Section 4: Confidence Intervals

Ricky Truong (rickytruong@college.harvard.edu),
Emily Xing (exing@college.harvard.edu)

1 Introduction

1.1 Logistics

Welcome! The goal of our weekly section is to review content from lecture, with emphasis on big-picture intuition, mathematical proof, and hands-on practice.

Specifically, we aim for our sections to be interactive, well-paced, inclusive, and fun! Please do not hesitate to reach out if you have any questions/feedback!

1.2 Office Hours

- Mondays, 7:30 - 9:30 PM in Adams D-Hall (Ricky).
- Fridays, 11 AM - 12 PM in Maxwell-Dworkin 2nd Floor (Emily).
- Saturdays, 10:30 – 11:30 AM in Cabot D-Hall (Emily).

2 Big Picture

We review maximum likelihood estimation with a famous example in statistics: the *German Tank Problem*.

Building on the notion of *asymptotics* from last week, we distinguish between *asymptotic distributions*, *exact distributions*, and *approximate distributions*.

So far, we've used *point estimators* (e.g., we estimate θ with 5), but we can arguably improve upon this with *interval estimators* (e.g., we estimate θ with $[4, 6]$).

We use the *known* distributions of *pivots* to construct *confidence intervals*. Depending on the type of distribution, we can build *exact* confidence intervals or *approximate* (i.e., nominal) confidence intervals. The approximate confidence intervals often result from the Central Limit Theorem, which requires a sufficiently large n but accommodates for non-Normal data (conversely, the exact confidence intervals often require the generous assumption of Normality).

3 Maximum Likelihood Estimation

Idea. As we've seen, the maximum likelihood estimator is incredibly powerful and enjoys many nice properties. Still, it is not always ideal. As illustrated in the *German Tank Problem*, it can sometimes result in a naive estimate. We conclude with a review of the multiple ways to find MLE (since calculus should be saved as a last resort).

3.1 Properties of MLE

Definition 1 (Maximum likelihood estimator). $\hat{\theta}_{\text{MLE}} = \arg\max_{\theta \in \Theta} \mathcal{L}(\theta; \vec{Y})$. Under regularity conditions, the MLE has the following properties:

• MLE is invariant, meaning if $\hat{\theta}_{\text{MLE}}$ is the MLE of θ, then $g(\hat{\theta}_{\text{MLE}})$ is the MLE of $g(\theta)$.

• MLE is consistent, meaning $\hat{\theta}_{\text{MLE}} \xrightarrow{p} \theta$.

• MLE is asymptotically unbiased, meaning $\lim_{n\to\infty} \text{Bias}[\hat{\theta}_{\text{MLE}}] = 0$.

• MLE is asymptotically efficient, meaning no other asymptotically unbiased estimator has a lower asymptotic variance.

• MLE is asymptotically Normal, meaning for i.i.d. $Y_1, \ldots, Y_n$, $\sqrt{n}(\hat{\theta}_{\text{MLE}} - \theta^*) \xrightarrow{d} \mathcal{N}\left(0, \frac{1}{\mathcal{I}_{Y_1}(\theta^*)}\right)$. This implies $\hat{\theta}_{\text{MLE}} \dot{\sim} \mathcal{N}\left(\theta^*, \frac{1}{n\mathcal{I}_{Y_1}(\theta^*)}\right)$ for large n.

• ☣: Notice the Fisher information in the Normal distribution is from Y_1, not $\vec{Y}$. For i.i.d. data, $\mathcal{I}_{\vec{Y}}(\theta^*) = n\mathcal{I}_{Y_1}(\theta^*)$. Keep the notation clear and consistent!

Concept Checker 1. Let $Y_1, \ldots, Y_n \overset{\text{i.i.d.}}{\sim} F_{Y;\mu}$ for some model $F_{Y;\mu}$ parameterized by μ, with $\mathbb{E}[Y_i] = \mu$ and $\text{Var}[Y_i] = \sigma^2 < \infty$. Assume regularity conditions hold. Suppose $\hat{\mu}_{\text{MLE}} = \bar{Y}$. How can we find $\mathcal{I}_{\vec{Y}}(\mu)$, the Fisher information for μ from $\vec{Y}$?

Solution

First, we know $\sqrt{n}(\bar{Y} - \mu) \xrightarrow{d} \mathcal{N}(0, \sigma^2)$ by CLT. Next, we know $\sqrt{n}(\hat{\mu}_{\text{MLE}} - \mu) \xrightarrow{d} \mathcal{N}\left(0, \frac{1}{\mathcal{I}_{Y_1}(\mu)}\right)$ as a property of the MLE since regularity conditions hold. Since $\hat{\mu}_{\text{MLE}} = \bar{Y}$, they have the same asymptotic distribution, so their limiting variances must match: $\mathcal{I}_{Y_1}(\mu) = \frac{1}{\sigma^2}$. For i.i.d. data, Fisher information is additive—i.e., $\mathcal{I}_{\vec{Y}}(\theta) = n\mathcal{I}_{Y_1}(\theta)$—so $\mathcal{I}_{\vec{Y}}(\mu) = \frac{n}{\sigma^2}$.

3.2 German Tank Problem

Definition 2 (Setup of the German Tank Problem)**.** Suppose there are θ tanks in existence. From our n captured tanks, we observe $y_1, \ldots, y_n$, corresponding to serial numbers in $\{1, \ldots, \theta\}$. Assume the captured tanks are a simple random sample (i.e., all subsets of size n are without replacement and equally likely). How can we estimate θ?

