Foundations of Statistical Inference

J. Berestycki & D. Sejdinovic

Department of Statistics
University of Oxford

MT 2019

Lecture 5 - Completenes, Lehmann-Scheffé Theorem, Method of Moments

Complete Sufficiency in EFs

Suppose that $Y_1, Y_2, \ldots, Y_n$ are from a k-parameter family

$$f(y;\theta) = h(y)\psi(\theta)\exp[\sum_{j=1}^{k} A_j(\theta)B_j(y)]$$

with natural parametrization

$$f(y;\phi) = h(y)\psi(\phi)\exp[\sum_{j=1}^{k} \phi_j B_j(y)]$$

so that the pdf/pmf for a sample of size n is

$$f(y_1,\ldots,y_n;\phi) = (\prod_{i-1}^{n} h(y_i))[\psi(\phi)]^n \exp[\sum_{j=1}^{k} \phi_j(\sum_i B_j(y_i))]$$

Complete Sufficiency in EFs

Lemma (6)

Then the vector

$$\mathbf{T}(y_1, \ldots, y_n) = (\sum_i B_1(y_i), \ldots, \sum_i B_n(y_i))$$

is

- *minimal sufficient for* θ *and* ϕ
- *complete sufficent for* η *and* ϕ *if* $\phi = A(\theta)$ *is a one-to-one function of* θ *and* Φ *contains a* k*-dimensional rectangle.*

Comment We showed before that MLEs for exponential families were functions of this same statistic. So if MLE is unbiased, it is UMVUE.

Summary

CRLB (How good can you get?)

1. $\hat{\theta}$ unbiased estimator of θ. $V(\hat{\theta}) \geq I_\theta^{-1}$
2. $\exists \hat{\theta}$ attains CRLB $\Leftrightarrow \frac{\partial \ell}{\partial \theta} = I_\theta(\hat{\theta} - \theta) \Rightarrow X \in$ expo. family.
3. If $\tilde{\theta}$ attains CRLB and $\hat{\theta}$ MLE, then $\tilde{\theta} = \hat{\theta}$ (if $\hat{\theta}$ is in the interior of Θ)

Rao-Blackwell (How to be better?)

1. $\hat{\theta}_T = E[\hat{\theta}|T]$ where T sufficient for θ and $\hat{\theta}$ unbiased, then $\hat{\theta}_T = f(T)$ and $V(\hat{\theta}_T) \leq V(\hat{\theta})$.
2. If $\exists$ MVUE and T is suff. then $\exists h(T)$ which is MVUE

Lehmann-Scheffé (How to be the best?)

1. T is complete suff. and $h(T)$ unbiased, then it is MVUE
2. If T is complete suff. and $\hat{\theta}$ is unbiased, then $\hat{\theta}_T$ is MVUE.

Quiz time!

- If a MVUE $\hat{\theta}$ exists, does it always attains the CRLB? See next example
- In what kind of situation is it the case that $\nexists$ an MVUE? Make example with $\hat{\theta}_1$ and $\hat{\theta}_2$
- If the MLE is an unbiased estimator and solves $\partial \ell / \partial \theta = 0$, is it always the MVUE (under reg conditions)? Does it attains CRLB?
- Can a MLE have a lower variance than CRLB? Higher variance? See next example
- In Rao-Blackwell, if $\hat{\theta}_T = \hat{\theta}$ for any sufficient T, is it the MVUE? Try to compare to Lehmann-Scheffé

Example 13

Let $X_1, \ldots, X_n$ be a random sample from $\mathsf{N}(\mu, \sigma^2)$.

Then we know the MLEs are $\widehat{\mu} = \bar{X}$, $\widehat{\sigma^2} = \frac{1}{n}\sum_{i=1}^n (X_i - \bar{X})^2$.

$\widehat{\mu}$ is unbiased, but $\widehat{\sigma^2}$ is biased.

Exercise The minimal sufficient complete statistic is $\left(\sum_{i=1}^n X_i, \sum_{i=1}^n X_i^2\right)$.

