STA 114: Statistics

Notes 6. Confidence intervals

Guaranteeing minimum coverage

We saw that coverage probabilities can be used in choosing the cut-off level c of an ML interval $B_c(x) = \{\theta : \ell_x(\theta) \ge \ell_x(\hat{\theta}_{\text{MLE}}(x)) - c^2/2\}$. This argument, however, was presented only from the perspective of a single parameter value of interest (for the opinion poll example, $p_0 = 0.5$). It can be substantially extended.

Let $A(x)$ be an interval procedure for a statistical model $X \sim f(x|\theta)$, $\theta \in \Theta$ and let $\gamma(A)$ denote the minimum coverage probability of A across Θ, i.e.,

$$\gamma(A) = \min_{\theta \in \Theta} \gamma(A; \theta).$$

Because $\gamma(A;\theta_0) \ge \gamma(A)$ for any $\theta_0 \in \Theta$, using $A(x)$ guarantees at least $\gamma(A)$ probability of capturing the true value of the parameter, no matter what this true value is.

In the opinion poll example, if we could establish that $\gamma(B_{1.96}) = 0.95$, then the guarantee (one in twenty chance of a report mismatched with the truth) we gave for the true value $p_0 = 0.5$ also applies for every true value $p_0 \in [0, 1]$. That is if the college had only 25% supporters, then at most only 5% of our infinitely many researchers would fail to include it in their reported intervals. Of course this is subject to establishing $\gamma(B_{1.96}) = 0.95$, we'll see this calculations shortly, after we have dealt with the normal and the uniform model.

The quantity $\gamma(A)$ is called the confidence coefficient, or the confidence level of an interval procedure A. If $\gamma(A) = 0.95$, then A is often referred to a 95%-confidence interval (in short 95%-CI) for θ. Similarly, if $\gamma(A) = 0.9$ then A is a 90%-CI, and so on.

Confidence interval for the normal mean (with known variance)

Consider the model $X_i \overset{\text{IID}}{\sim} \mathsf{Normal}(\mu, \sigma^2)$, $\mu \in (-\infty, \infty)$, σ fixed. Then, $B_c(x) = \bar{x} \mp c\sigma/\sqrt{n}$. For any μ_0,

$$\begin{aligned}\gamma(B_c; \mu_0) &= P_{[X|\mu_0]}\left(\bar{X} - c\frac{\sigma}{\sqrt{n}} \le \mu \le \bar{X} + c\frac{\sigma}{\sqrt{n}}\right) \\ &= P_{[X|\mu_0]}\left(-c \le \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} \le c\right).\end{aligned}$$

But, if $X_i \overset{\text{IID}}{\sim} \mathsf{Normal}(\mu_0, \sigma^2)$ then $\bar{X} \sim \mathsf{Normal}(\mu_0, \sigma^2/n)$ and so the above probability is exactly $\Phi(c) - \Phi(-c)$ where $\Phi(z)$ is the cumulative distribution function (CDF) of the standard normal distribution, i.e., $\Phi(z)$ is the area under the standard normal bell curve between $-\infty$ and z. Because the standard normal bell curve is symmetric around 0, the area below the curve between $-\infty$ to $-z$ is same as the area under the curve between z and ∞. So $\Phi(-z) = 1 - \Phi(z)$. And therefore

$$\gamma(B_c; \mu_0) = 2\Phi(c) - 1$$

and we have this identity for every $\mu_0 \in (-\infty, \infty)$. So,

$$\gamma(B_c) = \min_{\mu_0 \in (-\infty,\infty)} \gamma(B_c; \mu_0) = 2\Phi(c) - 1.$$

If we want to choose a c so that B_c is a 95%-CI for μ, then we must equate $2\Phi(c) - 1 = 0.95$. From the z-tables (or from any statistics software), we find $c = 1.96$.

Similarly for any fraction $\alpha \in (0, 1)$, if we want to choose c to make B_c a $100(1-\alpha)$%-CI of μ, then we must have

$$2\Phi(c) - 1 = 1 - \alpha, \text{ and so, } c = \Phi^{-1}(1 - \alpha/2).$$

We can evaluate the last quantity in `R` by `qnorm(1 - alpha / 2)`. You can verify that a 90%-CI is given by $B_{1.64}$ and a 99%-CI is given by $B_{2.58}$.

Example (Lactic acid concentration, Contd.)**.** For our cheese data with $n = 10$, $\bar{x} = 1.379$, a 95%-CI produces the interval $[1.17, 1.59]$ for the overall concentration, a 90%-CI produces $[1.21, 1.55]$ and a 99%-CI produces $[1.11, 1.65]$.

