Ch 9. Interval Estimation

Intro

Definition

Let $X_1, \cdots X_n$ have joint pdf/pmf $f(\mathbf{x} : \theta)$, $\theta \in \Theta$. Let $L(\mathbf{X})$ and $U(\mathbf{X})$ be two statistics such that $L(\mathbf{X}) \leq U(\mathbf{X})$ with probability 1.

1. The random interval $I(\mathbf{X}) = [L(\mathbf{X})\ ,\ U(\mathbf{X})]$ is called an interval estimator for θ.
2. $I(\mathbf{X}) = (-\infty\ ,\ U(\mathbf{X})]$ is said to be a one-sided upper interval estimator for θ.
3. $I(\mathbf{X}) = [L(\mathbf{X})\ ,\ \infty)$ is said to be a one-sided lower interval estimator for θ.
4. The coverage probability of an interval estimator $I(\mathbf{X})$ is defined as $P_\theta[I(\mathbf{X}) \ni \theta]$.
5. The confidence coefficient of $I(\mathbf{X})$ is defined as $\inf_{\theta \in \Theta} P_\theta[I(\mathbf{X}) \ni \theta]$.

Example 1

Let $X_1, \ldots, X_n$ be i.i.d. $U(0, \theta)$, and let $Y = \max(X_1, \ldots, X_n)$. We are interested in an interval estimator of θ.

Consider two candidates: $[aY, bY]$, $1 \le a < b$ and $[Y + c, Y + d]$, $0 \le c < d$. Recall that θ must be larger than Y.

For the first interval,

$$P_\theta\left(\theta \in [aY, bY]\right) = P_\theta\left(aY \le \theta \le bY\right)$$

$$= P_\theta\left(\frac{1}{b} \le \frac{Y}{\theta} \le \frac{1}{a}\right)$$

$$= P_\theta\left(\frac{1}{b} \le T \le \frac{1}{a}\right), \text{ where } T = \frac{Y}{\theta}.$$

We have shown that $f_Y(y) = ny^{n-1}/\theta^n$, $0 < y < \theta$ and $f_T(t) = nt^{n-1}$, $0 < t < 1$.

We then have

$$P_\theta\left(\frac{1}{b} \le T \le \frac{1}{a}\right) = \int_{1/b}^{1/a} nt^{n-1}dt = \left(\frac{1}{a}\right)^n - \left(\frac{1}{b}\right)^n.$$

Thus the coverage probability of this interval is independent of θ. Also, for $\theta \ge d$

$$P_\theta(Y + c \le \theta \le Y + d) = P_\theta\left(\frac{Y+c}{\theta} \le 1 \le \frac{Y+d}{\theta}\right)$$

$$= P_\theta\left(1 - \frac{d}{\theta} \le T \le 1 - \frac{c}{\theta}\right)$$

$$= \int_{1-d/\theta}^{1-c/\theta} nt^{n-1}dt = \left(1 - \frac{c}{\theta}\right)^n - \left(1 - \frac{d}{\theta}\right)^n.$$

In this case the coverage probability depends on θ, and

$$\lim_{\theta\to\infty}\left(1-\frac{c}{\theta}\right)^n-\left(1-\frac{d}{\theta}\right)^n=0,$$

Thus the confidence coefficient of this interval is 0.

▷ Example: $X_1, \cdots, X_n \overset{iid}{\sim} N(\theta, 1)$

$$H_0 : \theta = \theta_0 \quad vs \quad H_1 : \theta \neq \theta_0$$

The LRT of size α is

$$\phi(\mathbf{x}) = \begin{cases} 1, & |\sqrt{n}(\bar{x} - \theta_0)| \geq z_{1-\alpha/2}, \\ 0, & \text{elsewhere.} \end{cases}$$

$$\Longrightarrow P_{\theta_0}\{\bar{X} - z_{1-\alpha/2}/\sqrt{n} \leq \theta_0 \leq \bar{X} + z_{1-\alpha/2}/\sqrt{n}\} =$$

Theorem

Let $X_1, \cdots, X_n$ have joint pdf/pmf $f(\mathbf{x} : \theta)$, $\theta \in \Theta$. For each $\theta_0 \in \Theta$, let $A(\theta_0)$ denote the acceptance region of a size α simple test for testing

$$H_0 : \theta = \theta_0 \quad vs \quad H_1 : \theta \neq \theta_0$$

Define a set $C(\mathbf{x}) = \{\theta_0 \in \Theta : \mathbf{x} \in A(\theta_0)\}$. Then $C(\mathbf{X})$ is a confidence set with confidence coefficient $1 - \alpha$.

