CSE 312

Foundations of Computing II

Lecture 25: Biased Estimation, Confidence Interval

PAUL G. ALLEN SCHOOL OF COMPUTER SCIENCE & ENGINEERING

Stefano Tessaro

tessaro@cs.washington.edu

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

Distribution $\mathbb{P}(x|\theta)$ → Independent samples $x_1, \ldots, x_n$ from $\mathbb{P}(x|\theta)$ → **Algorithm** → $\hat{\theta}$

Parameter estimate

θ = 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 $x_1, \dots., x_n$ is

$$L(x_1, \dots., x_n | \theta) = \prod_{i=1}^{n} f(x_i | \theta)$$

Normal outcomes $x_1, \dots, x_n$

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

MLE estimator for **expectation**

$$\hat{\theta}_{\boldsymbol{\sigma}^2} = \frac{1}{n} \sum_{i=1}^{n} (x_i - \hat{\theta}_\mu)^2$$

MLE estimator for **variance**

Consistent Estimators & MLE

Distribution $\mathbb{P}(x|\theta)$ $\longrightarrow$ samples $X_1, \dots, X_n$ from $\mathbb{P}(x|\theta)$ $\longrightarrow$ **Algorithm** $\longrightarrow$ $\widehat{\Theta}_n$

Parameter estimate

θ = <u>unknown</u> parameter

Definition. An estimator is **unbiased** if $\mathbb{E}\left(\widehat{\Theta}_n\right) = \theta$ for all $n \geq 1$.

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

Theorem. MLE estimators are consistent.

(But not necessarily unbiased)

Example – Consistency

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

$$\widehat{\Theta}_{\sigma^2} = \frac{1}{n}\sum_{i=1}^{n}\left(X_i - \widehat{\Theta}_\mu\right)^2$$

Population variance – Biased!

$$S_n^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(X_i - \widehat{\Theta}_\mu\right)^2$$

Sample variance – Unbiased!

$\widehat{\Theta}_{\sigma^2}$ converges to same value as S_n^2, i.e., σ^2, as $n \to \infty$.

$\widehat{\Theta}_{\sigma^2}$ is "consistent"

Why is the estimator consistent, but biased?

linearity

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

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

…

Why is the estimator consistent, but biased?

linearity

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

…

$$= \left(1 - \frac{1}{n}\right)\sigma^2 = \frac{n-1}{n}\sigma^2 \quad \to \sigma^2 \text{ for } n \to \infty$$

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

Bessel's correction

Estimation – Confidence Intervals

Unbiasedness/consistency are not sufficient by themseleves.

- We want $\mathbb{P}(\widehat{\Theta}_n = \theta) = 1$
  - At least as $n \to \infty$
  - Note that $\widehat{\Theta}_n$ is continuous for Gaussian, so $\mathbb{P}(\widehat{\Theta}_n = \theta) = 0$
- <u>Relaxation:</u> Find <u>smallest</u> Δ such that $\mathbb{P}\left(\left|\widehat{\Theta}_n - \theta\right| \leq \Delta\right) \geq p$ for a given p
  - We say that Δ gives us the p-**confidence interval**
  - e.g., 95%-confidence interval means $p = 0.95$

Mean Estimator for Normal – Known Variance

Normal outcomes: $X_1, \dots, X_n$ iid according to $\mathcal{N}(\mu, \sigma^2)$, <u>known</u> σ^2.

$$\widehat{\Theta}_{\mu,n} = \frac{\sum_{i=1}^{n} X_i}{n}$$

Q: which distribution?

A: Normal!

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

Therefore: $\frac{\widehat{\Theta}_{\mu,n} - \mu}{\sigma/\sqrt{n}} \sim \mathcal{N}(0,1)$

Mean Estimator for Normal – Known Variance

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

$$\widehat{\Theta}_{\mu,n} = \frac{\sum_{i=1}^{n} X_i}{n} \qquad\qquad \frac{\widehat{\Theta}_{\mu,n} - \mu}{\sigma/\sqrt{n}} \sim \mathcal{N}(0,1)$$

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

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

E.g., $\Phi(-1.96) \approx 5\%$ → Estimate is within $\Delta = 1.96\sigma/\sqrt{n}$ of μ with probability $\approx 95\%$ (i.e., "Δ 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 σ^2.

