Math 459 Lecture 4

Todd Kuffner

Comments on Homework

- don't make your life harder - do you need to compute the normalizing constant to find the MLE and posterior mean?
- question 3 - have fun with it, any answer showing effort will be rewarded–there is not a single correct answer

Some Perspective

Why are we doing MLE on every problem? MLE is **NOT Bayesian**}

- posterior combines likelihood and prior
- we will compare Bayesian and ‘frequentist’ procedures – likelihood-based inference is the gold standard for frequentists

Point Estimators vs. Interval Estimators

For today's lecture, we are viewing θ as scalar unless specifically noted otherwise.

Point estimators specify a single value: $\hat{\theta}$ is an estimator of a single point θ in the parameter space.

- take data, do something to it, then spit out a 'best guess' of the unknown true value θ_0

Interval estimators specify a range (interval) of possible values: $L \leq \theta_0 \leq U$

Frequentist	confidence interval
Bayesian	credible interval

Random samples and realizations

Let $X^{(n)} = (X_1, \ldots, X_n)$ be a random sample (these are random variables).

- before we observe the data, each element of sample is a random variable, could take any value with probability specified by population distribution (the sampling model)
- any function of the random sample, say $q(X^{(n)})$, is also a random variable (a **statistic**)
- in frequentist statistics, we speak of the *sampling distribution* of a statistic, i.e. its distribution under *hypothetical* repeated sampling

A *particular* sample: the observed values $(x_1, \ldots, x_n)$ (these are not random variables).

- functions of $x^{(n)}$ are not random
- e.g.

$$q(X^{(n)}) = n^{-1} \sum_{i=1}^{n} X_i = \bar{X}$$

vs.

$$q(x^{(n)}) = n^{-1} \sum_{i=1}^{n} x_i = \bar{x}$$

Point estimators and estimates

Suppose we want to estimate $\theta = \mu$ (population mean) based on a random sample.

- if we don't know anything about the sampling model, natural choice: analogue estimator (sample analogue of population mean is sample mean)

$$\hat{\mu} = \bar{X}$$

- if we know (or assume) sampling model, may use MLE
- if we know (or assume) sampling model and prior, may use Bayesian estimator

Point estimates are the values of point estimators computed from a single observed sample.

Estimators are random variables; estimates are not.

Sampling distribution of estimator

- view θ as fixed; estimate it with point estimator, say $\hat{\theta} = q(X^{(n)})$
- imagine observing sample 1, calculating estimate $\hat{\theta^{[1]}} = q(x_1^{(n)})$
- imagine repeating the sampling procedure, get sample 2, $\hat{\theta^{[2]}} = q(x_2^{(n)})$
- do this a huge number of times, obtain $\{\theta^{[1]}, \theta^{[2]}, \theta^{[3]}, \ldots\}$
- compute *frequency* of each estimate $\hat{\theta} \Rightarrow$ sampling distribution of $\hat{\theta} = q(X^{(n)})$

Why Interval Estimators?

Point estimators not very useful by themselves.

- they tell us nothing about *precision* of the estimator
- to test hypotheses we need to know about precision

Example: observe $(x_1, x_2, x_3) = (4, 286, 10)$

- $\bar{x} = 300/3 = 100$
- $s^2 = (n-1)^{-1} \sum_{i=1}^{3} (x_i - \bar{x})^2 = 25,956$; $s = 161.11$

Also, does $\hat{\mu} = 100$ say anything about how probable it is that $\mu = 101$?

Frequentist Interval Estimators

A confidence interval is a *random interval*, i.e. it is a function of the random sample $X^{(n)}$.

- for a chosen $\alpha \in (0, 1)$, a $100(1-\alpha)$ **confidence interval** for θ, $L_F(X^{(n)}) \leq \theta \leq U_F(X^{(n)})$, has property that, ***whatever the true value of*** θ ***is***,

$$\Pr_\theta(L_F(X^{(n)}) \leq \theta \leq U_F(X^{(n)})) = 1 - \alpha$$

- $\Pr_\theta$: probability *under repeated sampling of* X, i.e. the frequentist probability under *hypothetically* repeatedly drawing samples $(X_1, \ldots, X_n)$ from the population distribution of X
- $X \sim f(x|\theta)$
- typically $X_i \stackrel{iid}{\sim} f(x|\theta)$, $i = 1, \ldots, n$
- L and U are random variables; θ is viewed as fixed

Bayesian Interval Estimators

A credible interval is also a *random interval*,
$L_B(X^{(n)}) \leq \theta \leq U_B(X^{(n)})$.

