Statistical Inference

Lecture 09b

ANU - RSFAS

Last Updated: Tue May 2 11:17:54 2017

Beyond Point Estimation - Interval Estimation

- Never be satisfied with a point estimate! We want to know something about the uncertainty!
- This leads to interval estimation.
- Construction methods for interval estimates:
  - parametric "exact" intervals
  - parametric asymptotic intervals
  - Bayesian intervals
  - non-parametric intervals
- Some general approaches (mostly for the frequentist parametric case(s)):
  - Inverting a test statistic
  - Pivotal Quantities
  - Pivoting the CDF

Interval Estimation

Definition: An interval estimate of a real-valued parameter θ is any pair of functions $L(\boldsymbol{x})$ and $U(\boldsymbol{x})$, of a sample that satisfy

$$L(\boldsymbol{x}) \leq U(\boldsymbol{x}) \quad \forall\, \boldsymbol{x} \in \mathcal{X}.$$

- If $\boldsymbol{X} = \boldsymbol{x}$ is observed, the inference $L(\boldsymbol{x}) \leq \theta \leq U(\boldsymbol{x})$ is made.
- The random interval $[L(\boldsymbol{X}), U(\boldsymbol{X})]$ is called an interval estimator.
- We can have one-sided estimates:

$$(-\infty, U(\boldsymbol{x})] \;\Rightarrow\; \theta \leq U(\boldsymbol{x})$$
$$[L(\boldsymbol{x}), \infty) \;\Rightarrow\; \theta \geq L(\boldsymbol{x})$$

Interval Estimation

- There is a strong relationship between hypothesis testing and interval estimation. In general, every confidence set corresponds to a test and vice versa.

Eg. $X_1, \ldots, X_n \overset{\text{iid}}{\sim} \text{normal}(\mu, \sigma^2)$, where σ^2 is known. Consider testing:

$$H_0 : \mu = \mu_0 \quad vs \quad \mu \neq \mu_0$$

$$R = \left\{ \left| \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} \right| \geq z_{\alpha/2} \right\}$$

- Now we know that under H_0 $P(R) = \alpha$. So the probability that H_0 is accepted is $1 - \alpha$:

$$P\left(-z_{\alpha/2} \leq \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} \leq z_{\alpha/2} \right) = 1 - \alpha$$

Interval Estimation

- Now fix α and determine the acceptance region. This is an interval estimator.

$$
\begin{aligned}
P\left(-z_{\alpha/2} \leq \frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \leq z_{\alpha/2}\right) &= 1-\alpha \\
P\left(-z_{\alpha/2}\left(\sigma/\sqrt{n}\right) \leq \bar{X}-\mu \leq z_{\alpha/2}\left(\sigma/\sqrt{n}\right)\right) &= 1-\alpha \\
P\left(-\bar{X}-z_{\alpha/2}\left(\sigma/\sqrt{n}\right) \leq -\mu \leq -\bar{X}+z_{\alpha/2}\left(\sigma/\sqrt{n}\right)\right) &= 1-\alpha \\
P\left(\bar{X}+z_{\alpha/2}\left(\sigma/\sqrt{n}\right) \geq \mu \geq \bar{X}-z_{\alpha/2}\left(\sigma/\sqrt{n}\right)\right) &= 1-\alpha \\
P\left(\bar{X}-z_{\alpha/2}\left(\sigma/\sqrt{n}\right) \leq \mu \leq \bar{X}+z_{\alpha/2}\left(\sigma/\sqrt{n}\right)\right) &= 1-\alpha
\end{aligned}
$$

Interval Estimation

- A $100(1-\alpha)\%$ confidence estimator for μ is:

$$[\bar{X} - z_{\alpha/2}\ (\sigma/\sqrt{n}),\ \bar{X} + z_{\alpha/2}\ (\sigma/\sqrt{n})]$$

- Remember, $\boldsymbol{X}$ is random not μ!!

Interval Estimation

Theorem: For each $\theta_0 \in \Theta$, let $A(\theta_0)$ be the acceptance region of a level α test of $H_0 : \theta = \theta_0$.

- For each $\boldsymbol{x} \in \boldsymbol{X}$, define a set $C(\boldsymbol{x})$ in the parameter space by

$$C(\boldsymbol{x}) = \{\theta_0 : \boldsymbol{x} \in A(\theta_0)\}$$

The random set $C(\boldsymbol{X})$ is a $1 - \alpha$ confidence set.