Confidence interval for the uniform model

Now consider the model $X_i \overset{\text{IID}}{\sim} \mathsf{Uniform}(0, \theta)$, $\theta \in (0, \infty)$. We know that for any fraction $k \in [0, 1]$, the ML interval $A_k(x) = \{\theta : L_x(\theta) \geq k L_x(\hat{\theta}_{\text{MLE}}(x))\}$ equals $[x_{\max}, x_{\max}/k^{1/n}]$. At any $\theta_0 \in (0, \infty)$, the coverage of this interval procedure is

$$\begin{aligned} \gamma(A_k; \theta_0) &= P_{[X|\theta_0]}(X_{\max} \leq \theta_0 \leq X_{\max}/k^{1/n}) \\ &= P_{[X|\theta_0]}(k^{1/n}\theta_0 \leq X_{\max} \leq \theta_0) \\ &= P_{[X|\theta_0]}(X_{\max} \geq k^{1/n}\theta_0) \end{aligned}$$

because $X_{\max} \leq \theta_0$ for sure if $X_i \overset{\text{IID}}{\sim} \mathsf{Uniform}(0, \theta_0)$. But,

$$\begin{aligned} P_{[X|\theta_0]}(X_{\max} \geq k^{1/n}\theta_0) &= 1 - P_{[X|\theta_0]}(X_{\max} < k^{1/n}\theta_0) \\ &= 1 - P_{[X|\theta_0]}(\text{every } X_i \leq k^{1/n}\theta_0) \\ &= 1 - P_{[X|\theta_0]}(X_1 \leq k^{1/n}\theta_0) \times \cdots \times P_{[X|\theta_0]}(X_n \leq k^{1/n}\theta_0) \\ &= 1 - \underbrace{k^{1/n} \times \cdots \times k^{1/n}}_{n \text{ many}} \\ &= 1 - k. \end{aligned}$$

Therefore, $\gamma(A_k; \theta_0) = 1 - k$ for every $\theta_0 \in (0, \infty)$, and consequently,

$$\gamma(A_k) = \min_{\theta_0 \in (0,\infty)} \gamma(A_k; \theta_0) = 1 - k.$$

Therefore a 95%-CI for θ is $A_{0.05}$, a 90%-CI for θ is $A_{0.1}$ and a 99%-CI for θ is $A_{0.01}$.

Use of confidence coefficient

Note that unlike the risk associated with an estimator, the confidence coefficient associated with an interval procedure does not help us compare two intervals in order to find the "better" one. The interval procedure with maximum confidence is the that reports the whole parameter space for all observed data (i.e., $A(x) = \Theta$ for all x,and so $\gamma(A) = 1$). But this is a stupid procedure. (One can compare two interval procedures A and B with the same confidence coefficient by looking at which one is shorter in length for most or all $x \in S$.)

The most practical use of confidence coefficient is to associate guarantees to a set of interval procedures that arise from a single way of evaluating "strength of evidence" of the parameter values. For ML intervals, the strength of evidence is given by the likelihood (or the log-likelihood) function, but different interval procedures A_k (or B_c) arise based on the cut-off point k (or c).

This concept goes beyond the ML principle and can be applied to any strength of evidence function $U_x(\theta)$. For example, if $T(x)$ is an estimator of θ and $g(t|\theta)$ denotes the pdf/pmf of $T(X)$ when $X \sim f(x|\theta)$, then one could construct $U_x(\theta) = g(T(x)|\theta), \theta \in \Theta$. This indeed is the likelihood function of θ for the model $T \sim g(t|\theta), \theta \in \Theta$, but this need not equal the original likelihood function $L_x(\theta) = f(x|\theta), \theta \in \Theta$. But once we have $U_x(\theta)$ we can obtain interval procedures based on it in the form $A'_k(x) = \{\theta : U_x(\theta) \geq k \max_\theta U_x(\theta)\}$ for any cut-off $k \in [0, 1]$. The choice of k can be calibrated by computing the confidence coefficient $\gamma(A'_k)$.

Example (CI based on median for normal data)**.** Consider $X_1, \cdots, X_n \overset{\text{IID}}{\sim} \mathsf{Normal}(\mu, \sigma^2)$, $\mu \in (-\infty, \infty)$, σ fixed. It is known that if $X_i \overset{\text{IID}}{\sim} \mathsf{Normal}(\mu, \sigma^2)$ then X_{med} is approximately distributed as $\mathsf{Normal}(\mu, \frac{\pi\sigma^2}{2n})$ (the larger the n, the better the approximation). Therefore we can generate median based interval procedures of the form $B'_c(x) = x_{\text{med}} \mp c\sigma\sqrt{\pi/(2n)}$ based on the surrogate model $X_{\text{med}} \sim \mathsf{Normal}(\mu, \frac{\pi\sigma^2}{2n})$, $\mu \in (-\infty, \infty)$, σ fixed. And we can state $\gamma(B'_c) \approx 2\Phi(c) - 1$, i.e., $\gamma(B'_{1.96}) = 0.95$, $\gamma(B'_{1.64}) = 0.90$ etc. $\square$