557: MATHEMATICAL STATISTICS II
INTERVAL ESTIMATION

For a random sample $X_1, \ldots, X_n$ from parametric probability model $f_{X|\theta}(x|\theta)$, an **interval estimator** for scalar parameter θ comprises a pair of statistics, $(L(\underset{\sim}{X}), U(\underset{\sim}{X}))$, such that for all $\underset{\sim}{x} \in \mathcal{X}$,

$$L(\underset{\sim}{x}) \leq U(\underset{\sim}{x}).$$

Specifically, the interval estimator for θ is the **random** closed interval $[L(\underset{\sim}{X}), U(\underset{\sim}{X})]$; for data $\underset{\sim}{x}$, the reported inference is that $L(\underset{\sim}{x}) \leq \theta \leq U(\underset{\sim}{x})$. Note that $L(.)$ and $U(.)$ may take the value infinity yielding **one-sided** intervals, and that **open** intervals may be used on occasion.

For interval estimator $[L(\underset{\sim}{X}), U(\underset{\sim}{X})]$, the following quantities describe the properties of the estimator:

- **Coverage Probability:** The **coverage probability** is defined as

  $$\gamma_{L,U}(\theta) = \Pr[\theta \in [L(\underset{\sim}{X}), U(\underset{\sim}{X})] | \theta] = \int_{\mathcal{X}_\theta} f_{\underset{\sim}{X}|\theta}(\underset{\sim}{x}|\theta)\, d\underset{\sim}{x}$$

  where

  $$\mathcal{X}_\theta \equiv \{\underset{\sim}{x} : L(\underset{\sim}{x}) \leq \theta \leq U(\underset{\sim}{x})\}$$

  defines a region in $\mathcal{X}$.

- **Confidence Coefficient or Level :** The **confidence coefficient or level** is defined as

  $$\inf_{\theta \in \Theta} \gamma_{L,U}(\theta)$$

  Often, the interval estimator is selected so that the confidence level is equal to some fixed target value, say $1-\alpha$. In practice, this is the only procedure that is possible, as in general the true value of θ is unknown, so that although the coverage probability function $\gamma_{L,U}(\theta)$ can be identified, the specific value at which it should be evaluated is not typically known.

Interval estimators are also referred to as **confidence intervals**. Note that, more generally **confidence sets** could be considered if the set $\underset{\sim}{x} \in \mathcal{X}$ is not a single interval, but is instead the union of intervals in $\mathbb{R}$. When constructing interval estimators, it may be necessary to use procedures that cannot be guaranteed to return a single interval.

4.1 Methods of Finding Interval Estimators

Four methods of constructing interval estimators are typically used:

(I) Procedures based on Hypothesis Test Rejection Regions : Consider a test $\mathcal{T}$ at level α of simple null hypothesis

$$H_0 \;:\; \theta = \theta_0$$

for $\theta_0 \in \Theta$. Let $\mathcal{A}(\theta_0) \equiv \mathcal{R}(\theta_0)'$ denote the **acceptance region** (the complement of the rejection region) associated with the test for each value of θ_0. For $\underset{\sim}{x} \in \mathcal{X}$, define

$$\mathcal{C}(\underset{\sim}{x}) \equiv \{\theta_0 \;:\; \underset{\sim}{x} \in \mathcal{A}(\theta_0)\}$$

to be a subset of Θ; for fixed data $\underset{\sim}{x}$, $\mathcal{C}(\underset{\sim}{x})$ is the set of θ_0 values that would not lead to rejection of the null hypothesis under $\mathcal{T}$. Then by construction

$$\Pr[\underset{\sim}{X} \notin \mathcal{A}(\theta_0) \mid \theta_0] = \Pr[\underset{\sim}{X} \in \mathcal{R}(\theta_0) \mid \theta_0] \leq \alpha \qquad \Longrightarrow \qquad \Pr[\underset{\sim}{X} \in \mathcal{A}(\theta_0) \mid \theta_0] \geq 1 - \alpha$$

