Chapter 3

Lecture 11

Unbiasedness has an appealing property, which we discuss here: Choose any estimate $t(s)$. Imagining for the moment that s is unknown but θ is provided, what is the best predictor for t?

Let λ be the prior; this determines M, as above. Regard t and g as elements of $L^2(M)$.

7. (t is an unbiased estimate of g) $\Leftrightarrow$ (for any choice of a probability λ on θ, g is the best (in MSE) predictor for t).

   Proof. If t is an unbiased estimate of g, then, for any λ, $E(t \mid \theta) = g$ – i.e., g is the projection of t to the subspace of functions in $L^2(M)$ which depend only on θ; or, equivalently, g is the best predictor of t in the sense of $||\cdot||_M$. Conversely, assume that each one-point set in Θ is measurable and take λ to be degenerate at a point θ. The assumption that g is the best predictor of t tells us that $g(\theta) = E(t \mid \theta)$ or, equivalently, that t is an unbiased estimate of g. □

Unbiased estimation; likelihood ratio

Choose and fix a $\theta \in \Theta$ and let $\delta \in \Theta$. Assume that P_δ is absolutely continuous with respect to P_θ on $\mathcal{A}$; then, by the Radon-Nikodym theorem, there exists an $\mathcal{A}$-measurable function $\Omega_{\delta,\theta}$ satisfying $0 \le \Omega_{\delta,\theta} \le +\infty$ and $dP_\delta = \Omega_{\delta,\theta} dP_\theta$ (i.e., $P_\delta(A) = \int_A \Omega_{\delta,\theta}(s) dP_\theta(s)$ for all $A \in \mathcal{A}$).

Note. Suppose that we begin with $dP_\delta(\theta) = \ell_\delta(s) d\mu(s)$ on S, where μ is given, and that we know that $P_\theta(A) = 0 \Rightarrow P_\delta(A) = 0$ (i.e., that P_δ is absolutely continuous with respect to P_θ). Then

$$\Omega_{\delta,\theta}(s) = \begin{cases} \ell_\delta(s)/\ell_\theta(s) & \text{if } 0 < \ell_\theta(s) < \infty \\ 1 & \text{if } \ell_\theta(s) = 0 \end{cases}$$

is an explicit formula for the likelihood ratio. In fact $\Omega_{\delta,\theta}$ can be defined arbitrarily on the set $\{s : \ell_\theta(s) = 0\}$.

In estimating g on the basis of s, let U_g be the class of all unbiased estimates of g. For an estimate $t \in U_g$, the risk function is given by $R_t(\theta) = E_\theta(t-g)^2 = \mathrm{Var}_\theta(t)$. Two questions arise immediately: What is the infimum (over U_g) of the variances at a given θ of the various estimates to g? Is it attained?

Remember that we fixed a $\theta \in \Theta$ above. Let $V_\theta = L^2(S, \mathcal{A}, P_\theta)$; then we assume throughout that

$$\{\Omega_{\delta,\theta} : \delta \in \Theta\} \subseteq V_\theta,$$

i.e., that $E_\theta(\Omega^2_{\delta,\theta}) < +\infty$. Let W_θ be the subspace of V_θ spanned by $\{\Omega_{\delta,\theta} : \delta \in \Theta\}$.

8. a. U_g is non-empty iff $U_g \cap W_\theta$ is non-empty.

We assume henceforth that U_g is non-empty. Then:

b. $U_g \cap W_\theta$ contains (essentially) only *one* estimate $\tilde{t}$.

c. $\tilde{t}$ is the orthogonal projection on W_θ of every $t \in U_g$.

d. $\mathrm{Var}_\theta(t) \geq \mathrm{Var}_\theta(\tilde{t})$ for all $t \in U_g$.

Note. The above means that $\tilde{t} \in U_g \cap W_\theta$ is the LMVUE of g at θ. $\tilde{t}$ often depends on θ, and this is the problem in practice.

Proof of (8). Note first that

1. $1 \in W_\theta$ (since $\Omega_{\theta,\theta} \equiv 1$).
2. For any t, $E_\delta(t) = \int_S t(s)dP_\delta(s) = \int_S t(s)\Omega_{\delta,\theta}(s)dP_\theta(s)$, so that $E_\delta(t) = (t, \Omega_{\delta,\theta})_\theta$, where $(\cdot,\cdot)_\theta$ is the inner product in $L^2(S, \mathcal{B}, P_\theta)$.

