

8.1: Introduction to Set Estimation

Basic Theory

The Basic Statistical Model

As usual, our starting point is a random experiment with an underlying sample space and a probability measure \mathbb{P} . In the basic statistical model, we have an observable random variable \mathbf{X} taking values in a set S . In general, \mathbf{X} can have quite a complicated structure. For example, if the experiment is to sample n objects from a population and record various measurements of interest, then

$$\mathbf{X} = (X_1, X_2, \dots, X_n) \quad (8.1.1)$$

where X_i is the vector of measurements for the i th object. The most important special case occurs when (X_1, X_2, \dots, X_n) are independent and identically distributed. In this case, we have a random sample of size n from the common distribution.

Suppose also that the distribution of \mathbf{X} depends on a parameter θ taking values in a parameter space Θ . The parameter may also be vector-valued, in which case $\Theta \subseteq \mathbb{R}^k$ for some $k \in \mathbb{N}_+$ and the parameter vector has the form $\theta = (\theta_1, \theta_2, \dots, \theta_k)$.

Confidence Sets

A *confidence set* is a subset $C(\mathbf{X})$ of the parameter space Θ that depends only on the data variable \mathbf{X} , and no unknown parameters. the *confidence level* is the smallest probability that $\theta \in C(\mathbf{X})$:

$$\min \{ \mathbb{P}[\theta \in C(\mathbf{X})] : \theta \in \Theta \} \quad (8.1.2)$$

Thus, in a sense, a confidence set is a *set-valued statistic*. A confidence set is an estimator of θ in the sense that we hope that $\theta \in C(\mathbf{X})$ with high probability, so that the confidence level is high. Note that since the distribution of \mathbf{X} depends on θ , there is a dependence on θ in the probability measure \mathbb{P} in the definition of confidence level. However, we usually suppress this, just to keep the notation simple. Usually, we try to construct a confidence set for θ with a prescribed confidence level $1 - \alpha$ where $0 < \alpha < 1$. Typical confidence levels are 0.9, 0.95, and 0.99. Sometimes the best we can do is to construct a confidence set whose confidence level is *at least* $1 - \alpha$; this is called a *conservative* $1 - \alpha$ confidence set for θ .

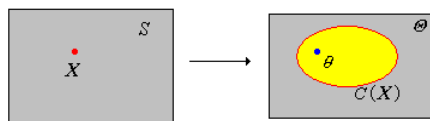


Figure 8.1.1: A set estimate that successfully captured the parameter

Suppose that $C(\mathbf{X})$ is $1 - \alpha$ level confidence set for a parameter θ . Note that when we run the experiment and observe the data \mathbf{x} , the *computed* confidence set is $C(\mathbf{x})$. The true value of θ is either in this set, or is not, and we will usually never know. However, by the law of large numbers, if we were to repeat the confidence experiment over and over, the proportion of sets that contain θ would converge to $\mathbb{P}[\theta \in C(\mathbf{X})] = 1 - \alpha$. This is the precise meaning of the term *confidence*. In the usual terminology of statistics, the random set $C(\mathbf{X})$ is the *estimator*; the deterministic set $C(\mathbf{x})$ based on an observed value \mathbf{x} is the *estimate*.

Next, note that the quality of a confidence set, as an estimator of θ , is based on two factors: the confidence level and the *precision* as measured by the “size” of the set. A good estimator has small size (and hence gives a precise estimate of θ) and large confidence. However, for a given \mathbf{X} , there is usually a tradeoff between confidence level and precision—increasing the confidence level comes only at the expense of increasing the size of the set, and decreasing the size of the set comes only at the expense of decreasing the confidence level. How we measure the “size” of the confidence set depends on the dimension of the parameter space and the nature of the confidence set. Moreover, the size of the set is usually random, although in some special cases it may be deterministic.

Considering the extreme cases may give us some insight. First, suppose that $C(\mathbf{X}) = \Theta$. This set estimator has maximum confidence 1, but no precision and hence it is worthless (we already *knew* that $\theta \in \Theta$). At the other extreme, suppose that $C(\mathbf{X})$ is a singleton set. This set estimator has the best possible precision, but typically for continuous distributions, would have confidence 0. In between these extremes, hopefully, are set estimators that have high confidence and high precision.

