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

Consistency of an estimator

① ⇏ ②

② ⇒ ① $②^c \Leftarrow ①^c$

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 $P(\lim_{n\to\infty} \hat{\theta}_n = \theta) = 1$.

③ ⇒ ①.

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

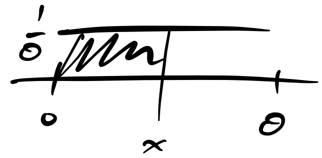
S-C covered in future courses. We will focus on W-C and MS-C.

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

yes.

$X_1, \dots, X_n \overset{iid}{\sim} \text{Unif}(0, \theta)$. Is $X_{(n)}$ WC for θ?

$f(x;\theta) = \frac{1}{\theta} \mathbb{1}(0 < x < \theta)$ $\qquad F_X(x;\theta) = \frac{x}{\theta}$ for $x \in (0,\theta)$



$$f_{X_{(n)}}(x) = n\left[F_X(x;\theta)\right]^{n-1} f_X(x;\theta)$$

$$= n\left[\frac{x}{\theta}\right]^{n-1} \frac{1}{\theta} \mathbb{1}(0 < x < \theta)$$

$$= \frac{n}{\theta^n} x^{n-1}.$$

For $\varepsilon > 0$,

$$P(|X_{(n)} - \theta| < \varepsilon) = P(\theta - \varepsilon < X_{(n)} < \theta + \varepsilon)$$

$$= P(\theta - \varepsilon < X_{(n)} \le \theta)$$

$$= \int_{\theta-\varepsilon}^{\theta} \frac{n}{\theta^n} x^{n-1}\, dx$$

$$= \frac{1}{\theta^n} \left[x^n\right]\Big|_{\theta-\varepsilon}^{\theta}$$

$$= \frac{1}{\theta^n}\left(\theta^n - (\theta - \varepsilon)^n\right)$$

$$= 1 - \left(\frac{\theta - \varepsilon}{\theta}\right)^n$$

$$\longrightarrow 1 \text{ as } n \to \infty.$$

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

$\text{MSE}\,\hat{\theta}_n = (\text{bias}\,\hat{\theta}_n)^2 + \text{Var}\,\hat{\theta}_n$

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

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

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

$\text{MSE}\,\hat{\theta}_n \to 0 \implies P(|\hat{\theta}_n - \theta| < \varepsilon) \to 1$ as $n \to \infty$
$\forall\, \varepsilon > 0.$

$$P(|\hat{\theta}_n - \theta| < \varepsilon) = P(|\hat{\theta}_n - \theta|^2 < \varepsilon^2)$$
$$= 1 - P(|\hat{\theta}_n - \theta|^2 \geq \varepsilon^2)$$
$$\geq 1 - \frac{\mathbb{E}\,|\hat{\theta}_n - \theta|^2}{\varepsilon^2}$$
$$= 1 - \frac{1}{\varepsilon^2}\,\text{MSE}\,\hat{\theta}_n$$

Markov's

For nonneg rv Y,

$$P(Y \geq \varepsilon) \leq \frac{\mathbb{E}\,Y}{\varepsilon}$$

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

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

$F_X(x;\lambda) = 1 - e^{-x/\lambda}$ $\quad f_X(x;\lambda) = \frac{1}{\lambda} e^{-x/\lambda}$

① $\hat{\lambda} = \bar{X}_n$. $\quad \mathbb{E}\hat{\lambda} = \lambda \implies \text{Bias}\,\hat{\lambda} = 0$ $\quad \hat{\lambda}$ is MS-C.

$\text{Var}\,\hat{\lambda} = \text{Var}\,\bar{X}_n = \frac{\lambda^2}{n} \longrightarrow 0.$ $\quad \implies \hat{\lambda}$ is W-C.

same as $\hat{\lambda} \xrightarrow{p} \lambda$.

