MTH211A: Theory of Statistics

Module 5: Methods of Interval Estimation

Sometimes providing a point estimate is not the ideal method of estimation, as it does not quantify the uncertainly related to the estimate. Rather, one may be interested in an interval (or a set), which *efficiently* captures the underlying parameter. For example, in a diagnostic test for a particular disease, one requires an interval (range) of the possible test results, with efficiently detects the occurrence or non-occurrence of a particular disease. In other words we require a random set which captures the underlying parameter with high probability (rather than a random point which is close to the underlying parameter in appropriate sense). This type of estimates are called interval estimates.

Definition 1 (Interval Estimate)**.** *An* interval estimate *for a real valued parameter θ is a pair of functions of sample observations $(L(\mathbf{x}), U(\mathbf{x}))$ that satisfy $L(\mathbf{x}) \leq U(\mathbf{x})$ for each point $\mathbf{x}$ in its support. Suppose $\mathbf{x}$ is a realization of a random vector $\mathbf{X}$, then the random interval $[L(\mathbf{X}), U(\mathbf{X})]$ is called a* interval estimator *of θ.*

Note: For some particular examples, $L(\mathbf{x})$ can be $-\infty$, or $U(\mathbf{X})$ can be ∞. In such cases we obtain a one sided interval estimate. Further, instead of closed interval, one may obtain an open, or a semi-closed interval estimate as well.

Definition 2 (Confidence Coefficient)**.** *The* confidence coefficient *of an interval estimator $[L(\mathbf{X}), U(\mathbf{X})]$ of a parameter θ, usually denoted by $(1-\alpha)$, is the probability that the random interval captures the true parameter θ, for any $\theta \in \Theta$. Notationally, the confidence coefficient is*

$$(1-\alpha) \coloneqq \inf_{\theta} P_{\theta\in\Theta}\left(L(\mathbf{X}) \leq \theta \leq U(\mathbf{X})]\right).$$

Interval estimators together with confidence coefficient are called confidence intervals.

Note: We can generalize the idea of confidence intervals to *confidence sets*. A random set $S(\mathbf{X})$ is said to be a confidence set for a parameter vector $\boldsymbol{\theta} \in \boldsymbol{\Theta} \subseteq \mathbb{R}^k$, with confidence coefficient $(1-\alpha)$, if $P_\theta\left(S(\mathbf{X}) \ni \theta\right) \geq (1-\alpha)$ for each $\theta \in \Theta$. A confidence interval can be interpreted as a special type of confidence set, where $S(\mathbf{X})$ is an interval.

Interpretation of Confidence Sets: A confidence set $S(\mathbf{X})$ with confidence coefficient $(1-\alpha)$ can be interpreted as follows: If repeated random samples, that is, repeated realizations of $\mathbf{X}$, are taken for a large (theoretically, infinite) number of times, then in $(1-\alpha)100\%$ cases, the realization of the confidence set, $S(\mathbf{x})$, will contain the true parameter $\boldsymbol{\theta}$.

Example 1. Let $X_1, \ldots, X_n$ be a random sample from `Normal`(μ, σ^2) distribution. Consider the interval estimate $[\bar{X}_n - c, \bar{X}_n + c]$ of μ for some constant $c \geq 0$. Find c such that the confidence coefficient is $(1-\alpha)$.

When σ^2 is known, then it can be seen that $c = \sigma\tau_{\alpha/2}/\sqrt{n}$ where $\tau_{\alpha/2}$ is the upper $\alpha/2$ point of the standard normal distribution. When σ^2 is unknown, then $C = C_{\mathbf{X}}$ is random and $C_{\mathbf{X}} = S_n^\star t_{n-1,\alpha/2}/\sqrt{n}$ where $t_{n-1,\alpha/2}$ is the upper $\alpha/2$ point of the t distribution with degrees of freedom $n-1$. [WHY?]