Definition 3 (Method of moments approach)**.** We begin by defining the first moment.

$$
\begin{aligned}
\mathbb{E}[Y_i] &= \frac{1}{\theta}\sum_{i=1}^{\theta} i && \text{by definition} \\
&= \frac{1}{\theta}\left(\frac{\theta(\theta+1)}{2}\right) && \text{by arithmetic series} \\
&= \frac{\theta+1}{2} && \text{by simplifying}
\end{aligned}
$$

This implies $\theta = 2\mathbb{E}[Y_i] - 1$, so by definition of MOM, our method of moments estimate is $\hat{\theta}_{\text{MOM}} = 2\bar{y} - 1$.

Definition 4 (Maximum likelihood approach)**.** We begin by writing the likelihood function.

$$\begin{aligned}
\mathcal{L}(\theta;\vec{y}) &= f_{\vec{Y}}(\vec{y};\theta) && \text{by definition}\\
&= P(Y_1 = y_1, \ldots, Y_n = y_n; \theta) && \text{for discrete random vector}\\
&= \begin{cases} \frac{1}{\binom{\theta}{n}}, & y_i \le \theta \,\forall\, i \in \{1,\ldots,n\} \\ 0, & \text{else} \end{cases} && \text{by simple random sample}\\
&= \frac{1}{\binom{\theta}{n}} \mathbb{1}\{y_{(n)} \le \theta\} && \text{by rewriting via order statistic}
\end{aligned}$$

For some intuition, this is saying the probability we observe our specific sample is just 1 out of the θ-choose-n possible combinations, with an indicator to ensure θ doesn't conflict with our data. This implies $\mathcal{L}(\theta;\vec{y})$ is 0 until $\theta \ge y_{(n)}$, after which point it is strictly decreasing on $\Theta = \mathbb{Z}^+$, so by definition of MLE, our maximum likelihood estimate is $\hat{\theta}_{\text{MLE}} = y_{(n)}$.

• Recall $\mathcal{L}(\theta;\vec{y})$ is a function of θ. E.g., it doesn't make sense for θ to be 1 if we capture $n = 3$ tanks with $y_{(3)} = 7$, so $\mathcal{L}(\theta = 1;\vec{y}) = \frac{1}{\binom{1}{3}}\mathbb{1}\{7 \le 1\} = 0$. But $\mathcal{L}(\theta = 7;\vec{y}) = \frac{1}{\binom{7}{3}}\mathbb{1}\{7 \le 7\} = \frac{1}{35}$ is well-defined. Notice the denominator increases (and thus the likelihood decreases) as we increase θ—e.g., $\mathcal{L}(\theta = 8;\vec{y}) = \frac{1}{\binom{8}{3}}\mathbb{1}\{7 \le 8\} = \frac{1}{56}$.

• Intuitively, this estimator seems naive. In the previous example, just because we observe a sample maximum of 7 doesn't mean the population maximum should be 7.

• Building on the MLE, we can define a new estimator: $\hat{\theta}_{\text{UB}} = (\frac{n+1}{n})\hat{\theta}_{\text{MLE}} - 1$, which we can show is unbiased.

3.3 How to Find MLE

1. Invariance: If $\theta = f(\lambda)$, then $\hat{\theta}_{\text{MLE}} = f(\hat{\lambda}_{\text{MLE}})$ by invariance.

• Hints: We know $\hat{\lambda}_{\text{MLE}}$.

2. German Tank Problem: If $\mathcal{L}(\theta;\vec{Y})$ strictly decreases, its first possible input is $\hat{\theta}_{\text{MLE}}$.

• Hints: Write out $\mathcal{L}(\theta;\vec{Y})$. There is an indicator in the likelihood, so it is not differentiable at the jump. θ must be discrete. θ is involved in the support.

3. NEF: $\hat{\mu}_{\text{MLE}} = \bar{Y}$, where $\mu = \mathbb{E}[Y]$ and $f_Y(y) = e^{\theta y - \psi(\theta)}h(y)$ (i.e., Y is a natural exponential family).[1]

4. Calculus: $\hat{\theta}_{\text{MLE}} = \arg\max_{\theta\in\Theta} \ell(\theta;\vec{Y})$ (i.e., set $\ell'(\theta;\vec{Y}) = 0$ and check $\ell''(\theta;\vec{Y}) < 0$).

Concept Checker 2. Which strategy would be most efficient for finding $\hat{\theta}_{\text{MLE}}$ for the following scenarios:

1. For $Y_1, \ldots, Y_n \overset{\text{i.i.d.}}{\sim} \text{Bern}(p)$, $\theta = P(Y_i = 0)$.
2. For i.i.d. data $Y_1, \ldots, Y_n$, $\mathcal{L}(\theta;\vec{Y}) = \left(\frac{5}{\theta^5}\right)^n \left(\prod_{i=1}^n e^{Y_i}\right)^5 \mathbb{1}\{Y_{(n)} \le \theta\}$, where $\theta \ge 0$.
3. For i.i.d. data $Y_1, \ldots, Y_n$, $f_{Y_i}(y) = \theta y^{\theta-1}$, where $0 < y < 1$ and $\theta > 0$.

[1] We will discuss this further later in the course.

Solution

1. We can apply invariance: $\hat{\theta}_{\text{MLE}} = 1 - \hat{P}(Y_i = 1)_{\text{MLE}} = 1 - \hat{p}_{\text{MLE}} = 1 - \bar{Y}$.
2. We can apply the German Tank Problem: $\hat{\theta}_{\text{MLE}} = Y_{(n)}$.
3. The other tricks don't apply, so we resort to calculus: $\hat{\theta}_{\text{MLE}} = -\frac{n}{\sum_{i=1}^{n} \log(Y_i)}$. A detailed solution is provided in "Section 2: Estimators."

