Asymptotics III: Bayes Inference and Large-Sample Tests

MIT 18.655

Dr. Kempthorne

Spring 2016
Outline

1. Asymptotics of Bayes Posterior Distributions
   - Consistency of Posterior Distribution
     - Asymptotic Normality of Posterior Distribution
     - Mutual Optimality of Bayes and MLE Procedures
   - Large Sample Tests
     - Likelihood Ratio Tests
     - Wald’s Large Sample Test
     - The Rao Score Test
Consistency of Posterior Distribution

Framework

- \(X_1, \ldots, X_n\) iid \(P_\theta, \theta \in \Theta\).
- \(\Theta\) (open) \(\subset \mathbb{R}\) or \(\Theta = \{\theta_1, \ldots, \theta_k\}\) finite.
- Regular model with identifiable \(\theta\).

Consistency: Finite \(\Theta\)

Posterior distribution of \(\theta\) given \(X_n = (X_1, \ldots, X_n)\):

\[
\pi(\theta' | X_n) \equiv P[\theta = \theta' | X_1, \ldots, X_n], \theta' \in \Theta.
\]

**Definition:** \(\pi(\cdot | X_n)\) is **consistent** if and only if for every \(\theta' \in \Theta\),

\[P_{\theta'}[|\pi(\theta' | X_n) - 1|] \geq \epsilon \to 0\]

for all \(\epsilon > 0\).

**Definition:** \(\pi(\cdot | X_n)\) is **a.s. (almost surely) consistent** if and only if for every \(\theta' \in \Theta\),

\[
\pi(\theta' | X_n) \xrightarrow{\text{a.s.} P_{\theta'}} 1.
\]
Consistency of Posterior Distribution

**Theorem 5.5.1** Let $\pi_j = P[\theta = \theta_j]$, $j = 1, \ldots, k$ denote the prior distribution of $\theta$. Then

$$\pi(\cdot \mid X_n)$$ is consistent iff $\pi_j > 0$, for all $\pi_j \in \Theta$.

**Proof:**

- Let $p(x \mid \theta)$ denote the density/pmf function of a single $X_i$.

  The posterior distribution is given by:

  $$\pi(\theta_j \mid X_1, \ldots, X_n) = P[\theta = \theta_j \mid X_1, \ldots, X_n] = \frac{\pi_j \prod_{i=1}^{n} p(X_i \mid \theta_j)}{\sum_{a=1}^{k} \pi_a \prod_{i=1}^{n} p(X_i \mid \theta_a)}$$

  If any $\pi_j = 0$, then $\pi(\theta_j \mid X_n) = 0$ for all $n$; i.e., the posterior is not consistent.

- Suppose all $\pi_j > 0$. For a fixed $j$, suppose $\theta_j$ is true, i.e., $\theta = \theta_j$.

  We show that

  $$\pi(\theta_j \mid X_n) \longrightarrow 1 \text{ and } \pi(\theta_a \mid X_n) \longrightarrow 0, \text{ for } a \neq j.$$
Proof (continued)

Evaluate the log of the posterior odds to the true \( \theta \):

\[
\log \left[ \frac{\pi(\theta_a \mid X_n)}{\pi(\theta_j \mid X_n)} \right] = \log \left[ \frac{\pi_a \prod_{i=1}^{n} p(X_i \mid \theta_a)}{\pi_j \prod_{i=1}^{n} p(X_i \mid \theta_j)} \right] \\
= \log \left[ \frac{\pi_a}{\pi_j} \right] + \log \left[ \frac{\prod_{i=1}^{n} p(X_i \mid \theta_a)}{\prod_{i=1}^{n} p(X_i \mid \theta_j)} \right] \\
= \log \left[ \frac{\pi_a}{\pi_j} \right] + \sum_{i=1}^{n} \log \left[ \frac{p(X_i \mid \theta_a)}{p(X_i \mid \theta_j)} \right] \\
= n \left( \frac{1}{n} \log \left[ \frac{\pi_a}{\pi_j} \right] \right) + \frac{1}{n} \sum_{i=1}^{n} \log \left[ \frac{p(X_i \mid \theta_a)}{p(X_i \mid \theta_j)} \right] \\
\rightarrow n \left( 0 + E \left[ \log \left[ \frac{p(X_1 \mid \theta_a)}{p(X_1 \mid \theta_j)} \right] \right] \right) \\
\rightarrow \begin{cases} 
0 & \text{if } a = j \\
-\infty & \text{if } a \neq j 
\end{cases}

