# Stat 710: Mathematical Statistics
# Lecture 4

Jun Shao

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

## Bayes estimators

In the frequentist approach, if a Bayes action $\delta(x)$ is a measurable function of $x$, then $\delta(X)$ is a nonrandomized decision rule.

It can be shown that $\delta(X)$ defined in Definition 4.1 (if it exists for $X = x \in A$ with $\int_\Theta P_\theta(A) d\Pi = 1$) also minimizes the Bayes risk

$$r_T(\Pi) = \int_\Theta R_T(\theta) d\Pi$$

over all decision rules $T$ (randomized or nonrandomized)

$R_T(\theta) = E[L(\theta, T(X))]$ is the risk function of $T$ (Chapter 2).

Thus, $\delta(X)$ is a Bayes rule defined in §2.3.2.

In an estimation problem, a Bayes rule is called a *Bayes estimator*.

Generalized Bayes risks, generalized Bayes rules (or estimators), and empirical Bayes rules (or estimators) can be defined similarly.

In view of the discussion in §2.3.2, even if we do not adopt the Bayesian approach, the method described in §4.1.1 can be used as a way of generating decision rules.

## Admissibility

Given $R_T(\theta) = E[L(T(X), \theta)]$, $T$ is $\Im$-admissible iff there is no $T_0 \in \Im$ with $R_{T_0}(\theta) \leq R_T(\theta)$ for all $\theta$ and $R_{T_0}(\theta) < R_T(\theta)$ for some $\theta$
Admissible = $\Im$-admissible with $\Im = \{$ all rules $\}$

Bayes rules are typically admissible: If $T$ is better than a Bayes rule $\delta$, then $T$ has the same Bayes risk as $\delta$ and is itself a Bayes rule.

## Theorem 4.2 (Admissibility of Bayes rules)

In a decision problem, let $\delta(X)$ be a Bayes rule w.r.t. a prior $\Pi$.
(i) If $\delta(X)$ is a unique Bayes rule, then $\delta(X)$ is admissible.
(ii) If $\Theta$ is a countable set, the Bayes risk $r_\delta(\Pi) < \infty$, and $\Pi$ gives positive probability to each $\theta \in \Theta$, then $\delta(X)$ is admissible.
(iii) Let $\Im$ be the class of decision rules having continuous risk functions. If $\delta(X) \in \Im$, $r_\delta(\Pi) < \infty$, and $\Pi$ gives positive probability to any open subset of $\Theta$, then $\delta(X)$ is $\Im$-admissible.

## Frequentist properties of Bayes rules/estimators

### Admissibility

Given $R_T(\theta) = E[L(T(X), \theta)]$, $T$ is $\Im$-admissible iff there is no $T_0 \in \Im$ with $R_{T_0}(\theta) \le R_T(\theta)$ for all $\theta$ and $R_{T_0}(\theta) < R_T(\theta)$ for some $\theta$
Admissible = $\Im$-admissible with $\Im = \{$ all rules $\}$

Bayes rules are typically admissible: If $T$ is better than a Bayes rule $\delta$, then $T$ has the same Bayes risk as $\delta$ and is itself a Bayes rule.

### Theorem 4.2 (Admissibility of Bayes rules)

In a decision problem, let $\delta(X)$ be a Bayes rule w.r.t. a prior $\Pi$.
(i) If $\delta(X)$ is a unique Bayes rule, then $\delta(X)$ is admissible.
(ii) If $\Theta$ is a countable set, the Bayes risk $r_\delta(\Pi) < \infty$, and $\Pi$ gives positive probability to each $\theta \in \Theta$, then $\delta(X)$ is admissible.
(iii) Let $\Im$ be the class of decision rules having continuous risk functions. If $\delta(X) \in \Im$, $r_\delta(\Pi) < \infty$, and $\Pi$ gives positive probability to any open subset of $\Theta$, then $\delta(X)$ is $\Im$-admissible.

## Frequentist properties of Bayes rules/estimators

### Admissibility

Given $R_T(\theta) = E[L(T(X),\theta)]$, $T$ is $\Im$-admissible iff there is no $T_0 \in \Im$ with $R_{T_0}(\theta) \le R_T(\theta)$ for all $\theta$ and $R_{T_0}(\theta) < R_T(\theta)$ for some $\theta$
Admissible = $\Im$-admissible with $\Im = \{$ all rules $\}$

Bayes rules are typically admissible: If $T$ is better than a Bayes rule $\delta$, then $T$ has the same Bayes risk as $\delta$ and is itself a Bayes rule.

