pi\sigma^2}{2}}=\frac{2}{\pi}\approx 0.64$$

Case 2:
$f(x)=\frac{1}{2\sigma}e^{-\frac{|x-\mu|}{\sigma}}$. Here $Var[X_i]=2\sigma^2$ and $f(\mu)=\frac{1}{2\sigma}$. So

$$ARE(\tilde{X}_n,\bar{X}_n)=\frac{2\sigma^2}{1/4\sigma^2}=\frac{2\sigma^2}{\sigma^2}=2$$

which is double the efficiency in case 1.

Asymptotic results for LRTs

Consider testing $H_0:\theta=\theta_0$ vs. $H_1:\theta\neq\theta_0$ using a likelihood ratio test. Since here, $\Theta_0=\{\theta_0\}$, we obtain the likelihood ratio as

$$\lambda(\mathbf{x}) = \frac{\sup\limits_{\theta\in\Theta_0} L(\theta)}{\sup\limits_{\theta\in\Theta_0} L(\theta)} = \frac{L(\theta_0)}{L(\hat{\theta})}$$

Theorem:

(Asymptotic distribution of the LRT–simple H_0) For testing $H_0 : \theta = \theta_0$ versus $H_1 : \theta \neq \theta_0$, suppose $x_1, \ldots, X_n$ are i.i.d. $f(x|\theta)$, $\hat{\theta}$ is the MLE for θ, and $f(x|\theta)$ satisfies the regularity conditions mentioned earlier (found in Miscellanea 10.6.2. in Casella and Berger). Then under H_0, as $n \to \infty$,

$$-2\log\lambda(\mathbf{X}) \xrightarrow{\mathcal{D}} \chi^2_1$$

Proof: We begin by expanding $\log L(\theta|\mathbf{x}) = l(\theta|\mathbf{x})$, where L is the likelihood function, in a Taylor series around $\hat{\theta}$:

$$l(\theta|\mathbf{x}) = l(\hat{\theta}|\mathbf{x}) + l'(\hat{\theta}|\mathbf{x})(\theta - \hat{\theta}) + l''(\hat{\theta}|\mathbf{x})\frac{(\theta-\hat{\theta})^2}{2!} + \ldots.$$

We can now substitute the expansion for $l(\theta_0|\mathbf{x})$ in $-2\log\lambda(\mathbf{x}) = -2l(\theta_0|\mathbf{x}) + 2l(\hat{\theta}|\mathbf{x})$, and use the fact that $l'(\hat{\theta}|\mathbf{x}) = 0$. Thus we have:

$$-2\log\lambda(\mathbf{x}) \approx \frac{(\theta-\hat{\theta})^2}{-l''(\hat{\theta}|\mathbf{x})}.$$

Since the denominator is the observed information $\hat{I}_n(\hat{\theta})$ and $\frac{1}{n}\hat{I}_n(\hat{\theta}) \to I(\theta_0)$ it follows from Slutsky's theorem and the theorem on the asymptotic efficiency of MLEs that $-2\log\lambda(\mathbf{X}) \xrightarrow{\mathcal{D}} \chi^2_1$