Suppose that $C_i(\mathbf{X})$ is a $1 - \alpha_i$ level confidence set for θ for $i \in \{1, 2, \dots, k\}$. If $\alpha = \alpha_1 + \alpha_2 + \dots + \alpha_k < 1$ then $C_1(\mathbf{X}) \cap C_2(\mathbf{X}) \cap \dots \cap C_k(\mathbf{X})$ is a conservative $1 - \alpha$ level confidence set for θ .

Proof

This follows from Bonferroni's inequality.

Real-Valued Parameters

In many cases, we are interested in estimating a real-valued parameter $\lambda = \lambda(\theta)$ taking values in an interval parameter space (a, b) , where $a, b \in \mathbb{R}$ with $a < b$. Of course, it's possible that $a = -\infty$ or $b = \infty$. In this context our confidence set frequently has the form

$$C(\mathbf{X}) = \{\theta \in \Theta : L(\mathbf{X}) < \lambda(\theta) < U(\mathbf{X})\} \quad (8.1.3)$$

where $L(\mathbf{X})$ and $U(\mathbf{X})$ are real-valued statistics. In this case $(L(\mathbf{X}), U(\mathbf{X}))$ is called a *confidence interval* for λ . If $L(\mathbf{X})$ and $U(\mathbf{X})$ are both random, then the confidence interval is often said to be *two-sided*. In the special case that $U(\mathbf{X}) = b$, $L(\mathbf{X})$ is called a *confidence lower bound* for λ . In the special case that $L(\mathbf{X}) = a$, $U(\mathbf{X})$ is called a *confidence upper bound* for λ .

Suppose that $L(\mathbf{X})$ is a $1 - \alpha$ level confidence lower bound for λ and that $U(\mathbf{X})$ is a $1 - \beta$ level confidence upper bound for λ . If $\alpha + \beta < 1$ then $(L(\mathbf{X}), U(\mathbf{X}))$ is a conservative $1 - (\alpha + \beta)$ level confidence interval for λ .

Proof

This follows immediately from (2).

Pivot Variables

You might think that it should be very difficult to construct confidence sets for a parameter θ . However, in many important special cases, confidence sets can be constructed easily from certain random variables known as *pivot variables*.

Suppose that V is a function from $S \times \Theta$ into a set T . The random variable $V(\mathbf{X}, \theta)$ is a *pivot variable* for θ if its distribution does not depend on θ . Specifically, $\mathbb{P}[V(\mathbf{X}, \theta) \in B]$ is constant in $\theta \in \Theta$ for each $B \subseteq T$.

The basic idea is that we try to combine \mathbf{X} and θ algebraically in such a way that we *factor out* the dependence on θ in the distribution of the resulting random variable $V(\mathbf{X}, \theta)$. If we know the distribution of the pivot variable, then for a given α , we can try to find $B \subseteq T$ (that does not depend on θ) such that $\mathbb{P}_\theta[V(\mathbf{X}, \theta) \in B] = 1 - \alpha$. It then follows that a $1 - \alpha$ confidence set for the parameter is given by $C(\mathbf{X}) = \{\theta \in \Theta : V(\mathbf{X}, \theta) \in B\}$.

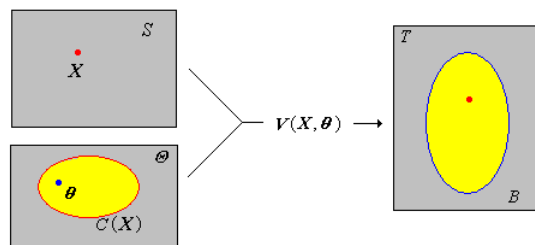


Figure 8.1.2: A confidence set constructed from a pivot variable

Suppose now that our pivot variable $V(\mathbf{X}, \theta)$ is real-valued, which for simplicity, we will assume has a continuous distribution. For $p \in (0, 1)$, let $v(p)$ denote the quantile of order p for the pivot variable $V(\mathbf{X}, \theta)$. By the very meaning of pivot variable, $v(p)$ does not depend on θ .