### Theorem 4.2 (Admissibility of Bayes rules)

In a decision problem, let $\delta(X)$ be a Bayes rule w.r.t. a prior $\Pi$.
(i) If $\delta(X)$ is a unique Bayes rule, then $\delta(X)$ is admissible.
(ii) If $\Theta$ is a countable set, the Bayes risk $r_\delta(\Pi) < \infty$, and $\Pi$ gives positive probability to each $\theta \in \Theta$, then $\delta(X)$ is admissible.
(iii) Let $\Im$ be the class of decision rules having continuous risk functions. If $\delta(X) \in \Im$, $r_\delta(\Pi) < \infty$, and $\Pi$ gives positive probability to any open subset of $\Theta$, then $\delta(X)$ is $\Im$-admissible.

# Frequentist properties of Bayes rules/estimators

## Admissibility

Given $R_T(\theta) = E[L(T(X), \theta)]$, $T$ is $\Im$-admissible iff there is no $T_0 \in \Im$ with $R_{T_0}(\theta) \leq R_T(\theta)$ for all $\theta$ and $R_{T_0}(\theta) < R_T(\theta)$ for some $\theta$
Admissible = $\Im$-admissible with $\Im = \{$ all rules $\}$

Bayes rules are typically admissible: If $T$ is better than a Bayes rule $\delta$, then $T$ has the same Bayes risk as $\delta$ and is itself a Bayes rule.

## Theorem 4.2 (Admissibility of Bayes rules)

In a decision problem, let $\delta(X)$ be a Bayes rule w.r.t. a prior $\Pi$.
(i) If $\delta(X)$ is a unique Bayes rule, then $\delta(X)$ is admissible.
(ii) If $\Theta$ is a countable set, the Bayes risk $r_\delta(\Pi) < \infty$, and $\Pi$ gives positive probability to each $\theta \in \Theta$, then $\delta(X)$ is admissible.
(iii) Let $\Im$ be the class of decision rules having continuous risk functions. If $\delta(X) \in \Im$, $r_\delta(\Pi) < \infty$, and $\Pi$ gives positive probability to any open subset of $\Theta$, then $\delta(X)$ is $\Im$-admissible.

Generalized Bayes rules or estimators are not necessarily admissible. Many generalized Bayes rules are limits of Bayes rules (Examples 4.3 and 4.7), which are often admissible.

## Theorem 4.3

Suppose that $\Theta$ is an open set of $\mathscr{R}^k$. In a decision problem, let $\Im$ be the class of decision rules having continuous risk functions. A decision rule $T \in \Im$ is $\Im$-admissible if there exists a sequence $\{\Pi_j\}$ of (possibly improper) priors such that (a) the generalized Bayes risks $r_T(\Pi_j)$ are finite for all $j$; (b) for any $\theta_0 \in \Theta$ and $\eta > 0$,

$$\lim_{j\to\infty} \frac{r_T(\Pi_j) - r_j^*(\Pi_j)}{\Pi_j(O_{\theta_0,\eta})} = 0,$$

where $r_j^*(\Pi_j) = \inf_{T\in\Im} r_T(\Pi_j)$ and $O_{\theta_0,\eta} = \{\theta \in \Theta : \|\theta - \theta_0\| < \eta\}$ with $\Pi_j(O_{\theta_0,\eta}) < \infty$ for all $j$.

Generalized Bayes rules or estimators are not necessarily admissible. Many generalized Bayes rules are limits of Bayes rules (Examples 4.3 and 4.7), which are often admissible.

## Theorem 4.3

Suppose that $\Theta$ is an open set of $\mathscr{R}^k$. In a decision problem, let $\Im$ be the class of decision rules having continuous risk functions. A decision rule $T \in \Im$ is $\Im$-admissible if there exists a sequence $\{\Pi_j\}$ of (possibly improper) priors such that (a) the generalized Bayes risks $r_T(\Pi_j)$ are finite for all $j$; (b) for any $\theta_0 \in \Theta$ and $\eta > 0$,

$$\lim_{j\to\infty} \frac{r_T(\Pi_j) - r_j^*(\Pi_j)}{\Pi_j(O_{\theta_0,\eta})} = 0,$$

where $r_j^*(\Pi_j) = \inf_{T\in\Im} r_T(\Pi_j)$ and $O_{\theta_0,\eta} = \{\theta \in \Theta : \|\theta - \theta_0\| < \eta\}$ with $\Pi_j(O_{\theta_0,\eta}) < \infty$ for all $j$.

