

Section 03: Confidence intervals

February 16, 2023.

R function of the day

- The function `which` returns the *index* of a vector is `TRUE`
- The derivative functions `which.max` and `which.min` return the index of the maximum and minimum value, respectively
 - If there are multiple, it returns the first

```
x <- c(5, 3, 4, 1, 1, 4, 5)
```

```
which(x == 3) # returns 2  
which.min(x) # returns 4  
which.max(x) # returns 1
```

Review

Asymptotics

- **Slutsky's theorem.** If X_1, X_2, \dots and Y_1, Y_2, \dots are sequences of r.v.s such that X_n converges to X in distribution and Y_n converges to c in probability, then

1. $X_n + Y_n \xrightarrow{d} X + c$
2. $X_n Y_n \xrightarrow{d} cX$
3. $X_n / Y_n \xrightarrow{d} X / c$ for $c \neq 0$

- **Continuous mapping theorem:** If X_1, X_2, \dots is a sequence of r.v.s and g is a continuous function, then

1. If $X_n \xrightarrow{d} X$, then $g(X_n) \xrightarrow{d} g(X)$
2. If $X_n \xrightarrow{p} X$, then $g(X_n) \xrightarrow{d} g(X)$

- **Delta method:** relates to convergence in distribution. Let g be differentiable such that $g'(\mu) \neq 0$.

$$\frac{\sqrt{n}(Y_n - \mu)}{\sigma} \xrightarrow{d} \mathcal{N}(0, 1)$$
$$\frac{\sqrt{n}(g(Y_n) - g(\mu))}{|g'(\mu)|\sigma} \xrightarrow{d} \mathcal{N}(0, 1)$$

- An estimator $\hat{\theta}$ is consistent for the estimand θ^* if $\hat{\theta} \xrightarrow{p} \theta^*$ as $n \rightarrow \infty$

Fisher info

- For continuously differentiable log-likelihood, the score function is defined as

$$s(\theta; \mathbf{Y}) = \frac{\partial \ell(\theta; \mathbf{Y})}{\partial \theta}$$

- **Regularity conditions**
 - Support of \mathbf{Y} does not depend on θ (when would this happen?)
 - Expectations and derivatives exist
- If the model is correctly specified and regularity conditions hold

$$\begin{aligned} E(s(\theta^*; \mathbf{Y})) &= 0 \\ \text{Var}(s(\theta^*; \mathbf{Y})) &= -E(s'(\theta^*; \mathbf{Y})) \end{aligned}$$

- **Fisher information** is defined as

$$\mathcal{I}_{\mathbf{Y}}(\theta^*) = \text{Var}(s(\theta^*; \mathbf{Y}))$$

- Often abbreviated $\mathcal{I}(\theta)$
- If model is well-specified, Fisher information is

$$\mathcal{I}(\theta) = -E(s'(\theta^*; \mathbf{Y})) = E(s(\theta^*; \mathbf{Y})^2)$$

- Fisher info for parameter transformation
 - Let parameter τ be $\tau = g(\theta)$
 - Let g be injective and differentiable

$$\mathcal{I}(\tau) = \frac{\mathcal{I}(\theta)}{g'(\theta)^2}$$

- Under regularity, if $\hat{\theta}$ is unbiased for θ , then

$$\text{Var}(\hat{\theta}) \geq \frac{1}{n\mathcal{I}_1(\theta^*)}$$

- \mathcal{I}_1 is Fisher info for one observation
- This is the **Cramer-Rao Lower Bound** for unbiased estimators (CRLB)
 - Lower bound on MSE for unbiased estimators
- The CRLB for estimators (no assumptions about bias), we have

$$\text{Var}(\hat{\theta}) \geq \frac{g'(\theta^*)^2}{n\mathcal{I}_n(\theta^*)}$$

- $g(\theta) = E(\hat{\theta})$

MLE properties

- The MLE is consistent w/ regularity conditions and correctly specified model
- Asymptotic distribution of MLE given regularity and correctly specified model is:

$$\sqrt{n}(\hat{\theta} - \theta^*) \xrightarrow{d} \mathcal{N}\left(0, \frac{1}{\mathcal{I}(\theta^*)}\right)$$

Interval estimation

Interval estimation will use techniques such as CLT, Delta method, etc.

