CSE 312

Autumn 2013

More on parameter estimation – Bias; and Confidence Intervals

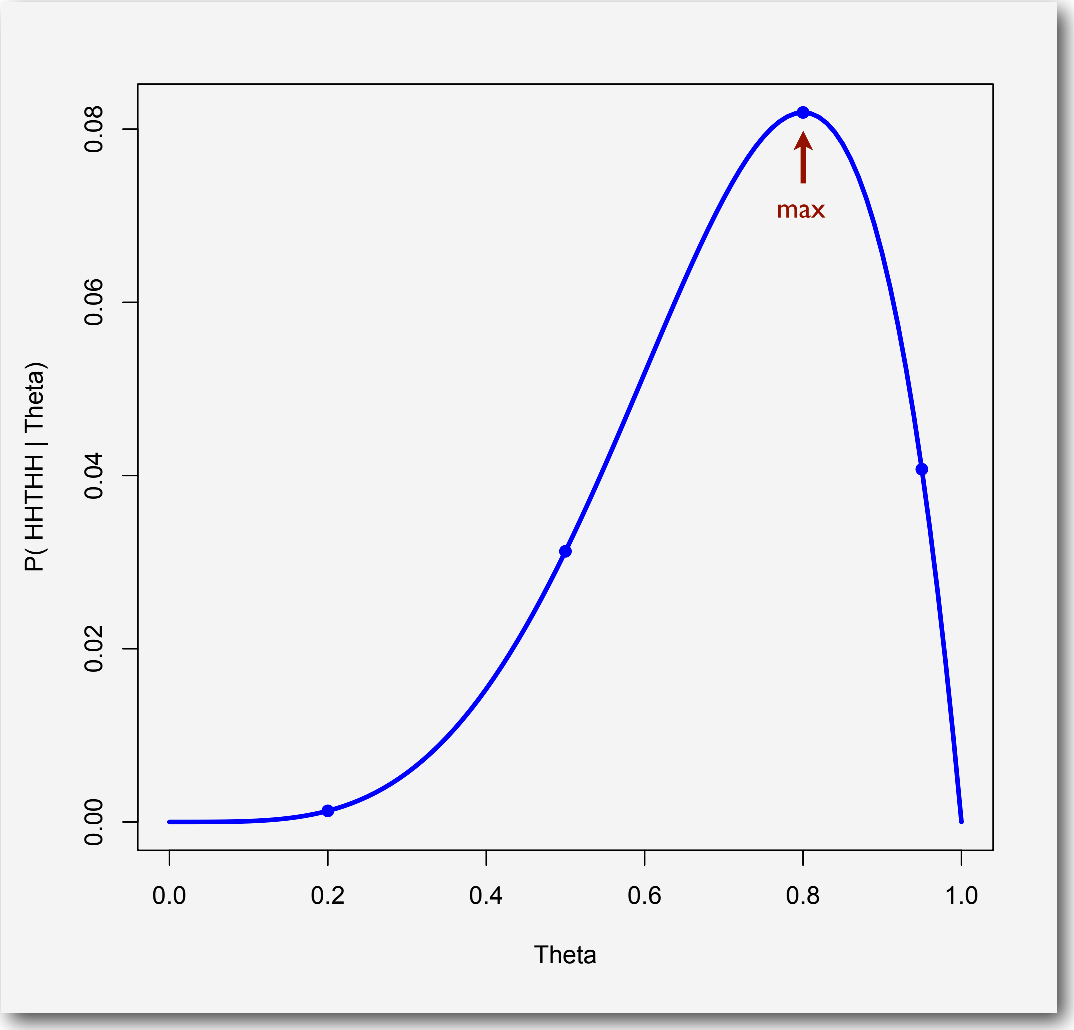
Bias

Recall

Likelihood Function

P(HHTHH | θ):
Probability of HHTHH,
given P(H) = θ:

