Mathematical Statistics Recitation 6

TA: Sonia Reilly

March 6, 2026

Lecture Review

Lecture 11

MLE is consistent for $X_1, \dots, X_n$ iid, and asymptotically normal with variance $\frac{1}{nI(\theta)}$.

Fisher information:

$$I(\theta) = -\mathbb{E}\left[\frac{\partial^2}{\partial\theta^2}\ell_\theta(X)\right] = \mathbb{E}\left[\left(\frac{\partial}{\partial\theta}\ell_\theta(X)\right)^2\right]$$

where $\ell_\theta(X)$ is the log likelihood of just one of the X_i's.

MLE satisfies both

$$\sqrt{nI(\theta_0)}(\hat{\theta}_n - \theta_0) \xrightarrow{d} N(0,1)$$
$$\sqrt{nI(\hat{\theta}_n)}(\hat{\theta}_n - \theta_0) \xrightarrow{d} N(0,1)$$

In the second, the estimator is substituted in because the true parameter θ_0 is unknown. This lets us build a $1-\alpha$ confidence interval for θ_0:

$$\theta_0 \in \left(\hat{\theta}_n \pm z_{\alpha/2}\frac{1}{\sqrt{nI(\hat{\theta}_n)}}\right).$$

Cramér-Rao Theorem: Any unbiased estimator $\hat{\theta}_n$ of θ given $X_1, \dots, X_n$ iid satisfies

$$\mathrm{Var}(\hat{\theta}_n) \geq \frac{1}{nI(\theta)},$$

so MLE is asymptotically more efficient, meaning it has lower asymptotic variance, than any unbiased estimator.

Lecture 12

Frequentist approach: θ is not random, but unknown. Result is estimator $\hat{\theta}$. (Everything we've done so far is frequentist.)

Bayesian approach: θ is random. Result is a distribution over θ (from which we can get estimates). Distribution represents our belief about what θ might be.

To find posterior distribution $\pi_{\theta|X^n}(\theta|X^n)$ (where $X^n = (X_1, \dots, X_n)$), pick some prior $\pi(\theta)$ to represent inital guess of what θ might be, then use Bayes' Law:

$$\pi_{\theta|X^n}(\theta|X^n) = \frac{\pi(\theta)f_\theta(X^n)}{\int \pi(\theta')f_{\theta'}(X^n)d\theta'} \propto \mathcal{L}_n(\theta)\pi(\theta)$$

Recall Bayes' Law from 1st recitation. This is the same!

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

Can make point estimates and CI's (now called "credible intervals") this way too

- Point estimates: posterior mode (= MAP, maximum a posteriori estimator), posterior mean, median, etc.
- These are all specific cases of the Bayesian Optimal Estimator:

$$\hat{\theta}(X^n) = \underset{c}{\operatorname{argmin}}\, \mathbb{E}_{\theta\sim\pi(\theta|X^n)}[L(\theta, c)]$$

  for various loss functions $L(\theta, \hat{\theta})$
- Credible interval: e.g., between 2.5% and 97.5% percentile of posterior

In $n \to \infty$ limit, Bayesian method $\to$ frequentist method:

- $\mathcal{L}_n(\theta)$ dominates
- $\pi(\theta)$ doesn't matter (given certain conditions)
- posterior mode and mean converge to $\hat{\theta}_{\text{MLE}}$
- posterior becomes close to Gaussian $N(\hat{\theta}_{MLE}, \frac{1}{nI(\hat{\theta}_{MLE})})$

Pros and cons:

- Pro: a way to use prior information, if you have any
- Pro: more information at small/medium n than just mean and variance
- Con: computation can be hard when posterior is not normal
- Con: depends on choice of prior, and sometimes you don't have much prior knowledge

Problems

1. True or False? Let $\pi(\theta) = \text{Unif}([0, 1])$ be a prior. The posterior mode of θ cannot fall outside of $[0, 1]$.

   Solution: True. Since $\pi(\theta) = 0$ outside $[0, 1]$, $f_{\theta|X}(\theta|X) \propto \mathcal{L}(\theta)\pi(\theta)$ is also 0 outside $[0, 1]$.

2. [8.50, continued from HW] Let $X_1, \dots, X_n$ be an i.i.d. sample from a Rayleigh distribution with parameter $\theta > 0$:

$$f(x \mid \theta) = \frac{x}{\theta^2} e^{-x^2/(2\theta^2)}, \qquad x \geq 0.$$

   For this problem, you may use the fact that $\mathbb{E}[X_i^2] = 2\theta^2$.

   On Homework 5 you found the MLE

$$\hat{\theta}_{\text{MLE}} = \sqrt{\frac{1}{2n}\sum_{i=1}^{n} X_i^2}.$$

   Now we are interested in the distribution of this estimator.

(a) Find the asymptotic variance of the maximum likelihood estimator.

(b) Find an approximate 95% confidence interval for θ.