So $\widehat{\mu}$ is MVUE and attains the CRLB with variance σ^2/n.

The sample variance $S^2 = \frac{1}{n-1}\sum_{i=1}^n (X_i - \bar{X})^2$ is unbiased and is a function of the minimal sufficient complete statistic so is MVUE with variance $2\sigma^4/(n-1)$ which is larger than the CRLB of $2\sigma^4/n$.

Method of moments

Definition

A moment estimator $\hat{\theta}$ is a value of θ which satisfies the system of moment equations :

$$\mathbb{E}_X[T_j(X_1, \cdots, X_n); \theta] = T_j(x_1, \cdots, x_n), \quad j = 1, \cdots, k \tag{1}$$

Particular case : sums if

$$T_j(X_1, \cdots, X_n) = \frac{\sum_{i=1}^n f_j(X_i)}{n}, \quad \& \, \mathbb{E}_X(f_j(X_i); \theta) = \mathbb{E}_X(f_j(X_1); \theta), \quad \forall i$$

Then (1) becomes $\hat{\theta} = \theta$ solution of

$$\forall j, \quad \mathbb{E}_X(f_j(X_1); \theta) = \frac{\sum_{i=1}^n f_j(x_i)}{n} \quad \text{(empirical mean)}$$

Typically $\theta \in \Theta \subset \mathbb{R}^k$ and

$$\theta \to H(\theta) = (\mathbb{E}_X(f_1(X_1);\theta), \cdots, \mathbb{E}_X(f_j(X_1);\theta)) \quad \text{is bijective}$$

so that

$$\hat{\theta}_n = H^{-1}(\sum_{i=1}^{n} F(x_i)/n), \quad F = (f_1, \cdots, f_k)$$

Comment Simple, easy to use. Often have good properties, but not guaranteed. Can provide good starting point for an iterative method for finding MLEs.

We do not need to restrict ourselves to the case of sums.
Example 14 $X_1, X_2, ..., X_n \stackrel{iid}{\sim} \mathcal{U}(0, \theta)$

$$L(\theta; x) = f(x; \theta) = \theta^{-n} \prod_{i=1}^{n} \mathbb{1}_{x_i \leq \theta}$$

Let $X_{(n)} = \max_i X_i$. Moment relation : $\hat{\theta}$ defined by θ s.t.

$$\mathbb{E}\left[X_{(n)}; \theta\right] = x_{(n)}$$

We have

$$\hat{\theta} = \frac{n+1}{n} X_{(n)}.$$

Proof of $\hat{\theta} = (1 + 1/n)X_{(n)}$.

The distribution of $X_{(n)} = \max_i X_i$ is obtained from the CDF

$$\mathbb{P}(X_{(n)} \leq y; \theta) = \left(\frac{y}{\theta}\right)^n, \quad 0 < y < \theta$$

(the probability all iid X_i fall in $(0, y)$) so

$$f_{X_{(n)}}(y; \theta) = \frac{ny^{n-1}}{\theta^n}, \; 0 < y < \theta$$

and

$$\mathbb{E}\left[X_{(n)}; \theta\right] = \frac{n}{n+1}\theta, \text{ so } \widehat{\theta} = \frac{n+1}{n}X_{(n)}$$

Ex 14 : properties of $\hat{\theta}$

Theorem

In Example 14 : $\hat{\theta}$ is unbiased and a minimal , complete sufficient statistics

We first check that $\widehat{\theta}$ is sufficient.

The distribution of $X \mid X_{(n)} = y$ is

$$f(x|X_{(n)} = y; \theta) = \frac{f(x;\theta)}{f_{X_{(n)}}(y;\theta)} = \frac{1}{ny^{n-1}}$$

which does not depend on θ.
$\widehat{\theta}$ is minimal sufficient since

$$\frac{L(\theta;x)}{L(\theta;y)} = \frac{\theta^{-n} I[x_{(n)} < \theta]}{\theta^{-n} I[y_{(n)} < \theta]}$$

does not depend on θ iff $x_{(n)} = y_{(n)}$ (Lehmann-Scheffé Theorem)

Finally, we can show that $X_{(n)}$ is complete.