Remark 1. *One could also choose an interval estimate of the type $[\bar{X}_n - c_1, \bar{X}_n + c_2]$ where $c_1, c_2 \geq 0$. Then any c_1, c_2, satisfying $\Phi\left(c_1\sqrt{n}/\sigma\right) - \Phi\left(-c_2\sqrt{n}/\sigma\right) = (1-\alpha)$ would provide a valid confidence interval with confidence coefficient $(1-\alpha)$, when σ^2 is known.* [WHY?]

Example 2. Let $X_1, \ldots, X_n$ be a random sample from `Normal`(μ, σ^2) distribution. Consider the interval estimate $[c_1 S_n^{\star 2}, c_2 S_n^{\star 2}]$ of σ^2 for some constants $0 < c_1 \leq c_2$. Find c_1, c_2 such that the confidence coefficient is $(1-\alpha)$,

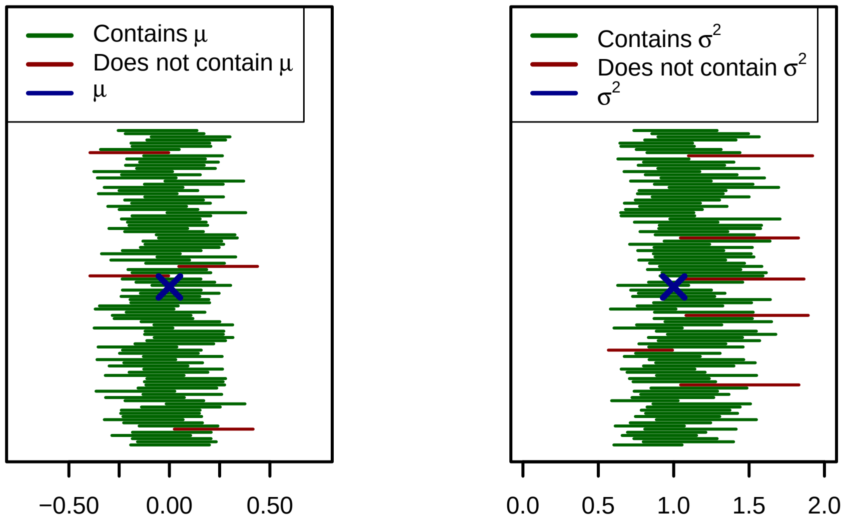

Figure 1: The true value of the parameter (indicated in blue) and a list of 100 interval estimates based on samples of size $n = 100$. An interval estimate is indicated in green if it captures the true parameter, and in red otherwise.

Using the result that $(n-1)\sigma^{-2}S_n^{\star 2} \sim \chi^2_{(n-1)}$, one can show that any c_1, c_2 satisfying

$$P\left(\frac{(n-1)}{c_2} < W \leq \frac{(n-1)}{c_1} \mid W \sim \chi^2_{(n-1)}\right) = (1-\alpha),$$

leads to a valid confidence interval. In particular, one may choose c_1, c_2 such that

$$P\left(W \geq \frac{(n-1)}{c_1} \mid W \sim \chi^2_{(n-1)}\right) = P\left(W \leq \frac{(n-1)}{c_2} \mid W \sim \chi^2_{(n-1)}\right) = \frac{\alpha}{2},$$

which leads to the interval $\left[(n-1)S_n^{\star 2}/\chi^2_{(n-1),1-\alpha/2}, (n-1)S_n^{\star 2}/\chi^2_{(n-1),\alpha/2}\right]$.

1 Methods of Finding Confidence Interval

1.1 Method of Pivots

Definition 3 (Pivot)**.** *Let $\mathbf{X} \sim f_{\mathbf{X}}(\cdot;\theta)$. A random variable $T(\mathbf{X};\theta)$ is called a pivot if the distribution of $T(\mathbf{X};\theta)$ does not depend on θ.*

Note: The pivot $T(\mathbf{X};\theta)$ itself may be a function of θ. However, the distribution of $T(\mathbf{X};\theta)$ must be free of θ.

