STAT 713 sp 2023 Lec 08 slides

Consistency of estimators

Karl B. Gregory

University of South Carolina

These slides are an instructional aid; their sole purpose is to display, during the lecture,
definitions, plots, results, etc. which take too much time to write by hand on the blackboard.
They are not intended to explain or expound on any material.

Table of Contents

1. Weak, mean square, and strong consistency

2. Consistency results for UMVUEs and MLEs

3. Newton-Raphson algorithm for computing the MLE

Consistency of an estimator

An estimator $\hat{\theta}_n$ of $\theta \in \Theta \subset \mathbb{R}$ is called

1. *weakly consistent (W-C)* if

$$\lim_{n \to \infty} P(|\hat{\theta}_n - \theta| < \varepsilon) = 1$$

   for every $\varepsilon > 0$ and every $\theta \in \Theta$. (Same as conv. in probability, $\hat{\theta}_n \xrightarrow{P} \theta$).

2. *mean square consistent (MS-C)* if $\lim_{n \to \infty} \text{MSE}\, \hat{\theta}_n \to 0$.

3. *strongly consistent (S-C)* if $P(\lim_{n \to \infty} \hat{\theta}_n = \theta) = 1$.

W-C means $\hat{\theta}_n \in (\theta - \varepsilon, \theta + \varepsilon)$ occurs w/prob $\to 1$ as $n \to \infty$, for any $\varepsilon > 0$.

S-C covered in future courses. We will focus on W-C and MS-C.

Exercise: Let $X_1, \ldots, X_n \overset{\text{ind}}{\sim} \text{Unif}(0, \theta)$. Is $X_{(n)}$ a weakly consistent estimator for θ?

Theorem (Mean square consistency implies weak consistency)

An estimator $\hat{\theta}_n$ is weakly consistent for θ if it is mean square consistent for θ.

Exercise: Prove the above.

Showing $\lim\limits_{n\to\infty} \text{MSE}\,\hat{\theta}_n = 0$ can be easier than showing $\lim\limits_{n\to\infty} P(|\hat{\theta}_n - \theta| < \varepsilon) = 1$.

To establish MS-C, and by implication, W-C, one can show

1. $\lim\limits_{n\to\infty} \text{Var}\,\hat{\theta}_n = 0$
2. $\lim\limits_{n\to\infty} \text{Bias}\,\hat{\theta}_n = 0$

The property $\lim\limits_{n\to\infty} \text{Bias}\,\hat{\theta}_n = 0$ is called *asymptotic unbiasedness*.

Exercise: For $X_1, \ldots, X_n \overset{\text{ind}}{\sim} \text{Exponential}(\lambda)$. Check mean square consistency of

1. $\hat{\lambda}_n = \bar{X}_n$.
2. $\tilde{\lambda}_n = nX_{(1)}$.

It is possible to have weak consistency without MS-C.

Example: Let $X_1, \ldots, X_n \overset{\text{ind}}{\sim} t_2$, $f_X(x) = (1/2)(1 + x^2)^{-3/2}$.

1. Since $\mathbb{E}|X_1| = 1$, can show $\bar{X}_n$ is S-C for $\mathbb{E}X_1$, from which W-C follows.
2. However, since $\mathbb{E}X^2$ does not exist, $\bar{X}_n$ is *not* MS-C for $\mathbb{E}X_1$.

Theorem (MS consistency of moments)

For $X_1, \ldots, X_n$ iid, let $\tau = \mathbb{E}h(X_1)$ and $\hat{\tau}_n = n^{-1} \sum_{i=1}^n h(X_i)$. If $\mathbb{E}h^2(X_1) < \infty$, then $\hat{\tau}_n$ is mean square consistent for τ.

Since MS-C $\implies$ W-C: If $\mathbb{E}h^2(X_1) < \infty$, then $\hat{\tau}_n$ is weakly consistent for τ.

We can also show $\mathbb{E}|h(X_1)| < \infty$ implies W-C if we use a S-C argument.

Exercise: Prove the result.

Exercise: Let $X_1, \ldots, X_n \overset{\text{ind}}{\sim} \text{Pareto}(\alpha, \beta)$, $f_X(x; \alpha, \beta) = \beta \alpha^\beta x^{-(\beta+1)} \mathbf{1}(x \geq \alpha)$.

Let $\tau = \mathbb{E} X_1^k$ and $\hat{\tau}_n = n^{-1} \sum_{i=1}^n X_i^k$. When do we have $\hat{\tau}_n \overset{\text{p}}{\longrightarrow} \tau$?