But, by construction, for any $\theta \in \Theta$, the coverage probability of $\mathcal{C}(\underset{\sim}{X})$ is

$$\Pr[\theta \in \mathcal{C}(\underset{\sim}{X}) \mid \theta] \equiv \Pr[\underset{\sim}{X} \in \mathcal{A}(\theta) \mid \theta] \geq 1 - \alpha$$

so therefore $\mathcal{C}(\underset{\sim}{X})$ is a $1-\alpha$ confidence set. Conversely, if $\mathcal{C}(\underset{\sim}{X})$ is a $1-\alpha$ confidence set, then we can reverse the argument to see that if

$$\mathcal{A}(\theta_0) \equiv \{\underset{\sim}{x} \,:\, \theta_0 \in \mathcal{C}(\underset{\sim}{x})\} \qquad \text{and} \qquad \mathcal{R}(\theta_0) \equiv \mathcal{A}(\theta_0)'$$

then

$$\Pr[\underset{\sim}{X} \in \mathcal{R}(\theta_0) \mid \theta_0] = \Pr[\underset{\sim}{X} \notin \mathcal{A}(\theta_0) \mid \theta_0] \equiv \Pr[\theta_0 \notin \mathcal{C}(\underset{\sim}{X}) \mid \theta_0] \leq \alpha$$

so the test of $H_0 \,:\, \theta = \theta_0$ based on rejection region $\mathcal{R}(\theta_0)$ is an α level test.

Note that a confidence set constructed by inverting the acceptance region of a test may be an interval, or the union of disjoint intervals of $\mathbb{R}$.

(II) Procedures based on Pivotal Quantities : A **pivotal quantity** is a random quantity $Q(\underset{\sim}{X}, \theta)$ whose distribution f_Q does not depend on θ. For any set $\mathcal{A}$, the probability

$$\Pr[Q(\underset{\sim}{X}, \theta) \in \mathcal{A}] = \int_{\mathcal{A}} f_Q(\underset{\sim}{x}) \, d\underset{\sim}{x}$$

does not depend on θ. An interval estimator for θ can be constructed by examining

$$\mathcal{C}(\underset{\sim}{x}) \equiv \{\theta \,:\, Q(\underset{\sim}{x}, \theta) \in \mathcal{A}\}.$$

(III) Procedures based on CDFs : Suppose that $T(\underset{\sim}{X})$ is a statistic with continuous cdf $F_{T|\theta}(t|\theta)$. Then, provided that $F_{T|\theta}$ exhibits **monotonicity** in θ, a confidence interval can be constructed directly.

- $F_{T|\theta}$ is **stochastically decreasing** in θ if for $\theta_1 < \theta_2$,

  $$F_{T|\theta}(t|\theta_1) < F_{T|\theta}(t|\theta_2) \qquad \forall t \in \mathcal{T}.$$

  For example, any location family model is stochastically decreasing in the location parameter θ; for instance, the logistic location family cdf is a decreasing function of θ for all t:

  $$F_{T|\theta}(t|\theta) = \frac{e^{(t-\theta)}}{1+e^{(t-\theta)}} = \frac{1}{1+e^{-(t-\theta)}} \qquad t \in \mathbb{R} \tag{1}$$

- $F_{T|\theta}$ is **stochastically increasing** in θ if for $\theta_1 < \theta_2$,

  $$F_{T|\theta}(t|\theta_1) > F_{T|\theta}(t|\theta_2) \qquad \forall t \in \mathcal{T}$$

  For instance, the Exponential(θ) distribution cdf is an increasing function of θ for all t:

  $$F_{T|\theta}(t|\theta) = 1 - e^{-\theta t} \qquad t \in \mathbb{R}^+ \tag{2}$$

Let $0 < \alpha_1, \alpha_2 < 1$ with $0 < \alpha_1 + \alpha_2 < 1$, and set $\alpha = \alpha_1 + \alpha_2$. In the monotonic case, define the quantities $\theta_L(t)$ and $\theta_U(t)$ by