4 Types of Distributions

Idea. The foundation for confidence intervals lies in *distributions*. We review the three main types of distributions and how they relate.

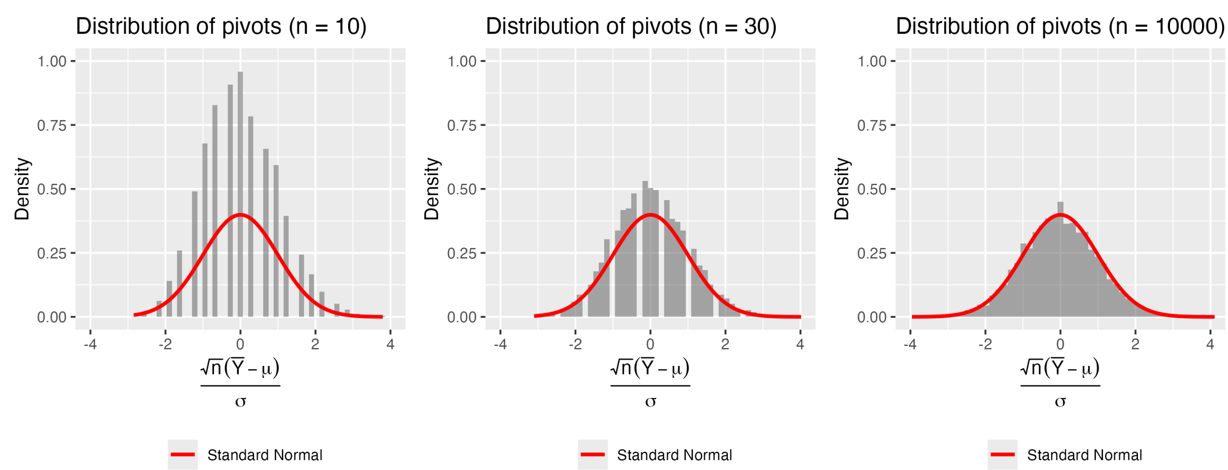

FIGURE 1: $\frac{\sqrt{n}(\bar{Y}-\mu)}{\sigma}$ (in gray) converges in distribution to $\mathcal{N}(0,1)$ (in red). If we "freeze" at $n = 30$, the approximate distribution is close to the asymptotic distribution.

Definition 5 (Exact distribution)**.** We can perfectly describe the behavior of the distribution (no approximation, no limit, no dependence on n).

- E.g., $Z \sim \mathcal{N}(0,1)$ is an exact distribution, so $Q(0.975) = 1.96$ (i.e., exactly 97.5% of the distribution falls to the left of 1.96).

Definition 6 (Asymptotic distribution)**.** The finite-sample distribution changes for each n, and the asymptotic distribution is the limit of these distributions as $n \to \infty$.

- E.g., $\frac{\sqrt{n}(\bar{Y}-\mu)}{\sigma} \xrightarrow{d} \mathcal{N}(0,1)$ is an asymptotic distribution, so $Q(0.975) \to 1.96$ as $n \to \infty$.

Definition 7 (Approximate distribution)**.** We "freeze" the finite-sample distribution at a sufficiently large n such that the distribution is close to (but not perfectly) the asymptotic distribution.

- $\frac{\sqrt{n}(\bar{Y}-\mu)}{\sigma} \dot{\sim} \mathcal{N}(0,1)$ is an approximate distribution for large n, so $Q(0.975) \approx 1.96$.

Example 1 (Approximations of pi)**.** As an analogy, let $\hat{\pi}_n$ be the nth digit of π. E.g., $\hat{\pi}_1 = 3$, $\hat{\pi}_2 = 3.1$, $\hat{\pi}_3 = 3.14$, and so on.

- Think of π as the asymptotic goal—$\hat{\pi}_n$ "converges to" π as $n \to \infty$.
- We can "freeze" at a sufficiently large n such that $\hat{\pi}_n$ is close to (but not perfectly) π.

5 Confidence Intervals

Idea. "But really confidence intervals are more like ring toss—the true value is fixed, and the interval might end up around it" - Ellie Murray.

We often estimate θ with $\hat{\theta}$, but why stop there? A *confidence interval* is a type of *interval estimator*. To construct one, we begin with the *known* distribution of a *pivot* that involves our estimand of interest. From there, we use algebra and quantiles to isolate the estimand in a probabilistic statement.

Confidence intervals have the benefit of capturing an additional "dimension" to the data: the *variability/uncertainty*. E.g., for $Y_1, \ldots, Y_n \overset{\text{i.i.d.}}{\sim} \text{Bern}(p)$, suppose we observe one dataset with only $n = 4$ values and another dataset with $n = 1000$ values. The first 95% CI may look like $[0.1, 0.9]$ while the second one may look like $[0.49, 0.51]$. Though both are centered at $\hat{p} = 0.5$, the latter is much more informative than the former, which makes sense with the uncertainty from a limited sample size.

