Homework Solutions: Lectures 7 & 8

Confidence Intervals & Bootstrap

Probability & Statistics Course

Problem 01 · Election Night Nail-Biter

Exercise

A poll of $n = 900$ likely voters finds that 52% support candidate A.

a) Build a 95% **Wald** CI for the true proportion p. Based on this interval, can you call the election for candidate A?

b) Now compute the **Wilson** CI for the same data. How does it differ from the Wald CI?

c) Imagine only $n = 20$ voters were polled and 11 (55%) said A. Recompute both Wald and Wilson CIs. Which one behaves better near the boundary, and why?

Solution

(a) Wald CI. Setup. We have $n = 900$ and $\hat{p} = 0.52$. The Wald CI for a proportion uses the formula:

$$\hat{p} \pm z_{\alpha/2} \cdot \text{SE}_{\text{Wald}}, \qquad \text{where} \quad \text{SE}_{\text{Wald}} = \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}.$$

This comes from the CLT: $\hat{p} \approx N\left(p,\ \frac{p(1-p)}{n}\right)$ for large n, and the Wald approach plugs in $\hat{p}$ for the unknown p in the SE.

Computation. For a 95% CI, we use $z_{0.025} = 1.96$ (the value cutting off 2.5% in each tail of the standard normal).

$$\text{SE}_{\text{Wald}} = \sqrt{\frac{0.52 \times 0.48}{900}} = \sqrt{\frac{0.2496}{900}} = \sqrt{0.0002773} = 0.01665.$$

The margin of error is:

$$\text{ME} = 1.96 \times 0.01665 = 0.03264.$$

So the CI is:

$$0.52 \pm 0.03264 = (0.52 - 0.03264,\ 0.52 + 0.03264) = \boxed{(0.4874,\ 0.5526)}.$$

Interpretation. The lower bound is $0.4874 < 0.5$. This means the interval **contains** 0.5, so the data is consistent with candidate A having exactly 50% (or less!) of the true vote. We **cannot** call the election for A at the 95% confidence level.

Intuition: 52% sounds like a lead, but with $n = 900$, the sampling error is about ± 3.3 percentage points – enough to swallow a 2-point lead.

(b) Wilson CI. Why Wilson? The Wald CI has a known weakness: it estimates the SE using $\hat{p}$, which can be a poor estimate of the true SE when p is near 0 or 1, or when n is small. The Wilson CI instead inverts the score test, which doesn't plug in $\hat{p}$ for the variance.

Formula. The Wilson CI is:

$$\tilde{p} \pm \frac{z}{1+z^2/n}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}+\frac{z^2}{4n^2}},$$

where the center is shifted:

$$\tilde{p} = \frac{\hat{p}+z^2/(2n)}{1+z^2/n}.$$

You can think of this as adding $z^2/2 \approx 2$ "pseudo-observations" split evenly between successes and failures, which shrinks $\hat{p}$ toward 0.5.

Computation. With $z = 1.96$:

$$z^2 = 3.8416, \qquad \frac{z^2}{n} = \frac{3.8416}{900} = 0.00427, \qquad \frac{z^2}{2n} = 0.00214.$$

The center:

$$\tilde{p} = \frac{0.52+0.00214}{1+0.00427} = \frac{0.52214}{1.00427} = 0.51992.$$

The margin term inside the square root:

$$\frac{0.52\times 0.48}{900} + \frac{3.8416}{4\times 900^2} = 0.0002773 + 0.0000012 = 0.0002785.$$

The margin:

$$\frac{1.96}{1.00427}\times\sqrt{0.0002785} = 1.9516 \times 0.01669 = 0.03258.$$

So:

$$\text{Wilson CI} = 0.51992 \pm 0.03258 = \boxed{(0.4873,\ 0.5525)}.$$

Comparison. At $n = 900$, the correction term $z^2/n \approx 0.004$ is negligible. The Wald and Wilson CIs are nearly identical. Both contain 0.5 – the Wilson CI does not change our conclusion.

(c) Small sample: $n = 20$, $\hat{p} = 11/20 = 0.55$. This is where the difference between Wald and Wilson becomes visible.

