Chapter 5

Lecture 16

Example 1(e). We have X_i iid $ae^{-b(x-\theta)^4}$, with $a, b > 0$ chosen so that this is a density and $\text{Var}_\theta(X_i) = 1$. Then

$$\ell_\theta(s) = \varphi_0(s)\exp\Big\{b\Big[\Big(4\sum_{i=1}^n X_i^3\Big)\theta - 6\Big(\sum_{i=1}^n X_i^2\Big)\theta^2 + 4\Big(\sum_{i=1}^n X_i\Big)\theta^3\Big] + A(\theta)\Big\},$$

which is not a one-parameter exponential family. It is called a "curved exponential family".

Sufficient conditions for the Cramér-Rao and Bhattacharya inequalities

As usual, we have $(S, \mathcal{A}, P_\theta)$, $\theta \in \Theta$, where Θ is an open subset of $\mathbb{R}^1$. μ is a fixed measure on S and $dP_\theta(s) = \ell_\theta(s)d\mu(s)$.

Condition 1. $\ell_\theta(s) > 0$ and $\delta \mapsto \ell_\delta(s)$ has, for each $s \in S$, a continuous derivative $\delta \mapsto \ell'_\delta(s)$. Let

$$\gamma_\theta^{(1)}(s) = \frac{\ell'_\theta(s)}{\ell_\theta(s)} = L'_\theta(s).$$

Condition 2. Given any $\theta \in \Theta$, we may find an $\varepsilon = \varepsilon(\theta) > 0$ such that $E_\theta(m_\theta^2) < +\infty$, where

$$m_\theta(s) = \sup_{|\delta-\theta|\le\varepsilon} |\gamma_\delta^{(1)}(s)|$$

– i.e., $m_\theta \in V_\theta$, which implies that $I(\theta) = E_\theta(\gamma_\theta^{(1)})^2 < +\infty$.

Condition 3. $I(\theta) > 0$.

12E Exact statement of Cramér-Rao inequality: Under conditions 1–3 above, if U_g is non-empty, then g is differentiable and

$$\text{Var}_\theta(t) \ge \frac{(g'(\theta))^2}{I(\theta)} \;\forall\theta \in \Theta, t \in U_g.$$

Proof.

i.

$$\Omega_{\delta,\theta} = \Omega_{\theta,\theta} + (\delta - \theta)\gamma_{\delta^*}^{(1)} = 1 + (\delta - \theta)\gamma_{\delta^*}^{(1)}$$

for some δ^* between θ and δ. By Condition 2, $\Omega_{\delta,\theta} \in V_\theta$ for $|\delta-\theta|$ sufficiently small.

ii.

$$\frac{\Omega_{\delta,\theta}(s) - 1}{\delta - \theta} - \gamma_\theta^{(1)}(s) = \gamma_{\delta^*}^{(1)}(s) - \gamma_\theta^{(1)}(s) \to 0$$

as $\delta \to \theta$ for all $s \in S$. (From Condition 1, $\gamma_\delta^{(1)}$ is continuous.) Also,

$$|\gamma_{\delta^*}^{(1)} - \gamma_\theta^{(1)}| \leq 2m_\theta \in V_\theta$$

and hence $E_\theta(\gamma_{\delta^*}^{(1)} - \gamma_\theta^{(1)})^2 \to 0$ as $\delta \to \theta$ (by dominated convergence) – i.e.,

$$\frac{\Omega_{\delta,\theta} - 1}{\delta - \theta} \xrightarrow{V_\theta} \gamma_\theta^{(1)}.$$

From this it follows that $\gamma_\theta^{(1)} \in W_\theta$.

iii. Choose $t \in U_g$. If we let $(\cdot,\cdot)$ and $||\cdot||$ be the inner product and norm, respectively, in V_θ, then $E_\delta(t) = E_\theta(t\Omega_{\delta,\theta}) = g(\delta)$ and so

$$(t, \Omega_{\delta,\theta} - 1) = g(\delta) - g(\theta) \Rightarrow (t - g(\theta), \Omega_{\delta,\theta} - 1) = g(\delta) - g(\theta)$$