## Proof

Suppose that $T$ is not $\Im$-admissible.
Then there exists $T_0 \in \Im$ such that $R_{T_0}(\theta) \leq R_T(\theta)$ for all $\theta$ and $R_{T_0}(\theta_0) < R_T(\theta_0)$ for a $\theta_0 \in \Theta$.
From the continuity of the risk functions, we conclude that

$$R_{T_0}(\theta) < R_T(\theta) - \varepsilon \quad \theta \in O_{\theta_0,\eta}$$

for some constants $\varepsilon > 0$ and $\eta > 0$.
Then, for any $j$,

$$\begin{aligned} r_T(\Pi_j) - r_j^*(\Pi_j) &\geq r_T(\Pi_j) - r_{T_0}(\Pi_j) \\ &\geq \int_{O_{\theta_0,\eta}} [R_T(\theta) - R_{T_0}(\theta)] d\Pi_j(\theta) \\ &\geq \varepsilon \Pi_j(O_{\theta_0,\eta}), \end{aligned}$$

which contradicts condition (b). Hence, $T$ is $\Im$-admissible.

While the proof of Theorem 4.3 is easy, the application of Theorem 4.3 is not easy.

## Example 4.6 (An application of Theorem 4.3)

Consider $X_1,...,X_n$ iid from $N(\mu,\sigma^2)$ with unknown $\mu$ and known $\sigma^2$
Loss = the squared error loss.
By Theorem 2.1, the risk function of any decision rule is continuous in $\mu$ if the risk is finite.
Apply Theorem 4.3 to the sample mean $\bar{X}$
Let $\Pi_j = N(0,j)$.
Since $R_{\bar{X}}(\mu) = \sigma^2/n$, $r_{\bar{X}}(\Pi_j) = \sigma^2/n$ for any $j$.
Hence, condition (a) in Theorem 4.3 is satisfied.
From Example 2.25, the Bayes estimator w.r.t. $\Pi_j$ is

$$\delta_j(X) = \frac{nj}{nj+\sigma^2}\bar{X}$$

Thus,

$$R_{\delta_j}(\mu) = \frac{\sigma^2 nj^2 + \sigma^4\mu^2}{(nj+\sigma^2)^2}$$

While the proof of Theorem 4.3 is easy, the application of Theorem 4.3 is not easy.

## Example 4.6 (An application of Theorem 4.3)

Consider $X_1,...,X_n$ iid from $N(\mu,\sigma^2)$ with unknown $\mu$ and known $\sigma^2$
Loss = the squared error loss.
By Theorem 2.1, the risk function of any decision rule is continuous in $\mu$ if the risk is finite.
Apply Theorem 4.3 to the sample mean $\bar{X}$
Let $\Pi_j = N(0,j)$.
Since $R_{\bar{X}}(\mu) = \sigma^2/n$, $r_{\bar{X}}(\Pi_j) = \sigma^2/n$ for any $j$.
Hence, condition (a) in Theorem 4.3 is satisfied.
From Example 2.25, the Bayes estimator w.r.t. $\Pi_j$ is

$$\delta_j(X) = \frac{nj}{nj+\sigma^2}\bar{X}$$

Thus,

$$R_{\delta_j}(\mu) = \frac{\sigma^2 nj^2 + \sigma^4\mu^2}{(nj+\sigma^2)^2}$$

and

$$r_j^*(\Pi_j) = \int R_{\delta_j}(\mu) d\Pi_j = \frac{\sigma^2 j}{nj + \sigma^2}.$$

For any $O_{\mu_0,\eta} = \{\mu : |\mu - \mu_0| < \eta\}$,

$$\Pi_j(O_{\mu_0,\eta}) = \Phi\left(\frac{\mu_0 + \eta}{\sqrt{j}}\right) - \Phi\left(\frac{\mu_0 - \eta}{\sqrt{j}}\right) = \frac{2\eta\Phi'(\xi_j)}{\sqrt{j}}$$

for some $\xi_j$ satisfying $(\mu_0 - \eta)/\sqrt{j} \leq \xi_j \leq (\mu_0 + \eta)/\sqrt{j}$, where $\Phi$ is the standard normal c.d.f. and $\Phi'$ is its derivative.
Since $\Phi'(\xi_j) \to \Phi'(0) = (2\pi)^{-1/2}$,

$$\frac{r_{\bar{X}}(\Pi_j) - r_j^*(\Pi_j)}{\Pi_j(O_{\mu_0,\eta})} = \frac{\sigma^4\sqrt{j}}{2\eta\Phi'(\xi_j)n(nj + \sigma^2)} \to 0$$

as $j \to \infty$.
Thus, condition (b) in Theorem 4.3 is satisfied.
Hence, Theorem 4.3 applies and the sample mean $\bar{X}$ is admissible.