For any $p \in (0, 1)$, a $1 - \alpha$ level confidence set for θ is

$$\{\theta \in \Theta : v(\alpha - p\alpha) < V(\mathbf{X}, \theta) < v(1 - p\alpha)\} \quad (8.1.4)$$

Proof

By definition, the probability of the event is $(1 - p\alpha) - (\alpha - p\alpha) = 1 - \alpha$.

The confidence set above corresponds to $(1-p)\alpha$ in the left tail and $p\alpha$ in the right tail, in terms of the distribution of the pivot variable $V(\mathbf{X}, \lambda)$. The special case $p = \frac{1}{2}$ is the *equal-tailed case*, the most common case.

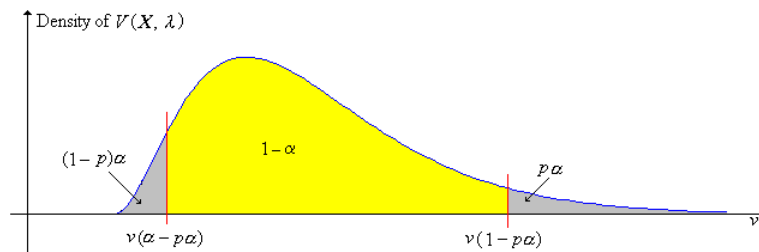


Figure 8.1.3: Distribution of the pivot variable showing $(1-p)\alpha$ is the left tail and $p\alpha$ is the right tail.

The confidence set (5) is decreasing in α and hence increasing in $1-\alpha$ (in the sense of the subset relation) for fixed p .

For the confidence set (5), we would naturally like to choose p that minimizes the size of the set in some sense. However this is often a difficult problem. The equal-tailed interval, corresponding to $p = \frac{1}{2}$, is the most commonly used case, and is sometimes (but not always) an optimal choice. Pivot variables are far from unique; the challenge is to find a pivot variable whose distribution is known and which gives tight bounds on the parameter (high precision).

Suppose that $V(\mathbf{X}, \theta)$ is a pivot variable for θ . If g is a function defined on the range of V and g involves no unknown parameters, then $U = g[V(\mathbf{X}, \theta)]$ is also a pivot variable for θ .

Examples and Special Cases

Location-Scale Families

In the case of location-scale families of distributions, we can easily find pivot variables. Suppose that Z is a real-valued random variable with a continuous distribution that has probability density function g , and no unknown parameters. Let $X = \mu + \sigma Z$ where $\mu \in \mathbb{R}$ and $\sigma \in (0, \infty)$ are parameters. Recall that the probability density function of X is given by

$$f_{\mu, \sigma}(x) = \frac{1}{\sigma} g\left(\frac{x - \mu}{\sigma}\right), \quad x \in \mathbb{R} \quad (8.1.5)$$

and the corresponding family of distributions is called the *location-scale family* associated with the distribution of Z ; μ is the *location parameter* and σ is the *scale parameter*. Generally, we are assuming that these parameters are unknown.

Now suppose that $\mathbf{X} = (X_1, X_2, \dots, X_n)$ is a random sample of size n from the distribution of X ; this is our observable outcome vector. For each i , let

$$Z_i = \frac{X_i - \mu}{\sigma} \quad (8.1.6)$$

The random vector $\mathbf{Z} = (Z_1, Z_2, \dots, Z_n)$ is a random sample of size n from the distribution of Z .

In particular, note that \mathbf{Z} is a pivot variable for (μ, σ) , since \mathbf{Z} is a function of \mathbf{X} , μ , and σ , but the distribution of \mathbf{Z} does not depend on μ or σ . Hence, any function of \mathbf{Z} will also be a pivot variable for (μ, σ) , (if the function does not involve the parameters). Of course, some of these pivot variables will be much more useful than others in estimating μ and σ . In the following exercises, we will explore two common and important pivot variables.

