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 $\{s : \ell_\theta(s) = 0\}$.