Stat 710: Mathematical Statistics
Lecture 7

Jun Shao

Department of Statistics
University of Wisconsin
Madison, WI 53706, USA

Theorem 4.14 (Admissibility in one-parameter exponential families)

Suppose that X has the p.d.f. $c(\theta)e^{\theta T(x)}$ w.r.t. a σ-finite measure ν, where $T(x)$ is real-valued and $\theta \in (\theta_-, \theta_+) \subset \mathscr{R}$.
Consider the estimation of $\vartheta = E[T(X)]$ under the squared error loss.
Let $\lambda \geq 0$ and γ be known constants and let

$$T_{\lambda,\gamma}(X) = (T + \gamma\lambda)/(1+\lambda).$$

Then a sufficient condition for the admissibility of $T_{\lambda,\gamma}$ is that

$$\int_{\theta_0}^{\theta_+} \frac{e^{-\gamma\lambda\theta}}{[c(\theta)]^{\lambda}} d\theta = \int_{\theta_-}^{\theta_0} \frac{e^{-\gamma\lambda\theta}}{[c(\theta)]^{\lambda}} d\theta = \infty,$$

where $\theta_0 \in (\theta_-, \theta_+)$.

Remarks

- Theorem 4.14 provides a class of admissible estimators.
- The reason why $T_{\lambda,\gamma}$ is considered is that it is often a Bayes estimator w.r.t. some prior; see Examples 2.25, 4.1, 4.7, and 4.8.
- Using this theorem and Theorem 4.13, we can obtain a class of minimax estimators.

Proof of Theorem 4.14: an application of the information inequality

By Theorem 2.1, $\vartheta' = \mathrm{Var}(T) = I(\theta)$ (Fisher information)
Suppose that there is an estimator δ of ϑ such that for all θ,

$$R_\delta(\theta) \le R_{T_{\lambda,\gamma}}(\theta) = [I(\theta) + \lambda^2(\vartheta - \gamma)^2]/(1+\lambda)^2.$$

Let $b_\delta(\theta)$ be the bias of δ. From the information inequality,

$$\mathrm{Var}(\delta) \ge \left[\frac{d}{d\theta}(b_\delta(\theta) + \vartheta)\right]^2 / I(\theta),$$

which is the same as

Remarks

- Theorem 4.14 provides a class of admissible estimators.
- The reason why $T_{\lambda,\gamma}$ is considered is that it is often a Bayes estimator w.r.t. some prior; see Examples 2.25, 4.1, 4.7, and 4.8.
- Using this theorem and Theorem 4.13, we can obtain a class of minimax estimators.

Proof of Theorem 4.14: an application of the information inequality

By Theorem 2.1, $\vartheta' = \mathrm{Var}(T) = I(\theta)$ (Fisher information)
Suppose that there is an estimator δ of ϑ such that for all θ,

$$R_\delta(\theta) \le R_{T_{\lambda,\gamma}}(\theta) = [I(\theta) + \lambda^2(\vartheta - \gamma)^2]/(1+\lambda)^2.$$

Let $b_\delta(\theta)$ be the bias of δ. From the information inequality,

$$\mathrm{Var}(\delta) \ge \left[\frac{d}{d\theta}(b_\delta(\theta) + \vartheta)\right]^2 / I(\theta),$$

which is the same as

$$R_\delta(\theta) \geq [b_\delta(\theta)]^2 + [I(\theta) + b'_\delta(\theta)]^2 / I(\theta).$$

Let $h(\theta) = b_\delta(\theta) - \lambda(\gamma - \vartheta)/(1+\lambda)$.
Then

$$[h(\theta)]^2 - \frac{2\lambda h(\theta)(\vartheta - \gamma) - 2h'(\theta)}{1+\lambda} + \frac{[h'(\theta)]^2}{I(\theta)} \leq 0,$$

which implies

$$[h(\theta)]^2 - \frac{2\lambda h(\theta)(\vartheta - \gamma) - 2h'(\theta)}{1+\lambda} \leq 0.$$

Let $a(\theta) = h(\theta)[c(\theta)]^\lambda e^{\gamma\lambda\theta}$.

$$a'(\theta) = h'(\theta)[c(\theta)]^\lambda e^{\gamma\lambda\theta} + h(\theta)\lambda[c(\theta)]^{\lambda-1}c'(\theta)e^{\gamma\lambda\theta} + h(\theta)[c(\theta)]^\lambda \gamma\lambda e^{\gamma\lambda\theta}$$