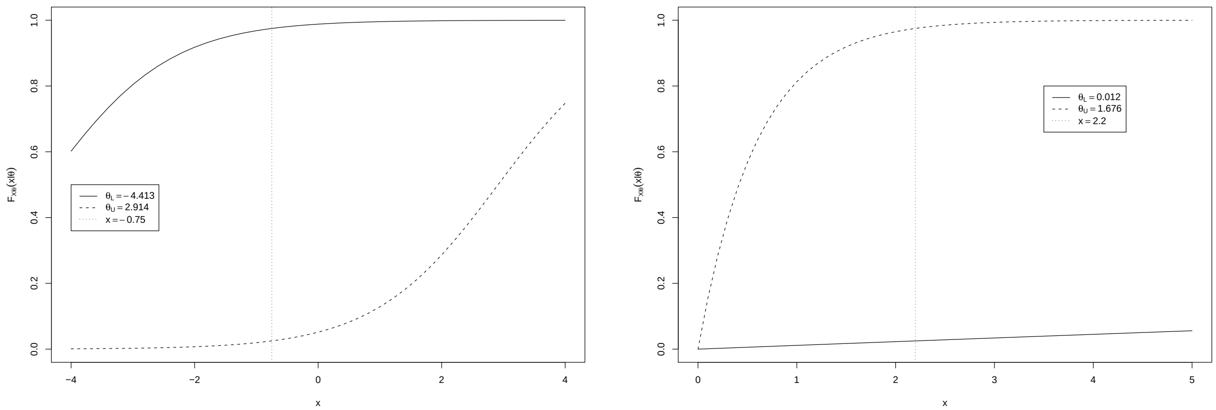
- if $F_{T|\theta}$ is **decreasing** in θ

  $$F_{T|\theta}(t|\theta_U(t)) = \alpha_1 \qquad\qquad F_{T|\theta}(t|\theta_L(t)) = 1 - \alpha_2$$

- if $F_{T|\theta}$ is **increasing** in θ

  $$F_{T|\theta}(t|\theta_U(t)) = 1 - \alpha_2 \qquad\qquad F_{T|\theta}(t|\theta_L(t)) = \alpha_1$$

Then the random interval $[\theta_L(T), \theta_U(T)]$ is a $1-\alpha$ confidence interval for θ. The plot in Figure 1 depicts the computation of the θ_U and θ_L quantities for specific value of t in the two models described above.



(a) Logistic Model: 95 % interval for parameter θ given that $t = -0.75$.

(b) Exponential Model: 95 % interval for parameter θ given that $t = 2.2$.

Figure 1: Inversion of the cdf in the monotone cases of the Logistic model (1) (decreasing in θ)) and Exponential model (2) (increasing in θ).

(IV) Bayesian Procedures : A Bayesian interval estimator can be constructed directly from the posterior distribution $\pi_{\theta|\underset{\sim}{X}}$; for data $\underset{\sim}{x}$, interval $[\theta_L(\underset{\sim}{x}), \theta_U(\underset{\sim}{x})]$ in Θ is a $1-\alpha$ **credible interval** if

$$\Pr[\theta \in [\theta_L(\underset{\sim}{x}), \theta_U(\underset{\sim}{x}) \mid \underset{\sim}{x}] = \int_{\theta_L(\underset{\sim}{x})}^{\theta_U(\underset{\sim}{x})} \pi_{\theta|\underset{\sim}{X}}(\theta|\underset{\sim}{x})\, d\theta = 1-\alpha$$

Typically, the $1-\alpha$ credible interval is not uniquely defined; for example, any interval corresponding to the α_1 and α_2 quantiles of the posterior distribution is a $1-\alpha$ credible interval if $\alpha_2 - \alpha_1 = 1-\alpha$.

4.2 Methods of Evaluating Interval Estimators

We seek methods for assessing the quality of an interval estimator that has coverage/confidence at least $1-\alpha$.

(I) **Length**

One criterion is the **length** of the interval; we choose the $1-\alpha$ interval that is as short as possible, that is, such that $(U(\underset{\sim}{X}) - L(\underset{\sim}{X}))$ is minimized in expectation. Now, if $f(x)$ is a unimodal pdf with mode $x^\star$, and

$$\int_a^b f(x)\, dx = 1-\alpha$$

with $f(a) = f(b) > 0$ so that $a \leq x^\star \leq b$, then $[a, b]$ is the shortest $1-\alpha$ probability interval for f. Furthermore, if f is symmetric $x^\star$, then it follows that the shortest interval containing probability $1-\alpha$ is the one that sets

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

yielding a **symmetric** interval about $x^\star$; for a $1-\alpha$ interval, we look for the $\alpha/2$ and $1-\alpha/2$ quantiiles of f. These results allows use to construct shortest intervals in each of the four construction methods above.