If $\mathbb{E}[h(X_{(n)}] = 0$ for all $\theta > 0$ then

$$\int_0^\theta h(y) \frac{ny^{n-1}}{\theta^n} dy = 0$$

for all $\theta > 0$ and hence

$$\int_0^\theta h(y) y^{n-1} dy = 0$$

so that

$$\int_0^\theta h^-(y) y^{n-1} dy = \int_0^\theta h^+(y) y^{n-1} dy$$

for all θ (where h^- and h^+ are negative/positive parts). We conclude that $h^- \equiv h^+$.

Since $\widehat{\theta}$ is not only sufficient but also complete and unbiased, it is MVUE.

We can't use the CRLB as the problem is non-regular.

Exercise $\ell(\theta; x) = -n \log(\theta)$ so

$$\mathbb{E}\left[\left(\frac{\partial \ell}{\partial \theta}\right)^2\right] = n^2/\theta^2$$

and the CRLB would be $\mathsf{Var}(\widehat{\theta}) \geq \theta^2/n^2$.

However, you can check (using $f_{X_{(n)}}(y; \theta)$ above) that

$$\mathsf{Var}(\widehat{\theta}) = \frac{\theta^2}{n(n+1)}$$

which is smaller than θ^2/n^2 for any $n \geq 1$.

This is not a contradiction, as $f(x; \theta)$ doesn't satisfy the regularity conditions (limits of x depend on θ).

We can find the MLE in this example

$$\widetilde{\theta} = X_{(n)}$$

but this estimator is

1. biased and the Uniform density does not satisfy the regularity condition needed for CRLB so it does not apply.
2. does not satisfy $\partial \ell / \partial \theta = 0$, so even if the CRLB did apply we cannot make the link between the MLE and the lower bound.

Exercise

Let $(X_1, \ldots, X_n)$ be i.i.d r.v.'s from $U[\theta - \frac{1}{2}, \theta + \frac{1}{2}]$ with $\theta \in \mathbb{R}$. Show that $T = (T_1, T_2) = (\min X_i, \max X_i)$ is sufficient but **not** complete.
Hint : choose $h(t_1, t_2) = t_2 - t_1 - \frac{n-1}{n+1}$

Example 15

A sample from $N(\mu, \sigma^2)$. The moment relations

$$\mathbb{E}\left[\sum_{i=1}^{n} X_i\right] = n\mu, \qquad \mathbb{E}\left[\sum_{i=1}^{n} X_i^2\right] = n(\mu^2 + \sigma^2)$$

lead to the estimating equations

$$\widehat{\mu} = \bar{X}, \qquad \widehat{\mu}^2 + \widehat{\sigma}^2 = n^{-1}\sum_{i=1}^{n} X_i^2.$$

We have seen that these are just the equations for the MLE's in this 2-dimensional exponential linear family, and solve to give the MLE

$$\widehat{\sigma}^2 = n^{-1}\sum_{i=1}^{n}(X_i - \bar{X})^2.$$

Example 16

A sample from Pois(λ). We have two possible moment relations

$$\mathbb{E}\left[\frac{1}{n}\sum_{i=1}^{n} X_i\right] = \lambda, \qquad \Rightarrow \widehat{\lambda}_1 = \frac{1}{n}\sum_{i=1}^{n} X_i$$

$$\mathbb{E}\left[\frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2\right] = \lambda, \qquad \Rightarrow \widehat{\lambda}_2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2$$

For a sample of size n we have
$A(\lambda) = \log\lambda$, $B(x) = \sum_{i=1}^{n} x_i$, $C(x) = -\log(\prod_{i=1}^{n} x_i!)$, $D(\lambda) = -n\lambda$

By Lemma 6 we have that $B(x) = \sum_{i=1}^{n} x_i$ is a sufficient complete statistic. Thus by the Lehmann-Scheffé Theorem we have that arithmetic mean $\widehat{\lambda}_1$ is MVUE. $\widehat{\lambda}_2$ is not a function of $B(x)$.