θ	$\theta^4(1-\theta)$
0.2	0.0013
0.5	0.0313
0.8	0.0819
0.95	0.0407

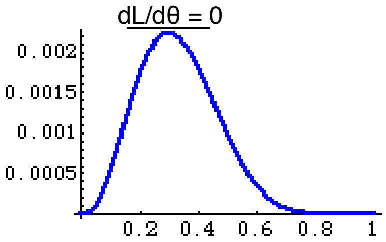


Recall

Example 1

n coin flips, $x_1, x_2, \ldots, x_n$; n_0 tails, n_1 heads, $n_0 + n_1 = n$;

θ = probability of heads



$$\begin{aligned}
L(x_1, x_2, \ldots, x_n \mid \theta) &= (1-\theta)^{n_0}\theta^{n_1} \\
\log L(x_1, x_2, \ldots, x_n \mid \theta) &= n_0 \log(1-\theta) + n_1 \log\theta \\
\frac{\partial}{\partial\theta} \log L(x_1, x_2, \ldots, x_n \mid \theta) &= \frac{-n_0}{1-\theta} + \frac{n_1}{\theta}
\end{aligned}$$

Setting to zero and solving:

$$\boxed{\hat{\theta} = \frac{n_1}{n}}$$

Observed fraction of successes in *sample* is MLE of success probability in *population*

(Also verify it's max, not min, & not better on boundary)

(un-) Bias

A desirable property: An estimator Y_n of a parameter θ is an *unbiased* estimator if

$$E[Y_n] = \theta$$

For coin ex. above, MLE is unbiased:

Y_n = fraction of heads = $(\Sigma_{1 \le i \le n} X_i)/n$,

(X_i = indicator for heads in i^{th} trial) so

$$E[Y_n] = (\Sigma_{1 \le i \le n} E[X_i])/n = n\,\theta/n = \theta$$

by linearity of expectation

Are all unbiased estimators equally good?

No!

E.g., “Ignore all but 1st flip; if it was H, let Y_n’ = 1; else Y_n’ = 0”

Exercise: show this is unbiased

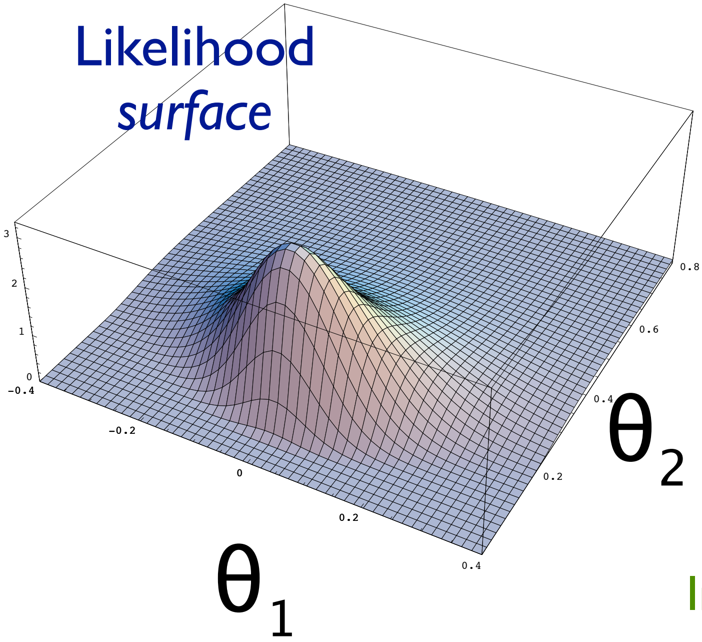
Exercise: if observed data has at least one H and at least one T, what is the likelihood of the data given the model with $\theta = Y_n$’ ?