② $f_{X_{(1)}}(x) = n\left[1 - F_X(x;\lambda)\right]^{n-1} f_X(x;\lambda)$

$$= n \left[e^{-x/\lambda} \right]^{n-1} \frac{1}{\lambda} e^{-x/\lambda}$$

$$= \frac{1}{(\lambda/n)} e^{-x/(\lambda/n)} \mathbb{1}(x > 0) \sim \text{Exponential}\left(\frac{\lambda}{n}\right).$$

For $\tilde{\lambda} = n X_{(1)}$, $\mathbb{E}\tilde{\lambda} = \mathbb{E}(n X_{(1)}) = n \frac{\lambda}{n} = \lambda$.

$$\text{Var}(\tilde{\lambda}) = \text{Var}(n X_{(1)}) = n^2 \text{Var}(X_{(1)}) = n^2 \left(\frac{\lambda}{n}\right)^2 = \lambda^2$$

$$\not\to 0$$

So $\tilde{\lambda}$ is not MS-C.

$\mathbb{E} X_1 = 1$ $\mathbb{E} X_1^2$ does not exist

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 $\mathbb{E}X_1$.

target — sample version

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.

$$\text{MSE } \hat{\tau}_n = (\text{Bias } \hat{\tau}_n)^2 + \text{Var } \hat{\tau}_n = \text{Var } \hat{\tau}_n = \frac{\text{Var } h(X_1)}{n} = \frac{\mathbb{E}h^2(X_1) - (\mathbb{E}h(X_1))^2}{n} \to 0$$

$$\mathbb{E}\hat{\tau}_n = \mathbb{E}\frac{1}{n}\sum_{i=1}^n h(X_i) = \mathbb{E}h(X_1) = \tau$$

$h(x) = x^k$

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

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

establish MS-C, which implies W-C.

$$\text{MSE } \hat{\tau}_n = \frac{\text{Var}(X_1^k)}{n} = \frac{\mathbb{E}X_1^{2k} - (\mathbb{E}X_1^k)^2}{n},$$

So I need $\mathbb{E}X_1^{2k} < \infty$. Then $\hat{\tau}_n \xrightarrow{p} \tau$.

Theorem (Helper results for establishing consistency)

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

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

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

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

$$\frac{n}{n-1}\hat{m}_2 + \frac{n}{n-1}\hat{m}_1^2 \to m_2 + m_1^2$$

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

1. $X_1, \dots, X_n \overset{\text{ind}}{\sim} \text{Uniform}(a, b)$.
2. $X_1, \dots, X_n \overset{\text{ind}}{\sim} \text{Normal}(\mu, \sigma^2)$.

Weak consistency $\hat{\theta}_n \xrightarrow{p} \theta$ as $n \to \infty$.

Mean-square consistency $\text{MSE}\ \hat{\theta}_n \to 0$ as $n \to \infty$

$$\text{MSE}\ \hat{\theta}_n = \left(\text{Bias}\ \hat{\theta}_n\right)^2 + \text{Var}\ \hat{\theta}_n$$

MS-C $\Rightarrow$ Weak Consistency

$\not\Leftarrow$

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

(i) Are MoMs estimators Weakly consistent?

$$\bar{a}_n = \hat{m}_1 - \sqrt{3}\sqrt{\hat{m}_2 - \hat{m}_1^2} = g_a(\hat{m}_1, \hat{m}_2)$$

$$\bar{b}_n = \hat{m}_1 + \sqrt{3}\sqrt{\hat{m}_2 - \hat{m}_1^2} = g_b(\hat{m}_1, \hat{m}_2)$$

$\hat{m}_1 = n^{-1} \sum_{i=1}^{n} X_i$

$\hat{m}_2 = n^{-1} \sum_{i=1}^{n} X_i^2$

$\text{MSE}\ \bar{a}_n \overset{?}{\to} 0$

$\text{MSE}\ \bar{b}_n \overset{?}{\to} 0$

$P(|\bar{a}_n - a| < \varepsilon) \to 1$ as $n \to \infty$ $\forall\, \varepsilon > 0$.

$$a = m_1 - \sqrt{3}\sqrt{m_2 - m_1^2} = g_a(m_1, m_2)$$

$$b = m_1 + \sqrt{3}\sqrt{m_2 - m_1^2} = g_b(m_1, m_2)$$

Question:

$$g_a(\hat{m}_1, \hat{m}_2) \overset{?}{\xrightarrow{p}} g_a(m_1, m_2)$$

$$g_b(\hat{m}_1, \hat{m}_2) \xrightarrow{p} g_b(m_1, m_2).$$

Know $\hat{m}_1 \xrightarrow{p} m_1$ and $\hat{m}_2 \xrightarrow{p} m_2$.