$\triangleleft$ Note:

1. $C(\mathbf{X})$ is not necessarily an interval.
2. One may need to consider one-sided test for one-sided confidence interval.

Example 2

▷ Example: $X_1, \cdots, X_n \overset{iid}{\sim} N(\mu, \sigma^2)$. μ and σ^2 are unknown. Find a $1-\alpha$ one-sided CI and two-sided CI for μ.

1. We invert the one-sided test

$$H_0 : \mu = \mu_0 \text{ against } H_1 : \mu < \mu_0$$

The test rejects *H0* if $\dfrac{\bar{x} - \mu_0}{s / \sqrt{n}} < -t_{n-1,\alpha}$.

Thus $A(\mu_0) = \left\{ \mathbf{x} : \bar{x} \geq \mu_0 - t_{n-1,\alpha} \dfrac{s}{\sqrt{n}} \right\}$.

Then

$$C^*(\mathbf{x}) = \{\mu_0 : \bar{x} \in C'\} = \left\{\mu_0 : \bar{x} + t_{n-1,\alpha} \frac{s}{\sqrt{n}} \geq \mu_0\right\}.$$

Then the random interval $C^*(\mathbf{X}) = \left(-\infty, \bar{X} + t_{n-1,\alpha} \frac{S}{\sqrt{n}}\right)$ is a $1-\alpha$ level confidence set for μ.

2. Consider the LRT of

$H_0 : \mu = \mu_0$ against $H_1 : \mu \neq \mu_0$, which rejects the null hypothesis if

$$\frac{\left|\bar{x} - \mu_0\right|}{s / \sqrt{n}} > t_{n-1,\alpha/2}.$$

So, we do not reject if

$$\frac{\left|\bar{x} - \mu_0\right|}{s / \sqrt{n}} \leq t_{n-1,\alpha/2} \Leftrightarrow -t_{n-1,\alpha/2} \frac{s}{\sqrt{n}} \leq \bar{x} - \mu_0 \leq t_{n-1,\alpha/2} \frac{s}{\sqrt{n}}$$

$$\Leftrightarrow \bar{x} - t_{n-1,\alpha/2} \frac{s}{\sqrt{n}} \leq \mu_0 \leq \bar{x} + t_{n-1,\alpha/2} \frac{s}{\sqrt{n}}.$$

This is the $1-\alpha$ level confidence interval for μ.

Example 3

Let $X_1, \ldots, X_n$ be i.i.d. exponential(θ). Consider the LRT of $H_0 : \theta = \theta_0$ against $H_1 : \theta \neq \theta_0$.

$$\lambda(\mathbf{x}) = \frac{\left(\frac{1}{\theta_0}\right)^n \exp\left[-\frac{\sum x_i}{\theta_0}\right]}{\sup_\theta \left(\frac{1}{\theta}\right)^n \exp\left[-\frac{\sum x_i}{\theta}\right]}$$

$$= \frac{\left(\frac{1}{\theta_0}\right)^n \exp\left[-\frac{\sum x_i}{\theta_0}\right]}{\left(\frac{1}{\bar{x}}\right)^n \exp\left[-\frac{\sum x_i}{\bar{x}}\right]}$$

$$= \left(\frac{\sum x_i}{n\theta_0} \right)^n \exp(n) \exp\left[-\frac{\sum x_i}{\theta_0} \right].$$

So, the region where we do not reject is

$$A(\theta_0) = \left\{ \mathbf{x} : \left(\frac{\sum x_i}{\theta_0} \right)^n \exp\left[-\frac{\sum x_i}{\theta_0} \right] \ge k \right\}.$$

The confidence set is

$$C(\mathbf{x}) = \left\{ \theta : \left(\frac{\sum x_i}{\theta} \right)^n \exp\left[-\frac{\sum x_i}{\theta} \right] \ge k \right\}.$$

Note: this curve will change for each sample, but will be unimodal with probability 1, with a maximum at $\theta = \overline{x}$.

