Consider estimators of a real-valued $\vartheta = g(\theta)$ based on a sample $X$ from $P_\theta$, $\theta \in \Theta$, under loss $L$ and risk $R_T(\theta) = E[L(T(X), \theta)]$.

**Minimax estimator**

A *minimax estimator* minimizes $\sup_{\theta \in \Theta} R_T(\theta)$ over all estimators $T$.

**Discussion**

- A minimax estimator can be very conservative and unsatisfactory. It tries to do as well as possible in the worst case.
- A unique minimax estimator is admissible, since any estimator better than a minimax estimator is also minimax.
- We should find an admissible minimax estimator.
- If a minimax estimator has some other good properties (e.g., it is a Bayes estimator), then it is often a reasonable estimator.
How to find a minimax estimator?
Candidates for minimax: estimators having constant risks

Theorem 4.11 (minimaxity of a Bayes estimator)
Let $\Pi$ be a proper prior on $\Theta$ and $\delta$ be a Bayes estimator of $\vartheta$ w.r.t. $\Pi$. Suppose $\delta$ has constant risk on $\Theta_{\Pi}$. If $\Pi(\Theta_{\Pi}) = 1$, then $\delta$ is minimax.
If, in addition, $\delta$ is the unique Bayes estimator w.r.t. $\Pi$, then it is the unique minimax estimator.

Proof
Let $T$ be any other estimator of $\vartheta$. Then

$$\sup_{\theta \in \Theta} R_T(\theta) \geq \int_{\Theta_{\Pi}} R_T(\theta) d\Pi \geq \int_{\Theta_{\Pi}} R_\delta(\theta) d\Pi = \sup_{\theta \in \Theta} R_\delta(\theta).$$

If $\delta$ is the unique Bayes estimator, then the second inequality in the previous expression should be replaced by $>$ and, therefore, $\delta$ is the unique minimax estimator.
Example 4.18

Let $X_1, \ldots, X_n$ be i.i.d. binary random variables with $P(X_1 = 1) = p$. Consider the estimation of $p$ under the squared error loss.

To find a minimax estimator by applying Theorem 4.11, we consider the Bayes estimator w.r.t. the beta distribution $B(\alpha, \beta)$ with known $\alpha$ and $\beta$ (Exercise 1):

$$\delta(X) = \frac{\alpha + n\bar{X}}{\alpha + \beta + n}.$$  

$$R_\delta(p) = \frac{[np(1-p) + (\alpha - \alpha p - \beta p)^2]}{\alpha + \beta + n}.$$  

To apply Theorem 4.11, we need to find values of $\alpha > 0$ and $\beta > 0$ such that $R_\delta(p)$ is constant.

It can be shown that $R_\delta(p)$ is constant if and only if $\alpha = \beta = \sqrt{n}/2$, which leads to the unique minimax estimator

$$T(X) = \frac{n\bar{X} + \sqrt{n}/2}{n + \sqrt{n}}.$$  

The risk of $T$ is $R_T = 1/[4(1 + \sqrt{n})^2]$. 
Example 4.18 (continued)

Comparing the risk of $T$ with that of $\tilde{X}$ ($R_{\tilde{X}} = p(1 - p)/n$), we find that $T$ has smaller risk if and only if

$$p \in \left( \frac{1}{2} - \frac{1}{2} \sqrt{1 - \frac{n}{(1+\sqrt{n})^2}}, \frac{1}{2} + \frac{1}{2} \sqrt{1 - \frac{n}{(1+\sqrt{n})^2}} \right).$$

For a small $n$, $T$ is better than $\tilde{X}$ for most of the range of $p$. When $n \to \infty$, the above interval shrinks toward $\frac{1}{2}$. Hence, for a large $n$, $\tilde{X}$ is better than $T$ for most of the range of $p$ (Figure 4.1).

Minimaxity depends strongly on the loss function.

Under the loss function $L(p, a) = (a - p)^2/[p(1 - p)]$, $\tilde{X}$ has constant risk and is the unique Bayes estimator w.r.t. the uniform prior on $(0, 1)$. By Theorem 4.11, $\tilde{X}$ is the unique minimax estimator.

