Mathematical Statistics Recitation 3

Section 002 (Prof. Niles-Weed)

TA: Sonia Reilly

February 13, 2026

Lecture Review

Lecture 7

Goal: estimate μ, σ^2 given i.i.d. X_i's with mean μ, variance σ^2.

Point Estimates

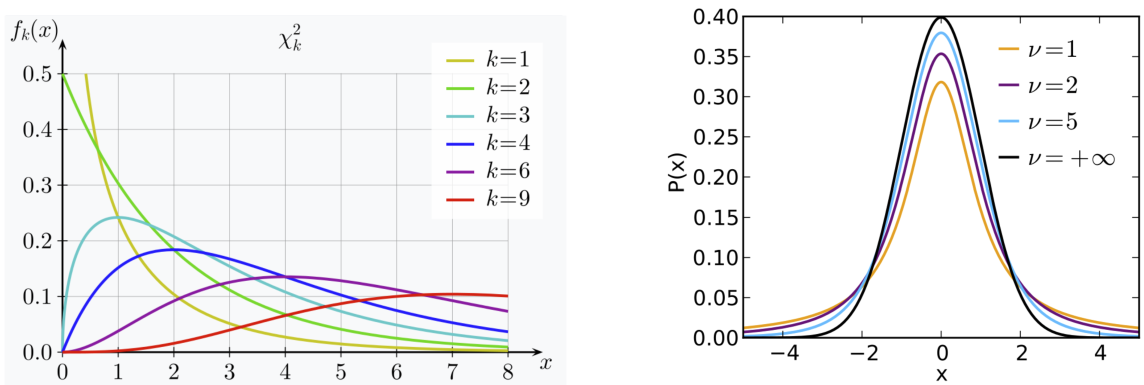
- Sample mean:
$$\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$$
- Sample variance:
$$S_n^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X}_n)^2$$
(or just use $\frac{1}{n}\sum_{i=1}^{n}(X_i - \mu)^2$ if μ known)

Both unbiased:
$$\mathbb{E}[\bar{X}_n] = \mu, \qquad \mathbb{E}[S_n^2] = \sigma^2,$$
i.e., estimate is equal in expectation to what it's estimating.

Confidence Intervals of μ

CLT also true with sample variance:
$$\frac{\sqrt{n}(\bar{X}_n - \mu)}{S_n} \xrightarrow{d} Z \sim N(0,1) \quad \text{as } n \to \infty.$$
Meaning that $\frac{S_n}{\sqrt{n}}$ is a good approximation to the stdev of $\bar{X}_n$ as $n \to \infty$, called "estimated standard error" ($\sigma/\sqrt{n}$ is standard error).

Define z_α by $P(Z > z_\alpha) = \alpha$, for $Z \sim N(0,1)$. From CLT,
$$\mathbb{P}\left(\left|\frac{\sqrt{n}(\bar{X}_n - \mu)}{S_n}\right| \le z_{\alpha/2}\right) = 1 - \alpha$$
so
$$\bar{X}_n - z_{\alpha/2}\frac{S_n}{\sqrt{n}} \le \mu \le \bar{X}_n + z_{\alpha/2}\frac{S_n}{\sqrt{n}}$$
with probability $1-\alpha$. This is a $1-\alpha$ confidence interval (CI) for μ.
Most common value: $z_{\alpha/2} = 1.96$ for 95% CI.



(a) Pdf of χ^2_k for various k (b) Pdf of t_ν for various ν

Figure 1: Distributions of χ^2_n and t distributions for various degrees of freedom. Credit: Wikipedia.

If X_i's are $\sim N(\mu, \sigma^2)$

We can find the exact distribution

$$\frac{\sqrt{n}(\bar{X}_n - \mu)}{S_n} \sim t_{n-1},$$

called the Student's t-distribution, defined by:

- χ^2_1 is the distribution of Z^2, where $Z \sim N(0,1)$.
- χ^2_n is the distribution of $\sum_{i=1}^n Z_i^2$, where $Z_i \sim N(0,1)$ i.i.d.
- t_n is the distribution of $\dfrac{Z}{\sqrt{U/n}}$, where $Z \sim N(0,1)$, $U \sim \chi^2_n$.

