STAB57: An Introduction to Statistics

Shahriar Shams

Week 5 (Large sample property of Score and MLE, Efficiency)

Winter 2023

Recap of Week 4

- Sufficient statistic
  - Factorization theorem
- Consistency
  - Using LLN
  - Using Slutsky's Lemma, Continuous mapping theorem
  - For large n, MLE is consistent ($\hat{\theta} \xrightarrow{P} \theta_0$)
- Score and Fisher information

Learning goals for this week

- Large sample property of MLE
  - Distribution of MLE (Rice Page 276-278)
- Efficiency
  - Cramer Rao Lower Bound(CRLB) (Rice page 298-302)

Section 1

Large Sample Property of MLE

Score (revisit)

- For the random variable X_i,

$$S(\theta|X_i) = \frac{\partial}{\partial\theta}\log f(X_i|\theta)$$

- For *iid* $(X_1, X_2, ..., X_n)$,

$$S(\theta|X_1, ..., X_n) = \sum_i S(\theta|X_i)$$

- Expected Score evaluated at $\theta = \theta_0$ is zero.

$$E[S(\theta|X)]\big|_{\theta=\theta_0} = 0 \implies E[S(\theta|X_1, X_2, ..., X_n)]\big|_{\theta=\theta_0} = 0$$

Fisher Information (revisit)

- For single obs.

$$I(\theta_0) = var[S(\theta|X)\big|_{\theta=\theta_0}]$$

- For *iid* $(X_1, X_2, ..., X_n)$,

$$nI(\theta_0) = var[S(\theta|X_1, X_2, ..., X_n)\big|_{\theta=\theta_0}]$$

Distribution of Score

- For a single obs X, we know,
  - Score is a random variable.
  - Its expectation is zero.
  - Its variance is $I(\theta_0)$

- $S(\theta|X_1, ..., X_n) = \sum_i S(\theta|X_i)$
- $S(\theta|X_1, X_2, ..., X_n)$ is the sum of n independent random variables each with the same mean and same variance.

- For large n, can we guess the distribution of $S(\theta|X_1, X_2, ..., X_n)$?

Distribution of Score using R

A little proof that we didn't do last week

$$I(\theta_0) = E\left[\left(\frac{\partial}{\partial\theta}\log f(X|\theta)\Big|_{\theta=\theta_0}\right)^2\right] = -E\left[\frac{\partial^2}{\partial\theta^2}\log f(X|\theta)\Big|_{\theta=\theta_0}\right]$$

- In words:
  Expectation of the **square** of the **first derivative** of the log-likelihood $\equiv$ **negative expectation** of its **second derivative**.

- Both the first and second derivative is evaluated at $\theta = \theta_0$

Subsection 1

Distribution of MLE

Distribution of MLE

CLAIM: Under "some conditions", as $n \to \infty$

$$\frac{(\hat{\theta} - \theta_0)}{\sqrt{1/nI(\theta_0)}} \xrightarrow{D} N(0,1)$$

Consequences of this claim:

- For large n, $E[\hat{\theta}] = \theta_0$
- For large n, $V[\hat{\theta}] = \frac{1}{nI(\theta_0)}$
- And we have a sampling distribution (though asymptotic) of MLE.

Sketch of the proof (Rice page 277)

Before trying to proof our claim let us introduce some notations to make our life easy.

- $l'(\theta)$= first derivative of the log-likelihood
- $l''(\theta)$= second derivative of the log-likelihood

Then we can write (for large n),

- $l'(\hat{\theta}) = 0$
- $E[l'(\theta_0)] = 0$
- $l'(\theta_0) \xrightarrow{D} N(0, nI(\theta_0)) \implies \frac{1}{n}l'(\theta_0) \xrightarrow{D} N(0, \frac{1}{n}I(\theta_0))$
- By LLN, $\frac{1}{n}l''(\theta_0) \xrightarrow{P} -I(\theta_0)$

Sketch of the proof (cont...)

Applying Taylor series on $l'(\hat{\theta})$

$$\begin{aligned}
l'(\hat{\theta}) &= l'(\theta_0) + (\hat{\theta} - \theta_0) l''(\theta_0) + \text{higher order terms} \\
l'(\hat{\theta}) &\approx l'(\theta_0) + (\hat{\theta} - \theta_0) l''(\theta_0) \\
0 &\approx l'(\theta_0) + (\hat{\theta} - \theta_0) l''(\theta_0) \\
\hat{\theta} - \theta_0 &\approx -\frac{l'(\theta_0)}{l''(\theta_0)} = \frac{(1/n) l'(\theta_0)}{-(1/n) l''(\theta_0)}
\end{aligned}$$

- numerator $\xrightarrow{D} N(0, \frac{1}{n} I(\theta_0))$
- denominator $\xrightarrow{P} I(\theta_0)$

This gives us

$$\hat{\theta} - \theta_0 \xrightarrow{D} N(0, \frac{1}{nI(\theta_0)}) \implies \frac{(\hat{\theta} - \theta_0)}{\sqrt{1/nI(\theta_0)}} \xrightarrow{D} N(0, 1)$$