Recall

Ex 3: $x_i \sim N(\mu, \sigma^2),\ \mu, \sigma^2$ both unknown

$$\ln L(x_1, x_2, \ldots, x_n | \theta_1, \theta_2) = \sum_{1 \le i \le n} -\frac{1}{2} \ln 2\pi\theta_2 - \frac{(x_i - \theta_1)^2}{2\theta_2}$$

$$\frac{\partial}{\partial\theta_1} \ln L(x_1, x_2, \ldots, x_n | \theta_1, \theta_2) = \sum_{1 \le i \le n} \frac{(x_i - \theta_1)}{\theta_2} = 0$$


Likelihood *surface*

$$\boxed{\hat{\theta}_1 = \left(\sum_{1 \le i \le n} x_i\right) / n = \bar{x}}$$

Sample mean is MLE of population mean, again

In general, a problem like this results in 2 equations in 2 unknowns. Easy in this case, since θ_2 drops out of the $\partial/\partial\theta_1 = 0$ equation

Recall

Ex. 3, (cont.)

$$\ln L(x_1, x_2, \ldots, x_n | \theta_1, \theta_2) = \sum_{1 \leq i \leq n} -\frac{1}{2} \ln 2\pi\theta_2 - \frac{(x_i - \theta_1)^2}{2\theta_2}$$

$$\frac{\partial}{\partial\theta_2} \ln L(x_1, x_2, \ldots, x_n | \theta_1, \theta_2) = \sum_{1 \leq i \leq n} -\frac{1}{2} \frac{2\pi}{2\pi\theta_2} + \frac{(x_i - \theta_1)^2}{2\theta_2^2} = 0$$

$$\boxed{\hat{\theta}_2 = \left(\sum_{1 \leq i \leq n} (x_i - \hat{\theta}_1)^2 \right) / n = \bar{s}^2}$$

Sample variance is MLE of *population* variance

Ex. 3, (cont.)

Bias? if $Y_n = (\Sigma_{1 \le i \le n} X_i)/n$ is the sample mean then

$E[Y_n] = (\Sigma_{1 \le i \le n} E[X_i])/n = n\,\mu/n = \mu$

so the MLE is an *unbiased* estimator of population mean

known μ

Similarly, $(\Sigma_{1 \le i \le n} (X_i-\mu)^2)/n$ is an unbiased estimator of σ^2.

Unfortunately, if μ is *unknown*, estimated *from the same data*, as above, $\hat{\theta}_2 = \sum_{1 \le i \le n} \frac{(x_i - \hat{\theta}_1)^2}{n}$ is a consistent, but *biased* estimate of population variance. (An example of *overfitting*.) Unbiased estimate (B&T p467):

$$\boxed{\hat{\theta}_2' = \sum_{1 \le i \le n} \frac{(x_i - \hat{\theta}_1)^2}{n-1}}$$

Roughly, $\lim_{n\to\infty}$ = correct

One Moral: MLE is a great idea, but not a magic bullet

More on Bias of $\hat{\theta}_2$

Biased? Yes. Why? As an extreme, think about n = 1. Then $\hat{\theta}_2 = 0$; probably an underestimate!

Also, consider n = 2. Then $\hat{\theta}_1$ is exactly between the two sample points, the position that *exactly minimizes* the expression for θ_2. Any other choices for θ_1, θ_2 make the likelihood of the observed data slightly *lower*. But it's actually pretty unlikely that two sample points would be chosen exactly equidistant from, and on opposite sides of the mean (p=0, in fact), so the MLE $\hat{\theta}_2$ systematically *underestimates* θ_2, i.e. is biased.

(But not by much, & bias shrinks with sample size.)

Confidence Intervals

A Problem With Point Estimates

Reconsider: estimate the mean of a normal distribution.

Sample $X_1, X_2, \ldots, X_n$

Sample mean $Y_n = (\Sigma_{1 \leq i \leq n} X_i)/n$ is an unbiased estimator of the population mean.

But with probability 1, it's wrong!

Can we say anything about *how* wrong?

E.g., could I find a value Δ s.t. I'm 95% confident that the true mean is within ±Δ of my estimate?

