7 Minimax Estimation

Previously we have looked at Bayes estimation when the overall measure of the estimation error is weighted over all values of the parameter space, with a positive weight function, a prior. Now we use the maximum risk as a relevant measure of the estimation risk,

$$\sup_\theta R(\theta, \delta),$$

and define the minimax estimator, if it exists, as the estimator that minimizes this risk, i.e.

$$\hat{\delta} \;=\; \arg\inf_\delta \sup_\theta R(\theta, \delta),$$

where the infimum is taken over all functions of the data.

To relate the minimax estimator with the Bayes estimator the Bayes risk

$$r_\Lambda \;=\; \int R(\theta, \delta_\lambda)\, d\Lambda(\theta)$$

is of interest.

Definition 16 *The prior distribution Λ is called least favorable for estimating $g(\theta)$ if $r_\Lambda \geq r_{\Lambda'}$ for every other distribution Λ' on Ω.*

When a Bayes estimator has a Bayes risk that attains the minimax risk it is minimax:

Theorem 11 *Assume Λ is a prior distribution and assume the Bayes estimator δ_Λ satisfies*

$$\int R(\theta, \delta_\Lambda)\, d\Lambda(\theta) \;=\; \sup_\theta R(\theta, \delta_\Lambda).$$

Then

- (i) *δ_Λ is minimax.*
- (ii) *If δ_Λ is the unique Bayes estimator, it is the unique mimimax estimator.*
- (iii) *Λ is least favorable.*

Proof. Let $\delta \neq \delta_\Lambda$ be another estimator. Then

$$\begin{aligned} \sup_\theta R(\theta, \delta) &\geq \int R(\theta, \delta)\, d\Lambda(\theta) \geq \int R(\theta, \delta_\Lambda)\, d\Lambda(\theta) \\ &= \sup_\theta R(\theta, \delta_\lambda), \end{aligned}$$

which proves that δ_Λ is minimax.

To prove (ii), uniqueness of the Bayes estimator means that no other estimator has the same Bayes risk, so $>$ replaces $\geq$ in the second inequality above, and thus no other estimator has the same maximum risk, which proves the uniqueness of the minimax estimator.

To show (iii), let $\Lambda' \neq \Lambda$ be a distribution on Ω, and let $\delta_\Lambda, \delta_{\Lambda'}$ be the corresponding Bayes estimators. The the Bayes risks are related by

$$\begin{aligned} r_{\Lambda'} &= \int R(\theta, \delta_{\Lambda'})\, d\Lambda'(\theta) \leq \int R(\theta, \delta_\Lambda)\, d\Lambda'(\theta) \leq \sup_\theta R(\theta, \delta_\Lambda) \\ &= r_\Lambda, \end{aligned}$$

where the first inequality follows since $\delta_{\Lambda'}$ is the Bayes estimator corresponding to Λ'. We have proven that δ_Λ is least favorable. □

Note 1 *The previous result says minimax estimators are Bayes estimators with respect to the least favorable prior.*

Two simple cases when the Bayes estimator's Bayes risk attains the minimax risk are given in the following.

Corollary 3 *Assume the Bayes estimator δ_Λ has constant risk, i.e. $R(\theta, \delta_\Lambda)$ does not depend on θ. Then δ_Λ is minimax.*

Proof. If the risk is constant the supremum over θ is equal to the average over θ so the Bayes and the minimax risk are the same, and the result follows from the previous theorem. □

Corollary 4 *Let δ_Λ be the Bayes estimator for Λ, and define*

$$\Omega_\Lambda = \{\theta \in \Omega : R(\theta, \delta_\Lambda) = \sup_{\theta'} R(\theta', \delta_\Lambda)\}.$$

Then if $\Lambda(\Omega_\Lambda) = 1$, δ_Λ is minimax.