From this result and the fact that

$$\vartheta = E[T(X)] = -c'(\theta)/c(\theta) \quad \text{(Theorem 2.1)}$$

we obtain

$$\frac{[a(\theta)]^2 e^{-\gamma\lambda\theta}}{[c(\theta)]^\lambda} + \frac{2a'(\theta)}{1+\lambda} \le 0.$$

Suppose that $a(\theta_0) < 0$ for some $\theta_0 \in (\theta_-, \theta_+)$.
Since $a'(\theta) \le 0$ for all θ, $a(\theta) < 0$ for all $\theta \ge \theta_0$
For $\theta > \theta_0$, the previous expression can be written as

$$\frac{d}{d\theta}\left[\frac{1}{a(\theta)}\right] \ge \frac{(1+\lambda)e^{-\gamma\lambda\theta}}{2[c(\theta)]^\lambda}.$$

Integrating both sides from θ_0 to θ gives

$$\frac{1+\lambda}{2}\int_{\theta_0}^{\theta} \frac{e^{-\gamma\lambda\theta}}{[c(\theta)]^\lambda} d\theta \le \frac{1}{a(\theta)} - \frac{1}{a(\theta_0)} \le -\frac{1}{a(\theta_0)}.$$

Letting $\theta \to \theta_+$, the left-hand side of the previous expression diverges to ∞ by the condition, which is impossible.
This shows that $a(\theta) \ge 0$ for all θ.

Similarly, we can show that $a(\theta) \leq 0$ for all θ.
Thus, $a(\theta) = 0$ for all θ.
This means that $h(\theta) = 0$ for all θ and

$$b_\delta(\theta) = \lambda(\gamma - \vartheta)/(1+\lambda)$$

$$b'_\delta(\theta) = -\lambda\vartheta'/(1+\lambda) = -\lambda I(\theta)/(1+\lambda)$$

by Theorem 2.1, $\vartheta' = I(\theta)$
Then

$$\begin{aligned} R_{T_{\lambda,\gamma}}(\theta) &= [I(\theta) + \lambda^2(\vartheta - \gamma)^2]/(1+\lambda)^2 \\ &= I(\theta)/(1+\lambda^2) + [b_\delta(\theta)]^2 \\ &= [I(\theta) + b'_\delta(\theta)]^2/I(\theta) + [b_\delta(\theta)]^2 \\ &\leq R_\delta(\theta) \end{aligned}$$

Hence, $R_{T_{\lambda,\gamma}}(\theta) = R_\delta(\theta)$.
This proves the admissibility of $T_{\lambda,\gamma}$.

Although the proof of this theorem is more complicated than that of Theorem 4.3, the application of Theorem 4.14 is typically easier.
To find minimax estimators, we may use the following result.

Corollary 4.3

Assume that X has the p.d.f. as described in Theorem 4.14 with $\theta_- = -\infty$ and $\theta_+ = \infty$.
(i) As an estimator of $\vartheta = E(T)$, $T(X)$ is admissible under the squared error loss and the loss $(a - \vartheta)^2/\operatorname{Var}(T)$.
(ii) T is the unique minimax estimator of ϑ under the loss $(a - \vartheta)^2/\operatorname{Var}(T)$.

Proof

(i) With $\lambda = 0$, the condition of Theorem 4.14 is clearly satisfied.
Hence, Theorem 4.14 applies under the squared error loss.
The admissibility of T under the loss $(a - \vartheta)^2/\operatorname{Var}(T)$ follows from the fact that T is admissible under the squared error loss and $\operatorname{Var}(T) \neq 0$.
(ii) This is a consequence of part (i) and Theorem 4.13.

Although the proof of this theorem is more complicated than that of Theorem 4.3, the application of Theorem 4.14 is typically easier.
To find minimax estimators, we may use the following result.

Corollary 4.3

Assume that X has the p.d.f. as described in Theorem 4.14 with $\theta_- = -\infty$ and $\theta_+ = \infty$.
(i) As an estimator of $\vartheta = E(T)$, $T(X)$ is admissible under the squared error loss and the loss $(a-\vartheta)^2/\operatorname{Var}(T)$.
(ii) T is the unique minimax estimator of ϑ under the loss $(a-\vartheta)^2/\operatorname{Var}(T)$.