To prove (a), suppose that U_g is non-empty. Let $t \in U_g$ and define $\tilde{t} = \pi t$, where $\pi = \pi_{W_\theta}$ is the orthogonal projection on W_θ. Then, for any $\delta \in \Theta$,

$$E_\delta(\tilde{t}) = (\tilde{t}, \Omega_{\delta,\theta})_\theta = (\pi t, \Omega_{\delta,\theta})_\theta = (t, \pi\Omega_{\delta,\theta})_\theta = (t, \Omega_{\delta,\theta})_\theta = E_\delta(t) = g(\delta).$$

To prove (b), suppose $t_1, t_2 \in U_g \cap W_\theta$; then

$$(t_1 - t_2, \Omega_{\delta,\theta})_\theta = E_\delta(t_1 - t_2) = g(\delta) - g(\delta) = 0 \ \forall \delta \in \Theta.$$

Hence $(t_1 - t_2) \perp \Omega_{\delta,\theta}$ for all $\delta \in \Theta$, and so $(t_1 - t_2) \perp W_\theta$; but $t_1 - t_2 \in W_\theta$, so

$$(t_1 - t_2) \perp (t_1 - t_2) \Rightarrow t_1 - t_2 = 0 \Rightarrow P_\theta(t_1 = t_2) = 1.$$

It follows by absolute continuity that $P_\delta(t_1 = t_2) = 1$ for all $\delta \in \Theta$.

(c) follows from (b) and the above construction.

(d) follows from (c) since $\tilde{t}$ is unbiased for g. □

Note. In verifying (8), please remember that, if $E_\delta(t) = g(\delta) = E_\delta(\tilde{t})$ for all $\delta \in \Theta$, then $\mathrm{Var}_\theta(t) = E_\theta(t^2) - g(\theta)^2$ and $\mathrm{Var}_\theta(\tilde{t}^2) = E_\theta(\tilde{t}^2) - g(\theta)^2$, so that $\mathrm{Var}_\theta(\tilde{t}) \leq \mathrm{Var}_\theta(t)$, with equality iff $t = \tilde{t}$.

Lecture 12

We may restate (8) as follows:

8'. a. For some $\tilde{t} \in W_\theta$, $E_\delta(\tilde{t}) = E_\delta(t)$ for all $\delta \in \Theta$ and $t \in U_g$.

b. $\pi_{W_\theta} t$ is such a $\tilde{t}$, and is the (essentially) unique such.

c. We have that

$$R_{\tilde{t}}(\theta) = E_\theta\big(\tilde{t} - g(\theta)\big)^2 = \mathrm{Var}_\theta(\tilde{t}) + \big[b_t(\theta)\big]^2 \le \mathrm{Var}_\theta(t) + \big[b_t(\theta)\big]^2 = R_t(\theta)$$

with equality iff $t = \tilde{t}$.

d. $\tilde{t}$ is (essentially) the only unbiased estimate of g which belongs to W_θ.

9. a. An estimate t is the locally MVUE of $g(\delta) := E_\delta(t)$ at θ iff t has finite variance at each δ and $t \in W_\theta$.

b. An estimate t is the UMVUE of $g(\theta) := E_\theta(t)$ iff $t \in \bigcap_{\theta\in\Theta} W_\theta$ (we assume that $\Omega_{\delta,\theta} \in L^2(P_\theta)$ for all $\theta, \delta \in \Theta$).

9(b) above raises the question: Can we describe $C := \bigcap_{\theta\in\Theta} W_\theta$? We know it contains the constant functions; does it contain any others?

10 (Lehman-Scheffé). Write

$$\tilde{V} = \bigcap_{\theta\in\Theta} V_\theta \cap \{v : E_\delta(v) = 0 \ \forall \delta \in \Theta\}.$$

If t has finite variance for each δ (i.e., $t \in \bigcap_{\theta\in\Theta} V_\theta$), then $t \in C$ iff, for each $\delta \in \Theta$, we have

$$E_\delta(tu) = 0 \ \forall u \in \tilde{V}.$$

Proof. Suppose that $t \in C$. Then $t \perp_\delta W_\delta^\perp$ for all $\delta \in \Theta$. Now, for all $u \in \tilde{V}$, u is an unbiased estimate of 0; from (8), we know that 0 is the projection of u to any W_δ. Since $u = 0 + u$, we must therefore have $u \in W_\delta^\perp$, so that $t \perp_\delta u$ for each δ – i.e., $E_\delta(tu) = 0$ for all δ.

Conversely, fix a $\theta \in \Theta$ and write $t = \pi t + u$, where $u = t - \pi t$ and $\pi = \pi_{W_\theta}$. Then $E_\delta(u) = 0$ for all δ and hence, by hypothesis, we have that

$$\begin{aligned} E_\theta(u^2) + E_\theta(u \cdot \pi t) = E_\theta\big((\pi t + u)u\big) &= E_\theta(tu) = 0 \\ &\Rightarrow E_\theta(u^2) = -E_\theta(u \cdot \pi t) = -(\pi t, u) = 0 \end{aligned}$$

– i.e., $u = 0$ a.e.(P_θ) and hence, by absolute continuity of each P_δ, $u = 0$ a.e.(P_δ) also for every $\delta \in \Theta$. This means that $t = \pi t = \pi_{W_\theta} t \Rightarrow t \in W_\theta$; since $\theta \in \Theta$ was arbitrary, this means that $t \in \bigcap_{\theta\in\Theta} W_\theta = C$ as desired. □

Example 1(d). We have $s = (X_1, \dots, X_n)$, with the X_i iid as $N(\theta, 1)$, and $\Theta = \{1, 2\}$. We have explicitly that

$$\ell_\theta(s) \propto e^{-\frac{n}{2}(\overline{X}-\theta)^2}$$

and

$$\Omega_{\delta,\theta}(s) = e^{n(\delta-\theta)\overline{X} - \frac{n}{2}(\delta^2-\theta^2)}.$$