Let $M(\mathbf{X})$ and $M(\mathbf{Z})$ denote the sample means of \mathbf{X} and \mathbf{Z} , respectively. Then $M(\mathbf{Z})$ is a pivot variable for (μ, σ) since

$$M(\mathbf{Z}) = \frac{M(\mathbf{X}) - \mu}{\sigma} \quad (8.1.7)$$

Let m denote the quantile function of the pivot variable $M(\mathbf{Z})$. For any $p \in (0, 1)$, a $1-\alpha$ confidence set for (μ, σ) is

$$Z_{\alpha, p}(\mathbf{X}) = \{(\mu, \sigma) : M(\mathbf{X}) - m(1-p\alpha)\sigma < \mu < M(\mathbf{X}) - m(\alpha-p\alpha)\sigma\} \quad (8.1.8)$$

The confidence set constructed above is a “cone” in the (μ, σ) parameter space, with vertex at $(M(\mathbf{X}), 0)$ and boundary lines of slopes $-1/m(1-p\alpha)$ and $-1/m(\alpha-p\alpha)$, as shown in the graph below. (Note, however, that both slopes might be negative or both positive.)

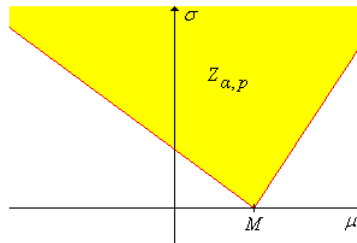


Figure 8.1.4: The confidence set for (μ, σ) constructed from M

The fact that the confidence set is unbounded is clearly not good, but is perhaps not surprising; we are estimating two real parameters with a single real-valued pivot variable. However, if σ is known, the confidence set defines a *confidence interval* for μ . Geometrically, the confidence interval simply corresponds to the horizontal cross section at σ .

$1 - \alpha$ confidence sets for (μ, σ) are

1. $Z_{\alpha,1}(\mathbf{X}) = \{(\mu, \sigma) : M(\mathbf{X}) - m(1 - \alpha)\sigma < \mu < \infty\}$
2. $Z_{\alpha,0}(\mathbf{X}) = \{(\mu, \sigma) : -\infty < \mu < M(\mathbf{X}) - m(\alpha)\sigma\}$

Proof

In the confidence set constructed above, let $p \uparrow 1$ and $p \downarrow 0$, respectively.

If σ is known, then (a) gives a $1 - \alpha$ confidence lower bound for μ and (b) gives a $1 - \alpha$ confidence upper bound for μ .

Let $S(\mathbf{X})$ and $S(\mathbf{Z})$ denote the sample standard deviations of \mathbf{X} and \mathbf{Z} , respectively. Then $S(\mathbf{Z})$ is a pivot variable for (μ, σ) and a pivot variable for σ since

$$S(\mathbf{Z}) = \frac{S(\mathbf{X})}{\sigma} \quad (8.1.9)$$

Let s denote the quantile function of $S(\mathbf{Z})$. For any $\alpha \in (0, 1)$ and $p \in (0, 1)$, a $1 - \alpha$ confidence set for (μ, σ) is

$$V_{\alpha,p}(\mathbf{X}) = \left\{ (\mu, \sigma) : \frac{S(\mathbf{X})}{s(1-p\alpha)} < \sigma < \frac{S(\mathbf{X})}{s(\alpha-p\alpha)} \right\} \quad (8.1.10)$$

Note that the confidence set gives no information about μ since the random variable above is a pivot variable for σ alone. The confidence set can also be viewed as a bounded confidence *interval* for σ .

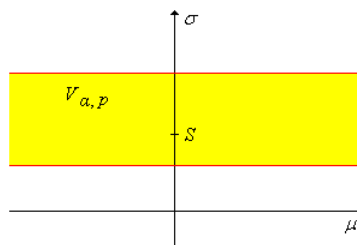


Figure 8.1.5: The confidence set for (μ, σ) constructed from S

$1 - \alpha$ confidence sets for (μ, σ) are

1. $V_{\alpha,1}(\mathbf{X}) = \{(\mu, \sigma) : S(\mathbf{X})/s(1 - \alpha) < \sigma < \infty\}$
2. $V_{\alpha,0}(\mathbf{X}) = \{(\mu, \sigma) : 0 < \sigma < S(\mathbf{X})/s(\alpha)\}$

Proof

In the confidence set constructed above, let $p \uparrow 1$ and $p \downarrow 0$, respectively.

The set in part (a) gives a $1 - \alpha$ confidence lower bound for σ and the set in part (b) gives a $1 - \alpha$ confidence upper bound for σ . We can intersect the confidence sets corresponding to the two pivot variables to produce conservative, bounded confidence sets.