(since $E_\theta(\Omega_{\delta,\theta} - 1) = 0$), whence

$$\left(t - g(\theta), \frac{\Omega_{\delta,\theta} - 1}{\delta - \theta}\right) = \frac{g(\delta) - g(\theta)}{\delta - \theta} \ \forall \delta \neq \theta.$$

From (ii) $\frac{g(\delta)-g(\theta)}{\delta-\theta}$ has a finite limit $(t - g(\theta), \gamma_\theta^{(1)})$ as $\delta \to \theta$. Thus g is differentiable and $g'(\theta) = (t - g(\theta), \gamma_\theta^{(1)})$, so that $|g'(\theta)| \leq ||t - g(\theta)||\ ||\gamma_\theta^{(1)}||$ – i.e., $\text{Var}_\theta(t) \geq \frac{[g'(\theta)]^2}{I(\theta)}$. □

Note. To know that $\int_S \ell'_\theta d\mu = 0 = \int_S \ell''_\theta d\mu$, it suffices to show that $\delta\ell''_\delta(s)$ exists and is continuous for each s and that

$$\int_S \{\max_{|\delta-\theta|\leq\varepsilon} |\ell''_\delta(s)|^2\} d\mu(s) < +\infty$$

for some $\varepsilon = \varepsilon(\theta) > 0$.

Note. Under Conditions 1–3, $\text{Span}\{1, \gamma_\theta^{(1)}\} = W_\theta^{(1)} \subseteq W_\theta$ and $1 \perp \gamma_\theta^{(1)}$ in V_θ. (Take $t \equiv 1$; then $(1, \Omega_{\delta,\theta}) \equiv 1$ and hence

$$\left(1, \frac{\Omega_{\delta,\theta} - 1}{\delta - \theta}\right) = 0 \ \forall \delta \neq \theta.$$

Letting $\delta \to \theta$, we have that $(1, \gamma_\theta^{(1)}) = 0$.)

Let k be a positive integer.

Condition 1_k. For each fixed s, $\theta \mapsto \ell_\theta(s)$ is positive and is k-times continuously differentiable.

Condition 2_k. Given any $\theta \in \Theta$, we may find an $\varepsilon = \varepsilon(\theta) > 0$ such that $E_\theta(m_\theta^2) < +\infty$, where

$$m_\theta(s) = \sup_{|\delta-\theta|\leq\varepsilon} |\gamma_\delta^{(k)}(s)|.$$

(From the above, we have that $1 \perp \gamma_\theta^{(j)}$ for $j = 1, \ldots, k$ – i.e., $E_\theta(\gamma_\theta^{(j)}) = 0$.)

Let $\Sigma_\theta^{(k)}$ be the covariance matrix of $\begin{pmatrix} \gamma_\theta^{(1)} \\ \vdots \\ \gamma_\theta^{(k)} \end{pmatrix}$.

Condition 3_k. $\Sigma_\theta^{(k)}$ is positive definite.

11E. If conditions 1_k–3_k hold and U_g is non-empty, then g is k-times continuously differentiable and

$$\mathrm{Var}_\theta(t) \geq b_k(\theta) \; \forall t \in U_g, \theta \in \Theta,$$

where $b_k(\theta) = h'(\theta)\left[\Sigma_\theta^{(k)}\right]^{-1} h(\theta)$ and $h(\theta) = \begin{pmatrix} g^{(1)}(\theta) \\ \vdots \\ g^{(k)}(\theta) \end{pmatrix}$ (Of course $g^{(j)} = \frac{d^j g}{d\theta^j}$.)

Proof (outline). $1, \gamma_\theta^{(1)}, \ldots, \gamma_\theta^{(k)} \in W_\theta$ and so $W_\theta^{(k)} \subseteq W_\theta$ and

$$\mathrm{Var}_\theta(t) \geq ||t_{\theta,k}^*||^2 - \left[g(\theta)\right]^2.$$

Lecture 17

Note.

i. $L'_\theta, L''_\theta, \ldots$ are derivatives of $\log_e \ell_\theta$, but $\gamma_\theta^{(1)} = \ell'_\theta/\ell_\theta, \gamma_\theta^{(2)} = \ell''_\theta/\ell_\theta, \ldots$ are *not* the same as $L'_\theta, L''_\theta, \ldots$.

