Chapter 9: Interval Estimation and Confidence Sets
Lecture 16: Confidence sets and credible sets

Confidence sets

We consider a sample X from a population indexed by $\theta \in \Theta \subset \mathscr{R}^k$.

We are interested in ϑ, a vector-valued function of θ with range Θ.

We now consider interval estimation or, more generally, set estimation, rather than a point estimator (a single value) of ϑ discussed in Chapter 7, to have some guarantee of capturing the true value of ϑ.

A set estimation is a statement that "$\vartheta \in C(x)$", where $C(x) \subset \Theta$ when the sample $X = x$.

$C(X)$ is called a set estimator of ϑ or a confidence set for ϑ.

$C(X)$ is a random set, whereas $C(x)$ is a realization of $C(X)$.

If ϑ is real-valued, we usually prefer the set estimator to be an interval and, hence, $C(X)$ is an interval estimator or confidence interval for ϑ.

A confidence interval can be expressed as $[L(X), U(X)]$ with a pair of statistics $L(X)$ and $U(X)$ satisfying $L(X) \leq U(X)$.

When X is discrete, we may want to consider open interval $(L(X), U(X))$ or $(L(X), U(X)]$ or $[L(X), U(X))$.

In some cases, we may take $L(X) = \vartheta_-$, which can be $-\infty$ or the inf of Θ; $(\vartheta_-, U(X)]$ is called a one-sided interval estimator or confidence interval or an upper confidence bound for ϑ.

Similarly, $[L(X), \vartheta_+)$ is a one-sided interval estimator or confidence interval or a lower confidence bound for ϑ.

Two key issues in interval or set estimation

Instead of estimating ϑ by a single point $\widehat{\vartheta}$, we must gain something!

(A) By giving up the precision in our estimate, we should gain some confidence or assurance, that our statement $\vartheta \in C(x)$ is correct.

(B) On the other hand, we should also choose an interval or a set $C(x)$ as "small" as possible in some sense, otherwise we could always choose $C(x) = \Theta$ (non-random) and be always correct.

We first address (A), and then turn to (B).

Do (A) and (B) sound like the two types of errors in testing hypotheses?

Definition.

The coverage probability of a set estimator or confidence set $C(X)$ is the probability that the random set $C(X)$ covers ϑ, i.e., $P_\theta(\vartheta \in C(X))$. The confidence coefficient of $C(X)$ is

$$\inf_{\theta \in \Theta} P_\theta(\vartheta \in C(X))$$

For $\alpha \in (0,1)$, we say that $C(X)$ is a $1-\alpha$ set estimator or confidence set if its confidence coefficient is $1-\alpha$; we say that $C(X)$ is level $1-\alpha$ set estimator or confidence set if its confidence coefficient $\geq 1-\alpha$.

We take infimum because we do not know which θ is the true value.

In some cases this does not matter because the coverage probability does not depend on any unknown quantity.

In other cases, however, the confidence coefficient may be much smaller than the true coverage probability.

