A Warning

• Statistics literature full of (somewhat redundant) jargon
• Don’t get confused when looking up extra materials
  – Do refer to the slides for what you actually need to know
Parameter Estimation – Workflow

\[
\mathbb{P}(x|\theta)
\]

Independent samples \(x_1, \ldots, x_n\) from \(\mathbb{P}(x|\theta)\)

Algorithm

\(\hat{\theta}\)

\(\theta = \text{unknown parameter}\)

**Maximum Likelihood Estimation (MLE).** Given data \(x_1, \ldots, x_n\), find \(\hat{\theta} = \hat{\theta}(x_1, \ldots, x_n)\) (“the MLE”) such that \(L(x_1, \ldots, x_n|\hat{\theta})\) is maximized!
Likelihood – Continuous Case

**Definition.** The likelihood of independent observations \( x_1, \ldots, x_n \) is

\[
L(x_1, \ldots, x_n | \theta) = \prod_{i=1}^{n} f(x_i | \theta)
\]

Normal outcomes \( x_1, \ldots, x_n \)

\[
\hat{\theta}_\mu = \frac{\sum_{i}^{n} x_i}{n}
\]

MLE estimator for expectation

\[
\hat{\theta}_\sigma^2 = \frac{1}{n} \sum_{i=1}^{n} (x_i - \hat{\theta}_1)^2
\]

MLE estimator for variance
**Definition.** An estimator is **unbiased** if $\mathbb{E}(\hat{\theta}_n) = \theta$ for all $n \geq 1$.

**Definition.** An estimator is **consistent** if $\lim_{n \to \infty} \mathbb{E}(\hat{\theta}_n) = \theta$.

**Theorem.** MLE estimators are consistent. (But not necessarily unbiased)
Example – Consistency

Normal outcomes $X_1, ..., X_n$ iid according to $\mathcal{N}(\mu, \sigma^2)$ \hspace{1cm} Assume: $\sigma^2 > 0$

\[
\hat{\Theta}_{\sigma^2} = \frac{1}{n} \sum_{i=1}^{n} (X_i - \hat{\Theta}_1)^2
\]

Population variance – Biased!

\[
\hat{\theta}_{\sigma^2} \text{ converges to same value as } S_n^2, \text{ i.e., } \sigma^2, \text{ as } n \to \infty.
\]

\[
\hat{\theta}_{\sigma^2} \text{ is “consistent”}
\]

\[
S_n^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \hat{\Theta}_1)^2
\]

Sample variance – Unbiased!
Why is the estimator consistent, but biased?

\[ \mathbb{E}(\hat{\Theta}_{\sigma^2}) = \frac{1}{n} \sum_{i=1}^{n} \mathbb{E} \left[ (X_i - \hat{\Theta}_1)^2 \right] = \frac{1}{n} \sum_{i=1}^{n} \mathbb{E} \left[ \left( X_i - \frac{1}{n} \sum_{j=1}^{n} X_j \right)^2 \right] \]

\[ = \frac{1}{n} \sum_{i=1}^{n} \mathbb{E} \left[ X_i^2 - \frac{2}{n} X_i \sum_{j=1}^{n} X_j + \frac{1}{n^2} \sum_{j=1}^{n} X_j \sum_{k=1}^{n} X_k \right] \]

...
Why is the estimator consistent, but biased?

\[ E(\hat{\Theta}_{\sigma^2}) = \frac{1}{n} \sum_{i=1}^{n} E\left[ (X_i - \hat{\Theta}_1)^2 \right] = \frac{1}{n} \sum_{i=1}^{n} E\left[ \left( X_i - \frac{1}{n} \sum_{j=1}^{n} X_j \right)^2 \right] \]

\[ \cdots \]

\[ = \left( 1 - \frac{1}{n} \right) \sigma^2 = \frac{n-1}{n} \sigma^2 \rightarrow \sigma^2 \text{ for } n \rightarrow \infty \]

Therefore: \[ E(S_n^2) = \frac{1}{n-1} \sum_{i=1}^{n} E\left[ (X_i - \hat{\Theta}_1)^2 \right] = \frac{n}{n-1} E(\hat{\Theta}_{\sigma^2}) = \sigma^2 \]

Bessel’s correction
Estimation – Confidence Intervals

Unbiasedness/consistency are not sufficient by themselves.