(II) **Optimality via Test Equivalence**

Suppose that $\mathcal{C}(\underset{\sim}{x})$ is a $1-\alpha$ confidence set constructed using the equivalence with a test rejection region, $\mathcal{R}_\theta$, or acceptance region $\mathcal{A}_\theta$. For true value θ, and any other value $\theta' \in \Theta$, define the probability of **false coverage**, $\varphi(\theta, \theta')$, by

$$\varphi(\theta,\theta') = \begin{cases} \Pr[\theta' \in C(\underset{\sim}{X})|\theta], \theta' \neq \theta & \text{if } C(\underset{\sim}{X}) \equiv [L(\underset{\sim}{X}), U(\underset{\sim}{X})] \\ \Pr[\theta' \in C(\underset{\sim}{X})|\theta], \theta' < \theta & \text{if } C(\underset{\sim}{X}) \equiv [L(\underset{\sim}{X}), \infty] \\ \Pr[\theta' \in C(\underset{\sim}{X})|\theta], \theta' > \theta & \text{if } C(\underset{\sim}{X}) \equiv [\infty, U(\underset{\sim}{X})] \end{cases}$$

The confidence set that minimizes $\varphi(\theta, \theta')$ across is termed the **uniformly most accurate (UMA)** confidence set, which can be shown to arise in parallel to UMP tests. Therefore, most UMA confidence sets are one-sided intervals, as this is the context in which UMP tests arise.

Theorem Suppose that $\underset{\sim}{X} \sim f_{\underset{\sim}{X}|\theta}(\underset{\sim}{x}|\theta)$, and consider a test of the hypotheses

$$\begin{aligned} H_0 &: \theta = \theta_0 \\ H_1 &: \theta > \theta_0 \end{aligned}$$

for $\theta_0 \in \Theta$. Let $\mathcal{T}^\star$ be the UMP level α test, let $\mathcal{A}^\star(\theta_0)$ be the acceptance region for $\mathcal{T}^\star$, and let $\mathcal{C}^\star(\underset{\sim}{x})$ be the corresponding $1-\alpha$ confidence set. Then for any other $1-\alpha$ confidence set $C(\underset{\sim}{X})$, and for all $\theta' < \theta$,

$$\Pr[\theta' \in \mathcal{C}^\star(\underset{\sim}{X})|\theta] \leq \Pr[\theta' \in \mathcal{C}(\underset{\sim}{X})|\theta]$$

Proof Let $\theta' < \theta$, and let $\mathcal{A}(\theta')$ be the acceptance region of the level α test of $H_0 : \theta = \theta'$ obtained by inverting $C(\underset{\sim}{x})$. Now, as $\mathcal{A}^\star(\theta')$ is the UMP acceptance region for testing

$$\begin{aligned} H_0 &: \theta = \theta' \\ H_1 &: \theta > \theta' \end{aligned}$$

we have

$$\Pr[\theta' \in \mathcal{C}^\star(\underset{\sim}{X})|\theta] \equiv \Pr \underset{\sim}{X} \in \mathcal{A}^\star(\theta')|\theta] \leq \Pr[\underset{\sim}{X} \in \mathcal{A}(\theta')|\theta] \equiv \Pr[\theta' \in \mathcal{C}(\underset{\sim}{X})|\theta]$$

as $\mathcal{T}^\star$ is UMP. ■

(III) **Bayesian Optimality**

If the posterior density $\pi_{\theta|\underset{\sim}{X}}$ is unimodal, then the previous result allows us to construct the shortest $1-\alpha$ credible interval as $\mathcal{H}(\underset{\sim}{x}) \equiv [L(\underset{\sim}{x}), U(\underset{\sim}{x})]$ defined by

$$\mathcal{H}(\underset{\sim}{x}) \equiv \{\theta : \pi_{\theta|\underset{\sim}{X}}(\theta|\underset{\sim}{x}) \geq k\}$$

so that

$$\int_{\mathcal{H}(\underset{\sim}{x})} \pi_{\theta|\underset{\sim}{X}}(\theta|\underset{\sim}{x})\, d\theta = 1 - \alpha$$

$\mathcal{H}$ is termed the **highest posterior density (HPD)** region

(IV) **Optimality via Loss Functions**

Once a suitable loss function for interval estimation is selected, the optimal interval estimator that minimizes the expected loss under the given probability model can be computed. The loss function should reflect both the **length** of the interval (shorter intervals are better) and the **accuracy** of the interval (the true value of θ should lie in the interval).