(Shannon’s Inequality gives \( E \left[ \log \left[ \frac{p(X_1 \mid \theta_a)}{p(X_1 \mid \theta_j)} \right] \right] < 0 \), for \( a \neq j \))
Outline

1. Asymptotics of Bayes Posterior Distributions
   - Consistency of Posterior Distribution
   - Asymptotic Normality of Posterior Distribution
   - Mutual Optimality of Bayes and MLE Procedures

2. Large Sample Tests
   - Likelihood Ratio Tests
   - Wald’s Large Sample Test
   - The Rao Score Test
Theorem 5.5.2 ("Bernstein/von Mises").

- \( \mathbf{X}_n = (X_1, \ldots, X_n) \) where the \( X_i \) are iid \( P_{\theta_0}, \theta_0 \in \Theta \).
- \( \hat{\theta}_n = \hat{\theta}_n(\mathbf{X}_n) \) is the MLE of \( \theta_0 \).
- Regularity conditions are satisfied such that
  \[
  \sqrt{n}(\hat{\theta}_n - \theta_0) \xrightarrow{L} N(0, I^{-1}(\theta_0)).
  \]
- The prior distribution on \( \Theta \) has density \( \pi(\cdot) \) which is continuous and positive at all \( \theta' \in \Theta \).
- Consider the scaled version of the posterior distribution:
  \[
  \mathcal{L} \left( \sqrt{n}(\theta - \hat{\theta}) \mid \mathbf{X}_n \right)
  \]

Under sufficient regularity conditions:
\[
\mathcal{L} \left( \sqrt{n}(\theta - \hat{\theta}) \mid \mathbf{X}_n \right) \longrightarrow N(0, I^{-1}(\theta_0))
\]
i.e.,
\[
\pi \left( \sqrt{n}(\theta - \hat{\theta}) \leq x \mid \mathbf{X}_n \right) \longrightarrow \Phi(x \sqrt{I(\theta_0)})
\]
Proof:

- To compute the asymptotic distribution of $\sqrt{n}(\theta - \hat{\theta}(X_n))$, define

$$t = \sqrt{n}(\theta - \hat{\theta}(X_n))$$

so that

$$\theta = \hat{\theta}(X_n) + \frac{t}{\sqrt{n}}.$$ 

- The posterior density of $t$ given $X_n$ is

$$q_n(t) \propto \pi(\hat{\theta}(X_n) + \frac{t}{\sqrt{n}}) \prod_{i=1}^{n} p(X_i \mid \hat{\theta}(X_n) + \frac{t}{\sqrt{n}})$$

$$= c_n^{-1} \pi(\hat{\theta}(X_n) + \frac{t}{\sqrt{n}}) \prod_{i=1}^{n} p(X_i \mid \hat{\theta}(X_n) + \frac{t}{\sqrt{n}})$$

where $c_n = \int_{-\infty}^{\infty} \pi(\hat{\theta}(X_n) + \frac{t}{\sqrt{n}}) \prod_{i=1}^{n} p(X_i \mid \hat{\theta}(X_n) + \frac{t}{\sqrt{n}})dt$.