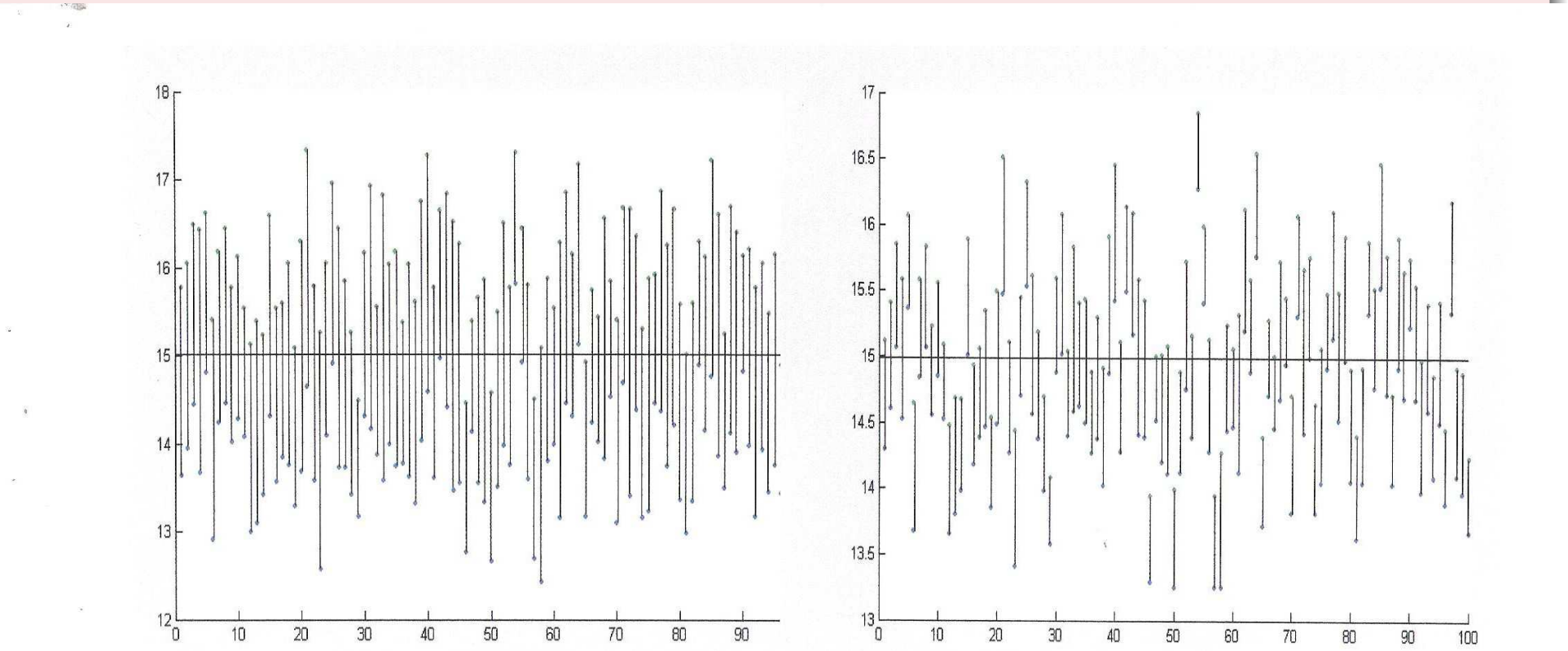
It is important to keep in mind that $C(X)$ is random, not the unknown parameter ϑ.

Interpretation of a confidence set or set estimator

The following figure (left or right) is a plot of 100 confidence intervals that are realizations of a $C(X)$ based on a sample X from a population with one unknown parameter θ whose true value is 15.

In the left figure, the confidence coefficient is 0.9 and 91 out of 100 intervals cover the true value.

In the right figure, the confidence coefficient is 0.5 and 50 intervals cover the true value.



Example.

Let $X_1, ..., X_n$ be iid from $N(\mu, \sigma^2)$ with unknown $\mu \in \mathscr{R}$ and known σ^2.

We consider an interval estimator of μ of the form $C(\bar{X}) = [\bar{X} - c, \bar{X} + c]$ with a constant $c > 0$, since $\bar{X}$ is sufficient and complete for μ and μ is a location parameter.

The coverage probability is

$$P_\mu(\mu \in C(\bar{X})) = P_\mu(\bar{X} - c \leq \mu \leq \bar{X} + c) = 1 - 2\Phi(-\sqrt{n}c/\sigma)$$

which does not depend on any unknown parameter.

Hence, the confidence coefficient of $C(\bar{X})$ is $1 - 2\Phi(-\sqrt{n}c/\sigma)$, an increasing function of c and converges to 1 as $c \to \infty$ or 0 as $c \to 0$.

We can choose a reasonable confidence coefficient with a constant $c > 0$.

Also, the confidence coefficient increases as the measure of variability $\sigma/\sqrt{n}$ decreases.

Consider now the same $C(\bar{X})$ in the situation where σ^2 is unknown.

We still have the same coverage probability, but now it depends on σ

$$P_{\mu,\sigma}(\mu \in C(\bar{X})) = 1 - 2\Phi(-\sqrt{n}c/\sigma)$$

Since σ varies on $(0,\infty)$, for any given c, the confidence coefficient of $C(\bar{X})$ is 0.

Example 9.1.6.

Let $X_1, ..., X_n$ be iid from $uniform(0,\theta)$ with unknown $\theta > 0$.

Let $Y = X_{(n)}$, the sufficient and complete statistic for θ.

We consider two interval estimators of θ based on Y.

The first one is $[aY, bY]$, where $1 \leq a < b$ are constants.

The second one is $[Y + c, Y + d]$, where $0 \leq c < d$ are constants. (Note that $P_\theta(Y \leq \theta) = 1$.)