The risk, however, of $T$ is $1/[4(1 + \sqrt{n})^2 p(1 - p)]$, which is unbounded.
A limit of Bayes estimators
In many cases a constant risk estimator is not a Bayes estimator (e.g., an unbiased estimator under the squared error loss), but a limit of Bayes estimators w.r.t. a sequence of priors. The next result may be used to find a minimax estimator.

**Theorem 4.12**
Let $\Pi_j, j = 1, 2, \ldots,$ be a sequence of priors and $r_j$ be the Bayes risk of a Bayes estimator of $\vartheta$ w.r.t. $\Pi_j$. Let $T$ be a constant risk estimator of $\vartheta$. If $\liminf r_j \geq R_T$, then $T$ is minimax.

The proof is similar to the proof of Theorem 4.11.

Although Theorem 4.12 is more general than Theorem 4.11 in finding minimax estimators, it does not provide uniqueness of the minimax estimator even when there is a unique Bayes estimator w.r.t. each $\Pi_j$. 
Let \( X_1, \ldots, X_n \) be i.i.d. \( \mathcal{N}(\mu, \sigma^2) \) where \( \sigma^2 \) is known. Consider the squared error loss and the priors of \( \mu \): \( \Pi_j = \mathcal{N}(0, j) \). It can be shown that \( \bar{X} \) as an estimator of \( \mu \) is minimax.

To discuss the minimaxity of \( \bar{X} \) in the case where \( \sigma^2 \) is unknown, we need the following lemma.

**Lemma 4.3**
Let \( \Theta_0 \) be a subset of \( \Theta \) and \( T \) be a minimax estimator of \( \vartheta \) when \( \Theta_0 \) is the parameter space. Then \( T \) is a minimax estimator if
\[
\sup_{\theta \in \Theta} R_T(\theta) = \sup_{\theta \in \Theta_0} R_T(\theta).
\]

**Proof**
If there is an estimator \( T_0 \) with \( \sup_{\theta \in \Theta} R_{T_0}(\theta) < \sup_{\theta \in \Theta} R_T(\theta) \), then
\[
\sup_{\theta \in \Theta_0} R_{T_0}(\theta) \leq \sup_{\theta \in \Theta} R_{T_0}(\theta) < \sup_{\theta \in \Theta} R_T(\theta) = \sup_{\theta \in \Theta_0} R_T(\theta),
\]
which contradicts the minimaxity of \( T \) when \( \Theta_0 \) is the parameter space. Hence, \( T \) is minimax when \( \Theta \) is the parameter space.
Example 4.19

Let $X_1, \ldots, X_n$ be i.i.d. from $N(\mu, \sigma^2)$ with unknown $\theta = (\mu, \sigma^2)$. Consider the estimation of $\mu$ under the squared error loss.

Suppose first that $\Theta = \mathbb{R} \times (0, c]$ with a constant $c > 0$. Let $\Theta_0 = \mathbb{R} \times \{c\}$. Then, $\bar{X}$ is a minimax estimator of $\mu$ when the parameter space is $\Theta_0$. By Lemma 4.3, $\bar{X}$ is also minimax when the parameter space is $\Theta$.

Consider next the case where $\Theta = \mathbb{R} \times (0, \infty)$, i.e., $\sigma^2$ is unbounded.

Let $T$ be any estimator of $\mu$. For any fixed $\sigma^2$,

$$\frac{\sigma^2}{n} \leq \sup_{\mu \in \mathbb{R}} R_T(\theta),$$

since $\sigma^2/n$ is the risk of $\bar{X}$ that is minimax when $\sigma^2$ is known. Letting $\sigma^2 \to \infty$, we obtain that $\sup_{\theta} R_T(\theta) = \infty$ for any estimator $T$.

Thus, minimaxity is meaningless (any estimator is minimax).
Theorem 4.13
Suppose that $T$ as an estimator of $\vartheta$ has constant risk and is admissible. Then $T$ is minimax.
If the loss function is strictly convex, then $T$ is the unique minimax estimator.