Wald:

$$\text{SE} = \sqrt{\frac{0.55\times 0.45}{20}} = \sqrt{\frac{0.2475}{20}} = \sqrt{0.012375} = 0.1112.$$

$$\text{CI} = 0.55 \pm 1.96\times 0.1112 = 0.55 \pm 0.2180 = \boxed{(0.3320,\ 0.7680)}.$$

Wilson:

$$\frac{z^2}{n} = \frac{3.8416}{20} = 0.1921, \qquad \frac{z^2}{2n} = 0.0960.$$

Now z^2/n is no longer negligible! The center shifts noticeably:

$$\tilde{p} = \frac{0.55+0.0960}{1+0.1921} = \frac{0.6460}{1.1921} = 0.5419.$$

The margin:

$$\frac{1.96}{1.1921}\sqrt{\frac{0.2475}{20}+\frac{3.8416}{1600}} = 1.6443 \times \sqrt{0.012375+0.002401} = 1.6443 \times 0.1216 = 0.1999.$$

$$\text{Wilson CI} = 0.5419 \pm 0.1999 = \boxed{(0.3421,\ 0.7418)}.$$

Key differences:

- The Wilson CI is **narrower** (width 0.400 vs 0.436) and **shifted toward 0.5** (center 0.542 vs 0.550).
- The Wald CI treats the normal approximation as exact and can overshoot [0,1] for extreme $\hat{p}$. For example, if $\hat{p} = 0.98$ with $n = 20$, the Wald CI would extend above 1 – nonsensical for a proportion.
- The Wilson CI always stays within $[0, 1]$ because the "pseudo-observations" anchor the interval. It has better **coverage** (i.e., actually covers p about 95% of the time) for small n.
- **Rule of thumb:** always prefer Wilson (or Agresti-Coull) over Wald for proportions, especially when $n < 100$ or $\hat{p}$ is near 0 or 1.

Problem 02 · A/B Test: Ship It or Wait?

Exercise

An e-commerce company runs an A/B test on their checkout page.

- Control ($n_1 = 200$): conversion rate $\hat{p}_1 = 0.08$ (8%).
- Treatment ($n_2 = 200$): conversion rate $\hat{p}_2 = 0.11$ (11%).

The product manager says "3% lift – ship it!"

a) Build a 95% CI for the difference $p_2 - p_1$.
   Recall: $\text{SE} = \sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}$.

b) Does the CI contain 0? What does this tell the PM?

c) How large would each group need to be (equal sizes) so that the expected CI width is narrow enough to exclude 0, assuming the true lift really is 3%?

Solution

(a) 95% CI for $p_2 - p_1$. Point estimate:

$$\hat{p}_2 - \hat{p}_1 = 0.11 - 0.08 = 0.03.$$

Standard error. Since the two groups are independent, the variance of the difference is the sum of the variances:

$$\text{Var}(\hat{p}_2 - \hat{p}_1) = \frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}.$$

We plug in the sample proportions:

$$\text{SE} = \sqrt{\frac{0.08 \times 0.92}{200} + \frac{0.11 \times 0.89}{200}}.$$

Computing each term:

$$\frac{0.08 \times 0.92}{200} = \frac{0.0736}{200} = 0.000368,$$
$$\frac{0.11 \times 0.89}{200} = \frac{0.0979}{200} = 0.000490.$$
$$\text{SE} = \sqrt{0.000368 + 0.000490} = \sqrt{0.000858} = 0.02928.$$

Confidence interval:

$$\text{CI} = 0.03 \pm 1.96 \times 0.02928 = 0.03 \pm 0.0574 = \boxed{(-0.0274,\ 0.0874)}.$$

(b) Does the CI contain 0? Yes. The interval goes from −2.7% to +8.7%. Since the CI includes 0 (and even negative values), we **cannot conclude** that the treatment is better than the control at the 95% confidence level.

