Stat 710: Mathematical Statistics
Lecture 34

Jun Shao

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

Credible sets

In Bayesian analysis, analogues to confidence sets are called *credible sets*.

Consider a sample X from a population in a parametric family indexed by $\theta \in \Theta \subset \mathscr{R}^k$ and dominated by a σ-finite measure.

Let $f_\theta(x)$ be the p.d.f. of X and $\pi(\theta)$ be a prior p.d.f. w.r.t. a σ-finite measure λ on $(\Theta, \mathscr{B}_\Theta)$.

Let

$$p_x(\theta) = f_\theta(x)\pi(\theta)/m(x)$$

be the posterior p.d.f. w.r.t. λ, where x is the observed X and $m(x) = \int_\Theta f_\theta(x)\pi(\theta)d\lambda$.

For any $\alpha \in (0,1)$, a level $1-\alpha$ credible set for θ is any $C \in \mathscr{B}_\Theta$ with

$$P_{\theta|x}(\theta \in C) = \int_C p_x(\theta)d\lambda \geq 1-\alpha.$$

Highest posterior density (HPD)

A level $1-\alpha$ *highest posterior density* (HPD) credible set for θ is defined to be the event

$$C(x) = \{\theta : p_x(\theta) \geq c_\alpha\},$$

where c_α is chosen so that $\int_{C(x)} p_x(\theta) d\lambda \geq 1-\alpha$.
When $p_x(\theta)$ has a continuous c.d.f., we can replace $\geq$ in the definitions for credible sets and HPD by $=$.
An HPD credible set is often an interval with the shortest length among all credible intervals of the same level (Exercise 40).

Discussions

The Bayesian credible sets and the confidence sets we have discussed so far are very different in terms of their meanings and interpretations, although sometimes they look similar.
In a credible set, x is fixed and θ is considered random and the probability statement is w.r.t. the posterior probability $P_{\theta|x}$.

Highest posterior density (HPD)

A level $1-\alpha$ *highest posterior density* (HPD) credible set for θ is defined to be the event

$$C(x) = \{\theta : p_x(\theta) \geq c_\alpha\},$$

where c_α is chosen so that $\int_{C(x)} p_x(\theta) d\lambda \geq 1-\alpha$.
When $p_x(\theta)$ has a continuous c.d.f., we can replace $\geq$ in the definitions for credible sets and HPD by $=$.
An HPD credible set is often an interval with the shortest length among all credible intervals of the same level (Exercise 40).

Discussions

The Bayesian credible sets and the confidence sets we have discussed so far are very different in terms of their meanings and interpretations, although sometimes they look similar.
In a credible set, x is fixed and θ is considered random and the probability statement is w.r.t. the posterior probability $P_{\theta|x}$.

Discussions

On the other hand, in a confidence set θ is nonrandom (although unknown) but X is considered random, and the confidence level is w.r.t. $P(\theta \in C(X))$, the probability related to the distribution of X.
The HPD set $C(X)$ is not necessarily a confidence set with confidence level $1-\alpha$.
When $\pi(\theta)$ is constant, which is usually an improper prior, the HPD credible set $C(x)$ is related to the idea of maximizing likelihood (a non-Bayesian approach introduced in §4.4; see also §7.3.2), since $p_x(\theta) = f_\theta(x)/m(x)$ is proportional to $f_\theta(x) = \ell(\theta)$, the likelihood function.
In such a case $C(X)$ may be a confidence set with confidence level $1-\alpha$.
More details about Bayesian credible sets can be found, for example, in Berger (1985, §4.3).