## Bias

For any estimator $T$ of $\vartheta$, its bias is $E(T) - \vartheta$

### A Bayes estimator is usually biased.

### Proposition 4.2

Let $\delta(X)$ be a Bayes estimator of $\vartheta = g(\theta)$ under the squared error loss. Then $\delta(X)$ is not unbiased unless the Bayes risk $r_\delta(\Pi) = 0$.

### Remarks

- $r_\delta(\Pi) = 0$ occurs usually in some trivial cases.
- Proposition 4.2 can be used to check whether an estimator can be a Bayes estimator w.r.t. some prior under the squared error loss.
- However, a generalized Bayes estimator may be unbiased; see, for instance, Examples 4.3 and 4.7.

## Bias

For any estimator $T$ of $\vartheta$, its bias is $E(T) - \vartheta$

## A Bayes estimator is usually biased.

## Proposition 4.2

Let $\delta(X)$ be a Bayes estimator of $\vartheta = g(\theta)$ under the squared error loss. Then $\delta(X)$ is not unbiased unless the Bayes risk $r_\delta(\Pi) = 0$.

## Remarks

- $r_\delta(\Pi) = 0$ occurs usually in some trivial cases.
- Proposition 4.2 can be used to check whether an estimator can be a Bayes estimator w.r.t. some prior under the squared error loss.
- However, a generalized Bayes estimator may be unbiased; see, for instance, Examples 4.3 and 4.7.

## Bias

For any estimator $T$ of $\vartheta$, its bias is $E(T) - \vartheta$

## A Bayes estimator is usually biased.

## Proposition 4.2

Let $\delta(X)$ be a Bayes estimator of $\vartheta = g(\theta)$ under the squared error loss. Then $\delta(X)$ is not unbiased unless the Bayes risk $r_\delta(\Pi) = 0$.

## Remarks

- $r_\delta(\Pi) = 0$ occurs usually in some trivial cases.
- Proposition 4.2 can be used to check whether an estimator can be a Bayes estimator w.r.t. some prior under the squared error loss.
- However, a generalized Bayes estimator may be unbiased; see, for instance, Examples 4.3 and 4.7.

## Bias

For any estimator $T$ of $\vartheta$, its bias is $E(T) - \vartheta$

## A Bayes estimator is usually biased.

## Proposition 4.2

Let $\delta(X)$ be a Bayes estimator of $\vartheta = g(\theta)$ under the squared error loss. Then $\delta(X)$ is not unbiased unless the Bayes risk $r_\delta(\Pi) = 0$.

## Remarks

- $r_\delta(\Pi) = 0$ occurs usually in some trivial cases.
- Proposition 4.2 can be used to check whether an estimator can be a Bayes estimator w.r.t. some prior under the squared error loss.
- However, a generalized Bayes estimator may be unbiased; see, for instance, Examples 4.3 and 4.7.

## Proof of Proposition 4.2

Suppose that $\delta(X)$ is unbiased, i.e., $E[\delta(X)|\vec{\theta}] = g(\vec{\theta})$. Conditioning on $\vec{\theta}$ and using Proposition 1.10, we obtain that

$$E[g(\vec{\theta})\delta(X)] = E\{g(\vec{\theta})E[\delta(X)|\vec{\theta}]\} = E[g(\vec{\theta})]^2.$$

Since $\delta(X) = E[g(\vec{\theta})|X]$, conditioning on $X$ and using Proposition 1.10, we obtain that

$$E[g(\vec{\theta})\delta(X)] = E\{\delta(X)E[g(\vec{\theta})|X]\} = E[\delta(X)]^2.$$

Then

$$r_\delta(\Pi) = E[\delta(X) - g(\vec{\theta})]^2 = E[\delta(X)]^2 + E[g(\vec{\theta})]^2 - 2E[g(\vec{\theta})\delta(X)] = 0.$$

## Consistency

An estimator $T$ is consistent for $\vartheta$ if $T \to_p \vartheta$ as $n \to \infty$

Bayes estimators are usually consistent and approximately unbiased.
When Bayes estimators have explicit forms, it is usually easy to check directly whether Bayes estimators are consistent and approximately unbiased (Examples 4.7-4.9).
Bayes estimators also have some other good asymptotic properties, which are studied in §4.5.3.