(ii) Are the MLEs weakly consistent?

$\hat{a}_n = X_{(1)}$

Bias $\hat{a}_n = \frac{b-a}{n+1} \to 0$

Var $\hat{a}_n = \frac{n(b-a)^2}{(n+1)^2(n+2)} \to 0$ as $n \to \infty$.

---

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

$$\hat{\sigma}^2_{mle} = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X}_n)^2$$

$$= \frac{1}{n}\sum_{i=1}^{n} X_i^2 - (\bar{X}_n)^2$$

$$= \hat{m}_2 - (\hat{m}_1)^2$$

$$= g(\hat{m}_1, \hat{m}_2)$$

$$\sigma^2 = m_2 - (m_1)^2 = g(m_1, m_2)$$

$\hat{\sigma}^2_{mle} \xrightarrow{P} \sigma^2$ b/c $\hat{m}_1 \xrightarrow{P} m_1$, $\hat{m}_2 \to m_2$.

$$S_n^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X}_n)^2$$

$$= \frac{n}{n-1}\frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X}_n)^2$$

$$= \frac{n}{n-1}\left(\frac{1}{n}\sum_{i=1}^{n} X_i^2\right) - \frac{n}{n-1}(\bar{X}_n)^2$$

$$= \frac{n}{n-1}\hat{m}_2 - \frac{n}{n-1}\hat{m}_1^2$$

$\frac{n}{n-1} \to 1$, $\hat{m}_2 \xrightarrow{P} m_2$; $\frac{n}{n-1} \to 1$, $\hat{m}_1^2 \xrightarrow{P} m_1^2$

Table of Contents

Theorem (Mean square consistency of the UMVUE)

Let $X_1, X_2, \dots$ 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 $X_1, \dots, X_n \overset{\text{ind}}{\sim}$ Geometric(θ), $\theta \in (0,1)$, $\tau = \theta(1-\theta)^{k-1}$.

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 $\tilde{\tau}_n = (1/\bar{X}_n)(1 - 1/\bar{X}_n)^{k-1}$. $= g(\bar{X}_n) \xrightarrow{p} g(m_1) = g(\frac{1}{\theta}) = \theta(1-\theta)^{k-1}$

$\hat{\theta}_n = \frac{1}{\bar{X}_n}$

How to show $\tilde{\tau}_n \xrightarrow{p} \tau$?

$P(|\tilde{\tau}_n - \tau| < \varepsilon) = \dots$

MSE $\tilde{\tau}_n$?

Exponential (λ), $\lambda \geq 0$

=> Weak consistency.

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.

Let θ_0 denote the true value of the parameter.

MLE $\hat{\theta}_n = \underset{\theta}{\operatorname{argmax}} \; \ell(\theta; \underline{X})$

$$= \underset{\theta}{\operatorname{argmax}} \sum_{i=1}^{n} \log f(X_i; \theta)$$

$$= \underset{\theta}{\operatorname{argmax}} \; \underbrace{\frac{1}{n} \sum_{i=1}^{n} \log f(X_i; \theta)}$$

$\xrightarrow{p}$ (actually "almost surely") $\mathbb{E}_{\theta_0} \log f(X_1; \theta)$

$\underbrace{\frac{1}{n} \sum_{i=1}^{n} \log f(X_i; \theta)}_{\text{maximizer of this, } \hat{\theta}_n}$ gets close to $\underbrace{\mathbb{E}_{\theta_0} \log f(X_1; \theta)}_{\text{maximizer of this, (which is } \theta_0)}$

gets close to

Show that $\mathbb{E}_{\theta_0} \log f(X_1; \theta)$ is maximized at θ_0.