Confidence Intervals for a Normal Mean

Assume X_i's are i.i.d. ~Normal(μ, σ^2)

Mean estimator $Y_n = (\Sigma_{1 \le i \le n} X_i)/n$ is a *random variable;* it has a distribution, a mean *and a variance.* Specifically,

$$Var(Y_n) = Var\left(\frac{\sum_{i=1}^{n} X_i}{n}\right) = \frac{1}{n^2}\sum_{i=1}^{n} Var(X_i) = \frac{n\sigma^2}{n^2} = \frac{\sigma^2}{n}$$

So, $\boxed{Y_n \sim \text{Normal}(\mu, \sigma^2/n),}$ $\therefore$ $\boxed{\frac{Y_n - \mu}{\sigma/\sqrt{n}} \sim \text{Normal}(0,1)}$

Confidence Intervals for a Normal Mean

X_i's are i.i.d. ~ Normal(μ, σ²)

Y_n ~ Normal(μ, σ² /n) $\frac{Y_n - \mu}{\sigma/\sqrt{n}}$ ~ Normal(0,1)

$$P\left(-z < \frac{Y_n - \mu}{\sigma/\sqrt{n}} < z\right) = 1 - 2\Phi(-z)$$

$$P\left(-z < \frac{\mu - Y_n}{\sigma/\sqrt{n}} < z\right) = 1 - 2\Phi(-z)$$

$$P\left(-z\sigma/\sqrt{n} < \mu - Y_n < z\sigma/\sqrt{n}\right) = 1 - 2\Phi(-z)$$

$$P\left(Y_n - z\sigma/\sqrt{n} < \mu < Y_n + z\sigma/\sqrt{n}\right) = 1 - 2\Phi(-z)$$

E.g., true μ within ±1.96σ/$\sqrt{n}$ of estimate ~ 95% of time

N.B: μ is fixed, not random; Y_n is random

C.I. of Norm Mean When σ^2 is Unknown?

