1 Notation

- (statistical model) $X \sim P_\theta$ with values in $\mathcal{X}$, where $\theta \in \Omega$ is the parameter.
- (loss and risk) $L(\theta, d)$ and $R(\theta, \delta) = \mathbb{E}_\theta[L(\theta, \delta(X))]$, where $\delta(X)$ is a statistic.

2 Unbiasedness

- We say that a statistic T is complete if the only function f such that $\mathbb{E}_\theta[f(T(X)] = 0$ for all θ is $f \equiv 0$.

U1 Assume T is a complete and sufficient statistic. Then for every function $g(\theta)$ that admits an unbiased estimator there is a (unique) UMVU given by the unique unbiased estimator that is a function of T. This is true for squared error loss, and any other convex loss function.

U2 (Cramér-Rao information bound.) Suppose that $\mathbb{P}_\theta$ has density f_θ with respect to some (dominating) measure on $\mathcal{X}$. Under some regularity conditions, for an estimator δ such that $\mathbb{E}_\theta[\delta(X)] = g(\theta)$, we have $\mathrm{Var}_\theta[\delta(X)] \geq [\dot{g}(\theta)]^2/I(\theta)$, where $I(\theta) = \mathbb{E}_\theta[(\dot{f}_\theta(X)/f_\theta(X))^2] = \mathrm{Var}_\theta[\dot{f}_\theta(X)/f_\theta(X)] = -\mathbb{E}_\theta[\partial^2 \log f_\theta(X)]$.

3 Equivariance

- We say that a transformation $g : \mathcal{X} \to \mathcal{X}$ leaves the model invariant if *(i)* it is one-to-one; *(ii)* if for any $\theta \in \Omega$, $X \sim \mathbb{P}_\theta$ implies $gX \sim \mathbb{P}_{\theta'}$ for some $\theta' \in \Omega$, which allows one to define $\bar{g} : \Omega \to \Omega$ that associates θ' to θ; *(iii)* $\bar{g}$ is one-to-one (here we assume the model is identifiable).
- We work with a group G of transformations, each leaving the model invariant, and define $\bar{G} = \{\bar{g} : g \in G\}$.
- We work with an estimand $h(\theta)$ that is equivariant with respect to $\bar{G}$. This allows one to define g^* for each $g \in G$ such that $h(\bar{g}\theta) = g^* h(\theta)$ for all $\theta \in \Omega$.
- We work with a loss function L that is invariant in that $L(\bar{g}\theta, g^* d) = L(\theta, d)$ for all θ, all d, and all g.
- We then focus on estimators that are equivariant in the sense that $\delta(gx) = g^*\delta(x)$ for all x and all g.

3.1 Location model

This model is of the form $X = (X_1, \dots, X_n) \sim f(x - \xi)$ where f is a given density on $\mathbb{R}^n$ and $\xi \in \mathbb{R}$ is unknown. (As usual, $x - \xi$ is understood coordinate-wise.) The transformations of interest are of the form $x \mapsto a + x$ where $a \in \mathbb{R}$. We want to estimate ξ and work with a loss of the form $L(\xi, d) = \rho(d - \xi)$. Let $Y = (Y_1, \dots, Y_{n-1})$ with $Y_i = X_i - X_n$.

E1 Let δ_0 be any equivariant statistic with finite risk, and suppose we may define $v^*(y) = \arg\min_{v \in \mathbb{R}} \mathbb{E}_0[\rho(\delta_0(X) - v) \mid Y = y]$. Then $\delta^*(x) = \delta_0(x) - v^*(y)$ is MRE.

E2 Under squared error loss, the MRE may be expressed as $\delta^*(x) = \int_{-\infty}^{\infty} v f(x - v) \mathrm{d}v / \int_{-\infty}^{\infty} f(x - v) \mathrm{d}v$.

E3 Under squared error loss, if an UMVUE exists and is equivariant, then it is MRE.

3.2 Scale model

This model is of the form $X = (X_1, \dots, X_n) \sim \tau^{-1} f(x/\tau)$ where f is a given density on $\mathbb{R}^n$ and $\tau > 0$ is unknown. The transformations of interest are of the form $x \mapsto bx$ where $b > 0$. We want to estimate τ and work with a loss of the form $L(\xi, d) = \gamma(d/\tau)$. Let $Z = (Z_1, \dots, Z_n)$ with $Z_i = X_i/X_n$ for $i \neq n$ and $Z_n = X_n/|X_n|$.