Proof

(i) With $\lambda = 0$, the condition of Theorem 4.14 is clearly satisfied.
Hence, Theorem 4.14 applies under the squared error loss.
The admissibility of T under the loss $(a-\vartheta)^2/\operatorname{Var}(T)$ follows from the fact that T is admissible under the squared error loss and $\operatorname{Var}(T) \neq 0$.
(ii) This is a consequence of part (i) and Theorem 4.13.

Although the proof of this theorem is more complicated than that of Theorem 4.3, the application of Theorem 4.14 is typically easier.
To find minimax estimators, we may use the following result.

Corollary 4.3

Assume that X has the p.d.f. as described in Theorem 4.14 with $\theta_- = -\infty$ and $\theta_+ = \infty$.
(i) As an estimator of $\vartheta = E(T)$, $T(X)$ is admissible under the squared error loss and the loss $(a-\vartheta)^2/\operatorname{Var}(T)$.
(ii) T is the unique minimax estimator of ϑ under the loss $(a-\vartheta)^2/\operatorname{Var}(T)$.

Proof

(i) With $\lambda = 0$, the condition of Theorem 4.14 is clearly satisfied.
Hence, Theorem 4.14 applies under the squared error loss.
The admissibility of T under the loss $(a-\vartheta)^2/\operatorname{Var}(T)$ follows from the fact that T is admissible under the squared error loss and $\operatorname{Var}(T) \neq 0$.
(ii) This is a consequence of part (i) and Theorem 4.13.

Example 4.20

Let $X_1, ..., X_n$ be i.i.d. from $N(0, \sigma^2)$ with an unknown $\sigma^2 > 0$ and let $Y = \sum_{i=1}^n X_i^2$.

Consider the estimation of σ^2.

The risk of $Y/(n+2)$ is a constant under the loss $(a - \sigma^2)^2/\sigma^4$.

We now apply Theorem 4.14 to show that $Y/(n+2)$ is admissible.

Note that the joint p.d.f. of X_i's is of the form $c(\theta)e^{\theta T(x)}$ with $\theta = -n/(4\sigma^2)$, $c(\theta) = (-2\theta/n)^{n/2}$, $T(X) = 2Y/n$, $\theta_- = -\infty$, and $\theta_+ = 0$.

By Theorem 4.14, $T_{\lambda,\gamma} = (T + \gamma\lambda)/(1+\lambda)$ is admissible under the squared error loss if, for some $c > 0$,

$$\int_{-\infty}^{-c} e^{-\gamma\lambda\theta} \left(\frac{-2\theta}{n}\right)^{-n\lambda/2} d\theta = \int_0^c e^{\gamma\lambda\theta}\theta^{-n\lambda/2} d\theta = \infty$$

This means that $T_{\lambda,\gamma}$ is admissible if $\gamma = 0$ and $\lambda = 2/n$, or if $\gamma > 0$ and $\lambda \geq 2/n$.

In particular, $2Y/(n+2)$ is admissible for estimating $E(T) = 2E(Y)/n = 2\sigma^2$, under the squared error loss.

Example 4.20 (continued)

It is easy to see that $Y/(n+2)$ is then an admissible estimator of σ^2 under the squared error loss and the loss $(a-\sigma^2)^2/\sigma^4$.
Hence $Y/(n+2)$ is minimax under the loss $(a-\sigma^2)^2/\sigma^4$.
Note that we cannot apply Corollary 4.3 directly since $\theta_+ = 0$.

Example 4.21

Let $X_1,...,X_n$ be i.i.d. from the Poisson distribution $P(\theta)$ with an unknown $\theta > 0$.
The joint p.d.f. of X_i's w.r.t. the counting measure is

$$(x_1!\cdots x_n!)^{-1}e^{-n\theta}e^{n\bar{x}\log\theta}$$

For $\eta = n\log\theta$, the conditions of Corollary 4.3 are satisfied with $T(X) = \bar{X}$.
Since $E(T) = \theta$ and $\mathrm{Var}(T) = \theta/n$, by Corollary 4.3, $\bar{X}$ is the unique minimax estimator of θ under the loss function $(a-\theta)^2/\theta$.

Example 4.20 (continued)

It is easy to see that $Y/(n+2)$ is then an admissible estimator of σ^2 under the squared error loss and the loss $(a-\sigma^2)^2/\sigma^4$.
Hence $Y/(n+2)$ is minimax under the loss $(a-\sigma^2)^2/\sigma^4$.
Note that we cannot apply Corollary 4.3 directly since $\theta_+ = 0$.