ii. Condition 2 in (12E) can be weakened slightly to:

Condition 2'. Given any $\theta \in \Theta$, we may find an $\varepsilon = \varepsilon(\theta) > 0$ such that

$$E\left[\frac{\max_{|\delta-\theta|\leq\varepsilon} |\ell'_\delta(s)|}{\ell_\theta}\right]^2 < +\infty.$$

and condition 2_k in (11E) can be weakened to:

Condition $2_k'$. Given any $\theta \in \Theta$, we may find an $\varepsilon = \varepsilon(\theta) > 0$ such that

$$E\left[\frac{\max_{|\delta-\theta|\le\varepsilon} |d^k\ell_\delta(s)/d\delta^k|}{\ell_\theta(s)}\right]^2 < +\infty.$$

iii. Suppose that U_g is non-empty; then (8) implies that the projection of any $t \in U_g$ to W_θ is the (fixed) $\tilde{t} \in U_g \cap W_\theta$. Also, $t^*_{\theta,k}$ is the projection of any $t \in U_g$ to $W_\theta^{(k)} = \text{Span}\{1, \gamma_\theta^{(1)}, \ldots, \gamma_\theta^{(k)}\} \subseteq W_\theta$ – i.e., $t^*_{\theta,k}$ is the (affine) "regression" of any $t \in U_g$ on $\{\gamma_\theta^{(1)}, \ldots, \gamma_\theta^{(k)}\}$. Thus

$$t^*_{\theta,k} = g(\theta) + \alpha_1\gamma_\theta^{(1)} + \cdots + \alpha_k\gamma_\theta^{(k)},$$

where $\alpha_1, \ldots, \alpha_k$ are determined as in our discussion of regression, and

$$b_k(\theta) = E_\theta(t^*_{\theta,k})^2 - \left[g(\theta)\right]^2 = \text{Var}_\theta(\alpha_1\gamma_\theta^{(1)} + \cdots + \alpha_k\gamma_\theta^{(k)})$$
$$= \left(\frac{dg}{d\theta}, \cdots, \frac{d^kg}{d\theta^k}\right)\left(\Sigma_\theta^{(k)}\right)^{-1}\left(\frac{dg}{d\theta}, \cdots, \frac{d^kg}{d\theta^k}\right)'$$

by the regression formula.

iv.

$$b_1(\theta) \le b_2(\theta) \le \cdots \le b_k(\theta) \le \cdots$$

(where $b_1(\theta)$ is the C-R bound) because $W_\theta^{(k)} \subseteq W_\theta^{(k+1)}$. If we define $b(\theta) := \lim_{k\to\infty} b_k(\theta)$, then

$$b(\theta) \le \text{Var}_\theta(\tilde{t}),$$

the actual lower bound at θ for an unbiased estimate of g. We have that $b(\theta) = \text{Var}_\theta(\tilde{t})$ iff $\tilde{t} \in \text{Span}\{1, \gamma_\theta^{(1)}, \gamma_\theta^{(2)}, \ldots\}$. This does hold for any g with non-empty U_g *if* the subspace spanned by $\{1, \gamma_\theta^{(1)}, \ldots, \gamma_\theta^{(k)}, \ldots\}$ is W_θ. This sufficient condition for $b_k \to b$ and $t^*_{\theta,k} \to \tilde{t}$ is plausible since, by the Taylor expansion,

$$\Omega_{\delta,\theta} = 1 + (\delta - \theta)\gamma_\theta^{(1)} + \frac{(\delta-\theta)^2}{2!}\gamma_\theta^{(2)} + \cdots.$$

It holds rigorously in the following case:

15. (One-parameter exponential family) Suppose that

$$\ell_\theta(s) = C(s)e^{A(\theta)+B(\theta)T(s)}$$

where $C(s) > 0$, T is a fixed statistic and B is a continuous strictly monotone function on $\Theta \subseteq \mathbb{R}$; then, under Condition (*) below, we have

a. $W_\theta^{(k)} = \text{Span}\{1, T, \ldots, T^k\}$ for $k = 1, 2, 3, \ldots$.
b. $\text{Span}\{1, T, T^2, \ldots\} = W_\theta$ (under θ).