Let $L(\mathbf{Y}), U(\mathbf{Y})$ be functions of the data such that $L(\mathbf{y}) \leq U(\mathbf{y})$ for all \mathbf{y} . The random interval $[L(\mathbf{Y}), U(\mathbf{Y})]$ is an **interval estimator** of a parameter θ if upon observing \mathbf{y} the inference $L(\mathbf{y}) \leq \theta \leq U(\mathbf{y})$.

The **coverage probability** of an interval estimator is the probability that the true parameter lies within the interval. The coverage probability is a function of θ .

An interval estimator with a coverage probability of at least $1 - \alpha$ for all θ is called a $100(1 - \alpha)$

Standard practice is to use $\alpha = 0.05$. This is arbitrary.

We can select CI's based on criteria such as:

- Shortest expected width: Minimizing the expectation of the difference between the upper and lower bound, i.e., $E(U(\mathbf{Y}) - L(\mathbf{Y}))$.
- Equal-tailed: Require that $P(\theta < U(\mathbf{Y})) = P(\theta > L(\mathbf{Y}))$
- Centered on estimator: Require the CI to be $[\hat{\theta} - C(\mathbf{Y}), \hat{\theta} + C(\mathbf{Y})]$

Constructing confidence intervals

We can construct CIs using pivotal quantities, asymptotics, and the bootstrap (Week 10).

Exact confidence interval

A **pivotal quantity** or **pivot** is a function of the data and parameter $Q(\mathbf{Y}, \theta)$ whose distribution does not depend on the parameter. Another name for a pivot is **score**.

An example of a pivot is the **z-score**, which appears in introductory plug-and-chug statistics courses. Suppose we have $Y_i \sim \mathcal{N}(\mu, \sigma^2)$ with known variance. We find that $\bar{Y} \sim \mathcal{N}(\mu, \sigma^2/n)$, which gives us the pivot

$$\frac{\sqrt{n}}{\sigma}(\bar{Y} - \mu) \sim \mathcal{N}(0, 1).$$

The quantity on the left is the *z-score of the mean*. Let Φ be the normal CDF and let $c = \Phi^{-1}(1 - \alpha/2)$. Then, we have that

$$\begin{aligned} 1 - \alpha &= P\left(-c < \frac{\sqrt{n}}{\sigma}(\bar{Y} - \mu) < c\right) \\ &= P\left(-\frac{c\sigma}{\sqrt{n}} < \bar{Y} - \mu < \frac{c\sigma}{\sqrt{n}}\right) \\ &= P\left(\bar{Y} - \frac{c\sigma}{\sqrt{n}} < \mu < \bar{Y} + \frac{c\sigma}{\sqrt{n}}\right) \end{aligned}$$

We often use standardizations or other properties of distributions to derive a pivot that is free of the parameter.

Generally, when determining $P(A \leq Q(\mathbf{Y}, \theta) \leq B) = 1 - \alpha$, we have that the pivot is a monotone function of θ and we can rearrange terms inside the probability. We typically use the quantiles for the known distribution as A and B .

Asymptotic confidence intervals

Using the pivot is convenient when we know the distribution well, but this is often not the case. For example, for i.i.d. Poisson r.v.s, the mean is not distributed Poisson.

We can consider asymptotic distributions of some estimator $\hat{\theta}$ using techniques such as CLT, delta method, and the asymptotic distribution of the MLE (the last one is applicable in this week's homework).

Practice

1 Discussion/concept check

1. If frequentists treat θ as fixed, where is the randomness involved in interval estimation?
2. It may be intuitive to say that there is a 95% probability that θ is between two values. Why is this incorrect? What is a better interpretation?
3. Why are 95% confidence intervals commonly used in science?
4. Are confidence intervals *frequentist* or *Bayesian*?

2 *t*-score

Suppose we have observations y_1, \dots, y_n from $Y_1, \dots, Y_n \stackrel{\text{i.i.d.}}{\sim} \mathcal{N}(\mu, \sigma^2)$, with both parameters unknown.