• We want $\mathbb{P}(\hat{\Theta}_n = \theta) = 1$
  – At least as $n \to \infty$
  – Note that $\hat{\Theta}_n$ is continuous for Gaussian, so $\mathbb{P}(\hat{\Theta}_n = \theta) = 0$

• Relaxation: Find smallest $\Delta$ such that $\mathbb{P}(|\hat{\Theta}_n - \theta| \leq \Delta) \geq p$
  for a given $p$
  – We say that $\Delta$ gives us the $p$-confidence interval
  – e.g., 95%-confidence interval means $p = 0.95$
Mean Estimator for Normal – Known Variance

Normal outcomes: $X_1, \ldots, X_n$ iid according to $\mathcal{N}(\mu, \sigma^2)$, known $\sigma^2$.

$$\hat{\theta}_{\mu,n} = \frac{\sum_i^n X_i}{n}$$

Q: which distribution?

A: Normal!

- Expectation $\frac{1}{n} (n \cdot \mu) = \mu$
- Variance $\frac{1}{n^2} n \sigma^2 = \frac{\sigma^2}{n}$

Therefore: $\frac{\hat{\theta}_{\mu,n} - \mu}{\sigma/\sqrt{n}} \sim \mathcal{N}(0,1)$
Mean Estimator for Normal – Known Variance

Normal outcomes: $X_1, \ldots, X_n$ iid according to $\mathcal{N} (\mu, \sigma^2)$, known $\sigma^2$.

\[
\hat{\Theta}_{\mu,n} = \frac{\sum_{i}^{n} X_i}{n} \quad \quad \hat{\Theta}_{\mu,n} - \mu \quad \frac{\sigma/\sqrt{n}}{\sigma/\sqrt{n}} \sim \mathcal{N}(0,1)
\]

\[
\mathbb{P}\left( -z < \frac{\hat{\Theta}_{\mu,n} - \mu}{\sigma/\sqrt{n}} < z \right) = \Phi(z) - \Phi(-z) = 1 - 2\Phi(-z)
\]

Equivalently: \(\mathbb{P}\left( |\hat{\Theta}_{\mu,n} - \mu| < z\sigma/\sqrt{n} \right) = 1 - 2\Phi(-z)\)

E.g., $\Phi(-1.96) \approx 5\% \rightarrow$ Estimate is within $\Delta = 1.96\sigma/\sqrt{n}$ of $\mu$ with probability $\approx 95\%$ (i.e., “$\Delta$ is the 95%-confidence interval”)
Mean Estimator for Normal – Unknown Variance

Normal outcomes: $X_1, \ldots, X_n$ iid according to $\mathcal{N}(\mu, \sigma^2)$, unknown $\sigma^2$.

$$P\left(\left|\hat{\Theta}_{\mu,n} - \mu\right| < \frac{z\sigma}{\sqrt{n}}\right) = 1 - 2\Phi(-z)$$

- Still true, but not that useful, as we cannot evaluate $\sigma$
- What about using $S_n = \sqrt{S_n^2}$ instead?

$$P\left(\left|\hat{\Theta}_{\mu,n} - \mu\right| < \frac{zS_n}{\sqrt{n}}\right) = 1 - 2\Phi(-z) ?$$

- Not true!

$$\frac{\hat{\Theta}_{\mu,n} - \mu}{\sigma/\sqrt{n}} \sim \mathcal{N}(0,1) \quad \frac{\hat{\Theta}_{\mu,n} - \mu}{S_n/\sqrt{n}} \sim \text{t-distribution with } n - 1 \text{ degrees of freedom}$$
Student’s t-Distribution

Parametrized by $\nu = \text{degrees of freedom}$

Notation: $\Psi_\nu(z)$ cdf of t-distribution with $\nu$ dof’s.

Student, "The probable error of a mean". Biometrika 1908.
"Student" was a pseudonym for William Gosset

• Worked for A. Guinness & Son
• Investigated e.g. brewing and barley yields
• Wasn’t allowed to publish with real name

Mean Estimator for Normal – Unknown Variance

Therefore: $\mathbb{P}(|\hat{\Theta}_{\mu,n} - \mu| < z S_n / \sqrt{n}) = 1 - 2\Psi_{n-1}(-z)$

E.g., $\Psi_{9}^{-1}(0.05) \approx 2.26$ → Estimate is within $2.26 S_n / \sqrt{n}$ of $\mu$ with probability $\approx 95\%$