Solution:

(a) The likelihood function for one of the X_i's is

$$\frac{X_i}{\theta^2}\exp\left(-\frac{X_i^2}{2\theta^2}\right).$$

The log-likelihood is

$$\ell_\theta(X_i) = \log X_i - 2\log\theta - \frac{1}{2\theta^2}X_i^2.$$

Differentiate twice with respect to θ:

$$\frac{d\ell}{d\theta} = -\frac{2}{\theta} + \frac{1}{\theta^3}X_i^2$$

$$\frac{d^2\ell}{d\theta^2} = \frac{2}{\theta^2} - \frac{3}{\theta^4}X_i^2.$$

Thus the Fisher information is

$$I(\theta) = -\mathbb{E}\left[\frac{d^2\ell}{d\theta^2}\right] = \frac{2}{\theta^2} - \frac{3}{\theta^4}(2\theta^2) = \frac{4}{\theta^2}.$$

Therefore, the asymptotic variance of the MLE is

$$\mathrm{Var}(\hat{\theta}_{\mathrm{MLE}}) \approx \frac{1}{nI(\theta)} = \frac{\theta^2}{4n}.$$

(b) So approximately,

$$\hat{\theta}_{MLE} \sim N\left(\theta, \frac{\theta^2}{4n}\right).$$

Since θ is unknown, we plug in $\hat{\theta}_{MLE}$ to estimate the standard error, so an approximate 95% confidence interval is

$$\hat{\theta}_{MLE} \pm 1.96\,\frac{\hat{\theta}_{MLE}}{2\sqrt{n}}.$$

Substituting the expression for $\hat{\theta}_{MLE}$ gives

$$\sqrt{\frac{1}{2n}\sum_{i=1}^n X_i^2} \;\pm\; 1.96\,\frac{1}{2\sqrt{n}}\sqrt{\frac{1}{2n}\sum_{i=1}^n X_i^2}.$$

3. [8.61] Laplace's rule of succession claims that when an event happens n times in a row and never fails to happen, the probability that the event will occur the next time is $(n+1)/(n+2)$. Show that this is the posterior mean of the probability θ of each event happening, given a certain simple choice of prior.

Solution:

Let $\theta \in [0,1]$ be the unknown probability of a single event happening. We assume a uniform prior, $\pi(\theta) = 1$ on $[0,1]$, 0 otherwise. Given data $X_1, \ldots, X_n$ consisting of n consecutive events occurring, the likelihood function is $p_\theta(X_1, \ldots, X_n) = \theta^n$.

Using Bayes' Theorem, the posterior distribution for θ is:

$$\pi(\theta \mid X_1, \ldots, X_n) = \frac{p_\theta(X_1, \ldots, X_n)\pi(\theta)}{\int_0^1 p_{\theta'}(X_1, \ldots, X_n)\pi(\theta')\, d\theta'} = \frac{\theta^n}{\int_0^1 \theta'^n\, d\theta'} = \frac{\theta^n}{\frac{1}{n+1}} = (n+1)\theta^n$$

on $[0,1]$, 0 otherwise.

The probability of the $(n+1)$-th event occurring is the expected value of θ under this posterior distribution:

$$\begin{aligned} E[\theta \mid X_1, \ldots, X_n] &= \int_0^1 \theta \cdot \pi(\theta \mid X_1, \ldots, X_n)\, d\theta \\ &= \int_0^1 \theta \cdot (n+1)\theta^n\, d\theta \\ &= (n+1)\int_0^1 \theta^{n+1}\, d\theta \\ &= (n+1)\left[\frac{\theta^{n+2}}{n+2}\right]_0^1 \\ &= \frac{n+1}{n+2}. \end{aligned}$$

4. [8.60] Let $X_1, \ldots, X_n$ be an i.i.d. sample from an exponential distribution with the density function

$$f(x|\tau) = \frac{1}{\tau}e^{-x/\tau}, \quad x \geq 0.$$

(a) Find the mle of τ.

(b) Show that the mle is unbiased, and find its exact variance.

(c) Is there any other unbiased estimate with smaller variance?

(d) Find the form of an approximate confidence interval for τ.

Solution:

(a) The log-likelihood function is:

$$\log \mathcal{L}(\tau) = \sum_{i=1}^n \ln f(x_i|\tau) = -n\ln\tau - \frac{1}{\tau}\sum_{i=1}^n x_i$$

Setting the first derivative to zero:

$$\frac{d}{d\tau}(\log \mathcal{L}(\tau)) = -\frac{n}{\tau} + \frac{1}{\tau^2}\sum x_i = 0 \implies \hat{\tau}_{MLE} = \bar{X}$$

(b) Find the expectation and variance:

$$\mathbb{E}[\hat{\tau}] = \mathbb{E}[\bar{X}] = \frac{1}{n}\sum \mathbb{E}[X_i] = \frac{1}{n}(n\tau) = \tau$$

$$\text{Var}(\hat{\tau}) = \text{Var}\left(\frac{1}{n}\sum X_i\right) = \frac{1}{n^2}(n\tau^2) = \frac{\tau^2}{n}.$$

(c) To find the Cramér-Rao lower bound, we derive the Fisher Information $I(\tau)$:

$$\ell_\tau(X) = -\ln\tau - \frac{X}{\tau}$$

$$\frac{d\ell}{d\tau} = -\frac{1}{\tau} + \frac{x}{\tau^2}, \quad \frac{d^2\ell}{d\tau^2} = \frac{1}{\tau^2} - \frac{2x}{\tau^3}$$

$$I(\tau) = -\mathbb{E}\left[\frac{d^2\ell}{d\tau^2}\right] = -\left(\frac{1}{\tau^2} - \frac{2\mathbb{E}[X]}{\tau^3}\right) = -\frac{1}{\tau^2} + \frac{2\tau}{\tau^3} = \frac{1}{\tau^2}$$

The Cramér-Rao lower bound for an unbiased estimator is:

$$\frac{1}{nI(\tau)} = \frac{\tau^2}{n}$$

Since this is the same as $\text{Var}(\hat{\tau})$, no other unbiased estimator has smaller variance.

(d) Using the normal approximation and substituting $\hat{\tau}$ for τ in the standard error:

$$\hat{\tau} \pm z_{\alpha/2}\frac{\hat{\tau}}{\sqrt{n}}$$