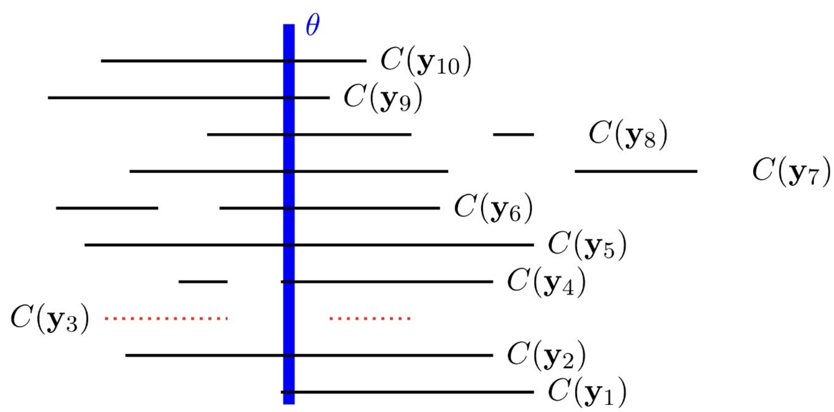

FIGURE 2: An illustration of confidence. For a 90% confidence interval and 100 randomly-generated datasets, we expect it to succeed (i.e., capture θ) 9 times out of 10.[2]

5.1 Fundamentals

Definition 8 (Interval estimator)**.** $C(\vec{Y}) = [L(\vec{Y}), U(\vec{Y})]$ is an interval estimator based on the random data $\vec{Y}$ for an estimand θ.

- The corresponding (fixed) interval estimate is $C(\vec{y}) = [L(\vec{y}), U(\vec{y})]$.

Definition 9 (Coverage probability)**.** $P(\theta \in C(\vec{Y}))$ (i.e., the probability θ is captured in the interval estimator).

- Though θ is fixed while $\vec{Y}$ is random in the Frequentist paradigm, we view coverage probability as a function of possible values for θ (like we do with likelihood).

[2] This figure is from *Introduction to Statistics: Inference, Description, Prediction, and Causality* by Joseph K. Blitzstein and Neil Shephard.

Definition 10 (Confidence interval)**.** $C(\vec{Y})$ is a $100(1-\alpha)\%$ confidence interval (CI) for θ if it has coverage probability of at least $(1-\alpha)$ for all possible values of θ (i.e., $P(\theta \in C(\vec{Y})) \geq 1-\alpha \; \forall \; \theta \in \Theta$).

• ☣: Again, θ is fixed while $\vec{Y}$ is random in the Frequentist paradigm, so it doesn't make sense to say "the probability of θ being between 0.5 and 0.7 is a 0.95" since θ is fixed and doesn't have a distribution!

5.2 Constructing Exact Confidence Intervals

Definition 11 (Statistic)**.** A random variable that is a function of the data $\vec{Y}$. Usually denoted as $T(\vec{Y})$, where the function T must not involve any unknown parameters.

• I.e., the statistic/random variable itself cannot involve unknown parameters, but its distribution may.

• E.g., $\bar{Y} \sim \mathcal{N}(\mu, \frac{\sigma^2}{n})$.

Definition 12 (Pivot)**.** A random variable whose exact distribution is known.

• I.e., the pivot/random variable itself may involve unknown parameters, but its distribution cannot.

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

• **Strategy**: Find a pivot (exact or asymptotic) that involves the estimand of interest, write a probabilistic statement using quantiles, and isolate the estimand.

Concept Checker 3. Which of the following random variables are pivots? Assume n is known.

1. $\frac{(n-1)\hat{\sigma}^2}{\sigma^2} \sim \chi^2_{n-1}$
2. $n\bar{Y} \sim \text{Gamma}(n, \lambda)$
3. $n\lambda\bar{Y} \sim \text{Gamma}(n, 1)$

Solution

Notice all random variables except $n\bar{Y} \sim \text{Gamma}(n, \lambda)$ have known distributions that we can describe (e.g., for $n = 2$, $Q_{\chi^2_1}(0.95) = 3.841$). Thus, all random variables except $n\bar{Y} \sim \text{Gamma}(n, \lambda)$ are pivots.

Concept Checker 4. For any random variable X where F_X is continuous and strictly increasing (i.e., F_X^{-1} exists), what does $P(Q_X(\frac{\alpha}{2}) \leq X \leq Q_X(1-\frac{\alpha}{2}))$ equal?

Solution

$$
\begin{aligned}
P(Q_X(\frac{\alpha}{2}) \leq X \leq Q_X(1-\frac{\alpha}{2})) &= P(X \leq Q_X(1-\frac{\alpha}{2})) - P(X \leq Q_X(\frac{\alpha}{2})) && \text{by splitting} \\
&= F_X(Q_X(1-\frac{\alpha}{2})) - F_X(Q_X(\frac{\alpha}{2})) && \text{by CDF} \\
&= F_X(F_X^{-1}(1-\frac{\alpha}{2})) - F_X(F_X^{-1}(\frac{\alpha}{2})) && \text{by quantiles} \\
&= 1 - \frac{\alpha}{2} - \frac{\alpha}{2} && \text{by algebra} \\
&= 1 - \alpha && \text{by simplifying}
\end{aligned}
$$

Example 2 (Exact pivot, Normal data, one parameter unknown)**.** For $Y_1, \ldots, Y_n \overset{\text{i.i.d.}}{\sim} \mathcal{N}(\mu, \sigma^2)$ with μ unknown and σ^2 known, $\bar{Y} \sim \mathcal{N}(\mu, \frac{\sigma^2}{n})$ by property of the Normal. This is an exact distribution. By standardizing, we have an exact pivot: $\frac{\bar{Y}-\mu}{\sigma/\sqrt{n}} = Z \sim \mathcal{N}(0,1)$. Suppose we want a $100(1-\alpha)\%$ confidence interval for μ.