Example 4.21

Let $X_1,...,X_n$ be i.i.d. from the Poisson distribution $P(\theta)$ with an unknown $\theta > 0$.
The joint p.d.f. of X_i's w.r.t. the counting measure is

$$(x_1!\cdots x_n!)^{-1}e^{-n\theta}e^{n\bar{x}\log\theta}$$

For $\eta = n\log\theta$, the conditions of Corollary 4.3 are satisfied with $T(X) = \bar{X}$.
Since $E(T) = \theta$ and $\mathrm{Var}(T) = \theta/n$, by Corollary 4.3, $\bar{X}$ is the unique minimax estimator of θ under the loss function $(a-\theta)^2/\theta$.

Exercise 37 (#4.83)

Let X be an observation from the distribution with Lebesgue density $\frac{1}{2}c(\theta)e^{\theta x-|x|}$, $|\theta| < 1$.
(i) Show that $c(\theta) = 1-\theta^2$.
(ii) Show that if $0 \le \alpha \le \frac{1}{2}$, then $\alpha X + \beta$ is admissible for estimating $E(X)$ under the squared error loss.

Solution

(i) Note that

$$\begin{aligned}
\frac{1}{c(\theta)} &= \frac{1}{2}\int_{-\infty}^{\infty} e^{\theta x-|x|}dx \\
&= \frac{1}{2}\left(\int_{-\infty}^{0} e^{\theta x+x}dx + \int_{0}^{\infty} e^{\theta x-x}dx\right) \\
&= \frac{1}{2}\left(\int_{0}^{\infty} e^{-(1+\theta)x}dx + \int_{0}^{\infty} e^{-(1-\theta)x}dx\right) \\
&= \frac{1}{2}\left(\frac{1}{1+\theta} + \frac{1}{1-\theta}\right) = \frac{1}{1-\theta^2}.
\end{aligned}$$

Exercise 37 (#4.83)

Let X be an observation from the distribution with Lebesgue density $\frac{1}{2}c(\theta)e^{\theta x-|x|}$, $|\theta|<1$.
(i) Show that $c(\theta)=1-\theta^2$.
(ii) Show that if $0\leq\alpha\leq\frac{1}{2}$, then $\alpha X+\beta$ is admissible for estimating $E(X)$ under the squared error loss.

Solution

(i) Note that

$$\begin{aligned}
\frac{1}{c(\theta)} &= \frac{1}{2}\int_{-\infty}^{\infty} e^{\theta x-|x|}dx \\
&= \frac{1}{2}\left(\int_{-\infty}^{0} e^{\theta x+x}dx+\int_{0}^{\infty} e^{\theta x-x}dx\right) \\
&= \frac{1}{2}\left(\int_{0}^{\infty} e^{-(1+\theta)x}dx+\int_{0}^{\infty} e^{-(1-\theta)x}dx\right) \\
&= \frac{1}{2}\left(\frac{1}{1+\theta}+\frac{1}{1-\theta}\right)=\frac{1}{1-\theta^2}.
\end{aligned}$$

Solution (continued)

(ii) Consider first $\alpha > 0$. Let $\alpha = (1+\lambda)^{-1}$ and $\beta = \gamma\lambda/(1+\lambda)$.

$$\int_{-1}^{0} \frac{e^{-\gamma\lambda\theta}}{(1-\theta^2)^\lambda} d\theta = \int_{0}^{1} \frac{e^{-\gamma\lambda\theta}}{(1-\theta^2)^\lambda} d\theta = \infty$$

if and only if $\lambda \geq 1$, i.e., $\alpha \leq \frac{1}{2}$.

Hence, $\alpha X + \beta$ is an admissible estimator of $E(X)$ when $0 < \alpha \leq \frac{1}{2}$.

Consider next $\alpha = 0$.

$$\begin{aligned} E(X) &= \frac{1-\theta^2}{2}\left(\int_{-\infty}^{0} xe^{\theta x + x} dx + \int_{0}^{\infty} xe^{\theta x - x} dx\right) \\ &= \frac{1-\theta^2}{2}\left(-\int_{0}^{\infty} xe^{-(1+\theta)} dx + \int_{0}^{\infty} xe^{-(1-\theta)x} dx\right) \\ &= \frac{1-\theta^2}{2}\left(\frac{1+\theta}{1-\theta} - \frac{1-\theta}{1+\theta}\right) = \frac{2\theta}{1-\theta^2}, \end{aligned}$$

which takes any value in $(-\infty, \infty)$.

Hence, the constant estimator β is an admissible estimator of $E(X)$.