Proof. The condition $\Lambda(\Omega_\Lambda) = 1$ means that the Bayes estimator has constant risk Λ-almost surely, and since the Bayes estimator is only determined modulo Λ-null sets, this is enough. □

Example 32 *Let $X \in \mathrm{Bin}(n, p)$. Let $\Lambda = B(a, b)$ be the Beta distribution for p. As previously established, via the conditional distribution of p given X (which is Beta also), the Bayes estimator of p is*

$$\delta_\Lambda(x) = \frac{a + x}{a + b + n},$$

with risk function, for quadratic loss,

$$\begin{aligned}
R(p,\delta_\Lambda) &= \mathrm{Var}(\delta_\Lambda) + (E(\delta_\Lambda)-p)^2 \\
&= \frac{1}{(a+b+n)^2}npq + \left(\frac{a+np-p(a+b+n)}{a+b+n}\right)^2 \\
&= \frac{1}{(a+b+n)^2}\left(npq+(aq-bp)^2\right) \\
&= \frac{a^2+(n-2a(a+b))p+((a+b)^2-n)p^2}{(a+b+n)^2}.
\end{aligned}$$

Since this function is a quadratic in p it is constant, as a function of p, if and only if the coefficients of p and p^2 are zero, i.e. iff

$$(a+b)^2 = n, \qquad 2a(a+b) = n,$$

which has a solution $a = b = \sqrt{n}/2$.

Thus

$$\delta_\Lambda(x) = \frac{x+\sqrt{n}/2}{n+\sqrt{n}}$$

is a constant risk Bayes estimator and therefore minimax for the (least favorable) distribution $\Lambda = B(\sqrt{n}/2, \sqrt{n}/2)$. The Bayes estimator is unique and therefore this is the unique minimax estimator for p, with respect to $\Lambda = B(\sqrt{n}/2, \sqrt{n}/2)$.

Now let Λ be an arbitrary distribution on $[0,1]$. Then the Bayes estimator of p is $\delta_\Lambda(x) = E(p|x)$, i.e.

$$\begin{aligned}
\delta_\Lambda(x) &= \int_0^1 p\, d\Lambda(p|x) = \int_0^1 p f(x|p)\frac{1}{f(x)}\, d\Lambda(p) \\
&= \frac{\int_1^1 pp^x(1-p)^{n-x}d\Lambda(p)}{\int_0^1 p^x(1-p)^{n-x}d\Lambda(p)}.
\end{aligned}$$

Using the power expansion

$$(1-p)^{n-x} = 1 + a_1 p + \ldots + a_{n-x}p^{n-x},$$

this becomes

$$\delta_\Lambda(x) = \frac{\int_0^1 p^{x+1} + a_1 p^{x+2} + \ldots + a_{n-x}p^{n+1}\, d\Lambda(p)}{\int p^x + a_1 p^{x+1} + \ldots + a_{n-x}p^n\, d\Lambda(p)},$$

which shows that the Bayes estimator depends on the distribution Λ only via the $n+1$ first moments of Λ. Therefore the least favorable distribution is not unique for estimating p in a $\mathrm{Bin}(n,p)$ distribution: Two priors with the same first $n+1$ moments will give the same Bayes estimator. □

Recall that when the loss function is convex, then for any randomized estimator there is a nonrandomized estimator which has at least as small a risk as the randomized, so then there is no need to consider randomized estimators.

The relation that was established between the minimax estimator and the Bayes estimator, and obtaining the prior Λ as least favorable distribution, is valid when Λ is proper prior. What happens when Λ is not proper? Sometimes the estimation problem at hand makes it natural to consider such an improper prior: One such situation is when estimating the mean in a Normal distribution with known variance, and the mean is unrestricted i.e. it is a real number. Then one could believe that the least favorable distribution is the Lebesgue measure on $\mathbb{R}$.

To model this, assume Λ is a fixed (improper) prior and let $\{\Lambda_n\}$ be a sequence of priors that in some sense approximate Λ:

Definition 17 *Let $\{\Lambda_n\}$ be a sequence of priors and let δ_n be the Bayes estimator corresponding to Λ_n with*

$$r_n = \int R(\theta, \delta_n)\, d\Lambda_n(\theta)$$

the Bayes risk. Define $r = \lim_{n \to} r_n$.

If $r_{\Lambda'} \leq r$ for every (proper) prior distribution Λ' then the sequence $\{\Lambda_n\}$ is called least favorable.

Now if δ is an estimator whose maximal risk attains this limiting Bayes risk, then δ is minimax and the sequence is least favorable:

Theorem 12 *Let $\{\Lambda_n\}$ be a sequence of priors and let $r = \lim_{n\to\infty} r_n$. Assume the estimator δ satisfies*

$$r = \sup_\theta R(\theta, \delta).$$

Then δ is minimax and the sequence $\{\Lambda_n\}$ is least favorable.