- for a chosen $\alpha \in (0,1)$, a $100(1-\alpha)$ **credible interval** for θ has the property that

$$\Pr_{\theta|X^{(n)}}(L_B(X^{(n)}) \leq \theta \leq U_B(X^{(n)}) \,|X^{(n)} = x^{(n)}) = 1 - \alpha$$

- $\Pr_{\theta|X^{(n)}}$ is the posterior probability, given the observed data, i.e. given that $X^{(n)} = x^{(n)}$
- when you condition on $X^{(n)} = x^{(n)}$, L_B and U_B are no longer random
- a Bayesian instead views θ as random

Interpretations

Confidence interval:

- if we were to repeat the sampling procedure many times, each time taking a sample of size n and computing the values of L_F and U_F, then the *frequency* with which θ is in the computed intervals is $1 - \alpha$
- it is a probability statement about the procedure; the probability, under repeated sampling, that the procedure produces a correct interval is $1 - \alpha$
- an interval from any particular sample is not random; it is either correct or incorrect, no probability involved

Credible interval:

- given the observed data, the probability that $\theta \in [L_B, U_B]$ is $1 - \alpha$
- it is a probability statement about the interval computed from a particular sample

Options for Credible Intervals

- narrowest interval having posterior probability $1 - \alpha$: **highest posterior density (HPD) interval**
- **equal-tailed interval** : $\Pr_{\theta|X^{(n)}}(\theta < L_B) = \Pr_{\theta|X^{(n)}}(\theta > U_B)$
- interval with posterior mean as center

Let X be duration (in minutes), that a Windows operating system will function before crashing. Want to learn about $E(X)$

- assume $X \sim \text{Exp}(\theta)$, $p(x|\theta) = \theta e^{-\theta x}$, $x > 0$, $\theta > 0$; sample $X^{(n)}$; $E(X) = 1/\theta$
- specify prior $p(\theta) = 1/\theta$
- posterior is

$$\pi(\theta|x^{(n)}) \propto p(\theta)L(\theta) = \theta^{-1}\theta^{n} \exp[-\theta \sum_{i=1}^{n} x_i]$$

Notice that $\theta|x^{(n)} \sim \text{Gamma}(\theta|n, \sum_{i=1}^{n} x_i)$ (rate is second parameter). With normalizing constant it is

$$\pi(\theta|x^{(n)}) = (\sum_{i=1}^{n} x_i)^{n}(\Gamma(n))^{-1}\theta^{n-1} \exp[-\theta \sum_{i=1}^{n} x_i]$$

HPD for θ

```
require(TeachingDemos)
```

```
## Loading required package: TeachingDemos
```

```
vals <- rgamma(100000,90,60)
emp.hpd(vals, conf=0.5)
```

```
## [1] 1.379540 1.590958
```

```
emp.hpd(vals, conf=0.95)
```

```
## [1] 1.198028 1.814985
```

```
emp.hpd(vals, conf=0.99999)
```

```
## [1] 0.9331015 2.3116640
```

Binomial example (from Charlie Geyer)

```r
x <- 17 ; n <- 22
alpha1 <- 1 / 2 ; alpha2 <- 1 / 2
pmean <- (x + alpha1) / (n + alpha1 + alpha2)
pmedian <- qbeta(0.5, x + alpha1, n - x + alpha2)
pmode <- (x + alpha1 - 1) / (n + alpha1 + alpha2 - 2)
pmode <- max(0, pmode)
pmode <- min(1, pmode)
print(pmean) ; print(pmedian) ; print(pmode)
```

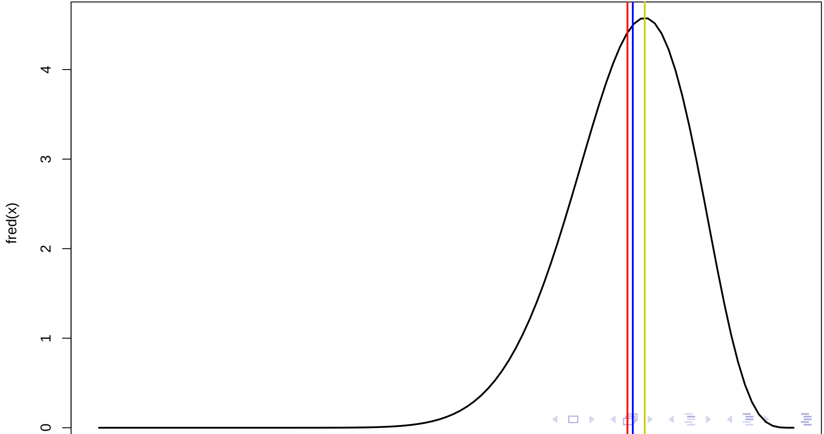
```
## [1] 0.7608696
```

```
## [1] 0.7685261
```

```
## [1] 0.7857143
```

