Mathematical Statistics Recitation 4

Section 002 (Prof. Niles-Weed)

TA: Sonia Reilly

February 20, 2026

[Started with review of Midterm 1 solutions.]

Lecture Review

Lecture 8

- Previously discussed samples from distributions, now samples from populations.
- Two types of sampling: with and without replacement
- Difference only matters if $n \gtrsim 0.05N$

Sampling with replacement

- Simpler math, but higher variance
- Each draw is independent
- Math is the same as sampling from a distribution, in particular:

$$\mathrm{Var}(\bar{X}_n) = \frac{\sigma^2}{n}$$

Sampling without replacement

- More complicated, but lower variance
- Called simple random sampling in the book
- Same mean and variance,

$$\mathbb{E}[X_i] = \mu, \quad \mathrm{Var}(X_i) = \sigma^2,$$

  but X_i's not independent.
- Different variance of sample mean:

$$\mathrm{Var}(\bar{X}_n) = \frac{\sigma^2}{n}\left(1 - \frac{n-1}{N-1}\right) < \frac{\sigma^2}{n} \quad \text{(because of nonzero covariances).}$$

- Can prove (though not directly from CLT, since that needs independence):

$$\frac{\bar{X}_n - \mu}{\sqrt{\operatorname{Var}(\bar{X}_n)}} \xrightarrow{d} N(0,1).$$

- Therefore

$$\mu \in \left[\bar{X}_n - 1.96\sqrt{\frac{\sigma^2}{n}\left(1 - \frac{n-1}{N-1}\right)}, \bar{X}_n + 1.96\sqrt{\frac{\sigma^2}{n}\left(1 - \frac{n-1}{N-1}\right)}\right]$$

is a 95% confidence interval for μ.

- But S_n^2 is no longer unbiased:

$$\mathbb{E}[S_n^2] = \frac{N}{N-1}\sigma^2 \neq \sigma^2.$$

- So for CI's we make both corrections and use

$$\frac{S_n^2}{n}\frac{\left(1 - \frac{n-1}{N-1}\right)}{\frac{N}{N-1}} = \frac{S_n^2}{n}\left(1 - \frac{n}{N}\right)$$

in place of $\frac{S_n^2}{n}$, so

$$\mu \in \left[\bar{X}_n - 1.96\sqrt{\frac{S_n^2}{n}\left(1 - \frac{n}{N}\right)}, \bar{X}_n + 1.96\sqrt{\frac{S_n^2}{n}\left(1 - \frac{n}{N}\right)}\right]$$

is a 95% CI.

Problems

1. Suppose instead of a large political poll, like in the last recitation, now we conduct a poll of 36 out of 71 people (sampling without replacement). 12 of those polled are in favor of the measure. Find a 95% confidence interval for the proportion that is in favor. (Challenge: for problems like these where $X_i \sim \operatorname{Ber}(p)$, find a general expression for the CI of p in terms of $\hat{p} = \bar{X}_n$.)

   Solution:

   We denote each person's poll response by $X_i \sim \operatorname{Ber}(p)$, where p is the proportion we are trying to estimate. The variance can be estimated by

$$\begin{aligned} S_n^2 &= \frac{1}{36-1}\left[36 \cdot \frac{2}{3} \cdot \left(0 - \frac{1}{3}\right)^2 + 36 \cdot \frac{1}{3} \cdot \left(1 - \frac{1}{3}\right)^2\right] \\ &= \frac{36}{35} \cdot \frac{6}{27} = \frac{8}{35}. \end{aligned}$$

   So the estimated standard error would be

$$\frac{S_n}{\sqrt{n}}\sqrt{1 - \frac{n}{N}} = \frac{S_n}{\sqrt{36}}\sqrt{\frac{35}{71}} = \frac{1}{3}\sqrt{\frac{2}{71}} \approx 0.0559.$$

So the CI would be ,

$$p \in \left[\frac{1}{3} - 1.96 \cdot 0.0559, \frac{1}{3} + 1.96 \cdot 0.0559\right] \approx [22\%, 44\%]$$

with probability 95%.

[Challenge: notice that for Bernoulli random variables

$$\frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X}_n)^2 = \hat{p}(1-\hat{p}),$$

where $\hat{p} = \bar{X}_n$ is the estimated p from the sample, so $S_n^2 = \frac{n}{n-1}\hat{p}(1-\hat{p})$. Therefore a 95% CI is

$$p \in \left[\hat{p} - 1.96\sqrt{\frac{\hat{p}(1-\hat{p})}{n-1}}\sqrt{1-\frac{n}{N}}, \hat{p} + 1.96\sqrt{\frac{\hat{p}(1-\hat{p})}{n-1}}\sqrt{1-\frac{n}{N}}\right].$$

This result is derived on page 212 of the textbook.]

2. **[7.36]** With simple random sampling, is $\bar{X}^2$ an unbiased estimator for μ^2? (Note the textbook uses "simple random sampling" to mean sampling without replacement.) If not, what is the bias?

   Solution:

$$\mathbb{E}[\bar{X}^2] = \mathrm{Var}(\bar{X}) + \mathbb{E}^2[\bar{X}] = \frac{\sigma^2}{n}\left(1 - \frac{n-1}{N-1}\right) + \mu^2.$$

   Bias is

$$\frac{\sigma^2}{n}\left(1 - \frac{n-1}{N-1}\right).$$

3. **[7.38]** Is $\frac{1}{n}\sum_{i=1}^{n} X_i^3$ an unbiased estimator for $\frac{1}{N}\sum_{i=1}^{N} x_i^3$, if we sample with replacement? What if we sample without replacement?

   Solution:

   X_i is a random variable that is uniformly chosen among the x_i's, whether or not we sample without replacement (recall the samples are still uniformly distributed, just not independent if there's no replacement), so X_i^3 has value x_j^3 with probability $1/N$ for any j. Therefore

$$\begin{aligned}
\mathbb{E}\left[\frac{1}{n}\sum_{i=1}^{n} X_i^3\right] &= \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}[X_i^3] \\
&= \frac{1}{n}\sum_{i=1}^{n}\left[\sum_{j=1}^{N} x_j^3\,\mathbb{P}(X_i = x_j)\right] \\
&= \frac{1}{n}\sum_{i=1}^{n}\left[\frac{1}{N}\sum_{j=1}^{N} x_j^3\right] \\
&= \frac{1}{N}\sum_{j=1}^{N} x_j^3,
\end{aligned}$$

   so it is an unbiased estimate.