c. W_θ is the space of all Borel functions f of T such that $E_\theta\big(f(T)\big)^2 < +\infty$.
d. If U_g is non-empty, then $b_k(\theta) \to b(\theta) = \mathrm{Var}_\theta(\tilde{t})$.
e. $\tilde{t} = E_\theta(t \mid T)$ for all $\theta \in \Theta$ and $t \in U_g$.
f. SUFFICIENCY OF T: Given any $A \subseteq S$, we may find an $h(T)$ independent of θ such that $h(T) = P_\theta(A \mid T)$ for all $\theta \in \Theta$.

Proof. (f) follows from (e) by defining $g(\theta) = P_\theta(A)$ and $t = I_A \in U_g$ and applying (c).

(e) follows from (c) since projection to W_θ is then the same as taking conditional expectation.

(d) follows from (a) and (b) and the above notes.

It now remains only to prove (a)–(c). To this end, let $\xi = B(\delta) - B(\theta)$. Then ξ is the parameter, and takes values in a neighborhood of 0. We have

$$\frac{dP_\xi}{dP_0}(s) = \frac{C(s)e^{A(\delta)+B(\delta)T(s)}}{C(s)e^{A(\theta)+B(\theta)T(s)}} = e^{\xi T(s)-K}.$$

Suppose that

Condition (*). $\xi = B(\delta) - B(\theta)$ takes all values in a neighborhood of 0 as δ varies in a neighborhood of θ.

Under this condition,

$$\int_S e^{\xi T(s)-K} dP_0(s) = \int_S dP_\xi(s) = 1$$

and hence the MGF of T exists for ξ in a neighborhood of 0, and

$$K = K(\xi) = \log_e \int e^{\xi T(s)} dP_0(s)$$

is the cumulant generating function of T under P_θ.

Thus the family of probabilities on S is $\{P_\xi : \xi \text{ in a neighborhood of } 0\}$, where $dP_\xi(s) = e^{\xi T(s)-K(\xi)} dP_0(s)$ – i.e., a one-parameter exponential family with ξ as the "natural" parameter and $T(s)$ as the "natural" statistic. $W_\theta = \mathrm{Span}\{\Omega_{\delta,\theta} : \delta \in \Theta\}$; the spanning set includes $\{e^{\xi T(s)-K(\xi)} : \xi \text{ in a neighbourhood of } 0\}$, so W_θ contains the subspace spanned by $\{e^{\xi T} : \xi \text{ in a neighborhood of } 0\}$. Now

$$\frac{e^{\eta T} - e^{\xi T}}{\eta - \xi} = e^{\xi T}\left(\frac{e^{(\eta-\xi)T} - 1}{\eta - \xi}\right) = e^{\xi T}\frac{(1 + (\eta-\xi)T + \frac{1}{2}(\eta-\xi)^2 T^2 e^{(\eta^*-\xi)T} - 1)}{\eta - \xi}$$

for some η^* between η and ξ. We have, however, that $\frac{1}{2}(\eta-\xi)T^2 e^{(\eta^*-\xi)T} \xrightarrow{L^2} 0$ since the MGFs of T exist around 0. Hence

$$Te^{\xi T} = \lim_{\eta\to\xi} \frac{1}{\eta-\xi}(e^{\eta T} - e^{\xi T}) \in W_\theta.$$

Similarly, $T^2e^{\xi T}, T^3e^{\xi T}, \ldots$ are in W_θ. Taking $\xi = 0$, we get $\{1, T, T^2, \ldots\} \subseteq W_\theta$, so that the subspace spanned by $\{1, T, T^2, \ldots\}$ is in W_θ; but this subspace is the subspace of all square-integrable Borel functions of T, so $\text{Span}\{1, T, T^2, \ldots\} = W_\theta$ actually, since each $\Omega_{\delta,\theta}$ is a (square-integrable Borel) function of T. □

Example 2. Here $s = (X_1, \ldots, X_N)$, N the total number of trials in a Bernoulli sequence, and $\ell_\theta(s) = \theta^{T(s)}(1-\theta)^{N(s)-T(s)}$, where T, the total number of successes, is $X_1 + X_2 + \cdots + X_N$. In general, this is a curved exponential family.