$$
\begin{aligned}
1-\alpha &= P\left(Q_Z(\frac{\alpha}{2}) \leq \frac{\bar{Y}-\mu}{\sigma/\sqrt{n}} \leq Q_Z(1-\frac{\alpha}{2})\right) && \text{by quantiles} \\
&= P\left((\frac{\sigma}{\sqrt{n}})Q_Z(\frac{\alpha}{2}) \leq \bar{Y}-\mu \leq (\frac{\sigma}{\sqrt{n}})Q_Z(1-\frac{\alpha}{2})\right) && \text{by algebra} \\
&= P\left((\frac{\sigma}{\sqrt{n}})Q_Z(\frac{\alpha}{2}) - \bar{Y} \leq -\mu \leq (\frac{\sigma}{\sqrt{n}})Q_Z(1-\frac{\alpha}{2}) - \bar{Y}\right) && \text{by algebra} \\
&= P\left(-((\frac{\sigma}{\sqrt{n}})Q_Z(\frac{\alpha}{2}) - \bar{Y}) \geq \mu \geq -((\frac{\sigma}{\sqrt{n}})Q_Z(1-\frac{\alpha}{2}) - \bar{Y})\right) && \text{by algebra} \\
&= P\left(\bar{Y} - (\frac{\sigma}{\sqrt{n}})Q_Z(1-\frac{\alpha}{2}) \leq \mu \leq \bar{Y} - (\frac{\sigma}{\sqrt{n}})Q_Z(\frac{\alpha}{2})\right) && \text{by simplifying}
\end{aligned}
$$

Thus, our $100(1-\alpha)\%$ confidence interval for μ is $\left[\bar{Y} - (\frac{\sigma}{\sqrt{n}})Q_Z(1-\frac{\alpha}{2}), \bar{Y} - (\frac{\sigma}{\sqrt{n}})Q_Z(\frac{\alpha}{2})\right]$. By the symmetry of the Normal (i.e., $Q_Z(\frac{\alpha}{2}) = -Q_Z(1-\frac{\alpha}{2})$), we can rewrite this as $\bar{Y} \pm (\frac{\sigma}{\sqrt{n}})Q_Z(1-\frac{\alpha}{2})$.

Example 3 (Exact pivot, Normal data, both parameters unknown)**.** For $Y_1, \ldots, Y_n \overset{\text{i.i.d.}}{\sim} \mathcal{N}(\mu, \sigma^2)$ with both μ and σ^2 unknown, define $\hat{\sigma} = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(Y_i - \bar{Y})^2}$.[3] By property of the t-distribution, we have an exact pivot: $\frac{\bar{Y}-\mu}{\hat{\sigma}/\sqrt{n}} \sim t_{n-1}$. Suppose we want a $100(1-\alpha)\%$ confidence interval for μ.

[3]Notice other estimators for σ exist. E.g., $\hat{\sigma} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(Y_i - \bar{Y})^2}$. The $n-1$ is necessary to construct our pivots.

$$\begin{aligned}
1-\alpha &= P\left(Q_{t_{n-1}}(\frac{\alpha}{2}) \leq \frac{\bar{Y}-\mu}{\hat{\sigma}/\sqrt{n}} \leq Q_{t_{n-1}}(1-\frac{\alpha}{2})\right) && \text{by quantiles}\\
&= P\left((\frac{\hat{\sigma}}{\sqrt{n}})Q_{t_{n-1}}(\frac{\alpha}{2}) \leq \bar{Y}-\mu \leq (\frac{\hat{\sigma}}{\sqrt{n}})Q_{t_{n-1}}(1-\frac{\alpha}{2})\right) && \text{by algebra}\\
&= P\left((\frac{\hat{\sigma}}{\sqrt{n}})Q_{t_{n-1}}(\frac{\alpha}{2}) - \bar{Y} \leq -\mu \leq (\frac{\hat{\sigma}}{\sqrt{n}})Q_{t_{n-1}}(1-\frac{\alpha}{2}) - \bar{Y}\right) && \text{by algebra}\\
&= P\left(-((\frac{\hat{\sigma}}{\sqrt{n}})Q_{t_{n-1}}(\frac{\alpha}{2}) - \bar{Y}) \geq \mu \geq -((\frac{\hat{\sigma}}{\sqrt{n}})Q_{t_{n-1}}(1-\frac{\alpha}{2}) - \bar{Y})\right) && \text{by algebra}\\
&= P\left(\bar{Y} - (\frac{\hat{\sigma}}{\sqrt{n}})Q_{t_{n-1}}(1-\frac{\alpha}{2}) \leq \mu \leq \bar{Y} - (\frac{\hat{\sigma}}{\sqrt{n}})Q_{t_{n-1}}(\frac{\alpha}{2})\right) && \text{by simplifying}
\end{aligned}$$

Thus, our $100(1-\alpha)\%$ confidence interval for μ is $\left[\bar{Y} - (\frac{\hat{\sigma}}{\sqrt{n}})Q_{t_{n-1}}(1-\frac{\alpha}{2}), \bar{Y} - (\frac{\hat{\sigma}}{\sqrt{n}})Q_{t_{n-1}}(\frac{\alpha}{2})\right]$. By the symmetry of the t-distribution (i.e., $Q_{t_{n-1}}(\frac{\alpha}{2}) = -Q_{t_{n-1}}(1-\frac{\alpha}{2})$), we can rewrite this as $\bar{Y} \pm (\frac{\hat{\sigma}}{\sqrt{n}})Q_{t_{n-1}}(1-\frac{\alpha}{2})$.