Proof
The risk of $T$ is $R_T$ (not depending on $\theta$).
By the admissibility of $T$, if there is another estimator $T_0$ with $\sup_{\theta} R_{T_0}(\theta) \leq R_T$, then $R_{T_0}(\theta) = R_T$ for all $\theta$.
This proves that $T$ is minimax.
If the loss function is strictly convex and $T_0$ is another minimax estimator, then

$$R_{(T + T_0)/2}(\theta) < (R_{T_0} + R_T)/2 = R_T$$

for all $\theta$ and, therefore, $T$ is inadmissible.
This shows that $T$ is unique if the loss is strictly convex.
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 $\sigma$-finite measure $\nu$, where $T(x)$ is real-valued and $\theta \in (\theta_-, \theta_+) \subset \mathcal{R}$.
Consider the estimation of $\vartheta = E[T(X)]$ under the squared error loss.
Let $\lambda \geq 0$ and $\gamma$ 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.
- Using this theorem and Theorem 4.13, we can obtain a class of minimax estimators.

Proof of Theorem 4.14:

By Theorem 2.1, $\vartheta' = \text{Var}(T) = I(\theta)$ (Fisher information)

Suppose that there is an estimator $\delta$ of $\vartheta$ such that for all $\theta$,

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

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

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

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

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

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

The above expression can be written as

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

Suppose that \(a(\theta_0) < 0\) for some \(\theta_0 \in (\theta_-, \theta_+)\). Since \(a'(\theta) \leq 0\) for all \(\theta\), \(a(\theta) < 0\) for all \(\theta \geq \theta_0\). For \(\theta > \theta_0\), integrating both sides from \(\theta_0\) to \(\theta\) gives

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

Letting \(\theta \to \theta_+\), the left-hand side diverges to \(\infty\) by the condition, which is impossible. This shows that \(a(\theta) \geq 0\) for all \(\theta\). Similarly, we can show that \(a(\theta) \leq 0\) for all \(\theta\).
Thus, \( a(\theta) = 0 \) for all \( \theta \).
This means that \( h(\theta) = 0 \) for all \( \theta \) and

\[
b_\delta(\theta) = \frac{\lambda(\gamma - \vartheta)}{1 + \lambda}
\]

\[
b'_\delta(\theta) = -\frac{\lambda \vartheta'}{1 + \lambda} = -\frac{\lambda I(\theta)}{1 + \lambda}
\]

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

Then

\[
RT_{\lambda, \gamma}(\theta) = \frac{I(\theta) + \lambda^2(\vartheta - \gamma)^2}{1 + \lambda}^2
\]

\[
= \frac{I(\theta)}{(1 + \lambda)} + [b_\delta(\theta)]^2
\]

\[
= \frac{[I(\theta) + b'_\delta(\theta)]^2}{I(\theta)} + [b_\delta(\theta)]^2
\]

\[
\leq R_\delta(\theta)
\]

Hence, \( RT_{\lambda, \gamma}(\theta) = R_\delta(\theta) \). This proves the admissibility of \( T_{\lambda, \gamma} \).
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 $(a - \vartheta)^2/\text{Var}(T)$.

(ii) $T$ is the unique minimax estimator of $\vartheta$ under the loss $(a - \vartheta)^2/\text{Var}(T)$.

**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 $(a - \vartheta)^2/\text{Var}(T)$ follows from the fact that $T$ is admissible under the squared error loss and $\text{Var}(T) \neq 0$.

(ii) This is a consequence of part (i) and Theorem 4.13.
Example 4.20

Let \( X_1, \ldots, X_n \) 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 \( \sigma^2 \).

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 \( \theta = -n/(4\sigma^2) \), \( c(\theta) = (-4\theta/n)^{n/2} \), \( T(X) = 2Y/n \), \( \theta_- = -\infty \), 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 \( c > 0 \),

\[
\int_{-\infty}^{-c} e^{-\gamma \lambda \theta} \left( \frac{-4\theta}{n} \right)^{-n\lambda/2} d\theta = \int_{c}^{\infty} e^{\gamma \lambda \theta} \theta^{-n\lambda/2} d\theta = \infty
\]

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 $\sigma^2$ 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 $\text{Var}(T) = \theta/n$, by Corollary 4.3, $\bar{X}$ is the unique minimax estimator of $\theta$ under the loss function $(a - \theta)^2/\theta$. 