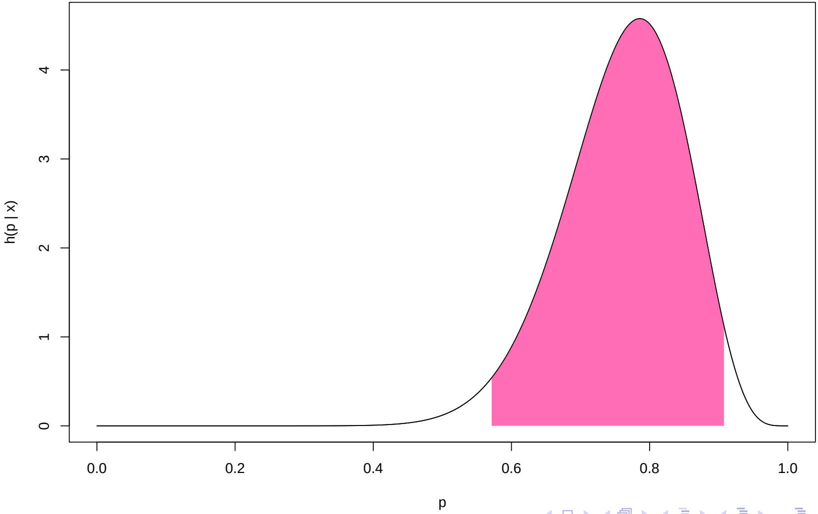
Plot

```
fred <- function(p) dbeta(p, x + alpha1, n - x + alpha2)
curve(fred, from = 0, to = 1, lwd = 2)
abline(v = pmean, col = "red", lwd = 2)
abline(v = pmedian, col = "blue", lwd = 2)
abline(v = pmode, col = "yellow3", lwd = 2)
```



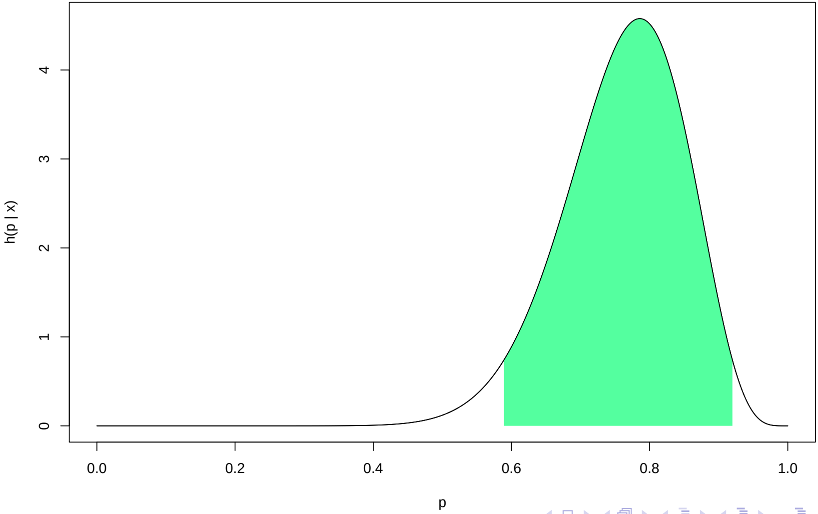
Equal-Tailed Interval

```
## [1] 0.5714 0.9076
```