Now, with the same setup, define $\hat{\sigma}^2 = \frac{1}{n-1}\sum_{i=1}^{n}(Y_i - \bar{Y})^2$. By Theorem 10.4.3 in the STAT 110 textbook, we have an exact pivot: $\frac{(n-1)\hat{\sigma}^2}{\sigma^2} \sim \chi^2_{n-1}$. Suppose we want a $100(1-\alpha)\%$ confidence interval for σ^2.

$$\begin{aligned}
1-\alpha &= P\left(Q_{\chi^2_{n-1}}(\frac{\alpha}{2}) \leq \frac{(n-1)\hat{\sigma}^2}{\sigma^2} \leq Q_{\chi^2_{n-1}}(1-\frac{\alpha}{2})\right) && \text{by quantiles}\\
&= P\left(\frac{1}{Q_{\chi^2_{n-1}}(\frac{\alpha}{2})} \geq \frac{\sigma^2}{(n-1)\hat{\sigma}^2} \geq \frac{1}{Q_{\chi^2_{n-1}}(1-\frac{\alpha}{2})}\right) && \text{by algebra}\\
&= P\left(\frac{(n-1)\hat{\sigma}^2}{Q_{\chi^2_{n-1}}(\frac{\alpha}{2})} \geq \sigma^2 \geq \frac{(n-1)\hat{\sigma}^2}{Q_{\chi^2_{n-1}}(1-\frac{\alpha}{2})}\right) && \text{by algebra}\\
&= P\left(\frac{(n-1)\hat{\sigma}^2}{Q_{\chi^2_{n-1}}(1-\frac{\alpha}{2})} \leq \sigma^2 \leq \frac{(n-1)\hat{\sigma}^2}{Q_{\chi^2_{n-1}}(\frac{\alpha}{2})}\right) && \text{by simplifying}
\end{aligned}$$

Thus, our $100(1-\alpha)\%$ confidence interval for σ^2 is $\left[\frac{(n-1)\hat{\sigma}^2}{Q_{\chi^2_{n-1}}(1-\frac{\alpha}{2})}, \frac{(n-1)\hat{\sigma}^2}{Q_{\chi^2_{n-1}}(\frac{\alpha}{2})}\right]$. Notice χ^2_{n-1} is not symmetric.

5.3 Constructing Approximate Confidence Intervals

Definition 13 (Asymptotic pivot)**.** A random variable whose asymptotic distribution is known.

- I.e., the pivot/random variable itself may involve unknown parameters, but its distribution cannot.
- E.g., $\frac{\bar{Y}-\mu}{\sigma/\sqrt{n}} \xrightarrow{d} \mathcal{N}(0,1)$.

Example 4 (Asymptotic pivot, non-Normal data, one parameter unknown)**.** Often, our data are not Normal, so we rely on the Central Limit Theorem. For i.i.d. $Y_1, \ldots, Y_n$ with $\mathbb{E}[Y_i] = \mu$ unknown and $\text{Var}[Y_i] = \sigma^2$ known, $\sqrt{n}(\bar{Y} - \mu) \xrightarrow{d} \mathcal{N}(0, \sigma^2)$ by the Central Limit Theorem. This is an asymptotic distribution. By standardizing, we have an asymptotic pivot: $\frac{\bar{Y}-\mu}{\sigma/\sqrt{n}} \xrightarrow{d} \mathcal{N}(0,1)$. Suppose we want a $100(1-\alpha)\%$ confidence interval for μ. We can "freeze" at a sufficiently large n such that $\frac{\bar{Y}-\mu}{\sigma/\sqrt{n}} \dot{\sim} \mathcal{N}(0,1)$.

$$\begin{aligned}
1-\alpha &\approx P\left(Q_Z(\frac{\alpha}{2}) \le \frac{\bar{Y}-\mu}{\sigma/\sqrt{n}} \le Q_Z(1-\frac{\alpha}{2})\right) && \text{by quantiles} \\
&= P\left((\frac{\sigma}{\sqrt{n}})Q_Z(\frac{\alpha}{2}) \le \bar{Y}-\mu \le (\frac{\sigma}{\sqrt{n}})Q_Z(1-\frac{\alpha}{2})\right) && \text{by algebra} \\
&= P\left((\frac{\sigma}{\sqrt{n}})Q_Z(\frac{\alpha}{2}) - \bar{Y} \le -\mu \le (\frac{\sigma}{\sqrt{n}})Q_Z(1-\frac{\alpha}{2}) - \bar{Y}\right) && \text{by algebra} \\
&= P\left(-((\frac{\sigma}{\sqrt{n}})Q_Z(\frac{\alpha}{2}) - \bar{Y}) \ge \mu \ge -((\frac{\sigma}{\sqrt{n}})Q_Z(1-\frac{\alpha}{2}) - \bar{Y})\right) && \text{by algebra} \\
&= P\left(\bar{Y} - (\frac{\sigma}{\sqrt{n}})Q_Z(1-\frac{\alpha}{2}) \le \mu \le \bar{Y} - (\frac{\sigma}{\sqrt{n}})Q_Z(\frac{\alpha}{2})\right) && \text{by simplifying}
\end{aligned}$$

Thus, our approximate $100(1-\alpha)\%$ confidence interval for μ is $\left[\bar{Y} - (\frac{\sigma}{\sqrt{n}})Q_Z(1-\frac{\alpha}{2}), \bar{Y} - (\frac{\sigma}{\sqrt{n}})Q_Z(\frac{\alpha}{2})\right]$. Again, we can rewrite this as $\bar{Y} \pm (\frac{\sigma}{\sqrt{n}})Q_Z(1-\frac{\alpha}{2})$.