So

$$A(\theta) = \left\{ \mathbf{x} : a \le \frac{\sum x_i}{\theta} \le b \right\} \text{ with } b > a.$$

we need $a^n e^{-a} = b^n e^{-b} = k$ (1)

(High density region)

We also have a probability constraint.

$$P\left(a \le \frac{\sum x_i}{\theta} \le b \right) = 1 - \alpha$$

Note that $\sum x_i \sim Gamma(n,\theta)$,

$$2\sum X_i / \theta \ \sim Gamma(n,2) = \chi^2(2n)$$

$$\mathrm{F}(2b) - \mathrm{F}(2a) = 1-\alpha \qquad (2)$$

We can solve (1) and (2) numerically to determine the values of a and b.

CI : $\left\{\theta : \frac{1}{b}\sum x_i \le \theta \le \frac{1}{a}\sum x_i\right\}$.

For convenience, we could solve (2) and choose equal tail areas.

Definition

Let $X_1, \cdots, X_n$ have joint pdf/pmf $f(\mathbf{x} : \theta)$, $\theta \in \Theta$. A random variable $Y = Q(\mathbf{X} : \theta)$ is called a *pivotal quantity (PQ)* if the distribution of $Y = Q(\mathbf{X} : \theta)$ does not depend on θ.

▷ Example: $X_1, \cdots, X_n \overset{iid}{\sim} f(x : \theta)$. Consider the following families of distributions and statistics.

1. $f(x : \theta) = f_0(x - \theta)$
2. $f(x : \theta) = \frac{1}{\theta} f_0(x)$
3. $f(x : \theta) = \frac{1}{\theta_2} f_0[(x - \theta_1)/\theta_2]$

$$\bar{X}_n - \theta, \quad \bar{X}_n/\theta, \quad (\bar{X}_n - \theta_1)/\theta_2$$

Example 3

▷ Example: $X_1, \cdots, X_n \overset{iid}{\sim} \exp(\lambda)$.

$$T = \sum X_i \sim \text{Gamma}(\qquad , \qquad)$$

Find a $(1-\alpha)\%$ confidence interval of λ.

If $X_1,\ldots,X_n$ i.i.d. $\exp(\theta)$, then $\sum X_i/\theta \sim Gamma(n,1)$. We could also use $2\sum X_i/\theta$, which has a $Gamma(n,\ 2\ )=\chi^2_{2n}$ distribution. As before,

$$C\left(\sum x_i\right)=\left\{\theta : a \le \frac{\sum x_i}{\theta} \le b\right\}$$

$$=\left\{\theta : \frac{1}{b}\sum x_i \le \theta \le \frac{1}{a}\sum x_i\right\}.$$

We need $P\left\{a \le \frac{\sum x_i}{\theta} \le b\right\} = 1-\alpha$.

We could choose the “high density region”, as we did with the LRT, or we could choose equal tail areas.

Example 4

Let $X_1, \ldots, X_n$ be i.i.d. $N\left(\mu, \sigma^2\right)$, where both parameters are unknown.

$$\frac{\left|\overline{x} - \mu_0\right|}{s / \sqrt{n}} = t_{n-1}$$

is a pivot for this location-scale family.

So we choose a, b so that $P\left(a \le T_{n-1} \le b\right) = 1 - \alpha$.

Here the "high density region" and equal tail region match since the distribution of the pivot is symmetric. So the $1-\alpha$ confidence region is

$$\left\{ \mu : \overline{x} - t_{n-1,\alpha/2} \frac{s}{\sqrt{n}} \le \mu \le \overline{x} + t_{n-1,\alpha/2} \frac{s}{\sqrt{n}} \right\}.$$

If we want an interval for σ^2 we should consider the sufficient statistic S^2.

$$\frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{n-1} \text{ is a pivot.}$$

We have $P\left(a \le \chi^2_{n-1} \le b\right) = 1-\alpha$ so the confidence interval is

$$\left\{\sigma^2 : \frac{(n-1)S^2}{b} \le \sigma^2 \le \frac{(n-1)S^2}{a}\right\}.$$

But as before, how do we choose a,b? We could use equal tails, i.e. have equal area under the tails, but is that optimal?