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

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

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

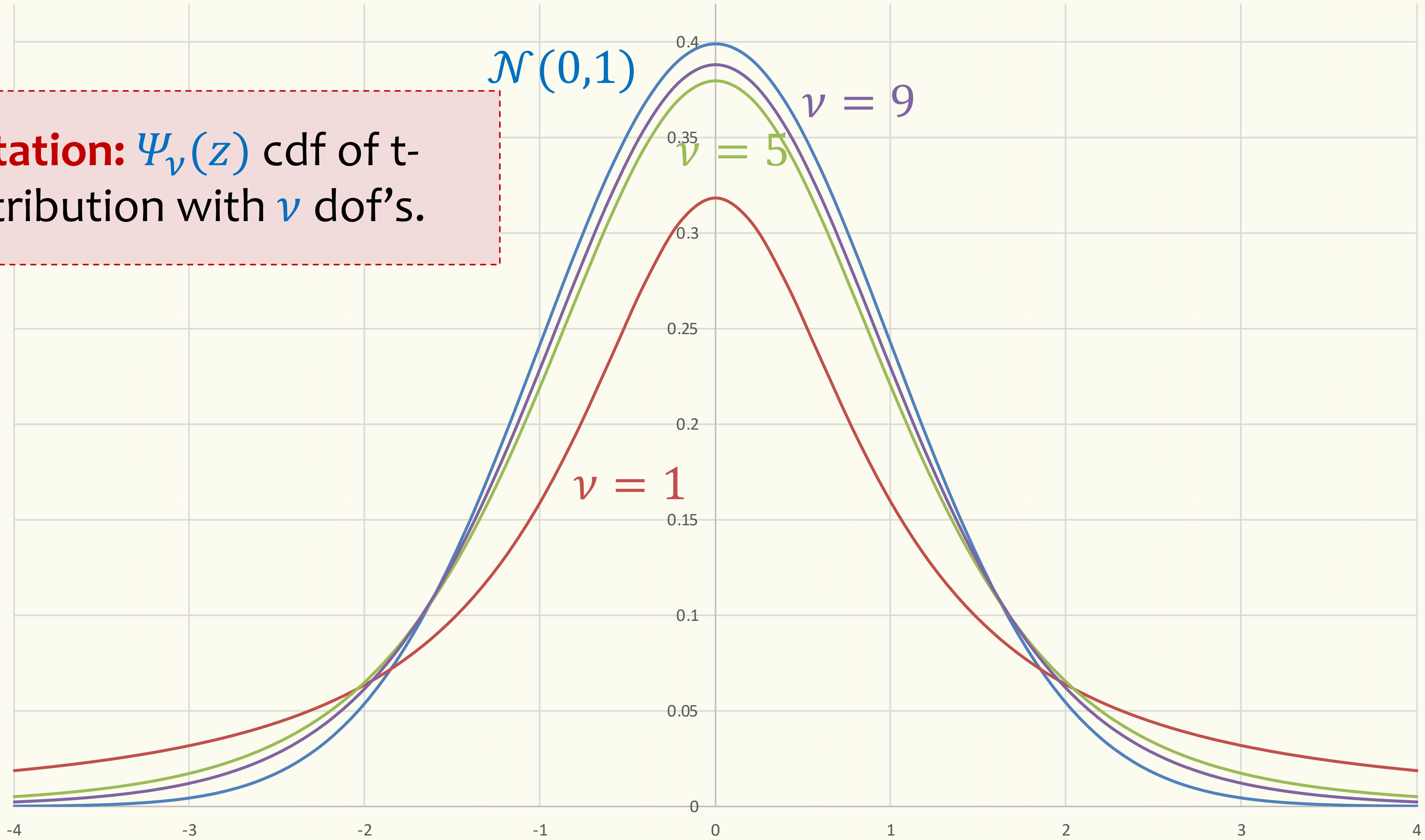
- Not true!

$\dfrac{\widehat{\Theta}_{\mu,n} - \mu}{\sigma/\sqrt{n}} \sim \mathcal{N}(0,1)$ $\qquad$ $\dfrac{\widehat{\Theta}_{\mu,n} - \mu}{S_n/\sqrt{n}} \sim$ t-distribution with $n - 1$ degrees of freedom

Student's t-Distribution

Parametrized by ν = degrees of freedom

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



Student,"The probable error of a mean". Biometrika 1908.

Student?

”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

Source: Wikipedia

Mean Estimator for Normal – Unknown Variance

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

E.g., $\Psi_9^{-1}(0.05) \approx 2.26$ → Estimate is within $2.26S_n/\sqrt{n}$ of μ with probability $\approx 95\%$