Example 5 (Asymptotic pivot, non-Normal data, both parameters unknown)**.** The previous example assumed σ^2 is known, which isn't always practical. Often, σ^2 is unknown, but we can estimate it. If our estimator is consistent (i.e., $\hat{\sigma}^2 \xrightarrow{p} \sigma^2$), we remarkably arrive at a very similar approximate confidence interval as before!

For i.i.d. $Y_1, \ldots, Y_n$ with both $\mathbb{E}[Y_i] = \mu$ and $\text{Var}[Y_i] = \sigma^2$ unknown, $\sqrt{n}(\bar{Y} - \mu) \xrightarrow{d} \mathcal{N}(0, \sigma^2)$ by the Central Limit Theorem. This is an asymptotic distribution. By standardizing, we have an asymptotic pivot: $\frac{\bar{Y}-\mu}{\sigma/\sqrt{n}} \xrightarrow{d} \mathcal{N}(0,1)$.

$$\begin{aligned}
\hat{\sigma}^2 &\xrightarrow{p} \sigma^2 && \text{by consistency} \\
\frac{\sigma}{\hat{\sigma}} &\xrightarrow{p} \frac{\sigma}{\sigma} && \text{by CMT} \\
\frac{\sigma}{\hat{\sigma}} &\xrightarrow{p} 1 && \text{by simplifying} \\
\left(\frac{\bar{Y}-\mu}{\sigma/\sqrt{n}}\right)\left(\frac{\sigma}{\hat{\sigma}}\right) &\xrightarrow{d} (Z)(1) && \text{by Slutsky's} \\
\frac{\bar{Y}-\mu}{\hat{\sigma}/\sqrt{n}} &\xrightarrow{d} (Z)(1) && \text{by simplifying} \\
\frac{\bar{Y}-\mu}{\hat{\sigma}/\sqrt{n}} &\xrightarrow{d} \mathcal{N}(0,1) && \text{by property of Normal}
\end{aligned}$$

This is our new asymptotic pivot: $\frac{\bar{Y}-\mu}{\hat{\sigma}/\sqrt{n}} \xrightarrow{d} \mathcal{N}(0,1)$. Suppose we want a $100(1-\alpha)\%$ confidence interval for μ. We can "freeze" at a sufficiently large n such that $\frac{\bar{Y}-\mu}{\hat{\sigma}/\sqrt{n}} \dot{\sim} \mathcal{N}(0,1)$.

$$\begin{aligned}
1-\alpha &\approx P\left(Q_Z(\frac{\alpha}{2}) \le \frac{\bar{Y}-\mu}{\hat{\sigma}/\sqrt{n}} \le Q_Z(1-\frac{\alpha}{2})\right) && \text{by quantiles} \\
&= P\left((\frac{\hat{\sigma}}{\sqrt{n}})Q_Z(\frac{\alpha}{2}) \le \bar{Y}-\mu \le (\frac{\hat{\sigma}}{\sqrt{n}})Q_Z(1-\frac{\alpha}{2})\right) && \text{by algebra} \\
&= P\left((\frac{\hat{\sigma}}{\sqrt{n}})Q_Z(\frac{\alpha}{2}) - \bar{Y} \le -\mu \le (\frac{\hat{\sigma}}{\sqrt{n}})Q_Z(1-\frac{\alpha}{2}) - \bar{Y}\right) && \text{by algebra} \\
&= P\left(-((\frac{\hat{\sigma}}{\sqrt{n}})Q_Z(\frac{\alpha}{2}) - \bar{Y}) \ge \mu \ge -((\frac{\hat{\sigma}}{\sqrt{n}})Q_Z(1-\frac{\alpha}{2}) - \bar{Y})\right) && \text{by algebra} \\
&= P\left(\bar{Y} - (\frac{\hat{\sigma}}{\sqrt{n}})Q_Z(1-\frac{\alpha}{2}) \le \mu \le \bar{Y} - (\frac{\hat{\sigma}}{\sqrt{n}})Q_Z(\frac{\alpha}{2})\right) && \text{by simplifying}
\end{aligned}$$

Thus, our approximate $100(1-\alpha)\%$ confidence interval for μ is $\left[\bar{Y} - (\frac{\hat{\sigma}}{\sqrt{n}})Q_Z(1-\frac{\alpha}{2}), \bar{Y} - (\frac{\hat{\sigma}}{\sqrt{n}})Q_Z(\frac{\alpha}{2})\right]$. Again, we can rewrite this as $\bar{Y} \pm (\frac{\hat{\sigma}}{\sqrt{n}})Q_Z(1-\frac{\alpha}{2})$.

6 Practice Problems

Problem 1. Let $Y_1, \ldots, Y_n$ be i.i.d. from some parametric model $F_{\vec{Y};\theta}$, where $\hat{\theta}_{\text{MLE}}$ is the MLE for θ. Assume n is sufficiently large for the Central Limit Theorem to apply. Suppose we are interested in a 95% confidence interval for θ.

(a) Propose an asymptotic pivot.

(b) Construct an approximate 95% confidence interval for θ.