If $\alpha, \beta, p, q \in (0, 1)$ with $\alpha + \beta < 1$ then $Z_{\alpha,p} \cap V_{\beta,q}$ is a conservative $1 - (\alpha + \beta)$ confidence set for (μ, σ) .

Proof

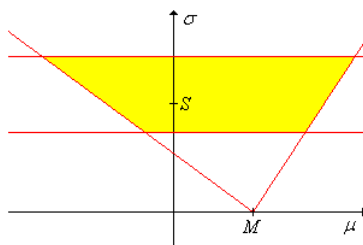


Figure 8.1.6: The bounded confidence set for (μ, σ) constructed from (M, S)

The most important location-scale family is the family of normal distributions. The problem of estimation in the normal model is considered in the next section. In the remainder of this section, we will explore another important scale family.

The Exponential Distribution

Recall that the exponential distribution with scale parameter $\sigma \in (0, \infty)$ has probability density function $f(x) = \frac{1}{\sigma} e^{-x/\sigma}$, $x \in [0, \infty)$. It is the scale family associated with the standard exponential distribution, which has probability density function $g(x) = e^{-x}$, $x \in [0, \infty)$. The exponential distribution is widely used to model random times (such as lifetimes and “arrival” times), particularly in the context of the Poisson model. Now suppose that $\mathbf{X} = (X_1, X_2, \dots, X_n)$ is a random sample of size n from the exponential distribution with unknown scale parameter σ . Let

$$Y = \sum_{i=1}^n X_i \quad (8.1.11)$$

The random variable $\frac{2}{\sigma} Y$ has the chi-square distribution with $2n$ degrees of freedom, and hence is a pivot variable for σ .

Note that this pivot variable is a multiple of the variable M constructed above for general location-scale families (with $\mu = 0$). For $p \in (0, 1)$ and $k \in (0, \infty)$, let $\chi_k^2(p)$ denote the quantile of order p for the chi-square distribution with k degrees of freedom. For selected values of k and p , $\chi_k^2(p)$ can be obtained from the special distribution calculator or from most statistical software packages.

Recall that

1. $\chi_k^2(p) \rightarrow 0$ as $p \downarrow 0$
2. $\chi_k^2(p) \rightarrow \infty$ as $p \uparrow 1$

For any $\alpha \in (0, 1)$ and any $p \in (0, 1)$, a $1 - \alpha$ confidence interval for σ is

$$\left(\frac{2Y}{\chi_{2n}^2(1-p\alpha)}, \frac{2Y}{\chi_{2n}^2(\alpha-p\alpha)} \right) \quad (8.1.12)$$

Note that

1. $2Y/\chi_{2n}^2(1-\alpha)$ is a $1 - \alpha$ confidence lower bound for σ .
2. $2Y/\chi_{2n}^2(\alpha)$ is a $1 - \alpha$ confidence lower bound for σ .

Of the two-sided confidence intervals constructed above, we would naturally prefer the one with the smallest length, because this interval gives the most information about the parameter σ . However, minimizing the length as a function of p is computationally difficult. The two-sided confidence interval that is typically used is the *equal tailed* interval obtained by letting $p = \frac{1}{2}$:

$$\left(\frac{2Y}{\chi_{2n}^2(1-\alpha/2)}, \frac{2Y}{\chi_{2n}^2(\alpha/2)} \right) \quad (8.1.13)$$

The lifetime of a certain type of component (in hours) has an exponential distribution with unknown scale parameter σ . Ten devices are operated until failure; the lifetimes are 592, 861, 1470, 2412, 335, 3485, 736, 758, 530, 1961.

1. Construct the 95% two-sided confidence interval for σ .
2. Construct the 95% confidence lower bound for σ .
3. Construct the 95% confidence upper bound for σ .

Answer

1. (769.1, 2740.1)
2. 836.7
3. 2421.9

This page titled [8.1: Introduction to Set Estimation](#) is shared under a [CC BY 2.0](#) license and was authored, remixed, and/or curated by [Kyle Siegrist \(Random Services\)](#) via [source content](#) that was edited to the style and standards of the LibreTexts platform.