Example 3. Let $X_1, \cdots, X_n$ be a random sample from a location family with location parameter θ, i.e., $X_i = W_i + \theta$ where $W_i,\ i = 1, \ldots, n$, are iid from a distribution free of θ. Then any function of $X_i - \theta;\ i = 1, \ldots, n$, is a pivot.

Example 4. Let $X_1, \cdots, X_n$ be a random sample from a scale family with scale parameter θ, i.e., $X_i = \theta W_i$ where $W_i,\ i = 1, \ldots, n$, are iid from a distribution free of θ. Then any function of $X_i/\theta;\ i = 1, \ldots, n$, is a pivot.

Example 5. If X has a continuous distribution, then the distribution function $F_X(\cdot;\theta)$ has a $\texttt{Uniform}(0,1)$ distribution. When n iid samples $X_1, \cdots, X_n$ are available, then one may take $T(\mathbf{X};\theta) = -\sum_{i=1}^n \log F_X(X_i;\theta)$ as a pivot. It can be shown that $T(\mathbf{X};\theta)$ distributed as a $\texttt{Gamma}(n,1)$ distribution. [WHY?]

A pivot may yield a confidence interval when supported with some additional features. The following theorem provides a set of sufficient conditions for a pivot to yield a confidence interval.

Theorem 1. *Let $T(\mathbf{X};\theta)$ be a pivot such that for each fixed θ, $T(\mathbf{X};\theta)$ is a statistic and as a function of θ, $T(\mathbf{X};\theta)$ is strictly monotone at each $\mathbf{x} \in \mathcal{X}^n \subseteq \mathbb{R}^n$. Let $\mathcal{T} = \{t = T(\mathbf{X};\theta) : \mathbf{x} \in \mathcal{X}^n \text{ and } \theta \in \Theta\} \in \mathbb{R}$ be the range of $T(\mathbf{X};\theta)$, and for each $t \in \mathcal{T}$ and $\mathbf{x} \in \mathcal{X}^n$ the equation $T(\mathbf{x};\theta) = t$ is solvable with respect to θ. Then one can construct a confidence interval for θ at any level.* [Proof]

Remark 2. *A sufficient condition for the equation $T(\mathbf{x};\theta) = t$ to be solvable is T is continuous and strictly monotone w.r.t. θ.*

For example, let $F_S(S(\mathbf{x});\theta)$ be the cdf of a (continuous) statistic $S(\mathbf{X})$. If $F_S(s;\theta)$ is strictly increasing in θ then for any $\alpha \in (0,1)$ one may choose α_1, α_2 such that $\alpha_1 + \alpha_2 = \alpha$, and the confidence interval $[L(\mathbf{X}), U(\mathbf{X})]$ such that

$$F_S(S(\mathbf{x}); L(\mathbf{x})) = \alpha_1, \quad \text{and} \quad F_S(S(\mathbf{x}); U(\mathbf{x})) = 1 - \alpha_2, \quad \text{for each } \mathbf{x}.$$

Remark 3. *The monotonicity assumption in the above theorem ensures that the confidence set obtained from the pivot $T(\mathbf{X},\theta)$ is of interval type. In case all the other assumptions in Theorem 1 are satisfied, except the monotonicity assumption, then one would still obtain a confidence set, but it may not be of interval type.*

Example 6. Let $X_1, \cdots, X_n$ be a random sample from location exponential distribution, with location parameter θ and scale parameter 1. Then obtain a $(1-\alpha)100\%$ confidence interval based on the complete sufficient statistic $X_{(1)}$ of θ.

1.2 Method of Test Inversion

We will learn this method after *Module 6: Testing of Hypothesis*, if time permits.

2 Method of Evaluating Confidence Intervals

For a particular parameter θ, there may exist many confidence intervals (CIs), due to different approaches, different pivots, or test procedures. Among them, it is desirable to obtain the CI, which has shortest length and largest confidence coefficient. Often the confidence coefficient is set to a pre-assigned level. In that case, it is desirable to obtain the CI, which has shortest length among all CIs with same confidence coefficient.