Proof. To prove the minimaxity: Let δ' be any other estimator. Then

$$\sup_\theta R(\theta, \delta') \geq \int R(\theta, \delta')\, d\Lambda(\theta) \geq r_n,$$

for every n. Since the right hand side converges to $r = \sup_\theta R(\theta, \delta)$ this implies that

$$\sup_\theta R(\theta, \delta') \geq \sup_\theta R(\theta, \delta),$$

and thus δ is minimax.

To prove that $\{\Lambda_n\}$ is least favorable: Let Λ be any (proper) prior distribution and let δ_Λ be the corresponding Bayes estimator. Then

$$\begin{aligned} r_\Lambda &= \int R(\theta, \delta_\Lambda)\, d\Lambda(\theta) \leq \int R(\theta, \delta)\, d\Lambda(\theta) \\ &\leq \sup_\theta R(\theta, \delta) = r, \end{aligned}$$

and thus $\{\Lambda_n\}$ is least favorable. □

Note 2 *Uniqueness of the Bayes estimators* δ_n *does not imply uniqueness of the minimax estimator, since in that case the strict inequality in* $r_n = \int R(\theta, \delta_n)\, d\Lambda_n(\theta) < \int R(\theta, \delta')\, d\Lambda(\theta)$ *is transformed to* $r \leq \int R(\theta, \delta')\, d\Lambda(\theta)$ *under the limit operation.*

To calculate the Bayes risk r_n for δ_n, the following is useful.

Lemma 9 *Let* δ_Λ *be the Bayes estimator of* $g(\theta)$ *corresponding to* Λ*, and assume quadratic loss. Then the Bayes risk is given by*

$$r_\Lambda = \int \mathrm{Var}(g(\Theta)|x)\, dP(x).$$

Proof. The Bayes estimator is for quadratic loss given by $\delta_\Lambda(x) = E(g(\Theta)|x)$, and thus the Bayes risk is

$$\begin{aligned}
r_\Lambda &= \int R(\theta, \delta_\Lambda) d\Lambda(\theta) = \text{...(Fubinis theorem)...} = \\
&= \int E([\delta_\Lambda(x) - g(\Theta)]^2|x)\, dP(x) \\
&= \int E([E(g(\Theta)|x) - g(\Theta)]^2|x)\, dP(x) \\
&= \int \mathrm{Var}(g(\Theta)|x)\, dP(x).
\end{aligned}$$

□

Note 3 *If the conditional variance* $\mathrm{Var}(g(\Theta)|x)$ *is independent of* x*, the Bayes risk can be obtained as* $\mathrm{Var}(g(\Theta)|x)$*.*

Example 33 *Let* $X = (X_1, \ldots, X_n)$ *be an i.i.d sequence with* $X_1 \in N(\theta, \sigma^2)$*. We want to estimate* $g(\theta) = \theta$ *and assume quadratic loss. Assume* σ^2 *is known. Then* $\delta = \bar{X}$ *is minimax: To prove this we will find a sequence a Bayes estimators* δ_n *whose Bayes risk satisfies* $r_n \to \sigma^2/2 = \sup_\theta R(\theta, \delta)$*; the last equality holds since the risk of* $\bar{X}$ *does not depend on* θ*.*

Let θ *be distributed according to the prior* $\Lambda = N(\mu, b^2)$*. Then the Bayes estimator is*

$$\delta_\Lambda(x) = \frac{n\bar{x}/\sigma^2 + \mu/b^2}{n/\sigma^2 + 1/b^2},$$

with posterior variance

$$\mathrm{Var}(\Theta|x) = \frac{1}{n/\sigma^2 + 1/b^2},$$

which, since it is independent of x*, implies the Bayes risk*

$$r_\Lambda = \frac{1}{n/\sigma^2 + 1/b^2}.$$

If $b \to \infty$*, then* $r_\Lambda = r_{\Lambda_b} \to \sigma^2/n$*. Thus* $\delta = \bar{X}$ *is minimax.* □

Lemma 10 *Let $\mathcal{F}_1 \subset \mathcal{F}$ be sets of distributions. Let $g(F)$ be an estimand (a functional) defined on $\mathcal{F}$. Assume δ_1 is minimax over $\mathcal{F}_1$. Assume*

$$\sup_{F \in \mathcal{F}} R(F, \delta_1) \quad = \quad \sup_{F \in \mathcal{F}_1} R(F, \delta_1).$$

Then δ_1 is minimax over $\mathcal{F}$.