Suff to show

$$\mathbb{E}_{\theta_0} \log f(X_1; \theta_0) - \mathbb{E}_{\theta_0} \log f(X_1; \theta) \geq 0 \quad \forall \theta.$$

$$\mathbb{E}_{\theta_0} \log f(X_1; \theta_0) - \mathbb{E}_{\theta_0} \log f(X_1; \theta)$$

$$= \mathbb{E}_{\theta_0} \log \left(\frac{f(X_1; \theta_0)}{f(X_1; \theta)} \right)$$

$$= \int_{\mathbb{R}} \log \left(\frac{f(x; \theta_0)}{f(x; \theta)} \right) f(x; \theta_0) \, dx$$

$$= \int_{\mathbb{R}} \log\left(\frac{f(x;\theta_0)}{f(x;\theta)} \right) \frac{f(x;\theta_0)}{f(x;\theta)} \, f(x;\theta) \, dx$$

$$= \mathbb{E}_{\theta} \Bigg[\underbrace{\log\left(\left(\frac{f(x_i;\theta_0)}{f(x_i;\theta)} \right) \right) \frac{f(x_i;\theta_0)}{f(x_i;\theta)}}_{\log(z)\, z} \Bigg]$$

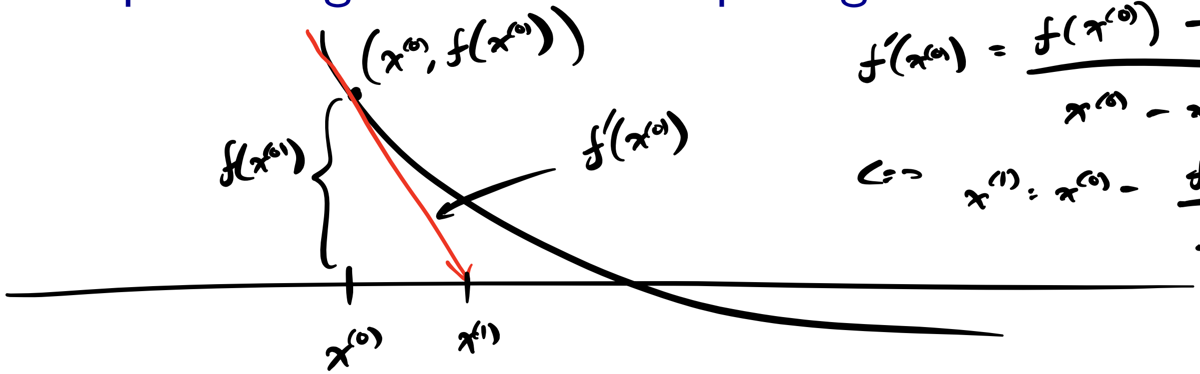
$$\geq \log\left(\mathbb{E}_{\theta}\left(\frac{f(x_i;\theta_0)}{f(x_i;\theta)} \right) \right) \mathbb{E}_{\theta}\left(\frac{f(x_i;\theta_0)}{f(x_i;\theta)} \right) \quad \text{By Jensen's ineq.}$$

$$= \log(1) \cdot 1$$

$$= 0.$$

$$\mathbb{E}_{\theta}\left(\frac{f(x_i;\theta_0)}{f(x_i;\theta)} \right) = \int_{\mathbb{R}} \left(\frac{f(x;\theta_0)}{\cancel{f(x;\theta)}} \right) \cancel{f(x;\theta)} \, dx = \int_{\mathbb{R}} f(x;\theta_0) \, dx = 1$$