However, the shortest length CI may not necessarily exist. In Theorem 2, we will demonstrate a procedure for obtaining the shortest interval based on a particular random variable $T(\mathbf{X},\theta)$ (pivot or test procedure), under suitable conditions. However, this procedure does not guarantee that the CI obtained would be the shortest one among all possible CIs . There may exist some other random variable $T^\star(\mathbf{X},\theta)$, which might lead to a better CI.

Theorem 2. *Let X be a continuous random variable with unimodal pdf $f_X(\cdot)$. If the interval $[a,b]$ satisfies*

i. $\int_a^b f_X(x)dx = 1 - \alpha$,

ii. $f_X(b) = f_X(a) > 0$, *and*

iii. $a \le x^\star \le b$, *where $x^\star$ is the mode of $f_X(\cdot)$,*

then $[a,b]$ is the shortest among all intervals that satisfy i.

The proof is omitted due to time constraint. Interested students may read the proof of Theorem 9.3.2. in Casella Berger.

Example 7. Revisit Example 1.

Example 8. Let $X_1, \ldots, X_n$ be a random sample of size n from double exponential distribution with pdf

$$f_X(x;\theta) = \frac{\theta}{2} \exp\{-\theta|x|\}, \quad x \in \mathbb{R}.$$

Observe that the distribution $W = \theta X$ is free of θ. We may choose the pivot in such a way that it is a function of the complete sufficient statistic $\sum_{j=1}^n |X_i|$. Let

$T(\mathbf{X};\theta) = \sum_{i=1}^n |W_i| = \theta \sum_{i=1}^n |X_i|$ be the chosen pivot. Clearly $|W_i|$ follows an exponential distribution with parameter 1, and $\sum_i |W_i|$ follows a $\texttt{Gamma}(n,1)$ distribution. We find $\lambda_i, i = 1,2$ in such a way that

$$P\left(\lambda_1 \le T(\mathbf{X};\theta) \le \lambda_2 \mid T(\mathbf{X};\theta) \sim \texttt{Gamma}(n,1)\right) = P\left(\frac{\lambda_1}{\sum_i |X_i|} \le \theta \le \frac{\lambda_2}{\sum_i |X_i|}\right) = (1-\alpha),$$

and length of the CI $(\lambda_2 - \lambda_1)/\sum_i |X_i|$ is minimized. For a given realization $\mathbf{x}$, minimizing the length of the CI is equivalent to minimizing $(\lambda_2 - \lambda_1)$, and therefore Theorem 2 is applicable here.

Example 9. Let $X_1, \ldots, X_n$ be a random sample of size n from double exponential distribution with pdf

$$f_X(x;\theta) = \frac{1}{2\theta} \exp\left\{-\frac{|x|}{\theta}\right\}, \quad x \in \mathbb{R}.$$

Observe that, Theorem 2 is not applicable here. As in the example above, we consider the pivot is $T(\mathbf{X};\theta) = \sum_{i=1}^n |W_i| = \sum_{i=1}^n |X_i|/\theta$, which has a $\texttt{Gamma}(n,1)$ distribution. Now we need to find $\lambda_i, i = 1,2$, such that

$$P\left(\lambda_1 \le T(\mathbf{X};\theta) \le \lambda_2 \mid T(\mathbf{X};\theta) \sim \texttt{Gamma}(n,1)\right) = P\left(\frac{\sum_i |X_i|}{\lambda_2} \le \theta \le \frac{\sum_i |X_i|}{\lambda_1}\right) = (1-\alpha), \tag{1}$$

and the length of the CI $(\lambda_1^{-1} - \lambda_2^{-1})\sum_i |X_i|$ is minimized. In such cases, one may minimize the length w.r.t. (λ_1, λ_2) subject to the constraint (1).

Instead of minimizing the length of CI, one may also minimize the expected length of the CI. The procedure for finding the shortest length CI, is also applicable for finding the shortest expected length CI.