In Example 2(a), since $N \equiv n$ (a constant),

$$\ell_\theta = e^{n \log_e(1-\theta) + T \log_e(\theta/(1-\theta))},$$

so that T is sufficient and any function of T is the UMVUE of its expected value. $C = \bigcap_{\theta \in \Theta} W_\theta$ is the set of all estimates of the form $f(T)$. The C-R bound b_1 is attained essentially only for $g(\theta) = -A'(\theta)/B'(\theta) = \theta$, i.e., for $g(\theta) = \alpha + \beta\theta$. The k^{th} Bhattacharya bound b_k is attained iff $g(\theta)$ is a polynomial of degree $k \leq n$. If $k > n$, then $b_k = b_n = b$.

Lecture 18

Note. In the context of (15), it is sometimes necessary to look at the distribution of the (sufficient) statistic T. Suppose that we have found the distribution function of T for a particular θ – say F_θ; then F_δ is given by

$$dF_\delta(x) = e^{[B(\delta)-B(\theta)]x + [A(\delta - A(\theta)]} dF_\theta(x),$$

where $x = T(s)$ (so that the distributions of T are a one-parameter exponential family with statistic the identity). (Please check, by computing, that $P_\delta(T \leq x) =: F_\delta(x) = \cdots$.)

Example 2(a).

Homework 4

1. U_g is non-empty iff g is a polynomial of degree $\leq n$ (in the case of Example 2(a)).

W_θ does not depend on θ; it is the class of all functions of $\overline{X}$, and hence an estimate is a UMVUE of its expected value iff it is a function of $\overline{X}$.

$$\text{Var}_\theta(\overline{X}) = \frac{\theta}{n} - \frac{\theta^2}{n} =: \sigma^2(\theta).$$

We will show that $\sigma^2(\theta)$ has a UMVUE when $n \geq 2$. This UMVUE should be a function of $\overline{X}$. $\frac{\theta}{n}$ may be estimated by $\frac{\overline{X}}{n}$. How about θ^2? Let

$$t = \begin{cases} 1 & \text{if } X_1 \text{ and } X_2 = 1 \\ 0 & \text{otherwise;} \end{cases}$$

then $E_\theta t = \theta^2$. We know that the projection to W_θ, which is $E_\theta(t \mid T)$, will give $\tilde{t}$ for $g(\theta) = \theta^2$. (Taking $E_\theta(t \mid T)$ is called "Blackwellization".)

$$E_\theta(t \mid T = k) = \frac{P_\theta(t = 1, T = k)}{P_\theta(T = k)}$$
$$= \frac{P_\theta(X_1 = 1 = X_2, \text{ exactly } k-2 \text{ successes in subsequent } n-2 \text{ trials})}{P_\theta(T = k)}$$
$$= \frac{\theta^2 \binom{n-2}{k-2} \theta^{k-2}(1-\theta)^{n-k}}{\binom{n}{k}\theta^k(1-\theta)^{n-k}} = \frac{\binom{n-2}{k-2}}{\binom{n}{k}} = \frac{k(k-1)}{n(n-1)},$$

which is independent of θ, as expected. Thus

$$\tilde{t} = \frac{T(T-1)}{n(n-1)},$$

which is the UMVUE of θ^2, and therefore $\sigma^2(\theta)$ may be estimated by

$$\frac{\overline{X}}{n} - \frac{\overline{X}}{n}\left(\frac{n\overline{X}-1}{n-1}\right) = \frac{\overline{X}}{n}\left[1 - \frac{n\overline{X}-1}{n-1}\right],$$

which is a function of $\overline{X}$ and hence is the UMVUE of $\sigma^2(\theta)$.