What to tell the PM:

- The observed 3% lift is **not statistically significant**. The data is consistent with anything from a 2.7% *decrease* to an 8.7% increase.
- This does not mean the treatment doesn't work – it means we don't have **enough data** to tell. With only 200 users per group, the SE ($\approx 3\%$) is as large as the effect we're trying to detect.
- Recommendation: either run the test longer to collect more data, or accept the risk and ship with the understanding that the lift might be zero.

(c) Required sample size. Goal: Find n such that the margin of error $\text{ME} = z \cdot \text{SE} < 0.03$, so the CI is narrow enough that if the true lift is 3%, the lower bound will be above 0.

Derivation:

$$\text{ME} = z \cdot \sqrt{\frac{p_1(1-p_1) + p_2(1-p_2)}{n}} < 0.03.$$

Squaring both sides and solving for n:

$$z^2 \cdot \frac{p_1(1-p_1) + p_2(1-p_2)}{n} < 0.03^2$$
$$n > \frac{z^2 \cdot \left[p_1(1-p_1) + p_2(1-p_2)\right]}{0.03^2}.$$

Plugging in:

$$p_1(1-p_1) + p_2(1-p_2) = 0.0736 + 0.0979 = 0.1715.$$
$$n > \frac{3.8416 \times 0.1715}{0.0009} = \frac{0.6588}{0.0009} = 732.1.$$

Rounding up:

$$\boxed{n \geq 733 \text{ per group}}.$$

That's 3.7× the original 200. The total experiment would need $2 \times 733 = 1{,}466$ users.

Note: This is a simplified calculation. A proper power analysis (e.g., targeting 80% power at $\alpha = 0.05$) would give a somewhat different number, but the order of magnitude is the same.

Problem 03 · Bootstrap vs Formula

Exercise

Generate $n = 30$ observations from $\text{Exp}(\lambda = 1)$ in Python (set `np.random.seed(509)`).

a) Compute the **analytical** SE of $\bar{X}$ (since $\sigma = 1/\lambda = 1$, this is $\sigma/\sqrt{n}$).

b) Compute the **bootstrap** SE with $B = 10{,}000$ resamples.

c) Build three 95% CIs for $\mu = 1/\lambda$:

   (i) Normal-theory: $\bar{X} \pm z_{0.025} \cdot \text{SE}$

   (ii) Bootstrap percentile: $[\hat{\theta}^*_{0.025},\ \hat{\theta}^*_{0.975}]$

   (iii) t-interval: $\bar{X} \pm t_{n-1,0.025} \cdot \frac{s}{\sqrt{n}}$

   Compare widths. Do all three contain the true $\mu = 1$?

d) Repeat (a)-(c) for the **sample median** instead of the mean. Which method **cannot** give you a formula-based SE?

Solution

See the accompanying Jupyter notebook (`hw_07_08_solutions.ipynb`) for all code and plots.

The data. With seed 509, we get $n = 30$ exponential draws. Key sample statistics:

$$\bar{X} = 0.8061, \qquad s = 0.8601, \qquad \text{median} = 0.5374.$$

(The true mean is $\mu = 1$ and the true median is $\ln 2 \approx 0.693$. Our sample happens to be on the low side – that's random sampling for you.)

(a) Analytical SE. For $X_1, \ldots, X_n \overset{\text{i.i.d.}}{\sim} \text{Exp}(\lambda = 1)$, we know:

- $E[X] = 1/\lambda = 1$
- $\text{Var}(X) = 1/\lambda^2 = 1$, so $\sigma = 1$

The standard error of the sample mean is:

$$\text{SE}(\bar{X}) = \frac{\sigma}{\sqrt{n}} = \frac{1}{\sqrt{30}} = \boxed{0.1826}.$$

This uses the true (known) σ. In practice we rarely know σ, which is why parts (b) and (c) exist.

(b) Bootstrap SE. Algorithm:

1. For $b = 1, \ldots, B = 10{,}000$:

   (a) Draw a resample of size $n = 30$ **with replacement** from the original data.

   (b) Compute $\bar{X}^*_b$ = mean of the resample.