Since Y has pdf ny^{n-1}/θ^n, $0 < y < \theta$,

$$P_\theta(aY \leq \theta \leq bY) = P\left(\frac{\theta}{b} \leq Y \leq \frac{\theta}{a}\right) = \frac{n}{\theta^n}\int_{\theta/b}^{\theta/a} y^{n-1}dy = a^{-n} - b^{-n}$$

This does not depend on θ and hence it is the confidence coefficient of $[aY, bY]$.

If we choose a and b to have confidence coefficient $1-\alpha$, then

$$a^{-n} - b^{-n} = 1 - \alpha$$

We still have one free constant (a or b) to choose, which will be discussed later.

On the other hand,

$$P_\theta(Y + c \le \theta \le Y + d) = P(\theta - d \le Y \le \theta - c) = \frac{n}{\theta^n} \int_{\theta-d}^{\theta-c} y^{n-1} dy$$

$$= \left(1 - \frac{c}{\theta}\right)^n - \left(1 - \frac{d}{\theta}\right)^n$$

which depends on θ; in fact,

$$\inf_{\theta>0} P_\theta(Y + c \le \theta \le Y + d) = \lim_{\theta\to\infty} \left[\left(1 - \frac{c}{\theta}\right)^n - \left(1 - \frac{d}{\theta}\right)^n\right] = 0$$

i.e., the confidence coefficient of $[Y + c, Y + d]$ is 0.

This example indicates the importance of choosing the right form of the interval estimator.

Credible sets

As we discussed previously, a confidence set $C(X)$ covers the true value of ϑ with respect to the probability from repeated sampling (the randomness of X as a sample).

Once we observe $X = x$, it is not correct to say that ϑ in $C(x)$ with probability $1 - \alpha$.

On the other hand, in the Bayesian approach θ as well as ϑ are treated as random vectors and it is appropriate to state that ϑ is in a set C with certain probability (typically with respect to the posterior distribution of $\vartheta | X = x$).

Such a set C, which also depends on x and can be denoted as C_x, is called a credible set.

Similar to the confidence coefficient or level, we want to find a credible set C_x satisfying $P(\vartheta \in C_x | X = x) \geq 1 - \alpha$ for a given small positive constant α.

Such a set is then referred to as a $1 - \alpha$ credible set.

If $\pi(\theta|x)$ is the posterior pdf of θ given $X = x$, then a $1-\alpha$ credible set C_x for $\vartheta \in \Theta$ is any $C_x \subset \Theta$ such that

$$P(\vartheta \in C_x|x) = \int_{C_x} \pi(\theta|x)d\theta = 1-\alpha$$

Note that if $\vartheta \neq \theta$, then this probability is actually the marginal posterior probability regarding ϑ.

If ϑ is discrete, then we replace the integral by a sum in the above expression, but we may also have to replace $= 1-\alpha$ by $\geq 1-\alpha$ since the equality may not be achieved.

Similar to the classical approach, when ϑ is univariate, a credible interval is preferred; i.e., we want to find L_x and U_x such that

$$P(L_x \leq \vartheta \leq U_x) = 1-\alpha$$

Although the interpretation and construction of the Bayes credible set are more straightforward than those of a classical confidence set, they come with additional assumptions: the Bayesian approach requires more input than the classical approach.

Example 9.2.16.

The following confidence interval is constructed in Example 9.2.15 (which will be discussed in lecture 16) when a random sample $X = (X_1, ..., X_n)$ is taken from $Poisson(\theta)$ with unknown $\theta > 0$:

$$\left[\frac{1}{2n}\chi^2_{2(a+t),1-\alpha/2},\ \frac{1}{2n}\chi^2_{2(a+t),\alpha/2}\right]$$

where $\chi^2_{k,\alpha}$ is the $100(1-\alpha)$th percentile of the chi-square with degrees of freedom k.

We now construct a $1-\alpha$ credible interval for θ and compare it with the confidence interval.

If we choose the conjugate prior, $Gamma(a, b)$, then the posterior pdf of θ is $Gamma(a+t, (n+b^{-1})^{-1})$, where $t = \sum_{i=1}^{n} x_i$, and the posterior pdf for $[2(nb+1)/b]\theta$ is the chi-square with degrees of freedom $2(a+t)$ (assuming that a is an integer).

There are many ways to construct a $1-\alpha$ credible interval $[L_x, U_x]$; one simple way is to split the α equally between the upper and lower endpoints, which produces the $1-\alpha$ credible interval

$$\left[\frac{b}{2(nb+1)}\chi^2_{2(a+t),1-\alpha/2},\ \frac{b}{2(nb+1)}\chi^2_{2(a+t),\alpha/2}\right]$$

Figure 9.2.3 shows the 90% credible and confidence intervals.