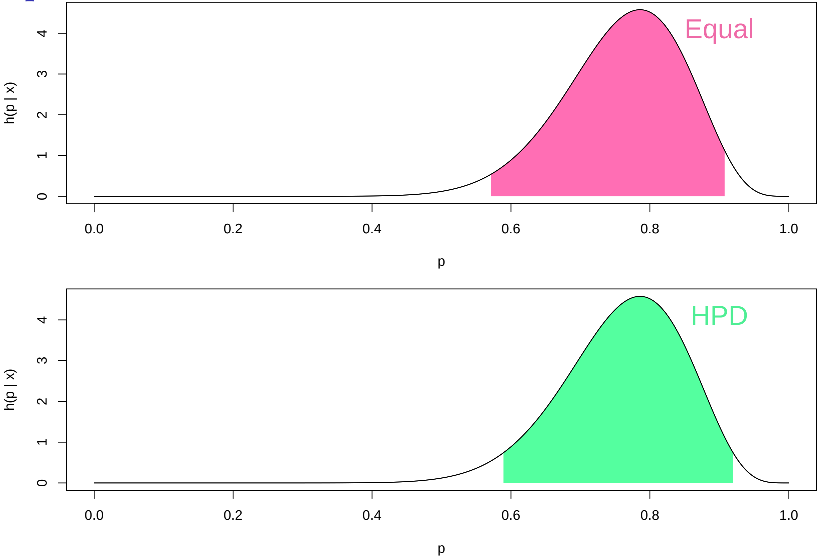


HPD Interval

```
## [1] 0.5892 0.9198
```

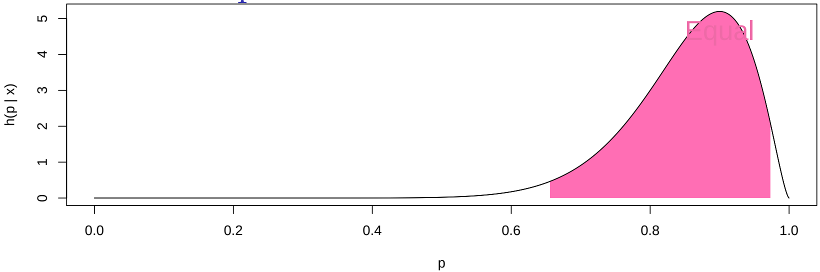


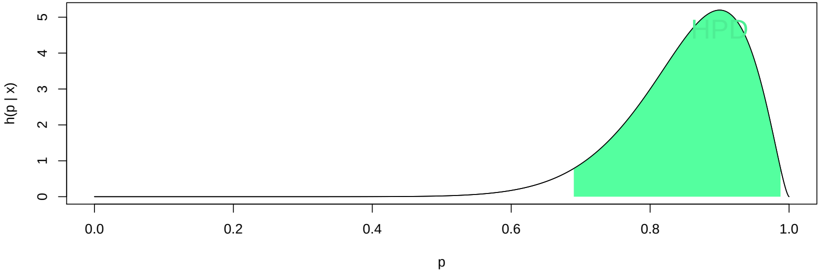
Comparison




```
## [1] 0.5714 0.9076
```

More skewed comparison






```
## [1] 0.6558 0.9731
```

Binomial with Beta Prior revisited

Suppose we use $\bar{x} = \sum_{i=1}^{n} x_i$ as an estimate for p.

- we know by CLT that as $n \to \infty$,

$$\Pr_p(\bar{X} - 1.96\sqrt{[\bar{X}(1-\bar{X})/n]} \le p \le \bar{X} + 1.96\sqrt{[\bar{X}(1-\bar{X})/n]}) = 0.95$$

Let's compare confidence intervals with credible intervals.

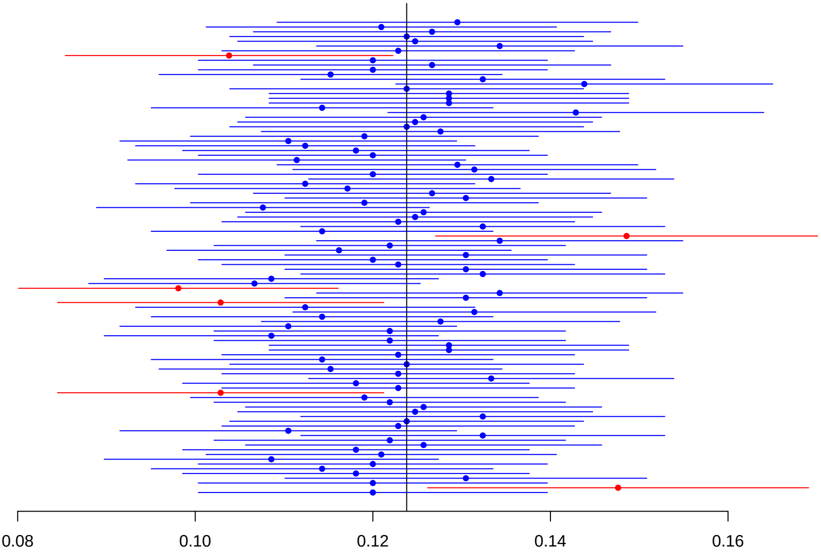
Generating samples, assume p=empirical (130/1050)

```r
xbar <- 130  ; n <- 1050 ; ns <- 100
M <- matrix(rbinom(n*ns, size=1, prob=xbar/n), nrow=n)
```

Compute mean, and confidence intervals for all ns samples

```
fIC=function(x)
  mean(x)+c(-1,1)*1.96*sqrt(mean(x)*(1-mean(x)))/sqrt(n)
IC=t(apply(M,2,fIC))
MN=apply(M,2,mean)
```

Plot



Equal-Tailed Credible Intervals