1. Interval Estimates

A point estimate $T(X)$ for a parameter $\theta \in \Theta$ on the basis of an observation $X \in \mathcal{X}$ from some known family of distributions $X \sim f_n(x \mid \theta)$ is nearly useless without some notion of its (likely) accuracy; a common approach is to offer *interval estimates*, set-valued statistics U with the properties that:

1. $\theta \in U(X)$, with "high probability;"
2. $U(X)$ is "small."

For one-dimensional parameters $\theta \in \Theta \subset \mathbb{R}$, for example, it is common to estimate θ with an *interval* of the form $U_X = [L_X, R_X]$ and require that $\mathsf{P}\Big[\theta \in [L_X, R_X]\Big] \geq 1 - \alpha$ for some small $\alpha > 0$ and that the interval length $[R_X - L_X]$ be as small as possible. Each of the three schools of statistical inference, *Likelihoodist*, *Bayesian*, and *Frequentist*, offers a different way of finding interval estimates. As usual, denote by $\hat{\theta}$ the Maximum Likelihood Estimator $\hat{\theta}_n = \hat{\theta}_n(X) = \text{argmax} f_n(x \mid \theta)$.

1.1. Likelihoodist Intervals

The Likelihoodist approach is to choose a small number $\rho \in (0, 1)$ and set

$$U(X) = \{\theta \in \Theta : \frac{f_n(x \mid \theta)}{f_n(x \mid \hat{\theta}_n)} \geq \rho\},$$

the set of points with likelihood at least $100\rho\%$ of the maximum possible value. There is no probabilistic interpretation of this set (*interval*, in the common case of one-dimensional unimodal densities), but to the Likelihoodist $U(X)$ contains all the values of θ supported by the data at least $100\rho\%$ as much as $\hat{\theta}$.

The method may be implemented in SPlus as follows. To illustrate, let's suppose we have Binomial data, with $y = 8$ successes in $n = 10$ tries and wish to estimate the success probability θ.

First fix the value of $\rho =$`rho` desired (I'll use $\rho = 0.10$ in the example) and find a lower bound A and upper bound B for the range of values of θ that might be in $U(X)$ (we'll need $U(X) \subset [A, B]$; obviously in the example $A = 0$ and $B = 1$ will work), and construct `theta <- seq(A,B,,10001)`; this divides the interval into 10,000 equal subintervals. Now estimate the Likelihoodist range by evaluating the likelihood at each point in this range, `lik`

<- dbinom(y, n, theta), and evaluating the expression range(theta[lik > rho * max(lik)]). In the present example the result is 0.4703 0.9708, indicating that points θ in the range from 0.47 to 0.97 have a a likelihood at least 10% of the maximum value, theta[lik == max(lik)]=0.8.

For large n the DeMoivre-Laplace limit theorem (special case of the Central Limit Theorem) tells us that $Y \sim \mathsf{Bi}(n, \theta)$ will have an approximately normal $\mathsf{No}(\mu, \sigma^2)$ distribution, with mean $\mu = n\theta$ and variance $\sigma^2 = n\theta(1-\theta)$, hence the maximum likelihood estimator $\hat{\theta}_n = y/n$ will have an approximate $\hat{\theta}_n \approx \mathsf{No}(\theta, \theta(1-\theta)/n)$ distribution $\approx n\hat{\theta}_n(1-\hat{\theta}_n)$ for

θ of the interval will approximately satisfy

$$\begin{aligned} \rho &= e^{-n(\theta-\hat{\theta}_n)^2/2\hat{\theta}_n(1-\hat{\theta}_n)} \\ \theta &= \hat{\theta} \pm \sqrt{-2\hat{\theta}(1-\hat{\theta}) \log \rho / n}, \end{aligned}$$

where $\hat{\theta} = y/n$ is the maximum likelihood estimate for θ.

1.2. Bayesian Credible Intervals

The Bayesian approach to set and interval estimation is to fix some small number $\alpha \in (0, 1)$ and, upon observing $X \sim f_n(x \mid \theta)$, construct a set $U(x)$ satisfying the Bayesian posterior probability bound

$$\mathsf{P}\Big[\theta \in U(x) \mid X = x\Big] \geq 1 - \alpha \tag{1}$$

In any number of dimensions the "HPD Region" is the set

$$U(x) = \{\theta \in \Theta : f_n(x \mid \theta) \geq c_\alpha(x)\},$$

where $c_\alpha(x)$ is chosen as large as possible without violating Eqn(1). In one dimension a simpler alternative is the symmetric or "equal tail" interval of the form $U(x) = [L_x, R_x]$ with L_x and R_x chosen to satisfy the symmetric requirements $\mathsf{P}[\theta < L_x \mid X = x] \leq \alpha/2$ and $\mathsf{P}[\theta > R_x \mid X = x] \leq \alpha/2$; in the binomial example above, with a Jeffreys ("arcsin" or $\mathsf{Be}(.5,.5)$) prior, this leads to L_x =qbeta(alpha/2, y+.5, n-y+.5) and R_x =qbeta(1-alpha/2, y+.5, n-y+.5), or L_x =qbeta(0.025, 8.5, 2.5) = 0.4972255 and R_x =qbeta(0.975, 8.5, 2.5) = 0.9559406 in our $y = 8$, $n = 10$ example with $\alpha = 0.05$ probability of failing to bracket θ with L_x, R_x.