Proof. Assume δ_1 is minimax over $\mathcal{F}_1$. Since for every δ,

$$\sup_{F \in \mathcal{F}_1} R(F, \delta) \quad \leq \quad \sup_{F \in \mathcal{F}} R(F, \delta)$$

we get

$$\sup_{F \in \mathcal{F}} R(F, \delta_1) = \sup_{F \in \mathcal{F}_1} R(F, \delta_1) = \inf_{\delta} \sup_{F \in \mathcal{F}_1} R(F, \delta) \leq \inf_{\delta} \sup_{F \in \mathcal{F}} R(F, \delta).$$

Thus we have equality in the last inequality, and thus δ_1 is minimax over $\mathcal{F}$. $\square$

7.1 Minimax estimation in exponential families

Recall that the standard approach to estimation in exponential families is to make a restriction to the unbiased estimator, and in this restricted family find the estimator (if it exists) that has smallest risk (or variance for quadratic loss), uniformly over the parameter space (the UMVU), or locally for a fixed parameter (the LMVU). If the UMVU estimator exists it might be inadmissable, i.e. there might be another estimator which has uniformly smaller risk but that is biased.

For Bayes estimators this can not happen.

Theorem 13 *Assume the Bayes estimator δ_Λ is unique, P_θ-a.s. for every $\theta \in \Omega$. Then δ_Λ is admissible.*

Proof. Assume δ_Λ is inadmissible and so it is is dominated by some δ' so that

$$\int R(\theta, \delta') \, d\Lambda(\theta) \quad \leq \quad \int R(\theta, \delta_\Lambda) \, d\Lambda(\theta).$$

Then δ' is also Bayes and uniqueness implies that $\delta' = \delta_\Lambda$ P_θ-a.s., i.e. δ_Λ is admissible. $\square$

Example 34 *Assume $X = (X_1, \ldots, X_n)$ is a random sample from a $N(\theta, \sigma^2)$ distribution with σ^2 known. We have previously shown that $\delta = \bar{X}$ is minimax for estimating θ. To investigate whether it is admissible define instead the estimator*

$$\delta_{a,b} \quad = \quad a\bar{X} + b.$$

Is this estimator admissible?

Let $\Lambda = N(\mu, \tau^2)$ be a prior distribution for θ. Then as previously shown the unique Bayes estimator of θ is

$$\delta_\Lambda(X) = \frac{n\tau^2}{\sigma^2 + n\tau^2}\bar{X} + \frac{\sigma^2}{\sigma^2 + n\tau^2}\mu.$$

Therefore δ_Λ is admissible.

So if the factor a satisfies $0 < a < 1$, for (the fixed n and σ^2 at hand) there is a prior parameter τ such that $n\tau^2/(\sigma^2 + n\tau^2) = a$. Thus for $0 < a < 1$ the estimator $\delta_{a,b}$ is admissible. □

What happens with the admissibility of $a\bar{X} + b$ for the other possible values of a, b?

Lemma 11 *Assume $X \in N(\theta, \sigma^2)$. Then the estimator $\delta(X) = aX + b$ is inadmissible whenever*

$$\begin{array}{rl} (i) & a > 1, \\ (ii) & a < 0, \\ (iii) & a = 1 \text{ and } b \neq 0. \end{array}$$

Proof. The risk of δ is

$$\begin{aligned} R(\theta, \delta) &= E(aX + b - \theta)^2 \\ &= a^2\sigma^2 + ((a-1)\theta + b)^2 =: \rho(a, b). \end{aligned}$$

Thus

$$\rho(a, b) \geq a^2\sigma^2 > \sigma^2,$$

when $a > 1$, so then δ is dominated by X, which has risk σ^2, which proves (i).

Furthermore, when $a < 0$,

$$\begin{aligned} \rho(a, b) &\geq ((a-1)\theta + b)^2 \\ &= (a-1)^2\left(\theta + \frac{b}{a-1}\right)^2 \\ &> \left(\theta + \frac{b}{a-1}\right)^2 \\ &= \rho\left(0, -\frac{b}{a-1}\right), \end{aligned}$$

so then δ is dominated by the constant estimator $-b/(a-1)$, which proves (ii).'