2. The bootstrap SE is the standard deviation of the B bootstrap means:

$$\text{SE}_{\text{boot}} = \text{sd}(\bar{X}_1^*, \ldots, \bar{X}_B^*).$$

Result: $\text{SE}_{\text{boot}} = 0.1531$.
Comparison with analytical SE:

$$\frac{\text{SE}_{\text{boot}}}{\text{SE}_{\text{analytical}}} = \frac{0.1531}{0.1826} = 0.84.$$

The bootstrap SE is about 16% smaller. This makes sense: the analytical SE uses $\sigma = 1$ (the population value), while the bootstrap SE is based on the sample, which has $s = 0.86 < 1$. The bootstrap adapts to the particular sample you drew.

(c) Three 95% CIs for $\mu = 1$. We need two critical values:

$$z_{0.025} = 1.96, \qquad t_{29,0.025} = 2.0452.$$

(The t critical value is larger because of the heavier tails with 29 degrees of freedom.)

(i) Normal-theory CI (uses known $\sigma = 1$):

$$\bar{X} \pm z_{0.025} \cdot \frac{\sigma}{\sqrt{n}} = 0.8061 \pm 1.96 \times 0.1826 = 0.8061 \pm 0.3579 = (0.4483,\ 1.1640).$$

(ii) Bootstrap percentile CI:
Take the 2.5th and 97.5th percentiles of the 10,000 bootstrap means:

$$(\hat{\theta}^*_{0.025},\ \hat{\theta}^*_{0.975}) = (0.5288,\ 1.1249).$$

(iii) t-interval (uses sample s instead of known σ):

$$\bar{X} \pm t_{29,0.025} \cdot \frac{s}{\sqrt{n}} = 0.8061 \pm 2.0452 \times \frac{0.8601}{\sqrt{30}} = 0.8061 \pm 2.0452 \times 0.1570 = 0.8061 \pm 0.3212 = (0.4850,\ 1.1273).$$

Summary:

Method	**95% CI**	**Width**	**Contains $\mu = 1$?**
Normal-theory (z, known σ)	(0.4483, 1.1640)	0.716	Yes
Bootstrap percentile	(0.5288, 1.1249)	0.596	Yes
t-interval (uses s)	(0.4850, 1.1273)	0.642	Yes

Discussion:

- **Normal-theory is widest.** It uses $\sigma = 1$, but in this sample $s = 0.86$, so it overestimates the spread.
- **t-interval is middle.** It uses the smaller s, but compensates with a bigger critical value ($t_{29} = 2.045 > z = 1.96$). This is the t-distribution's built-in penalty for estimating σ.
- **Bootstrap is narrowest.** It makes no distributional assumptions – it directly estimates the sampling distribution from the data. For skewed data like the exponential, this can be an advantage.
- All three contain $\mu = 1$, which is reassuring. In repeated experiments, about 95% of such intervals would contain the truth.

(d) Bootstrap for the sample median. **The key insight:** There is **no simple closed-form SE** for the median. In theory, for a continuous distribution with density f, the asymptotic SE of the sample median is:

$$\text{SE(median)} = \frac{1}{2f(m)\sqrt{n}},$$

where m is the population median. But this requires knowing $f(m)$ – the density at the median – which we usually don't know. For Exp(1), we could compute $f(\ln 2) = e^{-\ln 2} = 0.5$, giving $\text{SE} = 1/(2 \times 0.5 \times \sqrt{30}) = 0.1826$. But this defeats the purpose – the whole point is that we usually don't know the distribution.

Bootstrap to the rescue. We repeat the same bootstrap procedure but compute the *median* of each resample instead of the mean.

Results:

- Sample median: 0.5374
- Bootstrap SE of the median: 0.1968
- Bootstrap 95% percentile CI: (0.2412, 0.9499)
- True median of Exp(1): $\ln 2 = 0.6931$ – **inside** the CI.

Which method cannot give a formula-based SE? The **normal-theory** and **t-interval** both require an SE formula, which doesn't exist for the median (without knowing the density). Only the **bootstrap** works here. This is the bootstrap's greatest strength: it gives you standard errors and CIs for *any* statistic, even when no analytical formula exists.