E4 In the present setting, let δ_0 be any equivariant statistic with finite risk, and suppose we may define $w^*(z) =$

$\arg\min_{w>0} \mathbb{E}_1[\gamma(\delta_0(X)/w) \mid Z = z]$. Then $\delta^*(x) = \delta_0(x)/w^*(z)$ is MRE.

E5 When $L(\tau, d) = (d/\tau - 1)^2$, the MRE may be expressed as $\delta^*(x) = \int_0^\infty w^n f(wx)\mathrm{d}w / \int_0^\infty w^{n+1} f(wx)\mathrm{d}w$.

3.3 Location-scale model

This model is of the form $X = (X_1, \ldots, X_n) \sim \tau^{-n} f((x - \xi)/\tau)$ where f is a given density on $\mathbb{R}^n$ and $\xi \in \mathbb{R}$ and $\tau > 0$ are both unknown. The transformations of interest are of the form $x \mapsto a + bx$ where $a \in \mathbb{R}$ and $b > 0$. When our goal is to estimate ξ, we work with a loss of the form $L(\xi, \tau; d) = \rho((d - \xi)/\tau)$. When our goal is to estimate τ, we work with a loss of the form $L(\xi, \tau; d) = \gamma(d/\tau)$.

4 Average risk

- The setting here is that $\Theta \sim \Lambda$ and $X \mid \Theta = \theta \sim P_\theta$.
- (average risk) For a prior Λ and estimator δ, we denote by $r(\Lambda, \delta) = \int R(\theta, \delta)\Lambda(\mathrm{d}\theta) = \mathbb{E}[L(\Theta, \delta(X))]$ its average risk with respect to Λ. We also let δ_Λ denote a Bayes estimator (when one exists) and $r_\Lambda = r(\Lambda, \delta_\Lambda)$ the average risk of that estimator, also called the Bayes risk.

B1 Let δ_0 be any statistic with finite average risk, and suppose we may define $\delta^*(x) = \arg\min_d \mathbb{E}[L(\Theta, d) \mid X = x]$. Then δ^* is Bayes.

B2 Under loss $L(\theta, d) = w(\theta)(d - h(\theta))^2$, the Bayes estimator is $\delta_\Lambda(x) = \mathbb{E}[w(\theta)h(\theta) \mid X = x] / \mathbb{E}[w(\theta) \mid X = x]$.

B3 In a Bayesian setting, suppose the loss is strictly convex (in d) and that Q denotes the marginal of X. Then the Bayes estimator is unique if the Bayes risk is finite and, for any measurable set A, $Q(A) = 0$ implies $P_\theta(A) = 0$ for all $\theta \in \Omega$.

5 Maximum risk and admissibility

5.1 Minimaxity

- (maximum risk) For an estimator δ we let $\bar{R}(\delta) = \sup_{\theta \in \Omega} R(\theta, \delta)$, which is its maximum risk.

M1 Suppose that Λ is a prior such that $r_\Lambda = \bar{R}(\delta_\Lambda)$. Then δ_Λ is minimax, and uniquely so if it is unique Bayes.

M2 If an estimator is Bayes for some prior and has constant risk, it is minimax.

M3 If, for an estimator δ, we can find a sequence of priors (Λ_k) such that $\liminf_k r_{\Lambda_k} \geq \bar{R}(\delta)$, then δ is minimax.

M4 Consider $\Omega_0 \subset \Omega$. If an estimator is minimax over Ω_0 and achieves its maximum risk at some $\theta \in \Omega_0$, then this estimator is also minimax over Ω.

5.2 Admissibility

A1 A unique Bayes estimator is admissible.

A2 (Karlin's theorem) Suppose $X \sim f_\theta$, where $f_\theta(x) = \beta(\theta)e^{\theta T(x)}$ with respect to some underlying measure. Let $\Omega = [\theta_*, \theta^*]$ denote the natural parameter space. Suppose $L(\theta, d) = (d - h(\theta))^2$, where $h(\theta) = \mathbb{E}_\theta(T)$. Then, for $a \geq 0$ and $b \in \mathbb{R}$, a sufficient condition for $\frac{1}{1+a}T + \frac{a}{1+a}b$ to be admissible is that

$$\int_{\theta_*}^0 K(\theta)\mathrm{d}\theta = \infty \text{ and } \int_0^{\theta^*} K(\theta)\mathrm{d}\theta = \infty$$

where $K(\theta) = e^{-ba\theta}\beta(\theta)^{-a}$.

A3 If an estimator has constant risk and is admissible, it is minimax.

A4 If an estimator is unique minimax, it is admissible.