Some claims about MLE

- MLE is asymptotically unbiased
- MLE is function of sufficient statistic
- MLE is consistent
- MLE is asymptotically efficient (will revisit after we have covered efficiency)

- Most importantly,

$$\frac{(\hat{\theta} - \theta_0)}{\sqrt{1/nI(\theta_0)}} \xrightarrow{D} N(0, 1)$$

Section 2

Efficient Estimator

Efficiency (Rice-P298)

- Let T_1 and T_2 be two different estimators of θ
- Efficiency of T_1 relative to T_2 is defined as

$$eff(T_1, T_2) = \frac{var[T_2]}{var[T_1]}$$

- $eff(T_1, T_2) > 1 \implies T_1$ has smaller variance $\implies T_1$ is more efficient
- This comparison is meaningful when T_1 and T_2 are both unbiased or both have the same bias.

Lower bound of the variance of an unbiased estimator

(Rice-P300)

- There is a famous inequality that provides a **lower bound for the variance** of all the **unbiased estimators**.
- In other words it gives a lower bound of the MSE (since Bias=0)

- The estimator whose variance achieves this lower bound is said to be efficient.

Cramer-Rao Inequality

- Let $X_1, X_2, ...X_n$ be *i.i.d.* with density $f_{\theta_0}(x)$
- $T = t(X_1, X_2, ...X_n)$ be an **unbiased** estimator of θ_0.
- Then under some assumptions on $f_{\theta_0}(x)$,

$$var[T] \geq \frac{1}{nI(\theta_0)}$$

- $\frac{1}{nI(\theta_0)}$ is also known as the Cramer-Rao lower bound (CRLB)

Proof of Cramer-Rao inequality

- Let Z be the score evaluated at $\theta = \theta_0$

$$Z = S(\theta|X_1, X_2, ..., X_n)|_{\theta=\theta_0}$$

- Immediately we can write, $E[Z] = 0$ and $var[Z] = nI(\theta_0)$
- Correlation coefficient between two variable T and Z is defined as

$$\rho[T, Z] = \frac{cov[T, Z]}{\sqrt{var[T]var[Z]}}$$

  which is bounded between -1 and 1.
- Then,

$$\rho^2[T, Z] \leq 1$$

$$\frac{(cov[T, Z])^2}{var[T] * var[Z]} \leq 1$$

$$\implies var[T] \geq \frac{(cov[T, Z])^2}{var[Z]}$$

Proof of Cramer-Rao inequality (cont...)

Continuing from last slide

$$\begin{aligned} var[T] &\geq \frac{(cov[T,Z])^2}{var[Z]} \\ \implies var[T] &\geq \frac{(cov[T,Z])^2}{nI(\theta_0)} \end{aligned}$$

we just need to show that $cov[T,Z] = 1$ (show)

Comments on CRLB

- CRLB gives us the lower bound of variance for all the unbiased estimators.
- In other words if you have several unbiased estimators, none of them will have a variance lower than $\frac{1}{nI(\theta_0)}$
- So if we can find an unbiased estimator whose variance is $\frac{1}{nI(\theta_0)}$, we know that we have the efficient one.

Note: We showed that for large n, MLE is unbiased and has a variance of $\frac{1}{nI(\theta_0)} \implies$ MLE is asymptotically efficient.

Example of calculating CRLB for $Poisson(\lambda)$

- Step 1: log-likelihood, $l(\lambda) = -n\lambda + \sum_{i=1}^{n} X_i ln\lambda + const.$
- Step 2: Score, $\frac{\partial l(\lambda)}{\partial \lambda} = -n + \sum_{i=1}^{n} X_i/\lambda$
- Step 3: $\frac{\partial^2 l(\lambda)}{\partial \lambda^2} = -\sum_{i=1}^{n} X_i/\lambda^2$
- Step 4: Fisher Information,
  $-E[\frac{\partial^2 l(\lambda)}{\partial \lambda^2}] = -E[-\sum_{i=1}^{n} X_i/\lambda^2] = 1/\lambda^2 E[\sum_{i=1}^{n} X] = n/\lambda$
- Step 5: Inverting the quantity from step 4, we get, CRLB $= \lambda/n$

Note:

- we would have done step 1-3 for MLE calculation anyway. So step 4 and 5 are extra.
- MLE of λ is $\bar{X}$ and $var[\bar{X}] = \lambda/n$ (you do it...)
- $\implies \bar{X}$ is the efficient estimator *out of all unbiased estimators.*

Some claims about MLE(revisit)

- MLE is asymptotically unbiased
- MLE is function of sufficient statistic
- MLE is consistent
- MLE is asymptotically efficient

Important distributional findings from this week

- Score evaluated at $\theta = \theta_0$,

$$l'(\theta_0) \xrightarrow{D} N(0, nI(\theta_0))$$

- Maximum likelihood estimator,

$$\hat{\theta} \xrightarrow{D} N(\theta_0, \frac{1}{nI(\theta_0)})$$

- we will learn another version of the asymptotic distribution of MLE next week.

Homework (Non-credit)

John A. Rice

Exercise 8: 7(c), 16(c), 17(d), 18(c), 47(c), 50(c), 52(c), 60(d, e)