Lecture 17: Consistency and the Law of Large Numbers

MATH 667-01
Statistical Inference
University of Louisville

November 16, 2017

Introduction

- We discuss consistent sequences of estimators described in Section 10.1 in Casella and Berger (2002)[1].
- We also describe convergence in probability and the Law of Large Numbers as described in Section 5.5.
- Finally in Section 10.1, a result on the consistency of MLEs is discussed under regularity assumptions given in Section 10.6.
- We also use Chebyshev's inequality from Section 3.6 to give basic proofs of a couple of the theorems.

[1] Casella, G. and Berger, R. (2002). *Statistical Inference, Second Edition*. Duxbury Press, Belmont, CA.

Consistency

- In this section, we consider the behavior of a sequence of estimators of a parameter θ in a parameter space Θ as $n \to \infty$.
- Let $X_1, X_2, \ldots$ be iid random variables with pdf/pmf $f(x|\theta)$.
- Then $W_n(X_1, \ldots, X_n), n = 1, 2, \ldots,$ is a sequence of estimators of θ based on a sample size n.
- *Definition L17.1* (Def 10.1.1 on p.468): A sequence of estimators $W_n = W_n(X_1, \ldots, X_n), n = 1, 2, \ldots,$ of θ is *(weakly) consistent* if, for every $\varepsilon > 0$ and every $\theta \in \Theta$,

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

- The condition above is equivalent to

$$\lim_{n\to\infty} P_\theta(|W_n - \theta| \geq \varepsilon) = 0.$$

- *Theorem L17.1* (Thm 3.6.1 on p.122): Let X be a random variable such that $\mathsf{E}[g(X)]$ exists and let $g(x)$ be a nonnegative function. Then for any $r > 0$,

$$P(g(X) \geq r) \leq \frac{\mathsf{E}[g(X)]}{r}.$$

- Here are two common special cases of this inequality:
  1. Markov's Inequality: $P(|X| \geq r) \leq \dfrac{\mathsf{E}[\,|X|\,]}{r}$
  2. Chebyshev's Inequality: $P(|X - \mu| \geq t\sigma) \leq \dfrac{1}{t^2}$

- *Theorem L17.2* (Thm 10.1.3 on p.469): If $\lim_{n\to\infty} \mathsf{Var}_\theta[W_n] = 0$ and $\lim_{n\to\infty} \mathsf{Bias}_\theta[W_n] = 0$ for every $\theta \in \Theta$, then $\{W_n\}$ is a consistent sequence of estimators of θ.
- *Proof of Theorem L17.2*: By Chebyshev's inequality,

$$P_\theta(|W_n - \theta| \geq \varepsilon) \leq \frac{\mathsf{E}_\theta\left[(W_n - \theta)^2\right]}{\varepsilon^2}$$

and

$$\mathsf{E}_\theta\left[(W_n - \theta)^2\right] = (\mathsf{Bias}_\theta[W_n])^2 + \mathsf{Var}_\theta[W_n] \to 0 + 0 = 0$$

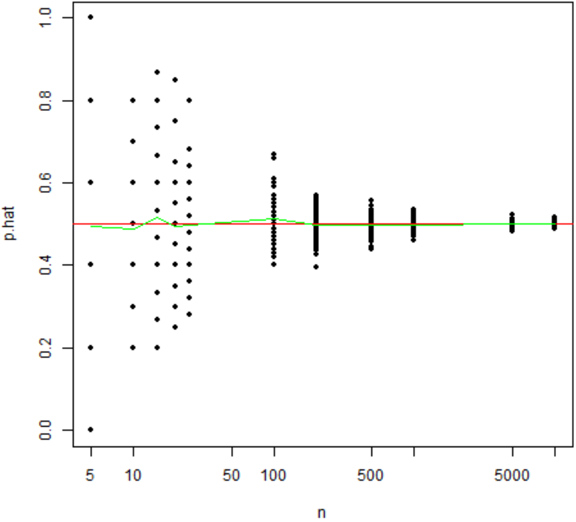
as $n \to \infty$.

- Simulation: Here we will examine the performance of the MLE of p in the model where $X_1, \ldots, X_n$ are iid Bernoulli(p) random variables. The MLE of p is $\hat{p}_n = \frac{1}{n} \sum_{i=1}^{n} X_i$ (see slide 7.12).
- Suppose the true value ofthe parameter is $p = 0.5$. For each n in $\{5, 10, 15, 20, 25, 100, 200, 500, 1000, 5000, 10000\}$, we simulate 100 data sets and plot $\hat{p}_n$ for each data set.
- As seen on slide 7, the bias and variance both appear to decrease towards 0 as n increases.