- Divide numerator and denominator of $q_n(t)$ by

$$\prod_{i=1}^{n} p(X_i \mid \hat{\theta}(X_n))$$
Proof (continued)

\[ q_n(t) = \frac{1}{c_n^{-1}} \pi(\hat{\theta}(X_n) + \frac{t}{\sqrt{n}}) \prod_{i=1}^{n} p(X_i | \hat{\theta}(X_n) + \frac{t}{\sqrt{n}}) \]

\[ = \frac{1}{c_n^{-1}} \pi(\hat{\theta} + \frac{t}{\sqrt{n}}) \exp\{\sum_{i=1}^{n} \log(p(X_i | \hat{\theta} + \frac{t}{\sqrt{n}}))\} \]

\[ = \frac{1}{d_n^{-1}} \pi(\hat{\theta} + \frac{t}{\sqrt{n}}) \exp\{\sum_{i=1}^{n} \ell(X_i | \hat{\theta} + \frac{t}{\sqrt{n}}) - \ell(X_i, \hat{\theta})\} \]

where

\[ d_n = \int_{-\infty}^{\infty} \pi(\hat{\theta} + \frac{t}{\sqrt{n}}) \exp\{\sum_{i=1}^{n} \ell(X_i | \hat{\theta} + \frac{t}{\sqrt{n}}) - \ell(X_i, \hat{\theta})\} \, dt \]

Claims

- \( d_n q_n(t) \xrightarrow{P_{\theta_0}} \pi(\theta_0) \exp\{-\frac{t^2I(\theta_0)}{2}\} \)
- \( d_n \xrightarrow{P_{\theta_0}} \pi(\theta_0) \int_{-\infty}^{\infty} \exp\{-\frac{s^2I(\theta_0)}{2}\} \, ds = \frac{\pi(\theta_0) \sqrt{2\pi}}{\sqrt{I(\theta_0)}} \)

which give:

\[ q_n \xrightarrow{P_{\theta_0}} \sqrt{I(\theta_0)} \phi(t \sqrt{I(\theta_0)}). \]

Theorem follows by Scheffe’s Theorem (B.7.6).
Limiting Posterior Distributions: Examples

Posterior Distribution of Normal Mean

- $X_1, \ldots, X_n$ iid $N(\theta_0, \sigma^2)$ with $\sigma^2$ known.
- Prior distribution: $\theta \sim N(\eta, \tau^2)$.
- Posterior distribution:
  \[
  \pi(\theta | X_n) = N(\eta_n, \tau_n^2),
  \]
  where
  \[
  \tau_n^{-2} = \tau^{-2} + \frac{n}{\sigma^2}
  \]
  \[
  \eta_n = w_n \eta + (1 - w_n) \bar{X}, \text{ with } w_n = \frac{\sigma^2}{n\tau^2 + \sigma^2}
  \]

Note:

- $\eta_n \rightarrow \hat{\theta} = \bar{X}, \tau_n^2 \rightarrow 0$, and $\bar{X} \xrightarrow{P_{\theta_0}} \theta$, so
  \[
  \pi(\theta | X_n) \xrightarrow{P_{\theta_0}} \text{ point-mass at } \theta = \theta_0.
  \]
- A posteriori,
  \[
  \sqrt{n}(\theta - \hat{\theta}) \sim N(\sqrt{n}w_n(\eta - \bar{X}), n\left(\frac{n}{\sigma^2} + \frac{1}{\tau^2}\right)^{-1})
  \]
  \[
  \longrightarrow N(0, I^{-1}(\theta_0)) = N(0, \sigma^2)
  \]
Limiting Posterior Distributions: Examples