The pivots that we have seen so far were mainly constructed using location and scale transformations.

However, it is not always the case.

We can pivot using the cdf of a sufficient statistic.

Theorem (See the theorem 2.1.10 for reference)

Suppose $T = T(\mathbf{X})$ *is a statistic calculated from* $X_1, \cdots, X_n$. *Assume* T *has a continuous distribution with cdf*

$$F(t:\theta) = P_\theta(T \le t).$$

Then

$$Q(T:\theta) = F(T:\theta)$$

is a PQ.

The idea is that given a formula for the cdf of the statistic, we transform the observed value by this formula. This transformation gives the pivot.

Thus, if $\alpha_1 + \alpha_2 = \alpha$, a non-rejection region for the hypothesis $H_0 : \theta = \theta_0$ is given

$$\left\{ t : \alpha_1 \le F_T\left(t \middle| \theta_0\right) \le 1 - \alpha_2 \right\},$$

with the associated confidence set is

$$\left\{ \theta : \alpha_1 \le F_T\left(t \middle| \theta\right) \le 1 - \alpha_2 \right\}.$$

To guarantee that the confidence set is an interval, we need $F_T\left(t \middle| \theta_0\right)$ to be monotone in θ.

Example 5

Let $X_1,\ldots,X_n$ be i.i.d. $f\left(x|\mu\right)=e^{-(x-\mu)}I\left[x>\mu\right]$

(location exponential).

Here $T = X_{(1)}$ is the sufficient statistic.

$f_T\left(t|\mu\right)=ne^{-n(t-\mu)}I\left[t>\mu\right]$ (location exponential with parameter $1/n$)

$$F_T\left(t|\mu\right)=\int_{-\infty}^{t} ne^{-n(y-\mu)}I\left[y>\mu\right]dy$$

$=\int_{\mu}^{t} ne^{-n(y-\mu)}dy=1-e^{-n(t-\mu)}$ (decreasing in μ).

Let t_0 be the observed value of t. Using

$$\left\{\mu : \alpha_1 < F\left(t_0 \middle| \mu\right) < 1-\alpha_2\right\}$$

$$\alpha_1 < 1-e^{-n(t_0-\mu)} < 1-\alpha_2$$

$$\Rightarrow \mu_L = t_0 + \frac{1}{n}\log\left(\alpha_2\right),$$

$$\mu_U = t_0 + \frac{1}{n}\log\left(1-\alpha_1\right).$$

Note that $\log\left(\alpha_2\right) < 0$ and $\log\left(1-\alpha_1\right) < 0$ so that both μ_L and μ_U are less than t_0.

▷ Example: $X_1, \cdots, X_n \overset{iid}{\sim} N(\mu, \sigma^2)$, σ^2 is known. Let $T(\mathbf{X}) = \bar{X}$. Then

$$Q(T : \mu) = F(T : \mu) = \Phi\left(\frac{T - \mu}{\sigma/\sqrt{n}}\right).$$

Methods of evaluating interval estimators

In set estimation two quantities compete against each other: the size of the set and the coverage probability.

Ideally, we would want to have a small size and a large coverage probability. However, coverage probability increases as the size of the set increases.

The coverage probability depends, in general, on the parameter, and thus we typically use the confidence coefficient to measure the performance in terms of coverage probability.

When it comes to confidence intervals, for a fixed confidence coefficient $1-\alpha$, we would like the shortest length confidence interval. How should we choose a, b?

Since the interval $\left[L(\mathbf{X}), U(\mathbf{X})\right]$ is random, we can minimize

$$E\left[U(\mathbf{X}) - L(\mathbf{X})\right]$$

subject to

$$P\left(a \le T(\mathbf{X}) \le b\right) = 1-\alpha.$$

CI: Length of CI vs Coverage probability

Definition

$f(x)$ is a unimodal pdf if $f(x)$ is nondecreasing for $x \leq x^*$ and nonincreasing for $x \geq x^*$ in which case x^* is the mode of the distribution.

Theorem(Shortest Interval based on unimodal distn.)

Let $f(x)$ be a unimodal pdf. If the interval $[a, b]$ satisfies

i. $\int_a^b f(x)dx = 1 - \alpha$
ii. $f(a) = f(b) > 0$
iii. $a \leq x^* \leq b$, *when x^* is a mode of $f(x)$*

Then no other interval satisfying (i) shorter than $[a, b]$.