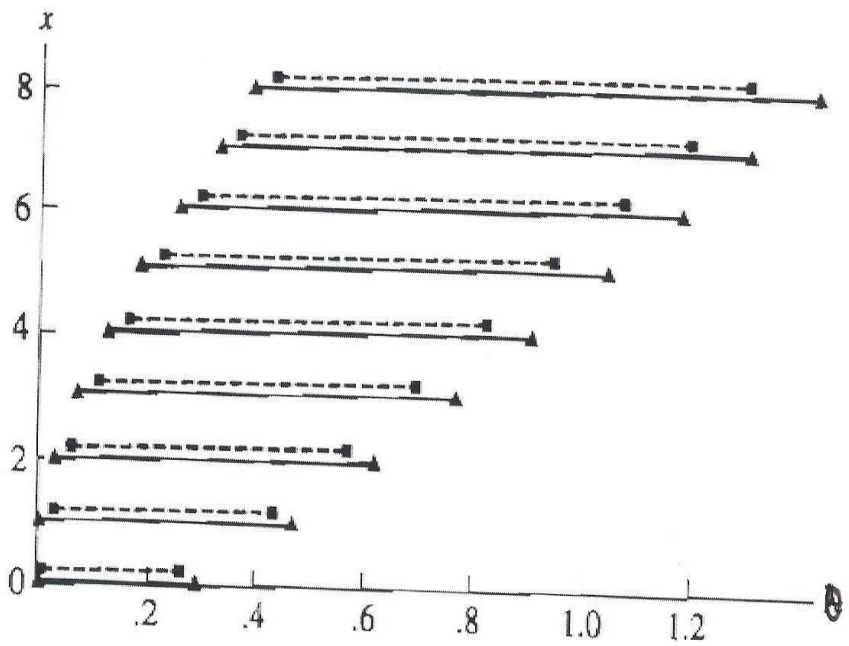


Figure 9.2.3. *The 90% credible intervals (dashed lines) and 90% confidence intervals (solid lines) from Example 9.2.16*

Difference between credible probability and coverage probability

It is important not to confuse credible probability with coverage probability.

- Credible probability is the experimenter's subjective beliefs, as expressed in the prior distribution and updated with the data to the posterior distribution.
- Coverage probability reflects the uncertainty in the sampling procedure, getting its probability from the objective mechanism of repeated experimental trials.
- A credible probability of 90% means that the experiment, upon combining prior knowledge with data, is 90% sure of "ϑ is in the credible set". It does not rely on whether the experiment will be repeated or not.
- A coverage probability of 90% means that in a long sequence of independent and identical trials, 90% of the realized confidence sets will cover the true value of ϑ. It does not ensure whether a confidence set based on a set of observed data covers the true parameter or not.

Example 9.2.17.

The $1-\alpha$ confidence and credible intervals in Examples 9.2.15-16 maintain their respective probability guarantees, but how do they perform under the other criterion?

First, we check the credible probability of the confidence interval derived in Example 9.2.15: with $\theta|x \sim Gamma(a+t,(n+b^{-1})^{-1})$,

$$P\left(\theta \in \left[\frac{1}{2n}\chi^2_{2t,1-\alpha/2},\ \frac{1}{2n}\chi^2_{2t,\alpha/2}\right]\middle| x\right)$$

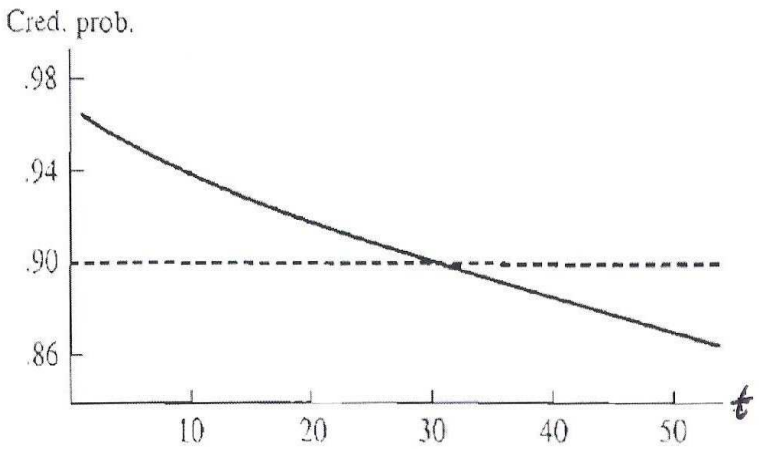


Figure 9.2.4. *Credible probabilities of the 90% credible intervals (dashed line) and 90% confidence intervals (solid line) from Example 9.2.16*

Figure 9.2.4 shows the credible probability of the confidence interval with $1-\alpha = 90\%$, which decreases as t increases.