Solution

Recall $\sqrt{n}(\hat{\theta}_{\text{MLE}}-\theta) \xrightarrow{d} \mathcal{N}\left(0, \frac{1}{\mathcal{I}_{Y_1}(\theta)}\right)$. Thus, $(\sqrt{n\mathcal{I}_{Y_1}(\theta)})(\hat{\theta}_{\text{MLE}}-\theta) = (\sqrt{\mathcal{I}_{\vec{Y}}(\theta)})(\hat{\theta}_{\text{MLE}}-\theta) \xrightarrow{d} \mathcal{N}(0,1)$. This is an asymptotic pivot. Since n is sufficiently large, $(\sqrt{\mathcal{I}_{\vec{Y}}(\theta)})(\hat{\theta}_{\text{MLE}}-\theta) \dot{\sim} \mathcal{N}(0,1)$.

$$
\begin{aligned}
0.95 &\approx P\left(Q_Z(0.025) \leq (\sqrt{\mathcal{I}_{\vec{Y}}(\theta)})(\hat{\theta}_{\text{MLE}} - \theta) \leq Q_Z(0.975)\right) && \text{by quantiles} \\
&= P\left(-1.96 \leq (\sqrt{\mathcal{I}_{\vec{Y}}(\theta)})(\hat{\theta}_{\text{MLE}} - \theta) \leq 1.96\right) && \text{by substituting} \\
&= P\left(\frac{-1.96}{\sqrt{\mathcal{I}_{\vec{Y}}(\theta)}} \leq \hat{\theta}_{\text{MLE}} - \theta \leq \frac{1.96}{\sqrt{\mathcal{I}_{\vec{Y}}(\theta)}}\right) && \text{by algebra} \\
&= P\left(\frac{-1.96}{\sqrt{\mathcal{I}_{\vec{Y}}(\theta)}} - \hat{\theta}_{\text{MLE}} \leq -\theta \leq \frac{1.96}{\sqrt{\mathcal{I}_{\vec{Y}}(\theta)}} - \hat{\theta}_{\text{MLE}}\right) && \text{by algebra} \\
&= P\left(\frac{1.96}{\sqrt{\mathcal{I}_{\vec{Y}}(\theta)}} - \hat{\theta}_{\text{MLE}} \geq \theta \geq \frac{-1.96}{\sqrt{\mathcal{I}_{\vec{Y}}(\theta)}} - \hat{\theta}_{\text{MLE}}\right) && \text{by algebra} \\
&= P\left(\hat{\theta}_{\text{MLE}} - \frac{1.96}{\sqrt{\mathcal{I}_{\vec{Y}}(\theta)}} \leq \theta \leq \hat{\theta}_{\text{MLE}} + \frac{1.96}{\sqrt{\mathcal{I}_{\vec{Y}}(\theta)}}\right) && \text{by simplifying}
\end{aligned}
$$

Since $\mathcal{I}_{\vec{Y}}(\theta)$ is theoretical, we use the "plug-in principle" and replace it with $\mathcal{I}_{\vec{Y}}(\hat{\theta}_{\text{MLE}})$. Thus, our approximate 95% confidence interval for θ is $\left[\hat{\theta}_{\text{MLE}} - \frac{1.96}{\sqrt{\mathcal{I}_{\vec{Y}}(\hat{\theta}_{\text{MLE}})}}, \hat{\theta}_{\text{MLE}} + \frac{1.96}{\sqrt{\mathcal{I}_{\vec{Y}}(\hat{\theta}_{\text{MLE}})}}\right]$. We can rewrite this as $\hat{\theta}_{\text{MLE}} \pm \frac{1.96}{\sqrt{\mathcal{I}_{\vec{Y}}(\hat{\theta}_{\text{MLE}})}}$.

Problem 2. Let $Y_1, \dots, Y_n \overset{\text{i.i.d.}}{\sim} \text{Expo}(\lambda)$. Unfortunately, we do not live in Asymptopia, so n is not large enough for the Central Limit Theorem to apply.

(a) Propose an exact pivot.

(b) Construct an exact 95% confidence interval for λ.

Solution

Recall for $X_1, \dots, X_n \overset{\text{i.i.d.}}{\sim} \text{Expo}(\lambda)$, $\sum_{i=1}^n X_i \sim \text{Gamma}(n, \lambda)$, $c\sum_{i=1}^n X_i \sim \text{Gamma}(n, \frac{\lambda}{c})$. Thus, $n\lambda\bar{Y} \sim \text{Gamma}(n, 1)$. This is an exact pivot.

$$
\begin{aligned}
0.95 &= P\left(Q_{\text{Gamma}(n,1)}(0.025) \leq n\lambda\bar{Y} \leq Q_{\text{Gamma}(n,1)}(0.975)\right) && \text{by quantiles} \\
&= P\left(\frac{Q_{\text{Gamma}(n,1)}(0.025)}{n\bar{Y}} \leq \lambda \leq \frac{Q_{\text{Gamma}(n,1)}(0.975)}{n\bar{Y}}\right) && \text{by algebra} \\
&= P\left(\frac{Q_{\text{Gamma}(n,1)}(0.025)}{\sum_{i=1}^n Y_i} \leq \lambda \leq \frac{Q_{\text{Gamma}(n,1)}(0.975)}{\sum_{i=1}^n Y_i}\right) && \text{by simplifying}
\end{aligned}
$$

Thus, our 95% confidence interval for λ is $\left[\frac{Q_{\text{Gamma}(n,1)}(0.025)}{\sum_{i=1}^n Y_i}, \frac{Q_{\text{Gamma}(n,1)}(0.975)}{\sum_{i=1}^n Y_i}\right]$. Notice $\text{Gamma}(n, 1)$ is not symmetric.