Choose $\theta = 1$; then

$$W_\theta = \operatorname{Span}\{\Omega_{11}, \Omega_{21}\} = \operatorname{Span}\{1, e^{n\overline{X}}\} = \{a + be^{n\overline{X}} : a, b \in \mathbb{R}\}.$$

Let $g(\delta) = \delta$. Since $\overline{X}$ is an unbiased estimate of g, we have a unique unbiased estimate of g in W_θ. Hence we want

$$\begin{aligned} E_1(a + be^{n\overline{X}}) &= 1 \\ E_2(a + be^{n\overline{X}}) &= 2 \end{aligned} \tag{*}$$

Since $\sqrt{n}(\overline{X} - \delta) \sim N(0, 1)$ for $\delta \in \Theta$, under δ, using the MGF of $N(0, 1)$, we have

$$E_\delta(e^{n\overline{X}}) = e^{n\delta} E_\delta(e^{\sqrt{n}\cdot\sqrt{n}(\overline{X}-\delta)}) = e^{n\delta + \frac{1}{2}n}$$

for any $\delta \in \Theta$. Solving (*), we find a and b ($b > 0$). Thus $a + be^{n\overline{X}}$ is LMVU for $E_\theta(X_1)$ at $\theta = 1$. This is not, however, a reasonable estimate. We already know that $\Theta = \{1, 2\}$, but this estimate takes values in $(-\infty, \infty)$. (Since Θ is not connected, we don't have Taylor's theorem here. Also, the LMVUE at $\theta = 2$ is a very different function of $\overline{X}$.) This is absurd. MSE is not suitable because g takes on only two values.

We try changing our parameter space to $\Theta = (\ell, u)$. Now

$$W_\theta = \operatorname{Span}\{\Omega_{\delta,\theta} : \ell < \delta < u\} = \operatorname{Span}\{e^{t\overline{X}} : t \text{ is sufficiently small}\}$$

(in the last set, 't is sufficiently small' means 'for t in a fixed neighbourhood of 0'). It can be shown that

$$W_\theta = \{f(\overline{X}) : f \text{ is a Borel function and } E_\theta f^2 < +\infty\}.$$

Proof (outline). Since $\frac{e^{t_1\overline{X}} - e^{t_2\overline{X}}}{t_1 - t_2} \in W_\theta$, we have that $\frac{d}{dt}e^{t\overline{X}} \in W_\theta$. Hence $\overline{X}e^{t\overline{X}} \in W_\theta$ for $|t|$ sufficiently small. Iterating this reasoning gives us that $\overline{X}^2 e^{t\overline{X}}, \dots, \overline{X}^i e^{t\overline{X}}, \dots \in W_\theta$ for $|t|$ sufficiently small (what "sufficiently small" means depends on i). Thus $1, \overline{X}, \overline{X}^2, \dots \in W_\theta$ and hence W_θ is as desired.

Since $\overline{X}$ is an unbiased estimate of $E_\delta(X_1)$ which belongs to W_θ, $\overline{X}$ is LMVU at θ, and hence $\overline{X}$ is the UMVUE; since $\overline{X}^2 - \frac{1}{n}$ is an unbiased estimate of $\left[E_\delta(X_1)\right]^2$, $\overline{X}^2 - \frac{1}{n}$ is the UMVUE for $\left[E_\delta(X_1)\right]^2$. (Here W_θ essentially does not depend on θ, and

$$C = \{f(\overline{X}) : f \text{ is Borel and } E_\theta f^2 < +\infty \ \forall \theta \in \Theta\}$$

by our above computation.)

Let $A \subseteq S$ be such that $P_1(A) \neq P_2(A)$(for example, $A = \{s : X_1(s) > 3/2\}$). Then $a + bI_A$ is an unbiased estimate of θ if a and b are chosen properly. Indeed, there are many unbiased estimates. To find the "best", we try to minimize variances, noting that

$$W_1 = \text{Span}\{\Omega_{11}, \Omega_{21}\} = \{a + be^{n\overline{X}} : a, b \in \mathbb{R}\}$$

is the class of all estimates which are unbiased for their own expected values and have minimum variance when $\theta = 1$ and hence that there is a $t_1 \in W_1$ such that $E_\delta(t_1) = \delta$ for $\delta = 1, 2$. (*Exercise*: What is t_1?).

Similarly, $W_2 = \{a + be^{-n\overline{X}} : a, b \in \mathbb{R}\}$ and there is a $t_2 \in W_2$ such that $E_\delta(t_2) = \delta$ for $\delta = 1, 2$. (*Exercise*: What is t_2?) $t_1 \neq t_2$, however; in fact, C is the set of all UMVUEs, which is just the set of constant functions.

As noted, the Neyman-Pearson theory implies that we should use $a + bI_A$ with $A = \{s : \overline{X} > c\}$ and $b > 0$. We should also restrict the estimation theory to a continuum of values (i.e., should have only connected Θ).