▷ Example: $X_1, \cdots, X_n \overset{iid}{\sim} N(\mu, \sigma^2)$. σ^2 is unknown.

since the t-distribution is unimodal, we may apply Theorem.

Now,the t-distribution is symmetric, so the shortest interval of μ

$$\left[\bar{X} - t_{\alpha/2,n-1}\frac{S}{\sqrt{n}}, \bar{X} + t_{\alpha/2,n-1}\frac{S}{\sqrt{n}}\right].$$

Let a and b be such that $P(a < \chi^2_{n-1} < b) = 1 - \alpha.$

A (1- α)100% confidence interval for σ2 is $\left[\frac{(n-1)S^2}{b}, \frac{(n-1)S^2}{a}\right].$

The length of the interval is $(n-1)S^2\left(\frac{1}{a} - \frac{1}{b}\right).$

(Exercise 9.52) the shortest interval ,a and b satisfy

$$a^2 g_n(a) = b^2 g_n(b) \quad \text{and} \quad G_n(b) - G_n(a) = 1 - \alpha,$$

Test-related optimality

Since there is a one-to-one correspondence between confidence sets and tests of hypotheses, there is some correspondence between optimality of tests and optimality of confidence sets.

However, test-related optimality typically doesn’t directly relate to the size of the confidence set but to the probability of covering false values.

Definition (Probability of false coverage)

For $\theta' \neq \theta$, $P_\theta[L(\mathbf{X}) \leq \theta' \leq U(\mathbf{X})]$

For $\theta' < \theta$, $P_\theta[L(\mathbf{X}) \leq \theta']$

For $\theta' > \theta$, $P_\theta[\theta' \leq U(\mathbf{X})]$

Definition

A $1-\alpha$ confidence interval with minimum probability of false coverage is called a *Uniformly Most Accurate (UMA)* $1-\alpha$ confidence interval.

Ch 9. Interval Estimation

Optimal theory for CI

Theorem (UMA CI based on UMP test)

Let $X_1, \cdots, X_n$ have a joint pdf/pmf $f(\mathbf{x} : \theta)$. Suppose that a UMP test of size α for testing $H_0 : \theta \leq \theta_0 \quad vs \quad H_1 : \theta > \theta_0$ exists and given as

$$\phi(\mathbf{x}) = \begin{cases} 1, & \mathbf{x} \notin A^*(\theta_0), \\ 0, & \mathbf{x} \in A^*(\theta_0). \end{cases}$$

Let $C^*(\mathbf{X})$ be the confidence interval obtained by inverting the UMP acceptance region. Then, for any other $1 - \alpha$ confidence region(set, interval),

$$P_\theta[\theta' \in C^*(\mathbf{X})] \leq P_\theta[\theta' \in C^*(\mathbf{X})],$$

for all $\theta' < \theta$. That is, inverting UMP test yields a UMA confidence region(set, interval).

As an example, let $X_1,\ldots,X_n$ be i.i.d. $N\left(\mu,\sigma^2\right)$, where σ^2 is known. The interval

$$C\left(\overline{x}\right)=\left(\mu:\mu\geq\overline{x}-z_{\alpha}\sigma/\sqrt{n}\right)$$

is a $1-\alpha$ UMA lower confidence bound since it can be obtained by inverting the UMP test of $H_0:\mu=\mu_0$ versus $H_1:\mu>\mu_0$.

The more common two-sided interval,

$$C\left(\overline{x}\right)=\left(\mu:\overline{x}-z_{\alpha/2}\sigma/\sqrt{n}\leq\mu\leq\overline{x}+z_{\alpha/2}\sigma/\sqrt{n}\right)$$

is not UMA, since it is obtained by inverting the two-sided non-rejection region from the test of $H_0:\mu=\mu_0$ versus $H_1:\mu\neq\mu_0$, for which no UMP test exists.

◃ Note: UMP unbiased test can be inverted to obtain UMA unbiased confidence region(set, interval).