Theorem (Helper results for establishing consistency)

1 If $\hat{\tau}_{1,n}, \ldots, \hat{\tau}_{k,n}$ are W-C for $\tau_1, \ldots, \tau_k$, resp., and g continuous, then

$$g(\hat{\tau}_{1,n}, \ldots, \hat{\tau}_{k,n}) \xrightarrow{p} g(\tau_1, \ldots, \tau_k).$$

2 If $\hat{\tau}_{1,n}, \ldots, \hat{\tau}_{k,n}$ are MS-C for $\tau_1, \ldots, \tau_k$, resp., and $a_{j,n} \to a_j$ and $b_n \to b$, then

$$\sum_{j=1}^{k} a_{j,n}\hat{\tau}_{j,n} + b_n \quad \text{is MS-C for} \quad \sum_{j=1}^{k} a_j \tau_j + b.$$

Exercise: Argue that the MoMs/MLEs are W-C for the parameters:

1 $X_1, \ldots, X_n \overset{\text{ind}}{\sim} \text{Uniform}(a, b)$.

2 $X_1, \ldots, X_n \overset{\text{ind}}{\sim} \text{Normal}(\mu, \sigma^2)$.

Table of Contents

1 Weak, mean square, and strong consistency

2 Consistency results for UMVUEs and MLEs

3 Newton-Raphson algorithm for computing the MLE

Theorem (Mean square consistency of the UMVUE)

Let $X_1, X_2, \ldots$ be iid and let $\hat{\tau}_n$ be the UMVUE for τ based on a sample of size n. Then $\hat{\tau}_n$ is mean square consistent.

Exercise: Go through proof.

Example: Let $X_1, \ldots, X_n \overset{\text{ind}}{\sim} \text{Geometric}(\theta)$, $\theta \in (0,1)$, $\tau = \theta(1-\theta)^{k-1}$.

Consider how to show that the following estimators of τ are MS-C or W-C:

1. The UMVUE

$$\hat{\tau}_n = \binom{T_n - k - 1}{n-2} \binom{T_n - 1}{n-1}^{-1} \mathbf{1}(T_n \geq n - 1 + k),$$

   where $T_n = \sum_{i=1}^n X_i$.

2. The MLE $\tilde{\tau}_n = (1/\bar{X}_n)(1 - 1/\bar{X}_n)^{k-1}$.

Theorem (Consistency of MLEs)

Let $X_1, \ldots, X_n$ be iid with cdf $F(x; \theta)$ and let $\hat{\theta}_n$ be the MLE for θ. Suppose

1. the support of $F(\cdot; \theta)$ does not depend on θ.
2. the score function exists and has finite mean.
3. the true value of θ lies in the interior of Θ.

Then $\hat{\theta}_n$ is strongly consistent for θ (implies $\hat{\theta}_n$ is W-C for θ).

A nice result since we cannot always write MLEs in closed form.

Exercise: Go through heuristics of proof.

Table of Contents

Newton-Raphson algorithm for computing the MLE

For data $\mathbf{X}$ with $S(\theta; \mathbf{X}) = \frac{\partial}{\partial \theta} \log f(\mathbf{X}; \theta)$ and $H(\theta; \mathbf{X}) = \frac{\partial^2}{\partial \theta \partial \theta^T} \log f(\mathbf{X}; \theta)$ and an initial value $\theta^{(0)}$ for $\hat{\theta}$, do

$$\theta^{(1)} \leftarrow \theta^{(0)} - [H(\theta^{(0)}; \mathbf{X})]^{-1} S(\theta^{(0)}; \mathbf{X})$$
$$\theta^{(0)} \leftarrow \theta^{(1)}$$

until there is little change between $\theta^{(0)}$ and $\theta^{(1)}$.

Exercise: Let $X_1, \ldots, X_n \overset{\text{ind}}{\sim} F(x; \mu) = 1/(1 + e^{-(x-\mu)})$, for $\mu \in \mathbb{R}$.

Write an algorithm for computing the MLE of μ.

```r
 6  ## MLE for logistic distribution location parameter
 7  set.seed(1)
 8  n <- 20
 9  U <- runif(n)
10  mu <- 2
11  X <- mu + log(U/(1-U)) # pass uniform(0,1) rvs through the quantile function
12
13  Fx <- function(x,mu){1/(1 + exp( - (x - mu)))}
14  Sx <- function(x,mu){ 2 * sum( Fx(x,mu) - 1/2 )}
15  Hx <- function(x,mu){ - 2 * sum(Fx(x,mu)*(1-Fx(x,mu)))}
16
17  mu0 <- 1
18  conv <- FALSE
19  while( conv == FALSE ){
20
21    mu1 <- mu0 - Sx(X,mu0) / Hx(X,mu0)
22
23    conv <- abs(mu1 - mu0) < 1e-4
24
25    mu0 <- mu1
26
27  }
28
29  mu.mle <- mu1
```