Table of Contents



$$f'(x^{(0)}) = \frac{f(x^{(0)}) - 0}{x^{(0)} - x^{(1)}}$$

$$\iff x^{(1)} = x^{(0)} - \frac{f(x^{(0)})}{f'(x^{(0)})}$$

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

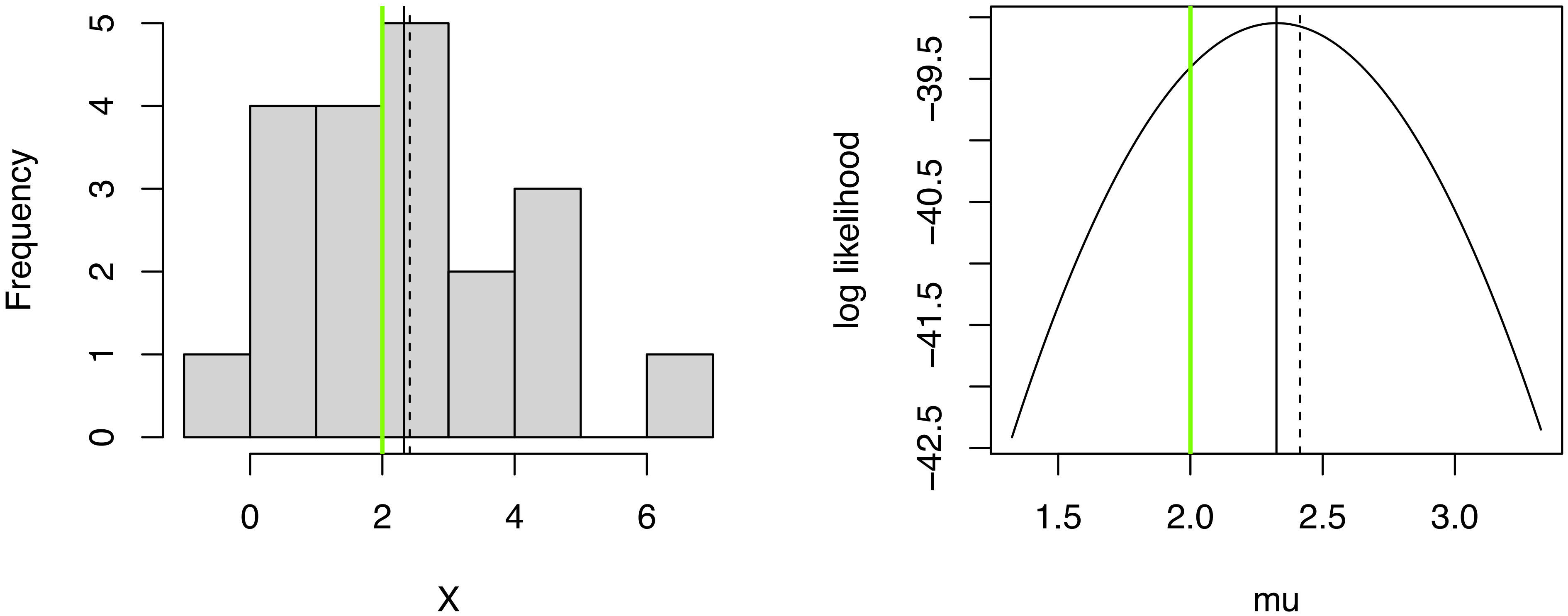
Hessian

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

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

Exercise: Let $X_1, \dots, X_n \overset{\text{ind}}{\sim} F(x; \mu) = 1/(1 + e^{-(x-\mu)})$, for $\mu \in \mathbb{R}$.
Write an algorithm for computing the MLE of μ.

```
## MLE for logistic distribution location parameter
set.seed(1)
n <- 20
U <- runif(n)
mu <- 2
X <- mu + log(U/(1-U)) # pass uniform(0,1) rvs through the quantile function

Fx <- function(x,mu){1/(1 + exp( - (x - mu)))}
Sx <- function(x,mu){ 2 * sum( Fx(x,mu) - 1/2 )}
Hx <- function(x,mu){ - 2 * sum(Fx(x,mu)*(1-Fx(x,mu)))}

mu0 <- 1
conv <- FALSE
while( conv == FALSE ){

  mu1 <- mu0 - Sx(X,mu0) / Hx(X,mu0)

  conv <- abs(mu1 - mu0) < 1e-4

  mu0 <- mu1

}

mu.mle <- mu1
```


Frequency
0
1
2
3
4
5
0
2
4
6
X
log likelihood
−42.5
−41.5
−40.5
−39.5
1.5
2.0
2.5
3.0
mu