In fact, this credible probability decreases to 0 as $t \to \infty$ for every fixed n unless $b = n^{-1}$.

The coverage probability of the credible interval derived in Example 9.2.16 is, with $T = \sum_{i=1}^{n} X_i \sim Poisson(n\theta)$,

$$P_\theta\left(\frac{b}{2(nb+1)}\chi^2_{2(a+T),1-\alpha/2} \le \theta \le \frac{b}{2(nb+1)}\chi^2_{2(a+T),\alpha/2}\right)$$

The credible interval does not perform much better when evaluated as a confidence interval.

Figure 9.2.5 suggests that the coverage probability of the credible interval tends to 0 as $\theta \to \infty$, which is in fact true.

In this problem, is there a situation the two approach agree?

Comparing two intervals, we find that they are exactly the same when $a = 0$ and $b = \infty$.

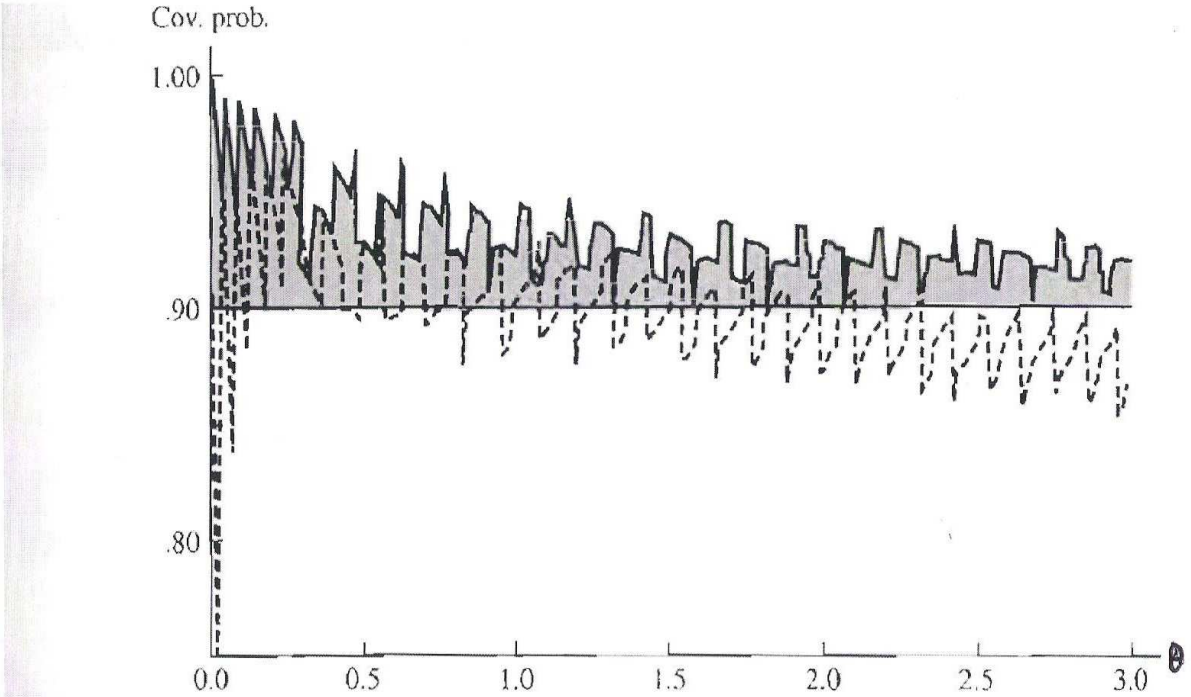


Figure 9.2.3. *Coverage probabilities of the 90% credible intervals (dashed lines) and 90% confidence intervals (solid lines) from Example 9.2.16*

Although $a = 0$ and $b = \infty$ are not valid values for the prior in the Bayesian approach, they are allowed in a generalized Bayesian analysis in which the prior does not have to be a proper distribution.

Or, we can think that the two intervals become the same as $a \to 0$ and $b \to \infty$.

The behavior exhibited in this example is somewhat typical: another example is given in Example 9.2.18.