Finally, the estimator $X - b$ has risk

$$\begin{aligned} r(a, b) &= \sigma^2 + b^2 \\ &> \sigma^2, \end{aligned}$$

and is therefore dominated by the estimator X, which proves (iii). □

Example 35 *(ctd.) The previous example and Theorem implies that $\delta = a\bar{X} + b$ is admissible for estimating θ in a $N(\theta, \sigma^2)$ distribution, when $0 < a < 1$. Also at a the estimator is $\delta = b$ which is the only estimator with zero risk so it is admissible also then.*

The estimator is inadmissible when $a < 0$ or $a > 1$ and when $a = 1, b \neq 0$. □

What happens when $a = 1, b = 0$, i.e. for the estimator $\delta(X) = \bar{X}$.

Example 36 *Assume $X = (X_1, \ldots, X_n)$ is a random sample from a $N(\theta, \sigma^2)$ distribution with $\sigma^2 = 1$, and assume we have squared error loss. Assume that $\delta = \bar{X}$ is inadmissible, with δ^* a dominating estimator.*

Thus, since δ is unbiased with variance $1/n$, we have

$$R(\theta, \delta^*) \quad \leq \quad \frac{1}{n},$$

for all θ, with strict inequality for at least one $\theta = \theta_0$. Since $R(\theta, \delta)$ is a continuous function of θ (it is a weighted mean of quadratic functions) the strict inequality holds in a neighbourhood $(\theta_1, \theta_2) \ni \theta_0$, i.e.

$$R(\theta, \delta^*) \quad \leq \quad \frac{1}{n} - \epsilon, \tag{5}$$

on (θ_1, θ_2), for some $\epsilon > 0$.

Let $\Lambda_\tau = N(0, \tau^2)$ and define

$$\begin{aligned} r_\tau &= \int R(\theta, \delta_\Lambda)\, d\Lambda_\tau(\theta) = (\textit{The Bayes risk for } \delta_\Lambda) \\ &= \frac{1}{n/\sigma^2 + 1/\tau^2} \\ &= \frac{\tau^2}{1 + n\tau^2}, \\ r^* &= \int R(\theta, \delta^*)\, d\Lambda_\tau(\theta). \end{aligned}$$

Then

$$\begin{aligned} \frac{1/n - r_\tau^*}{1/n - r_\tau} &= \frac{\frac{1}{\sqrt{2\pi}\tau} \int \left(\frac{1}{n} - R(\theta, \delta^*)\right) e^{-\tau^2/2\tau^2}\, d\tau}{\frac{1 + n\tau^2 - n\tau^2}{n(1+n\tau^2)}} \\ &\geq \frac{n(1 + n\tau^2)\epsilon}{\sqrt{2\pi}\tau} \int_{\theta_1}^{\theta_2} e^{-\theta^2/2\tau^2}. \end{aligned}$$

By monotone convergence the integral converges to $\int_{\theta_1}^{\theta 2} d\theta = \theta_2 - \theta_1$, which implies that

$$\frac{1/n - r_\tau^*}{1/n - r_\tau} \quad \to \quad +\infty,$$

as $\tau \to \infty$. But this implies that for some (large enough) τ_0 we have $r^*_{\tau_0} < r_{\tau_0}$, which contradicts the fact that δ_{τ_0} is the Bayes estimator.

Therefore (5) *can not hold, and thus $\delta = \bar{X}$ is admissible.* □

Karlin's theorem generalizes the obtained results to estimation of the mean in exponential families. Thus let X have probability density

$$p_\theta(x) \quad = \quad \beta(\theta)e^{\theta T(x)}, \tag{6}$$

with θ a real valued parameter and T a real valued function. The natural parameter space for this family is an interval $\Omega = (\theta_1, \theta_2)$ in the extended real line. Assume we have squared error loss. Let $\delta(X) = aT(X) + b$ be an estimator.

Then, when $a < 0$ and $a > 1$ the estimator is inadmissible for estimating $g(\theta) = E_\theta(T(X))$, the proof of which is analogous to the proof of Lemma 4. Also, for $a = 0$ the estimator is constant and is then admissible. What happens for $0 < a \leq 1$?