X_i's are i.i.d. normal, mean = μ, variance = σ^2 *unknown*

$Y_n = (\Sigma_{1 \leq i \leq n} X_i)/n$ is normal

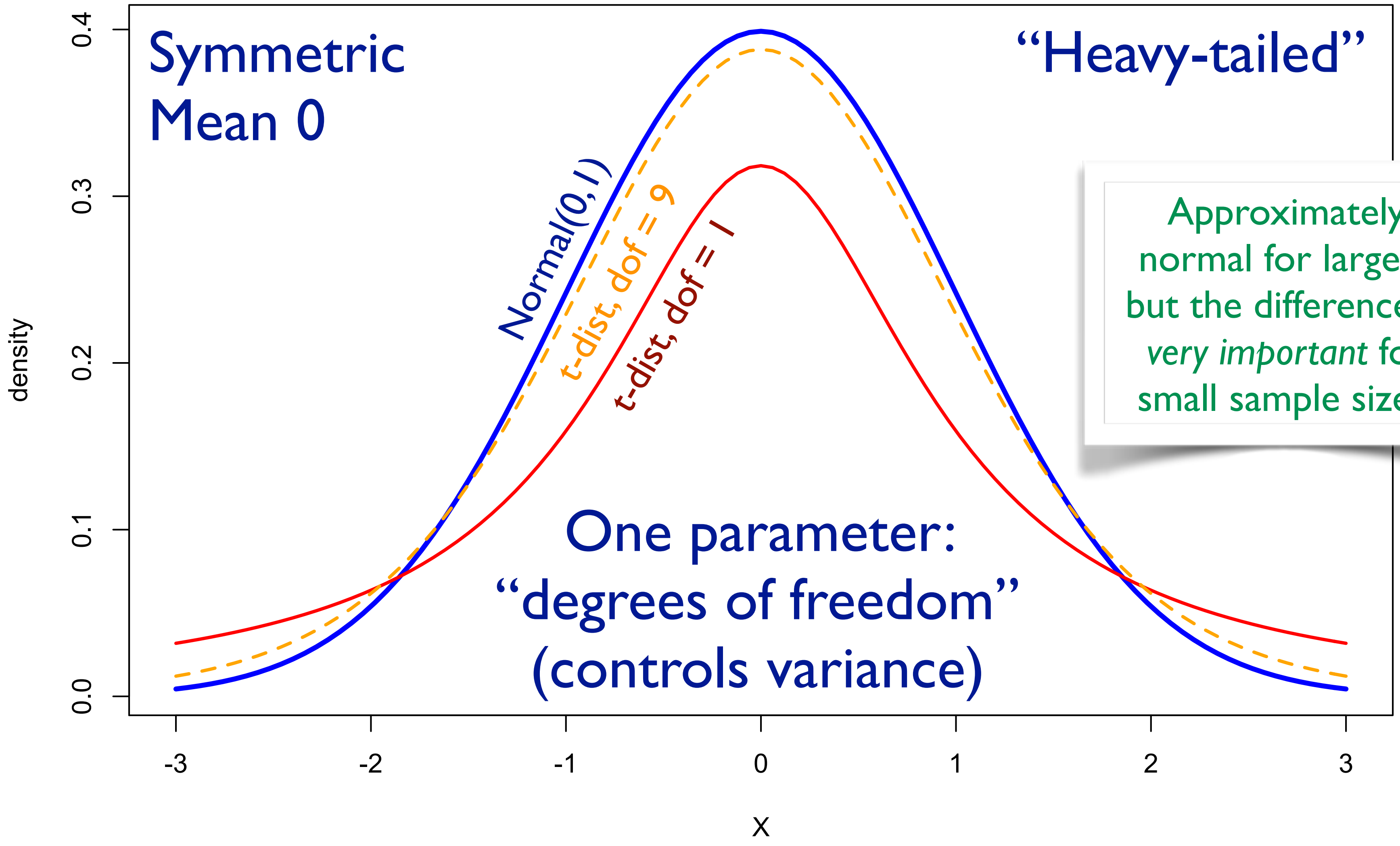
$(Y_n - \mu)/(\sigma / \sqrt{n})$ is std normal, but we don't know μ, σ

Let $S_n^2 = \Sigma_{1 \leq i \leq n} (X_i - Y_n)^2/(n-1)$, the unbiased variance est

$(Y_n - \mu)/(S_n / \sqrt{n})$?

Independent of μ, σ^2, but NOT normal:
"Students' t-distribution with n-1 degrees of freedom"

Student’s t-distribution



Symmetric
Mean 0

“Heavy-tailed”