Problem 04 · Bootstrap the Correlation

Exercise

Given these (x, y) pairs:

$$(1,2),\ (2,3),\ (3,5),\ (4,4),\ (5,7),\ (6,8),\ (7,6),\ (8,9),\ (9,10),\ (10,12).$$

a) Compute the Pearson correlation r.

b) Use the bootstrap ($B = 5{,}000$) to build a 95% percentile CI for the population correlation ρ.

c) Plot the bootstrap distribution of r^*. Is it symmetric? If not, why might that be?

d) Fisher's z-transform gives an analytical CI: transform $z = \frac{1}{2}\ln\frac{1+r}{1-r}$, build a normal CI using $\text{SE}(z) = \frac{1}{\sqrt{n-3}}$, then back-transform with $r = \frac{e^{2z}-1}{e^{2z}+1}$. Compare with your bootstrap CI.

Solution

(a) Pearson r. Recall that the Pearson correlation is:

$$r = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2} \cdot \sqrt{\sum_{i=1}^{n}(y_i - \bar{y})^2}}.$$

With $n = 10$, $\bar{x} = 5.5$, $\bar{y} = 6.6$:

$$\boxed{r = 0.9506}.$$

This is a strong positive linear association. (See notebook for the full computation.)

(b) Bootstrap 95% percentile CI. **Algorithm:**

1. For $b = 1, \ldots, B = 5{,}000$:

   (a) Draw indices $i_1, \ldots, i_{10}$ with replacement from $\{1, \ldots, 10\}$.

   (b) Compute $r_b^* = \text{cor}(x_{i_1}, \ldots, x_{i_{10}};\ y_{i_1}, \ldots, y_{i_{10}})$.
   Important: resample **pairs** (x_i, y_i) together, not x and y separately! Resampling them independently would destroy the dependence structure.

2. The 95% percentile CI is $[r^*_{(0.025)},\ r^*_{(0.975)}]$.

Result:

$$\boxed{\text{Bootstrap 95\% CI} = (0.8250,\ 0.9928)}.$$

The interval is entirely positive and quite narrow – we are confident the population correlation is high.

(c) Symmetry of the bootstrap distribution. **No, the distribution is strongly left-skewed** (skewness ≈ -5.0).

Why? The Pearson r is bounded in $[-1, 1]$. When the true ρ is near the upper boundary (here ≈ 0.95):

- There is very little room to go *higher* (only 0.05 from r to 1).
- There is a lot of room to go *lower* (0.95 from r to 0, and even further to -1).

This creates a long left tail and a compressed right tail – a "ceiling effect." The bootstrap faithfully reproduces this skewness.

Practical consequence: the percentile CI may not have exact 95% coverage for skewed distributions. This motivates both the Fisher z-transform (next part) and the BCa bootstrap (bias-corrected and accelerated), which adjusts for skewness.

(d) Fisher z-transform CI. **The idea.** Fisher (1921) discovered that the transformation

$$z = \frac{1}{2} \ln \frac{1 + r}{1 - r} = \text{arctanh}(r)$$

makes the sampling distribution of z approximately normal with $\text{SE}(z) = 1/\sqrt{n-3}$, regardless of the true ρ. So we:

1. Transform r to z-space.
2. Build a normal CI in z-space.
3. Transform back to r-space.

Step 1: Transform.

$$z = \frac{1}{2} \ln \frac{1 + 0.9506}{1 - 0.9506} = \frac{1}{2} \ln \frac{1.9506}{0.0494} = \frac{1}{2} \ln(39.49) = \frac{1}{2} \times 3.677 = 1.8384.$$

Step 2: Normal CI in z-space.