A1-a unbiased confidence interval is one for which

$$P_\theta(\theta' \in C(\boldsymbol{X})) \leq 1 - \alpha \quad \text{for all } \theta' \neq \theta.$$

◁ Note: UMP unbiased test can be inverted to obtain UMA unbiased confidence region(set, interval).

▷ Example: $X_1, \cdots, X_n \overset{iid}{\sim} N(\mu, \sigma^2)$. σ^2 is known.

Definition

$[L(\mathbf{x}), U(\mathbf{x})]$ is called a $(1-\alpha)100\%$ *credible set (or Bayesian interval)* if

$$1-\alpha = P[L(\mathbf{x}) < \theta < U(\mathbf{x}) | \mathbf{X} = \mathbf{x}]$$
$$= \begin{cases} \sum_\theta \pi(\theta|\mathbf{x}) & \text{discrete} \\ \int_\theta \pi(\theta|\mathbf{x}) d\theta & \text{continuous} \end{cases}$$

Note: Unlike confidence intervals, Bayes credible sets have a proper probability interpretation on a posteriori.

Example For a given value of θ, suppose that $X_1, \ldots, X_n$ are i.i.d. $N(\theta, 1)$. Suppose the prior distribution for θ is $N(\mu, \sigma^2)$. Find a Bayes credible set for θ.

$$\pi(\theta) = \frac{1}{\sqrt{2\pi}\sigma} e^{-(\theta-\mu)^2/(2\sigma^2)}$$

$$f(\boldsymbol{x}|\theta) = \left(\frac{1}{\sqrt{2\pi}}\right)^n \exp(-\sum_{i=1}^{n}(x_i - \theta)^2/2)$$

$$w(\boldsymbol{x})\pi(\theta|\boldsymbol{x}) = \left(\tfrac{1}{\sqrt{2\pi}}\right)^{n+1} \tfrac{1}{\sigma} \times$$

$$\exp\left[-\tfrac{1}{2}\left((\theta - \mu)^2/\sigma^2 + \textstyle\sum_{i=1}^{n}(x_i - \theta)^2\right)\right]$$

As a function of θ, $\pi(\theta|\boldsymbol{x})$ is proportional to

$$\exp\Big[-\frac{1}{2}\Big((\theta^2 - 2\mu\theta + \mu^2)/\sigma^2 + \sum_{i=1}^{n} x_i^2 - 2\theta n\bar{x} + n\theta^2\Big)\Big],$$

which is proportional to

$$\exp\left[-\frac{1}{2}\left((n + 1/\sigma^2)\theta^2 - 2\theta(n\bar{x} + \mu/\sigma^2)\right)\right].$$

Completing the square shows that the posterior distribution of θ given $\boldsymbol{x}$ is

$$N\left(\frac{n\bar{x} + \mu/\sigma^2}{n + \sigma^{-2}}, \frac{1}{n + \sigma^{-2}}\right).$$

If Y has this normal distribution, then a $(1-\alpha)$ Bayes credible set for θ is any set A such that

$$P(Y \in A) = 1 - \alpha.$$

The smallest such set is

$$\Big[\frac{n\bar{x}+\mu/\sigma^2}{n+\sigma^{-2}} - z_{\alpha/2}\frac{1}{\sqrt{n+\sigma^{-2}}},$$
$$\frac{n\bar{x}+\mu/\sigma^2}{n+\sigma^{-2}} + z_{\alpha/2}\frac{1}{\sqrt{n+\sigma^{-2}}}\Big].$$

Note:
As $\sigma \to \infty$, this interval tends to the "classical" confidence interval.

In general, the smallest $(1-\alpha)$ Bayes credible set is the *highest posterior density* (HPD) region, which is defined as follows:

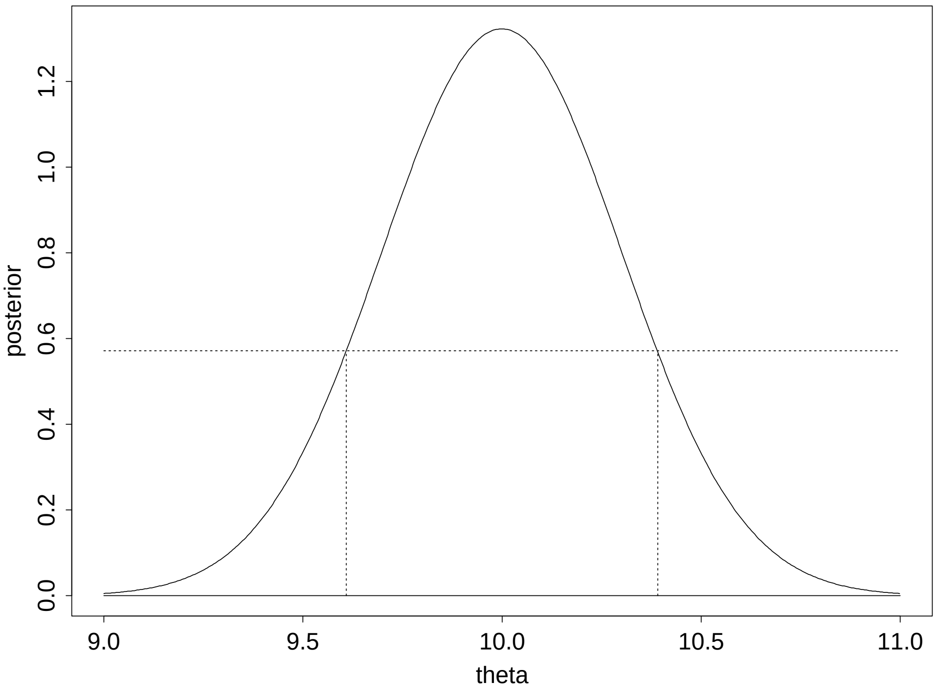
Given data $\boldsymbol{x}$, the $(1-\alpha)$ HPD region is

$$\{y : \pi(y|\boldsymbol{x}) \geq c\},$$

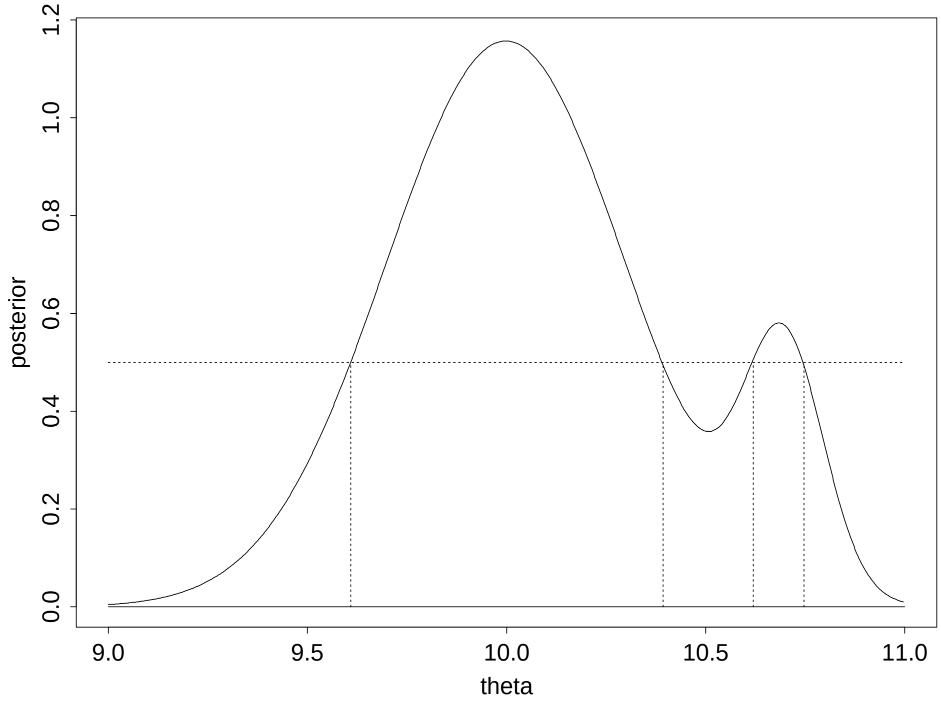
where c is such that

$$\int_{\{y:\pi(y|\boldsymbol{x})\geq c\}} \pi(y|\boldsymbol{x})\, dy = 1-\alpha.$$

An 80% HPD Region



HPD Region for a Bimodal Posterior



Homework: p452~p457

9.2, 9.4, 9.12, 9.13, 9.29, 9.37, 9.52