CSE 312

Autumn 2012

More on parameter estimation – Bias; and Confidence Intervals

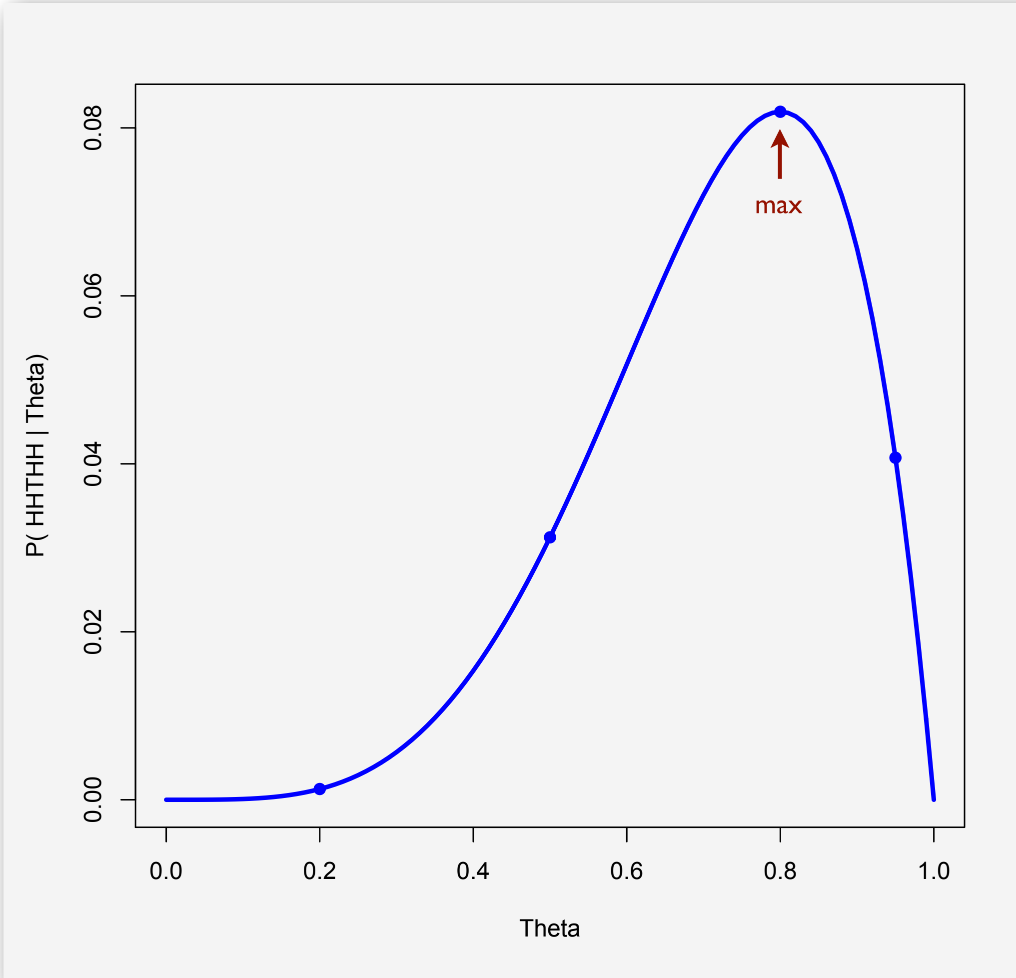
Bias

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

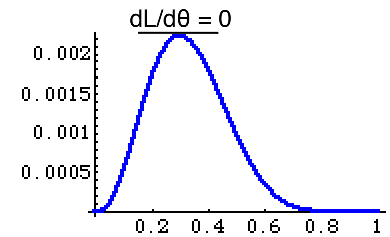


Recall

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 of a parameter θ is an *unbiased* estimator if

$$E[Y] = \theta$$

For coin ex. above, MLE is unbiased:

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

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

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

by linearity of expectation

Are all unbiased estimators equally good?

No!

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

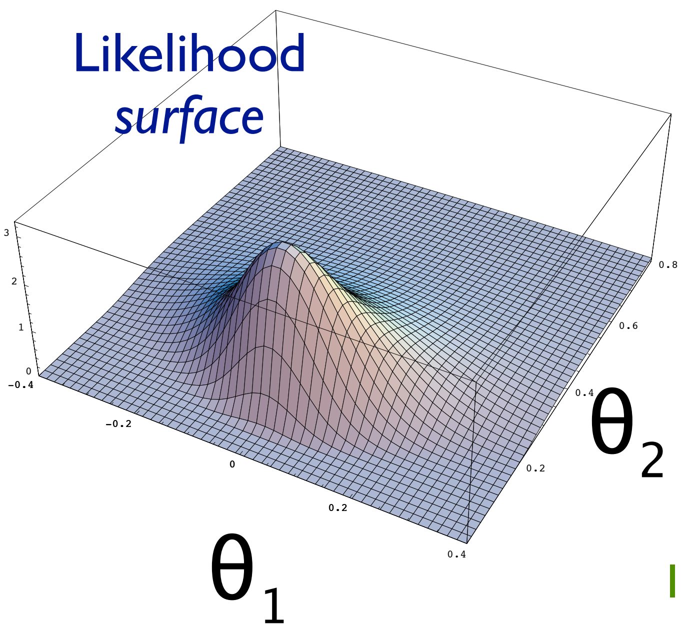
Exercise: show this is unbiased

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 $\theta = Y'$?

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

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

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



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

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.)

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

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

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

Sample variance is MLE of *population* variance

Ex. 3, (cont.)

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

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

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

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, so the MLE $\hat{\theta}_2$ systematically underestimates θ_2.

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

Confidence Intervals

A Problem With Point Estimates

Think again about estimating the mean of a normal distribution.

Sample $X_1, X_2, \ldots, X_n$

We showed sample mean $Y_n = (\Sigma_{1 \leq i \leq n} X_i)/n$ is an unbiased (and consistent) 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 $\pm\Delta$ of my estimate?

$Y_n = (\Sigma_{1 \le i \le n} X_i)/n$ is a *random variable*

It has a mean *and a variance*

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

$$\begin{aligned} Var(Y_n) &= Var((\Sigma_{1 \le i \le n} X_i)/n) = (1/n^2)\ \Sigma_{1 \le i \le n} Var(X_i) \\ &= (1/n^2)(n\ \sigma^2) = \sigma^2/n \end{aligned}$$

So, $Pr((\sqrt{n})|Y_n - \mu|/\sigma < z) = 2(1 - \Phi(z))$, $(z > 0)$

E.g., $Pr((\sqrt{n})|Y_n - \mu|/\sigma < 1.96) \approx 95\%$

I.e., true μ within $\pm 1.96\sigma/\sqrt{n}$ of estimate ~ 95% of time