Posterior Distribution of Success Probability in Bernoulli Trials

- $X_1, \ldots, X_n$ iid $Bernoulli(\theta_0)$.
- $S_n = \sum_{1}^{n} X_i \sim Binomial(n, \theta_0)$.
- Prior distribution: $\theta \sim Beta(r, s)$.
- Posterior distribution $\theta \mid S_n \sim Beta(r^*, s^*)$, where $r^* = S_n + r$, and $s^* = s + (n - S_n)$.
- By Problem 5.3.20, if $r^* \to \infty$ and $s^* \to \infty$ such that $r^*/(r^* + s^*) \to \theta_0 \in (0, 1)$, then the $Beta(r^*, s^*)$ r.v. $\theta$:
  $$P \left[ \sqrt{r^* + s^*} \frac{\theta - r^*/(r^* + s^*)}{\sqrt{\theta_0(1 - \theta_0)}} \right] \to N(0, 1).$$

This is easily shown to be equivalent to
  $$\sqrt{n} (\theta - \bar{X}) \overset{L}{\to} N(0, \theta_0(1 - \theta_0)) = N(0, I^{-1}(\theta_0))$$
Outline

1. Asymptotics of Bayes Posterior Distributions
   - Consistency of Posterior Distribution
   - Asymptotic Normality of Posterior Distribution
   - Mutual Optimality of Bayes and MLE Procedures

2. Large Sample Tests
   - Likelihood Ratio Tests
   - Wald’s Large Sample Test
   - The Rao Score Test
**Theorem 5.5.3** Under the conditions of the previous theorems, let $\hat{\theta}$ be the MLE of $\theta$ and let $\hat{\theta}^*$ be the median of the posterior distribution of $\theta$. Then

(i). From a frequentist point of view, i.e., given $P_\theta$:

$$\sqrt{n}(\hat{\theta}^* - \hat{\theta}) \xrightarrow{a.s.} 0, \text{ for all } \theta$$

$$\hat{\theta}^* = \theta + \frac{1}{n} \sum_{i=1}^{n} I^{-1}(\theta) \frac{\partial \ell}{\partial \theta}(X_i, \theta) + o_{P_\theta}(n^{-1/2})$$

$$\sqrt{n}(\hat{\theta}^* - \theta) \xrightarrow{L} N(0, I^{-1}(\theta)).$$

(ii). From a Bayesian point of view, i.e., for $\pi(\theta | X_1, \ldots, X_n)$:

$$E[\sqrt{n}(|\theta - \hat{\theta}|) - |\theta - \hat{\theta}^*| | X_1, \ldots, X_n] = o_P(1), \text{ and }$$

$$E[\sqrt{n}(|\theta - \hat{\theta}|) - |\theta| | X_1, \ldots, X_n] = \min_d \left( E[\sqrt{n}(|\theta - d|) - |\theta| | X_1, \ldots, X_n] \right) + o_P(1).$$
Significant Results

- Bayes estimates for a wide variety of loss functions and priors are asymptotically efficient in the sense being asymptotically unbiased with minimum asymptotic variance.
- Maximum-likelihood estimates are asymptotically equivalent in a Bayesian sense to the Bayes estimate for a variety of priors and loss functions.
  
  E.g., the Bayesian posterior median with \( L(\theta, d) = |\theta - d| \),
  
  the Bayesian posterior mean with \( L(\theta, d) = |\theta - d|^2 \).
Theorem 5.5.4 Under the conditions of the previous theorems, consider

- The **Bayes Credible Region**:
  \[ C_n(X_1, \ldots, X_n) = \{ \theta : \pi(\theta \mid X_1, \ldots, X_n) \geq c_n \} \]
  where \( c_n \) is chosen so that \( \pi(C_n \mid X_1, \ldots, X_n) = 1 - \alpha \).

- For \( \gamma : 0 < \gamma < 1 \), the level \( (1 - \gamma) \) **Asymptotically Optimal Interval Estimate** based on \( \hat{\theta} \), given by
  \[ \text{Interval}_n(\gamma) = [\hat{\theta} - d_n(\gamma), \hat{\theta} + d_n(\gamma)] \]
  where \( d_n(\gamma) = [\Phi^{-1}(1 - \gamma/2)] \times \left( \frac{1}{\sqrt{n} \sqrt{[I(\theta_0)]}} \right) \).