Example 7.11

Let $X_1,...,X_n$ be i.i.d. as $N(\theta,\sigma^2)$ with an unknown $\theta \in \mathscr{R}$ and a known σ^2.
Let $\pi(\theta)$ be the p.d.f. of $N(\mu_0,\sigma_0^2)$ with known μ_0 and σ_0^2.
Then, $p_x(\theta)$ is the p.d.f. of $N(\mu_*(x),c^2)$ (Example 2.25), where

$$\mu_*(x) = \frac{\sigma^2}{n\sigma_0^2+\sigma^2}\mu_0 + \frac{n\sigma_0^2}{n\sigma_0^2+\sigma^2}\bar{x} \qquad \text{and} \qquad c^2 = \frac{\sigma_0^2\sigma^2}{n\sigma_0^2+\sigma^2}$$

The HPD credible set is

$$\begin{aligned} C(x) &= \left\{\theta : e^{-[\theta-\mu_*(x)]^2/(2c^2)} \geq c_\alpha \sqrt{2\pi}c\right\} \\ &= \left\{\theta : |\theta-\mu_*(x)| \leq \sqrt{2}c[-\log(c_\alpha\sqrt{2\pi}c)]^{1/2}\right\}. \end{aligned}$$

Let Φ be the standard normal c.d.f.
The quantity $\sqrt{2}c[-\log(c_\alpha\sqrt{2\pi}c)]^{1/2}$ must be $cz_{1-\alpha/2}$, where $z_a = \Phi^{-1}(a)$, since it is chosen so that $P_{\theta|x}(C(x)) = 1-\alpha$ and $P_{\theta|x} = N(\mu_*(x),c^2)$.

Example 7.11 (continued)

Therefore,

$$C(x) = [\mu_*(x) - cz_{1-\alpha/2},\ \mu_*(x) + cz_{1-\alpha/2}].$$

If we let $\sigma_0^2 \to \infty$, which is equivalent to taking the Lebesgue measure as the (improper) prior, then $\mu_*(x) = \bar{x}$, $c^2 = \sigma^2/n$, and

$$C(x) = [\bar{x} - \sigma z_{1-\alpha/2}/\sqrt{n},\ \bar{x} + \sigma z_{1-\alpha/2}/\sqrt{n}],$$

which is the same as the confidence interval in Example 2.31 for θ with confidence coefficient $1-\alpha$.
Although the Bayesian credible set coincides with the classical confidence interval, which is frequently the case when a noninformative prior is used, their interpretations are still different.

Prediction sets

In some problems the quantity of interest is the future (or unobserved) value of a random variable ξ.
An inference procedure about a random quantity instead of an unknown nonrandom parameter is called *prediction*.
If the distribution of ξ is known, then a level $1-\alpha$ *prediction set* for ξ is any event C satisfying

$$P_\xi(\xi \in C) \geq 1-\alpha.$$

In applications, however, the distribution of ξ is usually unknown.
Suppose that the distribution of ξ is related to the distribution of a sample X from which prediction will be made.
For instance, $X = (X_1,...,X_n)$ is the observed sample and $\xi = X_{n+1}$ is to be predicted, where $X_1,...,X_n,X_{n+1}$ are i.i.d. random variables.
A set $C(X)$ depending only on the sample X is said to be a level $1-\alpha$ *prediction set* for ξ if

$$\inf_{P\in\mathscr{P}} P\big(\xi \in C(X)\big) \geq 1-\alpha,$$

where P is the joint distribution of (ξ, X) and $\mathscr{P}$ contains all possible P.

Construction of prediction sets

Prediction sets are very similar and closely related to confidence sets.
Hence, some methods for constructing confidence sets can be applied to obtained prediction sets.
For example, if $\Re(X,\xi)$ is a pivotal quantity in the sense that its distribution does not depend on P, then a prediction set can be obtained by inverting $c_1 \leq \Re(X,\xi) \leq c_2$.
The following example illustrates this idea.

Example 7.12

Many prediction problems encountered in practice can be formulated as follows.
The variable ξ to be predicted is related to a vector-valued covariate ζ (called predictor) according to $E(\xi|\zeta) = \zeta^\tau\beta$, where β is a p-vector of unknown parameters.
Suppose that at $\zeta = Z_i$, we observe $\xi = X_i$, $i = 1,...,n$, and X_i's are independent.