R code for simulation

```
> set.seed(126573)
> p=.5
> n=c(5,10,15,20,25,100,200,500,1000,5000,10000)
> repetitions=100
> p.hat=matrix(0,repetitions,length(n))
> for (i in 1:length(n)){
+  x=(1:n[i])/max(n)
+  for (r in 1:repetitions){
+   p.hat[r,i]=rbinom(1,size=n[i],prob=p)/n[i]
+  }
+ }
> plot(rep(n,repetitions),c(t(p.hat)),pch=19,cex=.7,
+ log="x",xlab="n",ylab="p.hat")
> abline(h=p,col="red")
> points(n,apply(p.hat,2,mean),type="l",col="green")
```

R Code for Simulation



Consistency

- *Example L17.1*: Suppose $X_1, \ldots, X_n$ are iid Bernoulli(p) random variables where $p \in (0,1)$. Show that the MLE $\hat{p}_n = \sum_{i=1}^n X_i/n$ is a consistent sequence of estimators of p.
- *Answer to Example L17.1*: This follows from *Theorem L17.2* since

$$\mathsf{E}[\hat{p}_n] = \frac{1}{n}\mathsf{E}\left[\sum_{i=1}^n X_i\right] = \frac{1}{n}\sum_{i=1}^n \mathsf{E}[X_i] = \frac{1}{n}\sum_{i=1}^n p = \frac{1}{n}np = p$$

and

$$\mathsf{Var}[\hat{p}_n] = \frac{1}{n^2}\mathsf{Var}\left[\sum_{i=1}^n X_i\right] = \frac{1}{n^2}\sum_{i=1}^n \mathsf{Var}[X_i] = \frac{1}{n^2}\sum_{i=1}^n p(1-p)$$

$$= \frac{1}{n^2}np(1-p) = \frac{1}{n}p(1-p) \to 0$$

as $n \to \infty$.

Convergence in Probability

- *Definition L17.2* (Def 5.5.1 on p.232): A sequence of random variables $X_1, X_2, \ldots$ *converges in probability* to a random variable X if, for every $\varepsilon > 0$, $\lim_{n\to\infty} P(|X_n - X| < \varepsilon) = 1$.
- *Theorem L17.3* (Thm 5.5.4 on p.233): Suppose that $X_1, X_2, \ldots$ converges in probability to a random variable X and that h is a continuous function. Then $h(X_1), h(X_2), \ldots$ converges in probability to $h(X)$
- *Example L17.2*: Suppose $X_1, \ldots, X_n$ are iid Bernoulli(p) random variables where $p \in (0,1)$. Show that $1/\hat{p}_n = n/\sum_{i=1}^n X_i$ is a consistent sequence of estimators of $1/p$.
- *Answer to Example L17.2*: *Example L17.1* shows that $\hat{p}_n$ converges in probability to p. Since $h(p) = 1/p$ is continuous on $(0,1)$, *Theorem L17.2* implies $1/\hat{p}_n$ converges in probability to $1/p$. By *Definition L17.1*, $1/\hat{p}_n$ is a consistent sequence of estimators of $1/p$.

Weak Law of Large Numbers

- *Theorem L17.4* (Thm 5.5.2 on p.232): Let $X_1, X_2, \dots$ be iid random variables with $\mathsf{E}[X_i] = \mu$ and $\mathsf{Var}[X_i] = \sigma^2 < \infty$. Then $\bar{X}_n = \sum_{i=1}^{n} X_i/n$ converges in probability to μ.
- *Proof of Theorem L17.4*: For every $\varepsilon > 0$, Chebyshev's inequality implies that

$$P\left(\left|\bar{X}_n - \mu\right| \geq \varepsilon\right) = P\left(\left(\bar{X}_n - \mu\right)^2 \geq \varepsilon^2\right)$$

$$\leq \frac{\mathsf{E}\left[\left(\bar{X}_n - \mu\right)^2\right]}{\varepsilon^2} = \frac{\mathsf{Var}[\bar{X}]}{\varepsilon^2} = \frac{\sigma^2}{n\varepsilon^2}.$$

- Therefore, it follows that

$$P(|\bar{X}_n - \mu| < \varepsilon) = 1 - P(|\bar{X}_n - \mu| \geq \varepsilon) \geq 1 - \frac{\sigma^2}{n\varepsilon^2} \to 1$$

as $n \to \infty$.