$$\text{SE}(z) = \frac{1}{\sqrt{n-3}} = \frac{1}{\sqrt{7}} = 0.3780.$$

$$z \in 1.8384 \pm 1.96 \times 0.3780 = 1.8384 \pm 0.7409 = (1.0976,\ 2.5792).$$

Step 3: Back-transform. Using $r = \tanh(z) = \frac{e^{2z}-1}{e^{2z}+1}$:

$$r_{\text{lo}} = \tanh(1.0976) = \frac{e^{2.1952}-1}{e^{2.1952}+1} = \frac{8.982-1}{8.982+1} = \frac{7.982}{9.982} = 0.7996,$$
$$r_{\text{hi}} = \tanh(2.5792) = \frac{e^{5.1584}-1}{e^{5.1584}+1} = \frac{173.6-1}{173.6+1} = \frac{172.6}{174.6} = 0.9886.$$

$$\boxed{\text{Fisher CI} = (0.7996,\ 0.9886)}.$$

Comparison:

Method	**95% CI**	**Width**
Bootstrap percentile	(0.825, 0.993)	0.168
Fisher z-transform	(0.800, 0.989)	0.189

The Fisher CI is wider, especially on the left (lower bound 0.80 vs 0.83). This is because the z-transform properly symmetrizes the distribution, producing a more honest interval that accounts for the skewness. With only $n = 10$, the Fisher CI is slightly more conservative – which is appropriate.

Both methods agree that ρ is strongly positive (> 0.8).

Problem 05 · When Bootstrap Breaks

Exercise

Consider $X_1, \ldots, X_n \overset{\text{i.i.d.}}{\sim} \text{Uniform}(0, \theta)$. The MLE is $\hat{\theta} = X_{(n)} = \max(X_i)$.

a) Show that $n(\theta - X_{(n)}) \xrightarrow{d} \text{Exp}(1/\theta)$.

   Hint: start from $P(X_{(n)} \le x) = (x/\theta)^n$ *and substitute* $u = n(\theta - x)$.

b) Using $n = 50$ and $\theta = 1$, simulate 10,000 samples. For each sample, also run $B = 1{,}000$ bootstrap resamples of $\hat{\theta}^*$. Compare the true distribution of $n(\theta - \hat{\theta})$ with the bootstrap distribution of $n(\hat{\theta} - \hat{\theta}^*)$. Do they match?

c) Explain why the bootstrap fails here. What is special about the convergence rate of $\hat{\theta}$?

   Hint: the bootstrap "works" when the convergence rate is $\sqrt{n}$. *Here it is* n.

Solution

(a) Limiting distribution of $n(\theta - X_{(n)})$. Step 1: CDF of the maximum.

Each $X_i \sim \text{Uniform}(0, \theta)$ has CDF $F(x) = x/\theta$ for $0 \le x \le \theta$.

The maximum $X_{(n)} = \max(X_1, \ldots, X_n)$ satisfies:

$$P(X_{(n)} \le x) = P(\text{all } X_i \le x) = \prod_{i=1}^{n} P(X_i \le x) = \left(\frac{x}{\theta}\right)^n, \qquad 0 \le x \le \theta.$$

(This uses independence: the max is $\le x$ iff every observation is $\le x$.)

Step 2: Change of variables.

Let $U = n(\theta - X_{(n)})$. We want $P(U \le u)$ for $u \ge 0$.
Since $U \le u \iff X_{(n)} \ge \theta - u/n$:

$$\begin{aligned}P(U \le u) &= P(X_{(n)} \ge \theta - u/n) \\ &= 1 - P(X_{(n)} < \theta - u/n) \\ &= 1 - \left(\frac{\theta - u/n}{\theta}\right)^n \\ &= 1 - \left(1 - \frac{u}{n\theta}\right)^n.\end{aligned}$$

Step 3: Take $n \to \infty$.
We use the fundamental limit $(1 - a/n)^n \to e^{-a}$ as $n \to \infty$:

$$\left(1 - \frac{u}{n\theta}\right)^n \xrightarrow{n\to\infty} e^{-u/\theta}.$$

Therefore:

$$P(U \le u) \to 1 - e^{-u/\theta}, \qquad u \ge 0.$$

This is the CDF of Exp(rate $= 1/\theta$) – an exponential with mean θ.

$$\boxed{n(\theta - X_{(n)}) \xrightarrow{d} \text{Exp}(1/\theta).}$$

□

Remark: For $\theta = 1$, the limit is Exp(1) with mean 1. This is already exact for finite n: $n(\theta - X_{(n)})$ has *exactly* CDF $1 - (1 - u/(n\theta))^n$ – the limiting Exp is just the asymptotic approximation.