Consider the odds ratio $g(\theta) = \frac{\theta}{1-\theta}$. This has no unbiased estimate. Since θ has MLE $\overline{X}$, $\hat{t}$, the MLE for this g, is $\frac{\overline{X}}{1-\overline{X}}$. Since $P_\theta(\overline{X} = 1) = \theta^n > 0$, we have $E_\theta(\hat{t}) = +\infty$, so the expectation breaks down. If, however, $I(\theta) = \frac{n}{\theta(1-\theta)}$ is large – i.e., n is large – then

$$\hat{t} = \overline{X} + \cdots + \overline{X}^n + \frac{\overline{X}^{n+1}}{1-\overline{X}} = \overline{X} + \cdots + \overline{X}^n + R_n,$$

where $R_n = \frac{\overline{X}^{n+1}}{1-\overline{X}}$. For each $\theta \in (0,1)$, R_n is very small with large probability, and

$$\frac{R_n}{\theta^{n+1}} \to \frac{1}{1-\theta}$$

in P_θ-probability as $n \to \infty$.

Example 2(b) (Negative binomial sampling). Here

$$\ell_\theta = \theta^k(1-\theta)^{N-k} = \exp\left\{k \log\frac{\theta}{1-\theta} + k\log(1-\theta)\cdot y\right\},$$

where $y = N/k$, so that

$$T = y, \quad A = k\log(\theta/(1-\theta)) \quad \text{and} \quad B = k\log(1-\theta),$$

and hence $-A'(\theta)/B'(\theta) = 1/\theta$. Thus $E_\theta(y) = 1/\theta$ and $\mathrm{Var}_\theta(y)$ is the C-R bound, and the C-R bound is attained only for $g(\theta) = a + b/\theta$.

Now assume $k \geq 3$. We know (even for $k \geq 2$) that $\frac{k-1}{N-1}$ is an unbiased estimate of θ. Since $\frac{k-1}{N-1} = \frac{k-1}{ky-1}$ is a function of y, it is in fact the UMVUE of θ.

Let $\sigma^2(\theta) = \mathrm{Var}_\theta(\frac{k-1}{N-1})$. Since $\tilde{t} = \frac{k-1}{N-1}$ is not a polynomial in y – in fact, $\tilde{t} \notin W_{\theta,k}\ \forall k$ – we have (for $g(\theta) = \theta$)

$$b_1(\theta) < b_2(\theta) < \cdots < b_{k+1}(\theta) < \sigma^2(\theta),$$

but $b_k(\theta) \to \sigma^2(\theta)$ as $k \to \infty$. We can, however, find a UMVUE for $\sigma^2(\theta)$ (without knowing what the b_ks are).

Suppose that we can find an unbiased estimate u of θ^2. Then $v = \tilde{t}^2 - u$ is an unbiased estimate of $\sigma^2(\theta)$ ($\sigma^2(\theta) = \mathrm{Var}_\theta(\tilde{t}) = E_\theta(\tilde{t}^2) - \theta^2$).

Let

$$t = \begin{cases} 1 & \text{if } X_1 = 1 = X_2 \\ 0 & \text{otherwise.} \end{cases}$$

Then (even at present) $E_\theta(t) = \theta^2$ and hence $u = E_\theta(t \mid N)$ (the Blackwellization of t) is the UMVUE of θ^2 (when $k \geq 3$).

$$\begin{aligned} E_\theta(t \mid N = m) &= \frac{P_\theta(X_1 = 1 = X_2, N = m)}{P_\theta(N = m)} \\ &= \frac{\theta^2 \binom{m-3}{k-3} \theta^{k-3}(1-\theta)^{m-k}\theta}{\binom{m-1}{k-1}\theta^{k-1}(1-\theta)^{m-k}\theta} = \frac{\binom{m-3}{k-3}}{\binom{m-1}{k-1}} = \frac{(k-1)(k-2)}{(m-1)(m-2)} \end{aligned}$$

– i.e., $u = \frac{(k-1)(k-2)}{(N-1)(N-2)}$ is the UMVUE of θ^2, so that the UMVUE of $\sigma^2(\theta)$ is

$$\left(\frac{k-1}{N-1}\right)^2 - \frac{(k-1)(k-2)}{(N-1)(N-2)} = \frac{(k-1)(N-k)}{(N-1)^2(N-2)}.$$

Homework 4

2. Does every polynomial in θ have an unbiased estimate? (Yes?) Does $\frac{\theta}{1-\theta}$ have an unbiased estimate? (No?)