Construction of prediction sets

Prediction sets are very similar and closely related to confidence sets.
Hence, some methods for constructing confidence sets can be applied to obtained prediction sets.
For example, if $\Re(X,\xi)$ is a pivotal quantity in the sense that its distribution does not depend on P, then a prediction set can be obtained by inverting $c_1 \le \Re(X,\xi) \le c_2$.
The following example illustrates this idea.

Example 7.12

Many prediction problems encountered in practice can be formulated as follows.
The variable ξ to be predicted is related to a vector-valued covariate ζ (called predictor) according to $E(\xi|\zeta) = \zeta^\tau \beta$, where β is a p-vector of unknown parameters.
Suppose that at $\zeta = Z_i$, we observe $\xi = X_i$, $i = 1,...,n$, and X_i's are independent.

Example 7.12 (continued)

Based on $(X_1,Z_1),...,(X_n,Z_n)$, we would like to construct a prediction set for the value of $\xi = X_0$ when $\zeta = Z_0 \in \mathscr{R}(Z)$, where Z is the $n \times p$ matrix whose ith row is the vector Z_i.

The Z_i's are either fixed or random observations (in the latter case all probabilities and expectations given in the following discussion are conditional on $Z_0, Z_1, ..., Z_n$).

Assume further that $X = (X_1,...,X_n) = N_n(Z\beta, \sigma^2 I_n)$ follows a normal linear model and is independent of $X_0 = N(Z_0^\tau \beta, \sigma^2)$.

Let $\widehat{\beta}$ be the LSE of β, $\widehat{\sigma}^2 = \|X - Z\widehat{\beta}\|^2/(n-r)$, and $\|Z_0\|_Z^2 = Z_0^\tau (Z^\tau Z)^- Z_0$, where r is the rank of Z.

Note that X_0 and $Z_0^\tau \widehat{\beta}$ are independently normal,

$$E(X_0 - Z_0^\tau \widehat{\beta}) = 0, \qquad \mathrm{Var}(X_0 - Z_0^\tau \widehat{\beta}) = \sigma^2(1 + \|Z_0\|_Z^2),$$

$(n-r)\widehat{\sigma}^2$ has the chi-square distribution χ^2_{n-r}, and X_0, $Z_0^\tau \widehat{\beta}$, and $\widehat{\sigma}^2$ are independent (Theorem 3.8).

Example 7.12 (continued)

Based on $(X_1,Z_1),...,(X_n,Z_n)$, we would like to construct a prediction set for the value of $\xi = X_0$ when $\zeta = Z_0 \in \mathscr{R}(Z)$, where Z is the $n \times p$ matrix whose ith row is the vector Z_i.
The Z_i's are either fixed or random observations (in the latter case all probabilities and expectations given in the following discussion are conditional on $Z_0, Z_1, ..., Z_n$).

Assume further that $X = (X_1,...,X_n) = N_n(Z\beta, \sigma^2 I_n)$ follows a normal linear model and is independent of $X_0 = N(Z_0^\tau \beta, \sigma^2)$.
Let $\widehat{\beta}$ be the LSE of β, $\widehat{\sigma}^2 = \|X - Z\widehat{\beta}\|^2/(n-r)$, and $\|Z_0\|_Z^2 = Z_0^\tau (Z^\tau Z)^- Z_0$, where r is the rank of Z.
Note that X_0 and $Z_0^\tau \widehat{\beta}$ are independently normal,

$$E(X_0 - Z_0^\tau \widehat{\beta}) = 0, \qquad \mathrm{Var}(X_0 - Z_0^\tau \widehat{\beta}) = \sigma^2(1 + \|Z_0\|_Z^2),$$

$(n-r)\widehat{\sigma}^2$ has the chi-square distribution χ^2_{n-r}, and X_0, $Z_0^\tau \widehat{\beta}$, and $\widehat{\sigma}^2$ are independent (Theorem 3.8).

Example 7.12 (continued)

Then

$$\Re(X, X_0) = \frac{X_0 - Z_0^\tau \widehat{\beta}}{\widehat{\sigma}\sqrt{1 + \|Z_0\|_Z^2}}$$

has the t-distribution t_{n-r} and, therefore, is a pivotal quantity.