Reparametrize the estimator as

$$\delta_{\lambda,\gamma}(x) \quad = \quad \frac{1}{1+\lambda}T(x) + \frac{\lambda}{1+\lambda}\gamma,$$

so that λ, γ replaces a, b, and $0 < a \leq 1$ is translated to $0 \leq \lambda < \infty$.

Theorem 14 *Assume that for some $\theta_0 \in (\theta_1, \theta_2)$*

$$\begin{aligned} \lim_{u \uparrow \theta_2} \int_{\theta_0}^{u} \frac{e^{-\gamma\lambda u}}{\beta(u)^\lambda}\, du &= \infty, \\ \lim_{u \downarrow \theta_1} \int_{u}^{\theta_0} \frac{e^{-\gamma\lambda u}}{\beta(u)^\lambda}\, du &= \infty. \end{aligned}$$

Then $\delta_{\lambda,\gamma}$ is admissible for estimating $g(\theta)$.

For a proof, see Lehmann [**?**].

Corollary 5 *If the natural parameter space of* (6) *is the whole real line $\Omega = (-\infty, \infty)$ then $T(X)$ is admissible for estimating $g(\theta)$.*

Proof. $T(X)$ corresponds to the estimator $\delta_{0,1}(X)$ i.e. $\lambda = 0, \gamma = 1$. Then the integrands in Karlin's theorem is the constant 1, so both integrals tend to ∞, and thus $T(X)$ is admissible. □

Example 37 *The natural parameter space is $\mathbb{R}$ for the Normal distribution with known variance, the Poisson distribution and the Binomial distribution. (Excercise).* □

Lemma 12 *If δ is admissible and has constant risk, then δ is minimax.*

Proof. If δ is not minimax, there is another estimator δ' such that

$$\sup_\theta R(\theta, \delta') \quad < \quad \sup_\theta R(\theta, \delta) = c,$$

so that for all θ we have $R(\theta, \delta') < c = R(\theta, \delta)$, which implies that δ is inadmissible. □

Lemma 13 *If δ is unique minimax, δ is admissible.*

Proof. Assume δ is inadmissible so that the estimator $\delta' \neq \delta$ satisfies $R(\theta, \delta') \leq R(\theta, \delta)$ for all θ. But then δ can not be minimax, since it is unique. □

Corollary 6 *Let*

$$L(\theta, d) \quad = \quad \frac{(d - g(\theta))^2}{\mathrm{Var}(T(X))}$$

be the loss function for estimating $g(\theta) = E_\theta(T(X))$, and assume the natural parameter space of (6) *is the real line. Then $\delta(x) = T(x)$ minimax. It is unique.*

Proof. The estimator $T(X)$ is admissible and has constant risk under the loss L, and is therefore minimax. It is unique since the loss function is strictly convex. □

Example 38 *Let $X \in Bin(n, p)$ and assume the loss function $L(p, d) = (p - d)^2/pq$. The natural parameter space of the Binomial distribution is $\mathbb{R}$, and the density is*

$$\begin{aligned} p(x) \quad &= \quad \binom{n}{x} p^x (1-p)^{n-x} \\ &= \quad \binom{n}{x} \left(\frac{p}{1-p}\right)^x \left(\frac{1}{1-p}\right)^n \end{aligned}$$

i.e. of the form (6) *with $\theta = p/(1-p)$, $\beta(\theta) = (1-p)^{-n}$ and $T(x) = x$. Thus $T(X) = X$ is unique minimax for estimating $E(X) = np$ and therefore $\delta(X) = X/n$ is unique minimax for estimating p. Since δ is unique minimax it is admissible.* □

7.2 Minimax estimation in group families

Assume $\mathcal{P}, L, g(\theta)$ is an invariant estimation problem for the transformation groups $\mathcal{G}, \bar{\mathcal{G}}, \mathcal{G}^*$.

Then typically there exists a MRE estimator, and it has constant risk. Recall also that a Bayes estimator with constant risk is minimax, and admissible if it is unique Bayes. Recall the relation between Bayes estimators and equivariant estimators that said that a Bayes estimator is almost equivariant. This implies that a unique Bayes estimator in the present setting is admissible

It is possible to obtain results on minimaxity and admissibility for Bayes estimators under improper priors, which we however refrain from.