Regularity Assumptions

- Here are some assumptions needed for asymptotic results regarding the MLE (p.516):
  - (A1) We observe $X_1, \ldots, X_n$, where $X_i \sim f(x|\theta)$ are iid.
  - (A2) The parameter is identifiable; that is, if $\theta \neq \theta'$, then $f(x|\theta) \neq f(x|\theta')$.
  - (A3) The densities $f(x|\theta)$ have common support, and $f(x|\theta)$ is differentiable in θ.
  - (A4) The parameter space Ω contains an open set ω of which the true parameter value θ_0 is an interior point.
  - (A5) For every $x \in \mathcal{X}$, the density $f(x|\theta)$ is three times differentiable with respect to θ, the third derivative is continuous in θ, and $\int f(x|\theta)\,dx$ can be differentiated three times under the integral sign.
  - (A6) For any $\theta_0 \in \Omega$, there exists a positive number c and a function $M(x)$ (both of which may depend on θ_0) such that

$$\left| \frac{\partial^3}{\partial \theta^3} \ln f(x|\theta) \right| \leq M(x) \quad \text{for all } x \in \mathcal{X}, \theta_0 - c < \theta < \theta_0 + c$$

    with $\mathsf{E}_{\theta_0}[M(X)] < \infty$.

- *Theorem L17.5* (Thm 10.1.6 on p.470): Let $X_1, \ldots, X_n$ be iid $f(x|\theta)$, and let $L(\theta|\boldsymbol{x}) = \prod_{i=1}^n f(x_i|\theta)$ be the likelihood function. Let $\hat{\theta}_n$ denote the MLE of θ. Let $\tau(\theta)$ be a continuous function of θ. Suppose assumptions (A1)–(A4) hold. Then for all $\varepsilon > 0$ and every $\theta \in \Theta$,

$$\lim_{n\to\infty} P_\theta(|\tau(\hat{\theta}_n) - \tau(\theta)| \geq \varepsilon) = 0.$$

  That is, $\left\{\tau(\hat{\theta}_n)\right\}$ is a consistent sequence of estimators of $\tau(\theta)$.

- *Example L17.3*: Suppose $X_1, \ldots, X_n$ is a random sample from a distribution with pdf

$$f(x|\theta) = e^{-(x-\theta)} I_{[\theta,\infty)}(x)$$

where $\theta \in \mathbb{R}$. Compute the MLE and show that it is a consistent estimator of θ.

- *Answer to Example L17.3*: The likelihood function

$$L(\theta; \boldsymbol{x}) = \prod_{i=1}^{n} e^{-(x_i-\theta)} I_{[\theta,\infty)}(x_i) = e^{-\sum_{i=1}^{n} x_i} e^{n\theta} I_{[\theta,\infty)}(x_{(1)}).$$

is a positive, increasing function of θ on $(-\infty, x_{(1)}]$ and equal to 0 on $(x_{(1)}, \infty)$. So, the MLE of θ is $\hat{\theta}_n = X_{(1)}$.

- *Answer to Example L17.3 continued*: Note that regularity condition (A3) is not satisfied.
- So, we determine the distribution of $\hat{\theta}_n - \theta$.
  For $t > 0$, the cdf of $\hat{\theta}_n - \theta$ is

$$
\begin{aligned}
P(\hat{\theta}_n - \theta \le t) &= P(\hat{\theta}_n \le t + \theta) \\
&= 1 - P(X_{(1)} > t + \theta) \\
&= 1 - P(X_1 > t + \theta, \ldots, X_n > t + \theta) \\
&= 1 - \prod_{i=1}^{n} P(X_i > t + \theta) \\
&= 1 - \prod_{i=1}^{n} \int_{t+\theta}^{\infty} e^{-(x-\theta)}\, dx \\
&= 1 - \prod_{i=1}^{n} \left[-e^{-(x-\theta)} \right]_{t+\theta}^{\infty}
\end{aligned}
$$

- *Answer to Example L17.3 continued*:

$$\begin{aligned} F(t) &= 1 - \prod_{i=1}^{n} \left[-e^{-(x-\theta)} \right]_{t+\theta}^{\infty} \\ &= 1 - \prod_{i=1}^{n} \left(e^{-t} \right) \\ &= 1 - e^{-nt}. \end{aligned}$$

Then the pdf of $\hat{\theta}_n - \theta$ is

$$f(t) = F'(t) = ne^{-nt} I_{(0,\infty)}(t)$$

so $\hat{\theta}_n - \theta$ is Exponential with mean $\frac{1}{n}$.

- *Answer to Example L17.3 continued*: Consequently, we have

$$\mathsf{E}[\hat{\theta}_n - \theta] = \frac{1}{n} \Rightarrow \mathsf{E}[\theta] = \theta + \frac{1}{n} \to \theta$$

  and

$$\mathsf{Var}[\hat{\theta}_n - \theta] = \left(\frac{1}{n}\right)^2 \Rightarrow \mathsf{Var}[\hat{\theta}_n] = \frac{1}{n^2} \to 0$$

  as $n \to \infty$.

- So, *Theorem L17.2* implies that $\hat{\theta}_n$ is a consistent estimator of θ.