Then, for every \( \epsilon > 0 \), and every \( \theta \):
\[
P_\theta [\text{Interval}_n(\alpha + \epsilon) \subset C_n(X_1, \ldots, X_n) \subset \text{Interval}_n(\alpha - \epsilon)] \longrightarrow 1 \]
Outline

1. Asymptotics of Bayes Posterior Distributions
   - Consistency of Posterior Distribution
   - Asymptotic Normality of Posterior Distribution
   - Mutual Optimality of Bayes and MLE Procedures

2. Large Sample Tests
   - Likelihood Ratio Tests
   - Wald’s Large Sample Test
   - The Rao Score Test
Likelihood Ratio Test Statistic

- \( \mathbf{X}_n = (X_1, \ldots, X_n) \) iid \( P_\theta, \theta \in \Theta \).
- Testing null vs alternative hypotheses:
  \[ H : \theta \in \Theta_0 \text{ vs } K : \theta \not\in \Theta_0. \]
- Likelihood ratio statistic:
  \[
  \lambda(x_n) = \frac{\sup_{\theta \in \Theta} p(x_n | \theta)}{\sup_{\theta \in \Theta_0} p(x_n | \theta)}
  \]
  Standard transformation:
  \[
  2 \log \lambda(x_n) = 2[\ell_n(\hat{\theta} | x_n) - \ell_n(\hat{\theta}_0 | x_n)]
  \]
  where \( \hat{\theta}(x_n) \) is the MLE (over all \( \Theta \)) and \( \hat{\theta}_0(x_n) \) is the MLE under \( H : \theta \in \Theta_0 \).

Theorem 6.3.1 Given suitable assumptions (e.g. Theorem 6.2.2), if \( \Theta \subset \mathbb{R}^r \), and \( H : \theta = \theta_0 \) is true, then

\[
2 \log \lambda(x) = 2[\ell_n(\hat{\theta} | x) - \ell_n(\theta_0)] \xrightarrow{\mathcal{L}} \chi^2_r,
\]
Theorem 6.2.2 Proof

- By Theorem 6.2.2. Given suitable assumptions, the MLE \( \hat{\theta}(x_n) \) satisfies
  \[
  \hat{\theta}(x_n) = \theta + \frac{1}{n} \sum_{i=1}^{n} I^{-1}(\theta)D\ell(X_i, \theta) + o_P(n^{-1/2})
  \]
  so that
  \[
  \sqrt{n}(\hat{\theta}(x_n) - \theta) \xrightarrow{L} N(0, I^{-1}(\theta)).
  \]

- The Taylor expansion of \( \ell_n(\theta) \) about \( \hat{\theta}(x_n) \) evaluated at \( \theta = \theta_0 \) gives
  \[
  2 \log \lambda(x) = 2[\ell_n(\hat{\theta} | x) - \ell_n(\theta_0 | X)]
  = n(\hat{\theta}(x_n) - \theta_0)^T I_n(\theta^*)(\hat{\theta}(x_n) - \theta_0)
  \]
  where \( I_n(\theta) = \| - \frac{1}{n} \sum_{i=1}^{n} \frac{\partial}{\partial \theta_k} \frac{\partial}{\partial \theta_j} \log p(X_i | \theta) \| \),
  the \( r \times r \) matrix: \( I_n(\theta) \xrightarrow{P_{\theta_0}} I(\theta_0) \).
With $V \sim N(0, I^{-1}(\theta_0))$,
\[
2 \log \lambda(x) = 2[\ell_n(\hat{\theta} | x) - \ell_n(\theta_0 | X)]
= n(\hat{\theta}(x_n) - \theta_0)^T I_n(\theta^*)(\hat{\theta}(x_n) - \theta_0)
\xrightarrow{\mathcal{L}} V^T I(\theta_0) V
\]
and by Corollary B.6.2
\[
V^T I(\theta_0) V \sim \chi^2_r.
\]