2.1 Sufficient statistic Find a 2-dimensional sufficient statistic for (μ, σ^2) . That is, simply the likelihood function enough so that you can express them in terms of 2 quantities S, T .

2.2 Distribution of sample mean Suppose σ^2 is known. Find a pivot that does not include the sample variance and use it to write an exact 95% CI for μ .

2.3 Distribution of sample variance The sample variance is $S^2 = \frac{1}{n-1} \sum_{i=1}^n (Y_i - \bar{Y})^2$. Can you recall the distribution of $(n-1)S^2/\sigma^2$ from Stat 110?

2.4 Sample mean and variance independence Show that the sample mean is independent of the sample variance. (Hint: Use MVN and the vector $(\bar{Y}, Y_1 - \bar{Y}, \dots, Y_n - \bar{Y})$. This part is difficult, so feel free to take this on faith.)

2.5 Pivot Find a pivotal quantity based on parts d and e.

2.6 Confidence interval Use your pivot to write an exact 95% CI for μ .

3 Randomized control trial (2020 HW 4.2)

The Physicians' Health Study was a randomized experiment, led by investigators at Harvard Medical School and Brigham and Women's Hospital, that was designed to study, among other things, the effect of taking aspirin on cardiovascular disease. A paper from the New England Journal of Medicine reporting on the study is posted along with this homework.

Of the 22,071 participants, 11,037 were assigned to the treatment group and 11,034 were assigned to the control group. Since the difference in sizes between the two groups is tiny, for simplicity let $n = 11035$ and (as an extremely good approximation) assume that each group has size n . Let Y_1 and Y_0 be the numbers of participants who had myocardial infarction (better known to non-doctors as a heart attack) in the treatment group and the control group, respectively, during the time period in which participants were observed.

It is natural to model $Y_j \sim \text{Bin}(n, p_j)$, where p_1 and p_0 are the theoretical probabilities of having myocardial infarction for a person who receives the treatment and a person who receives the placebo, respectively. Of course, in reality p_j depends on age and various other factors, but for simplicity we will regard it as a constant. Let's assume that the p_j are small enough that a Poisson approximation to the Binomial is good, and use the model

$$Y_j \sim \text{Pois}(np_j).$$

In biostatistics and epidemiology, the *relative risk* is the ratio of the probability of a particular outcome (e.g., heart attack) for people exposed to a particular treatment (e.g., aspirin) to the probability of the outcome for people not exposed to the treatment. Let our estimand be the relative risk

$$\theta = \frac{p_1}{p_0}.$$

The investigators conjectured that taking aspirin is beneficial for reducing the risk of a heart attack, and hoped that the data would provide strong evidence that $\theta < 1$.

3.1 Find the likelihood function $L(p_1, p_0; \mathbf{Y})$.

3.2 Find the MLE $\hat{\theta}$ of the relative risk θ .

3.3 Reparameterization Let's reparameterize from (p_1, p_0) to (θ, p_0) since θ is our parameter of interest. Find the likelihood function $L(\theta, p_0; \mathbf{Y})$.

For the remainder of this problem, take a *conditional* approach, conditioning on the total $T = Y_0 + Y_1$. Suppose that we observe $T = t$.

3.4 Conditional distribution Find the conditional distribution of $Y_1|(T = t)$. Use this to find the conditional likelihood $L(\theta; \mathbf{Y}|t)$. Briefly explain what the advantage of conditioning on T is.

3.5 Score and Fisher Find the score function and the Fisher information for θ , again with all calculations conditional on $T = t$.

3.6 Provide a Normal approximation for $\hat{\theta}|(T = t)$.

3.7 Confidence interval Give an approximate 95% confidence interval $C(\mathbf{Y})$ for θ using the result of (f). Show that if the conditional coverage $P(\theta \in C(\mathbf{Y})|T = t)$ is approximately 0.95 for all θ and all t , then the unconditional coverage $P(\theta \in C(\mathbf{Y}))$ is also approximately 0.95 for all θ .