Exercise Show that $\widehat{\lambda}_1$ attains the CRLB.

Method of moments asymptotics in the case of sums

Suppose the dimension of θ is k and $\bar{H} = n^{-1} \sum_{i=1}^{n} F(x_i)$ where $\mathbb{E}(\bar{H}) = H(\theta)$ and $\widehat{\theta}$ is the solution to $\bar{H} = H(\theta)$.

Theorem (Asymptotic normality of moment estimators)

Assume that $\theta \to H(\theta)$ is 1-1 and $\mathcal{C}^1$ on an open set Θ_0, that $\nabla H(\theta)$ is non singular for all $\theta \in \Theta_0$ and that

$$E_\theta \left(\|F(X)\|^2 \right) < +\infty \quad \forall \theta \in \Theta_0$$

then as $n \to \infty$, $\widehat{\theta}$ is asymptotically normally distributed under P_θ where

$$\sqrt{n}(\widehat{\theta} - \theta) \Rightarrow \mathcal{N}(0, V_\theta), \quad V_\theta = [\nabla H^{-1}(\theta)] \mathit{Var}[F(X); \theta]([\nabla H^{-1}(\theta)])^t$$

Proof : Delta method

Central limit Theorem on $\bar{H}$
$\mathbb{E}(\bar{H}) = H(\theta)$, so by the Central Limit Theorem,

$$\sqrt{n}(\bar{H} - H(\theta)) \Rightarrow \mathcal{N}(0, \mathsf{Var}[F(X);\theta])$$

Delta method Then writing $Z_n = \sqrt{n}(\bar{H} - H(\theta))$

$$\begin{aligned}\widehat{\theta} &= H^{-1}(\bar{H}) = H^{-1}(H(\theta) + Z_n/\sqrt{n}) \\ &= H^{-1}(H(\theta)) + \nabla H^{-1}(\theta)\frac{Z_n}{\sqrt{n}} + o_p(Z_n/\sqrt{n})\end{aligned}$$

so

$$\sqrt{n}(\widehat{\theta} - \theta) = \nabla H^{-1}(\theta) Z_n + o_p(1)$$

Example 17 - Exponential distribution

$$f(x;\theta) = \theta \exp(-\theta x),\ x > 0$$

$$\mathbb{E}[\bar{X}] = \theta^{-1},\ \widehat{\theta} = \bar{X}^{-1}$$

That is

$$F(X) = X,\ H(\theta) = \theta^{-1},\ \bar{H} = \bar{X}$$

$$\mathsf{Var}\big[X;\theta\big] = \theta^{-2}, H'(\theta) = -\theta^{-2}$$

Now

$$\sqrt{n}(\widehat{\theta} - \theta) \Rightarrow \mathcal{N}\left(0, \frac{V(f(X);\theta)}{(H'(\theta))^2}\right) \equiv \mathcal{N}(0,\theta^2)$$

Exponential families - Method of moments and MLEs

MLE are moment estimates because we are solving

$$\mathbb{E}\Big[\sum_{i=1}^{n} T(X_i)\Big] = \sum_{i=1}^{n} T(x_i)$$

for $\widehat{\theta}$. We showed earlier that

$$\mathbb{E}\Big[T(X);\theta\Big] = D'(\theta)$$

so

$$\mathbb{E}\Big[\sum_{i=1}^{n} T(X_i);\theta\Big] = nD'(\theta)$$

Now

$$H(\theta) = D'(\theta),\ H'(\theta) = D''(\theta) = V_X(T(X);\theta) = I_1(\theta)$$

$$\sqrt{n}(\widehat{\theta} - \theta) \Rightarrow \mathcal{N}(0, (D''(\theta))^{-1})$$