### Example 4.7

Let $X_1,...,X_n$ be i.i.d. from the exponential distribution $E(0,\theta)$ with an unknown $\theta > 0$.
Let the prior be such that $\omega = \theta^{-1}$ has the gamma distribution $\Gamma(\alpha,\gamma)$ with known $\alpha > 0$ and $\gamma > 0$.
Then the posterior of $\omega = \theta^{-1}$ is the gamma distribution $\Gamma(n+\alpha,(n\bar{X}+\gamma^{-1})^{-1})$, where $\bar{X}$ is the sample mean.

## Consistency

An estimator $T$ is consistent for $\vartheta$ if $T \to_p \vartheta$ as $n \to \infty$

Bayes estimators are usually consistent and approximately unbiased.
When Bayes estimators have explicit forms, it is usually easy to check directly whether Bayes estimators are consistent and approximately unbiased (Examples 4.7-4.9).
Bayes estimators also have some other good asymptotic properties, which are studied in §4.5.3.

## Example 4.7

Let $X_1, ..., X_n$ be i.i.d. from the exponential distribution $E(0, \theta)$ with an unknown $\theta > 0$.
Let the prior be such that $\omega = \theta^{-1}$ has the gamma distribution $\Gamma(\alpha, \gamma)$ with known $\alpha > 0$ and $\gamma > 0$.
Then the posterior of $\omega = \theta^{-1}$ is the gamma distribution $\Gamma(n + \alpha, (n\bar{X} + \gamma^{-1})^{-1})$, where $\bar{X}$ is the sample mean.

## Consistency

An estimator $T$ is consistent for $\vartheta$ if $T \to_p \vartheta$ as $n \to \infty$

Bayes estimators are usually consistent and approximately unbiased.
When Bayes estimators have explicit forms, it is usually easy to check directly whether Bayes estimators are consistent and approximately unbiased (Examples 4.7-4.9).
Bayes estimators also have some other good asymptotic properties, which are studied in §4.5.3.

## Example 4.7

Let $X_1,...,X_n$ be i.i.d. from the exponential distribution $E(0,\theta)$ with an unknown $\theta > 0$.
Let the prior be such that $\omega = \theta^{-1}$ has the gamma distribution $\Gamma(\alpha,\gamma)$ with known $\alpha > 0$ and $\gamma > 0$.
Then the posterior of $\omega = \theta^{-1}$ is the gamma distribution $\Gamma(n+\alpha,(n\bar{X}+\gamma^{-1})^{-1})$, where $\bar{X}$ is the sample mean.

## Example 4.7 (continued)

Consider first the estimation of $\theta = \omega^{-1}$.
The Bayes estimator of $\theta$ under the squared error loss is

$$\delta(X) = \frac{(n\bar{X} + \gamma^{-1})^{n+\alpha}}{\Gamma(n+\alpha)} \int_0^\infty \omega^{n+\alpha-2} e^{-(n\bar{X}+\gamma^{-1})\omega} d\omega = \frac{n\bar{X} + \gamma^{-1}}{n+\alpha-1}.$$

The bias of $\delta(X)$ is

$$\frac{n\theta + \gamma^{-1}}{n+\alpha-1} - \theta = \frac{\gamma^{-1} - (\alpha-1)\theta}{n+\alpha-1} = O\left(\frac{1}{n}\right).$$

It is also easy to see that $\delta(X)$ is consistent.
The UMVUE of $\theta$ is $\bar{X}$.
Since $\mathrm{Var}(\bar{X}) = \theta^2/n$, $r_{\bar{X}}(\Pi) > 0$ for any $\Pi$
Hence, $\bar{X}$ is not a Bayes estimator.
In this case, $\bar{X}$ is the generalized Bayes estimator w.r.t. the improper prior $\frac{d\Pi}{d\omega} = I_{(0,\infty)}(\omega)$ and is a limit of Bayes estimators $\delta(X)$ as $\alpha \to 1$ and $\gamma \to \infty$.

## Example 4.7 (continued)

Consider next the estimation of $e^{-t/\theta} = e^{-t\omega}$.
The Bayes estimator under the squared error loss is

$$\begin{aligned}\delta_t(X) &= \frac{(n\bar{X}+\gamma^{-1})^{n+\alpha}}{\Gamma(n+\alpha)}\int_0^\infty \omega^{n+\alpha-1}e^{-(n\bar{X}+\gamma^{-1}+t)\omega}d\omega \\ &= \left(1+\frac{t}{n\bar{X}+\gamma^{-1}}\right)^{-(n+\alpha)}.\end{aligned}$$

Again, this estimator is biased and it is easy to show that $\delta_t(X)$ is consistent as $n\to\infty$.
In this case, the UMVUE given in Example 3.3 is neither a Bayes estimator nor a limit of $\delta_t(X)$.