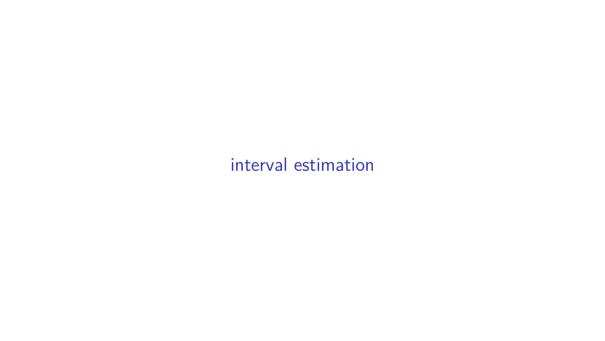
STA261 Lecture 8 — 2017-08-02

Neil Montgomery

Last edited: 2017-08-02 20:28



communicating uncertainty

Some population of interest is modeled by a random variable X, up to some parameter θ , whose value is not known. So a sample $\mathbf{X} = X_1, \dots, X_n$ i.i.d. with the same distribution of X will be gathered. We have extensively studied the problem of estimating θ .

It would be nice to communicate an amount of uncertainty along with the estimate itself, which will depend on:

- the population variance itself
- the estimator being used
- ▶ the sample size

confidence interval

A $(1 - \alpha) \cdot 100$ % confidence interval for θ is a pair of statistics $L(\mathbf{X})$ and $U(\mathbf{X})$ with these properties:

- 1. L(X) < U(X)
- 2. $P(\{L(X) > \theta)\} \cup \{U(X) < \theta)\}) = 1 \alpha$

It's not clear where such a pair of statistics might come from. But they are often easy to find.

Note: We can allow one of L or U to be a constant, including $\pm \infty$, resulting in a so-called "one-sided" confidence interval.

Note: α is arbitrary and can technically be anything between 0 and 1, but it's usually small and almost always 0.05.

pivots

A *pivot*, or *pivotal quantity*, for θ is a function $g(\mathbf{X}, \theta)$ whose distribution is the same for all θ .

Less formally, a pivot contains θ in its formula, but not in its distribution.

Example 8.0: $N(\mu, 1)$.

Examples 8.1: $N(\mu, \sigma^2)$ (with $\theta = \mu$, and then with $\theta = \sigma^2$)

Example 8.2: Gamma (α, λ) (with $\theta = \lambda$)

Example 8.3: X_{11}, \ldots, X_{1n} i.i.d $N(\mu_1, \sigma_1^2)$ and X_{21}, \ldots, X_{2m} i.i.d $N(\mu_2, \sigma_2^2)$ and $X_{1i} \perp X_{2j}$ (with $\theta = \sigma_1^2/\sigma_2^2$))

Example 8.4: Uniform $(0, \theta)$

from pivot to confidence interval

Since a pivot has θ in the formula, but not in the distribution, confidence intervals can be produces in great abundance.

Example 8.5: $N(\mu, 1)$.

Definition: z_{τ} is the solution of $P(Z > z_{\tau}) = \tau$ and $t_{\nu,\tau}$ is the solution of $P(t_{\nu} > t_{\nu,\tau}) = \tau$.

Examples 8.6: $N(\mu, \sigma^2)$ (with $\theta = \mu$, and then with $\theta = \sigma^2$)

Example 8.7: Uniform $(0, \theta)$

(observed) confidence intervals

Suppose you actually encounter some data, say from a $N(\mu, 1)$ distribution. Sample size n = 9.

The observed value of \overline{X} is what I'd call $\overline{x} = 2.87$.

To report your estimate for μ you might say "the 95% confidence interval is [2.22, 3.52]".

That's not strictly speaking a confidence interval, but we call them that anyway.

Just don't go thinking "there is a 95 chance that μ is inside [2.22, 3.52]"