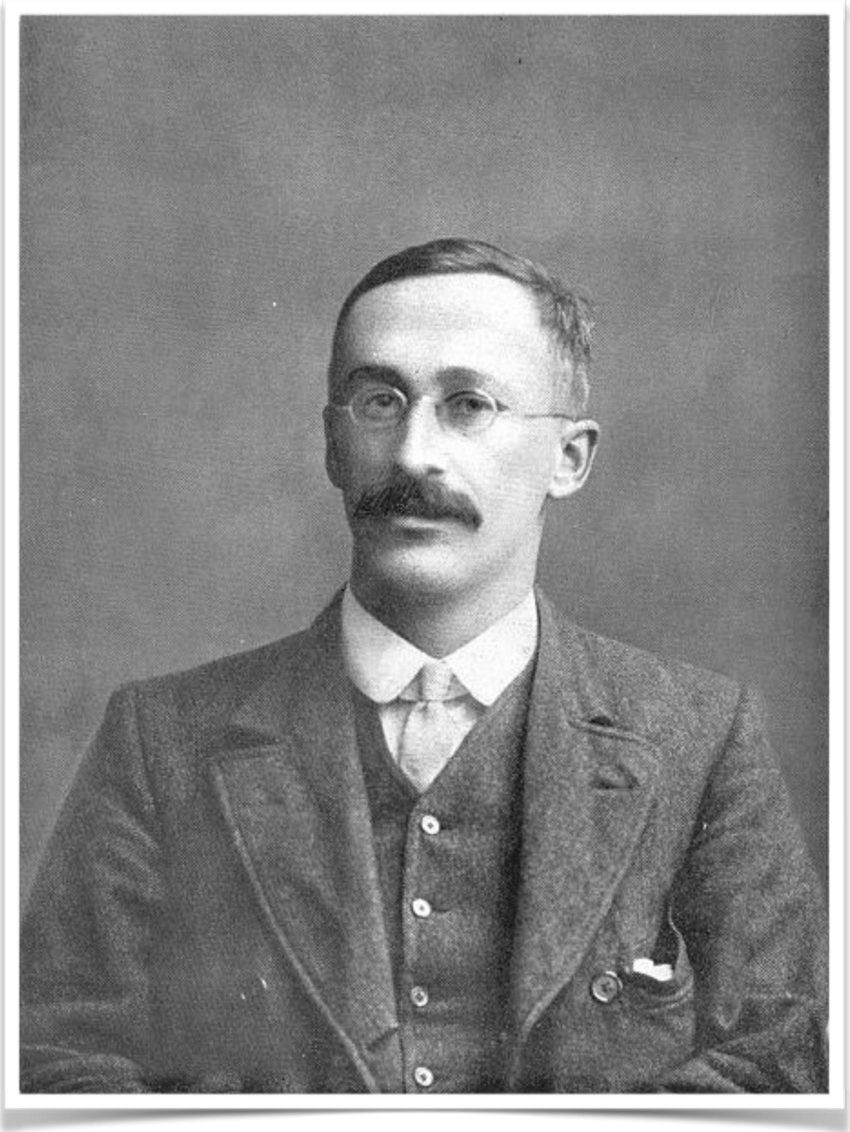
Approximately normal for large *n*, but the difference is *very important* for small sample sizes.

One parameter:
“degrees of freedom”
(controls variance)

William Gossett
aka
“Student”

Worked for A. Guinness & Son, investigating, e.g., brewing and barley yields. Guinness didn’t allow him to publish under his own name, so this important work is tied to his pseudonym…

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



June 13, 1876–October 16, 1937

Letting Ψ_{n-1} be the c.d.f. for the t-distribution with n-1 degrees of freedom, as above we have:

$$P\left(-z < \frac{Y_n - \mu}{S_n/\sqrt{n}} < z\right) = 1 - 2\Psi_{n-1}(-z)$$

$$P\left(-z < \frac{\mu - Y_n}{S_n/\sqrt{n}} < z\right) = 1 - 2\Psi_{n-1}(-z)$$

$$P\left(-zS_n/\sqrt{n} < \mu - Y_n < zS_n/\sqrt{n}\right) = 1 - 2\Psi_{n-1}(-z)$$

$$P\left(Y_n - zS_n/\sqrt{n} < \mu < Y_n + zS_n/\sqrt{n}\right) = 1 - 2\Psi_{n-1}(-z)$$

E.g., for n=10, 95% interval, use z ≈ 2.26, vs 1.96

What about non-normal

If $X_1, X_2, \ldots, X_n$ are *not* normal, you can still get approximate confidence intervals, based on the central limit theorem.

I.e., $Y_n = (\Sigma_{1 \leq i \leq n} X_i)/n$ is *approximately* normal with unknown mean and *approximate* variance $S_n^2 = \Sigma_{1 \leq i \leq n} (X_i - Y_n)^2/(n-1)$, and

$(Y_n - \mu)/(S_n / \sqrt{n})$ is *approximately* t-distributed, so

$$P\left(Y_n - zS_n/\sqrt{n} < \mu < Y_n + zS_n/\sqrt{n}\right) \approx 1 - 2\Psi_{n-1}(-z)$$