These are defined so that

$$\frac{S_n^2}{\sigma^2/(n-1)} \sim \chi^2_{n-1}, \quad \text{so} \quad \frac{\sqrt{n}(\bar{X}_n - \mu)}{S_n} \sim t_{n-1}.$$

$$\frac{\sqrt{n}(\bar{X}_n - \mu)}{S_n} = \frac{\sqrt{n}}{\sigma}(\bar{X}_n - \mu) \Big/ \sqrt{\frac{(n-1)S_n^2}{\sigma^2(n-1)}} \approx \frac{Z}{\sqrt{U/(n-1)}} \sim t_{n-1}.$$

$$t_{n-1} \xrightarrow[n\to\infty]{} N(0,1),$$

so this is consistent with the CLT.
So for $X_i \sim N(\mu, \sigma^2)$ i.i.d., we can make a confidence interval

$$\left[\bar{X}_n - t_{n-1,\alpha/2}\frac{S_n}{\sqrt{n}}, \bar{X}_n + t_{n-1,\alpha/2}\frac{S_n}{\sqrt{n}}\right]$$

even if n is not large!

Summary:

	X_i's $\sim N(\mu, \sigma^2)$	X_i's not normal
$n \leq 30$	$\dfrac{\sqrt{n}(X_n - \mu)}{S_n} \sim t_{n-1}$	no general theorem
$n \geq 30$	$\dfrac{\sqrt{n}(X_n - \mu)}{S_n} \sim t_{n-1} \approx N(0,1)$ (approx equivalent)	$\dfrac{\sqrt{n}(X_n - \mu)}{S_n} \sim N(0,1)$ (approx)

Problems

1. **[7.3]** Which of the following is a random variable?

   - a. population mean   not random
   - b. population size, N not random
   - c. sample size, n not random
   - d. sample mean   random
   - e. variance of sample mean   not random
   - f. largest value in sample   random
   - g. population variance   not random
   - h. estimated variance of sample mean   random

2. **[7.10]** True or false? If a sample from a population is large, a histogram of the values in the sample will be approximately normal, even if the population is not normal.

   Solution:

   False. This is true of the sample means, not the sample values. The sample values will have a similar distribution to the population.

3. **[7.20]** In one example in the textbook, a 95% confidence interval for μ was found to be (1.44, 1.76). Because μ is some fixed number, it either lies in this interval or it doesn't, so it doesn't make any sense to claim that $P(1.44 \leq \mu \leq 1.76) = .95$. What do we mean, then, by saying this is a "95% confidence interval?"

   Solution:

   The confidence interval is one realization of a random interval that depends on the random samples X_i. We mean that 95% of the time, the X_i's will give us a confidence interval that contains μ.

4. **[7.9]** In a simple random sample of 1,500 voters, 55% said they planned to vote for a certain proposition, and 45% said they would vote against it. The estimated margin of victory for the proposition is thus $55\% - 45\% = 10\%$. What is the standard error of this estimated margin? What is a 95% confidence interval for the margin?

   Solution:

$$X_i \sim \text{Ber}(p), \quad p = \mu \text{ (pop. mean) is unknown}$$

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i = 0.55 \quad \text{if } X_i = 1 \text{ indicates yes vote}$$

$$S_n^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2 = \frac{1}{1499}\left[1500(0.55)(1-0.55)^2 + 1500(0.45)(0-0.55)^2\right] = 0.2477$$

[Alternatively: for Bernoulli distribution, $\sigma^2 = p(1-p) = \mu(1-\mu)$, so

$$\sigma^2 \approx \frac{n}{n-1}\bar{X}_n(1-\bar{X}_n) = \frac{1500}{1499}0.55(0.45) = 0.2477.$$

This can be done when the X_i come from a distribution for which σ^2 and μ are directly related. The $\frac{n}{n-1}$ is required to make it an unbiased estimator (see the textbook for more info).]