A level $1-\alpha$ prediction interval for X_0 is then

$$\left[Z_0^\tau \widehat{\beta} - t_{n-r,\alpha/2}\widehat{\sigma}\sqrt{1 + \|Z_0\|_Z^2},\, Z_0^\tau \widehat{\beta} + t_{n-r,\alpha/2}\widehat{\sigma}\sqrt{1 + \|Z_0\|_Z^2} \right],$$

where $t_{n-r,\alpha}$ is the $(1-\alpha)$th quantile of the t-distribution t_{n-r}.

To compare prediction sets with confidence sets, let us consider a confidence interval for $E(X_0) = Z_0^\tau \beta$.

Using the pivotal quantity

$$\Re(X, Z_0^\tau \beta) = \frac{Z_0^\tau \beta - Z_0^\tau \widehat{\beta}}{\widehat{\sigma}\|Z_0\|_Z},$$

Example 7.12 (continued)

Then

$$\Re(X, X_0) = \frac{X_0 - Z_0^\tau \widehat{\beta}}{\widehat{\sigma}\sqrt{1 + \|Z_0\|_Z^2}}$$

has the t-distribution t_{n-r} and, therefore, is a pivotal quantity.
A level $1-\alpha$ prediction interval for X_0 is then

$$\left[Z_0^\tau \widehat{\beta} - t_{n-r,\alpha/2}\widehat{\sigma}\sqrt{1 + \|Z_0\|_Z^2}, Z_0^\tau \widehat{\beta} + t_{n-r,\alpha/2}\widehat{\sigma}\sqrt{1 + \|Z_0\|_Z^2} \right],$$

where $t_{n-r,\alpha}$ is the $(1-\alpha)$th quantile of the t-distribution t_{n-r}.

To compare prediction sets with confidence sets, let us consider a confidence interval for $E(X_0) = Z_0^\tau \beta$.
Using the pivotal quantity

$$\Re(X, Z_0^\tau \beta) = \frac{Z_0^\tau \beta - Z_0^\tau \widehat{\beta}}{\widehat{\sigma}\|Z_0\|_Z},$$

Example 7.12 (continued)

we obtain the following confidence interval for $Z_0^\tau \beta$ with confidence coefficient $1-\alpha$:

$$\left[Z_0^\tau \widehat{\beta} - t_{n-r,\alpha/2} \widehat{\sigma} \|Z_0\|_Z \,, Z_0^\tau \widehat{\beta} + t_{n-r,\alpha/2} \widehat{\sigma} \|Z_0\|_Z \right].$$

Since a random variable is more variable than its average (an unknown parameter), the prediction interval is always longer than the confidence interval, although each of them covers the quantity of interest with probability $1-\alpha$.

In fact, when $\|Z_0\|_Z^2 \to 0$ as $n \to \infty$, the length of the confidence interval tends to 0 a.s., whereas the length of the prediction interval tends to a positive constant a.s.

Because of the similarity between confidence sets and prediction sets, in the rest of this chapter we do not discuss prediction sets in detail. Some examples are given in Exercises 30 and 31.

Example 7.12 (continued)

we obtain the following confidence interval for $Z_0^\tau \beta$ with confidence coefficient $1-\alpha$:

$$\left[Z_0^\tau \widehat{\beta} - t_{n-r,\alpha/2} \widehat{\sigma} \|Z_0\|_Z \,,\, Z_0^\tau \widehat{\beta} + t_{n-r,\alpha/2} \widehat{\sigma} \|Z_0\|_Z \right].$$

Since a random variable is more variable than its average (an unknown parameter), the prediction interval is always longer than the confidence interval, although each of them covers the quantity of interest with probability $1-\alpha$.

In fact, when $\|Z_0\|_Z^2 \to 0$ as $n \to \infty$, the length of the confidence interval tends to 0 a.s., whereas the length of the prediction interval tends to a positive constant a.s.

Because of the similarity between confidence sets and prediction sets, in the rest of this chapter we do not discuss prediction sets in detail. Some examples are given in Exercises 30 and 31.