(b) Simulation: true vs bootstrap distribution. See the notebook for full code and plots. Here is what the simulation shows:

Setup: $n = 50$, $\theta = 1$. For each of 10,000 Monte Carlo samples:

1. Draw $X_1, \ldots, X_{50} \overset{\text{i.i.d.}}{\sim} \text{Uniform}(0, 1)$.
2. Compute $\hat{\theta} = \max(X_i)$ and record $n(\theta - \hat{\theta}) = 50(1 - \hat{\theta})$.
3. For the bootstrap: draw 1,000 resamples from $\{X_1, \ldots, X_{50}\}$ with replacement, compute $\hat{\theta}^* = \max(X_i^*)$ for each, and record $n(\hat{\theta} - \hat{\theta}^*)$.

Results:
The two distributions are **completely different**:

	True: $n(\theta - \hat{\theta})$	**Bootstrap:** $n(\hat{\theta} - \hat{\theta}^*)$
Shape	Smooth Exp(1) curve	Spike at 0 + decaying tail
Point mass at 0	No	Yes ($\approx 63\%$)
Mean	≈ 1.0	≈ 0.6

The bootstrap distribution has a massive spike at 0. This happens because $\hat{\theta}^* = \hat{\theta}$ whenever the resample includes the original maximum – which occurs with probability $1 - (1 - 1/n)^n \approx 1 - 1/e \approx 0.632$.

The true distribution has no such spike – $\hat{\theta}$ is never exactly equal to θ.

(c) Why does the bootstrap fail? There are three levels of explanation:

Level 1: Mechanical. The bootstrap resamples from $\{X_1, \ldots, X_n\}$, so $\hat{\theta}^* = \max(X_i^*) \leq \hat{\theta}$ **always**. The bootstrap world has a hard upper boundary at $\hat{\theta}$, just like the true world has a boundary at θ. But the bootstrap replaces θ with $\hat{\theta}$ and doesn't realize that $\hat{\theta} < \theta$ – it systematically underestimates the support.

Level 2: Convergence rate. The bootstrap is proven to work when the estimator converges at rate $\sqrt{n}$ (the "regular" case). Here:

- For the sample mean: $\sqrt{n}(\bar{X} - \mu) \xrightarrow{d} N(0, \sigma^2)$. Rate is $\sqrt{n}$. Bootstrap works.
- For the sample max: $n(\theta - X_{(n)}) \xrightarrow{d} \text{Exp}(1/\theta)$. Rate is n. Bootstrap **fails**.

The n-rate is "super-efficient" – the estimator converges so fast that the bootstrap can't keep up. Formally, the bootstrap approximation error is $O(1)$ instead of the usual $o(1)$.

Level 3: Boundary problem. $\hat{\theta}$ lives at the boundary of the parameter space (it's always the rightmost observation). The bootstrap can never generate $\hat{\theta}^* > \hat{\theta}$, so it can't explore what happens above $\hat{\theta}$ – which is exactly where the true θ lives.

Remedies:

- **Parametric bootstrap:** instead of resampling from the data, simulate from Uniform$(0, \hat{\theta})$. This correctly generates new maxima that can be less than $\hat{\theta}$ with the right distribution.
- **Exact methods:** use the known result $n(\theta - X_{(n)}) \sim \text{Exp}(1/\theta)$ to build a CI directly. Since $n(\theta - \hat{\theta})/\theta \xrightarrow{d} \text{Exp}(1)$, an approximate CI is $\left[\hat{\theta},\ \hat{\theta} + q_{0.95}/n\right]$ where $q_{0.95} = -\ln(0.05) \approx 3$ is the 95th percentile of Exp(1).
- **m-out-of-n bootstrap:** resample fewer than n observations (e.g., $m = n^{2/3}$). This slows down the effective convergence rate and can restore consistency. (Advanced topic.)