**Theorem 6.3.2** Given suitable assumptions (e.g. Theorem 6.2.2), if $\Theta \subset \mathbb{R}^r$, and $H : \theta \in \Theta_0$ with $\Theta_0$ of dimension $q < r$, then
\[
2 \log \lambda(x) = 2[\ell_n(\hat{\theta} | x) - \ell_n(\hat{\theta}_0 | X)] \xrightarrow{\mathcal{L}} \chi^2_{r-q}.
\]
Outline

1. Asymptotics of Bayes Posterior Distributions
   - Consistency of Posterior Distribution
   - Asymptotic Normality of Posterior Distribution
   - Mutual Optimality of Bayes and MLE Procedures

2. Large Sample Tests
   - Likelihood Ratio Tests
   - Wald’s Large Sample Test
   - The Rao Score Test
The Wald Test

The asymptotic level-\( \alpha \) Wald Test of the simple hypothesis

\[ H : \theta = \theta_0 \text{ vs } K : \theta \neq \theta_0 \]

rejects \( H \) when

\[ W_n(\theta_0) = n(\hat{\theta}(x_n) - \theta_0)^T I(\theta_0)(\hat{\theta}(x_n) - \theta_0) \geq C^*, \]

where the critical value \( C^* \) is such that \( P(\chi^2_r > C^*) = 1 - \alpha \).

- Under the assumptions of Theorem 6.2.2
  \[ \sqrt{n}(\hat{\theta}(x_n) - \theta) \xrightarrow{\mathcal{L}} N(0, I^{-1}(\theta)). \]

- By Slutsky’s theorem:
  \[ n(\hat{\theta}(x_n) - \theta)^T I(\theta)(\hat{\theta}(x_n) - \theta) \xrightarrow{\mathcal{L}} V^T I(\theta)V \]
  where \( V \sim N_r(0, I^{-1}(\theta)). \)

The Wald Test extends to apply to a composite null hypothesis \( H : \theta \in \Theta_0 \subset \mathbb{R}^q \). If the MLE \( \hat{\theta}_0(x_n) \) under the null is consistent, then it can replace \( \theta_0 \) in the Wald Test statistic which is asymptotically \( \chi^2_{r-q} \) under \( H \), where \( q \) is the dimensionality of \( \Theta_0 \).
Outline

1. Asymptotics of Bayes Posterior Distributions
   - Consistency of Posterior Distribution
   - Asymptotic Normality of Posterior Distribution
   - Mutual Optimality of Bayes and MLE Procedures

2. Large Sample Tests
   - Likelihood Ratio Tests
   - Wald’s Large Sample Test
   - The Rao Score Test
The Rao Score Test

- Simple hypothesis \( H : \theta = \theta_0 \).
- Apply the Central Limit Theorem to the maximum-likelihood contrast function, evaluated at \( \theta = \theta_0 \):
  \[
  \psi_n(\theta_0) = \frac{1}{n} \sum_{i=1}^{n} D_\theta \ell_n(\theta_0) \xrightarrow{\mathcal{L}} N(0, I(\theta_0)),
  \]
  when \( H \) is true.
- It follows that under \( H \)
  \[
  R_n(\theta_0) = n\psi_n^T(\theta_0)I^{-1}(\theta_0)\psi_n(\theta_0) \xrightarrow{\mathcal{L}} \chi^2_r.
  \]

The asymptotic level-\( \alpha \) Rao Score Test rejects \( H \) when

\[
R_n(\theta_0) \geq C^*
\]

where \( C^* : P(\chi^2_r > C^*) = 1 - \alpha \).

Notes:

- The Rao Score Test does not require the MLE!!
- Extension to composite null hypothesis \( H \) only requires MLE under \( H \) (see Theorem 6.3.5).
18.655 Mathematical Statistics
Spring 2016

For information about citing these materials or our Terms of Use, visit: http://ocw.mit.edu/terms.