Asymptotically the Beta is well approximated by the Normal with the same mean and variance, so approximately

$$\begin{aligned} L_x &\approx \texttt{th - Za * sqrt(th*(1-th)/(n+3))} \approx \hat\theta - z_{\alpha/2}\sqrt{\hat\theta(1-\hat\theta)/n} \\ R_x &\approx \texttt{th + Za * sqrt(th*(1-th)/(n+3))} \approx \hat\theta + z_{\alpha/2}\sqrt{\hat\theta(1-\hat\theta)/n}, \end{aligned}$$

where $\hat\theta = y/n$ is the maximum likelihood estimate, `th=(y+.5)/(n+1)` is the posterior mean and $z_{\alpha/2} = \texttt{Za} = \texttt{qnorm}(1 - \texttt{alpha}/2)$ is the usual normal quantile (approximately 1.96, for $\alpha = .05$). In our example this would give `0.5449 1.0005`, showing that $n = 10$ may be too small to justify a normal approximation.

1.3. Frequentist Confidence Intervals

A frequentist $100(1-\alpha)\%$ **confidence set** is a random set $U(X)$ with the property that

$$\mathsf{P}[\theta \in U(X) \mid \theta] \ge 1 - \alpha$$

for each $\theta \in \Theta$; notice that this is a probabilistic statement about the *set* $U(X)$, and not about the parameter θ. In one-dimensional problems ($\Theta \subset \mathbb{R}$), the "equal-tail" (or "symmetric") set is the interval $U(X) = [L_X, R_X]$ chosen to satisfy $\mathsf{P}[\theta < L_X \mid \theta] \le \alpha/2$ and $\mathsf{P}[R_X < \theta \mid \theta] \le \alpha/2$ for all $\theta \in \Theta$; for one-dimensional data $\mathcal{X} \subset \mathbb{R}$ with a monotone likelihood function (this includes the normal (with known variance), exponential, and Poisson means; Bernoulli and binomial probabilities; uniform $\mathsf{Un}[0, \theta]$; and many other examples), the task is to construct an increasing sequence of numbers $L_x \in \mathbb{R}$ for $x \in \mathcal{X}$ with the property that

$$\alpha/2 \ge \mathsf{P}[\theta < L_X \mid \theta] \ge \mathsf{P}[X \ge x \mid \theta]$$

for each $\theta \le L_x$; evidently the shortest allowable interval will be that with $L_x = \sup\{\theta : \mathsf{P}[X \ge x \mid \theta] \le \alpha/2\}$, and similarly $R_x = \inf\{\theta : \mathsf{P}[X \le x \mid \theta] \le \alpha/2\}$.

For the case of n Bernoulli trials x_i with success probability θ, for example, with observed sum x, we have $\mathsf{P}[X \ge x \mid \theta] = \texttt{1-pbinom(x-1,n,}\theta\texttt{)} = \texttt{pbeta(}\theta\texttt{,x,n-x+1)}$; thus $L_x = \sup\{\theta : \texttt{pbeta(}\theta\texttt{,x,n-x+1)} \le \alpha/2 = \texttt{qbeta(}\alpha/2\texttt{,x,n-x+1)}$ and similarly $R_x = 1 - L_{n-x} = \texttt{qbeta(}1-\alpha/2\texttt{,x+1,n-x)}$.

In the limit as $n \to \infty$ the normal approximation to the Beta leads to approximate confidence intervals of the form

$$\left[\frac{x}{n+1} - z_{\alpha/2} \sqrt{\frac{x(n-x+1)}{(n+1)^2\,(n+2)}}, \quad \frac{x+1}{n+1} + z_{\alpha/2} \sqrt{\frac{(x+1)(n-x)}{(n+1)^2\,(n+2)}} \right],$$

or (in a less accurate further approximation suggested by looking at the approximately normal distribution of $\hat{\theta} = x/n$),

$$\left[\hat{\theta} - z_{\alpha/2} \sqrt{\hat{\theta}(1-\hat{\theta})/n}, \quad \hat{\theta} + z_{\alpha/2} \sqrt{\hat{\theta}(1-\hat{\theta})/n} \right]$$

where $\hat{\theta} = x/n$ is the MLE for θ, identical to the asymptotic reference Bayesian interval above.