Estimated standard error:

$$\frac{S_n}{\sqrt{n}} = \sqrt{\frac{0.2477}{1500}} = 0.01285.$$

However, this is the standard error of $\bar{X}$, not of the estimated margin of victory, which is $2\bar{X} - 1$. If $\bar{X}$ has stdev $\approx \frac{S_n}{\sqrt{n}}$, then $2\bar{X} - 1$ has stdev $\approx \frac{2S_n}{\sqrt{n}}$, so the standard error of the estimated margin of victory is

$$2(0.01285) = \boxed{0.026}.$$

We know how to find a 95% CI on μ:

$$\bar{X}_n - 1.96\frac{S_n}{\sqrt{n}} \le \mu \le \bar{X}_n + 1.96\frac{S_n}{\sqrt{n}}$$

$$0.55 - 1.96(0.01285) \le \mu \le 0.55 + 1.96(0.01285),$$

$$0.525 \le \mu \le 0.575.$$

The true margin of victory is $2\mu - 1$, so our 95% CI is

$$2(0.525) - 1 \le 2\mu - 1 \le 2(0.575) - 1,$$

or

$$\boxed{(0.05, 0.15)},$$

between 5 and 15 percentage points.

5. **[7.19]** Using the central limit theorem, how should the constant k be chosen so that the interval $(-\infty, \bar{X} + k\frac{S_n}{\sqrt{n}})$ is a 90% confidence interval for μ — i.e., so that $P(\mu \le \bar{X} + k\frac{S_n}{\sqrt{n}}) = .9$? This is called a one-sided confidence interval. How should k be chosen so that $(\bar{X} - k\frac{S_n}{\sqrt{n}}, \infty)$ is a 95% one-sided confidence interval?

   Solution:

   From the CLT,

$$\frac{\sqrt{n}(\bar{X}_n - \mu)}{S_n} \xrightarrow{d} Z \sim N(0,1),$$

so

$$\begin{aligned}
\mathbb{P}\left(\mu \in \left(-\infty, \bar{X}_n + k\frac{S_n}{\sqrt{n}}\right)\right) &= \mathbb{P}\left(\mu \leq \bar{X}_n + k\frac{S_n}{\sqrt{n}}\right) \\
&= \mathbb{P}\left(\bar{X}_n - \mu \geq -k\frac{S_n}{\sqrt{n}}\right) \\
&= \mathbb{P}\left(\frac{\sqrt{n}(\bar{X}_n - \mu)}{S_n} \geq -k\right) \\
&\approx \mathbb{P}\left(Z \geq -k\right) \\
&= 1 - \mathbb{P}\left(Z \geq k\right) \\
&= 0.90
\end{aligned}$$

so $\mathbb{P}\left(Z \geq k\right) = 0.10 \implies k = z_{0.10} = 1.29$, since z_α is the value such that $P(Z \geq z_\alpha) = \alpha$. Similarly,

$$\begin{aligned}
\mathbb{P}\left(\mu \in \left(\bar{X}_n - k\frac{S_n}{\sqrt{n}}, \infty\right)\right) &= \mathbb{P}\left(\mu \geq \bar{X}_n - k\frac{S_n}{\sqrt{n}}\right) \\
&= \mathbb{P}\left(\bar{X}_n - \mu \leq k\frac{S_n}{\sqrt{n}}\right) \\
&= \mathbb{P}\left(\frac{\sqrt{n}(\bar{X}_n - \mu)}{S_n} \leq k\right) \\
&\approx \mathbb{P}\left(Z \leq k\right) \\
&= 1 - \mathbb{P}\left(Z \geq k\right) \\
&= 0.95
\end{aligned}$$

implies that $k = z_{0.05} = 1.65$. So a one-sided $1 - \alpha$ confidence interval (of either side) should have $k = z_\alpha$.