- Conversely, let $C(\boldsymbol{X})$ be a $1 - \alpha$ confidence set for any $\theta \in \Theta_0$,

$$A(\theta_0) = \{\boldsymbol{x} : \theta_0 \in C(\boldsymbol{x})\}$$

Then $A(\theta_0)$ be the acceptance region of a level α test of $H_0 : \theta = \theta_0$

$$A(\theta_0) \Longleftrightarrow C(\boldsymbol{x})$$

Interval Estimation

Eg. $X_1, \ldots, X_n \overset{\text{iid}}{\sim} \text{normal}(\mu, \sigma^2)$, where σ^2 is unknown. Consider testing:

$$H_0 : \mu = \mu_0 \quad vs \quad \mu \leq \mu_0$$

- Based on a Likelihood Ratio Test we can find a rejection region of:

$$R = \left\{ \frac{\bar{X} - \mu_0}{s/\sqrt{n}} \leq -t_{n-1,\alpha} \right\}$$

Interval Estimation

- This leads to an acceptance region of:

$$
\begin{aligned}
A(\mu_0) &= \left\{ \boldsymbol{x} : \frac{\bar{X} - \mu_0}{s/\sqrt{n}} \geq -t_{n-1,\alpha} \right\} \\
&= \{ \bar{x} \geq \mu_0 - t_{n-1,\alpha}\, (s/\sqrt{n}) \}
\end{aligned}
$$

- This leads to a $(1-\alpha)$ upper bound confidence set for μ :

$$
\begin{aligned}
C(\boldsymbol{x}) &= \{ \mu_0 : \bar{x} + t_{n-1,\alpha}\, (s/\sqrt{n}) \geq \mu_0 \} \\
&= (-\infty, \bar{x} + t_{n-1,\alpha}\, (s/\sqrt{n})]
\end{aligned}
$$

Interval Estimation

Example: Suppose that 2.6, 1.2 and 4.9 are a random sample from a normal distribution whose mean is zero and whose variance σ^2 is unknown. Derive and compute a central 99% confidence interval for σ^2.

- Approach 1:

$$\left(\frac{X_i - \mu}{\sigma}\right)^2 = \left(\frac{X_i}{\sigma}\right)^2 \sim Z^2 = \chi^2_1$$

$$\sum_{i=1}^{3}\left(\frac{X_i}{\sigma}\right)^2 = \frac{1}{\sigma^2}\sum_{i=1}^{3} X_i^2 \sim \chi^2_3$$

- Let $Y = \sum_{i=1}^{3} X_i^2$.

$$
\begin{aligned}
P\left(\chi^2_{1-\alpha/2,3} \le \frac{Y}{\sigma^2} \le \chi^2_{\alpha/2,3}\right) &= 1-\alpha \\
P\left(\frac{1}{\chi^2_{1-\alpha/2,3}} \ge \frac{\sigma^2}{Y} \ge \frac{1}{\chi^2_{\alpha/2,3}}\right) &= 1-\alpha \\
P\left(\frac{Y}{\chi^2_{\alpha/2,3}} \le \sigma^2 \le \frac{Y}{\chi^2_{1-\alpha/2,3}}\right) &= 1-\alpha
\end{aligned}
$$

$$
\left[\frac{Y}{\chi^2_{\alpha/2,3}} \quad , \quad \frac{Y}{\chi^2_{1-\alpha/2,3}}\right]
$$

$$
\left[\frac{32.21}{12.8381} \quad , \quad \frac{32.21}{0.0717212}\right]
$$

$$
[2.51 \quad , \quad 449]
$$

- Note: R does probability to the left for quantiles while C&B does probability to the right.

```
qchisq(0.01/2, 3)
```

```
## [1] 0.07172177
```

```
qchisq(1-0.01/2, 3)
```

```
## [1] 12.83816
```

- Approach 2:

$$\frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{n-1}$$

$$\begin{aligned} P\left(\chi^2_{1-\alpha/2,2} \leq \frac{(n-1)S^2}{\sigma^2} \leq \chi^2_{\alpha/2,2}\right) &= 1-\alpha \\ P\left(\frac{(n-1)S^2}{\chi^2_{\alpha/2,2}} \leq \sigma^2 \leq \frac{(n-1)S^2}{\chi^2_{1-\alpha/2,2}}\right) &= 1-\alpha \end{aligned}$$

$$\left[\frac{(n-1)S^2}{\chi^2_{\alpha/2,2}} \quad , \quad \frac{(n-1)S^2}{\chi^2_{1-\alpha/2,2}}\right]$$
$$\left[\frac{(2)3.49}{10.5966} \quad , \quad \frac{(2)3.49}{0.0100251}\right]$$
$$[0.659 \quad , \quad 696]$$

- Approach 3:

$$\left(\frac{\bar{X}-\mu}{\sigma/\sqrt{n}}\right)^2 = \left(\frac{\bar{X}}{\sigma/\sqrt{n}}\right)^2 = \frac{n\bar{X}^2}{\sigma^2} \sim Z^2 = \chi^2_1$$

$$\begin{aligned} P\left(\chi^2_{1-\alpha/2,1} \leq \frac{n\bar{X}^2}{\sigma^2} \leq \chi^2_{\alpha/2,1}\right) &= 1-\alpha \\ P\left(\frac{n\bar{X}^2}{\chi^2_{\alpha/2,1}} \leq \sigma^2 \leq \frac{n\bar{X}^2}{\chi^2_{1-\alpha/2,1}}\right) &= 1-\alpha \end{aligned}$$

$$\left[\frac{n\bar{X}^2}{\chi^2_{\alpha/2,1}} \quad , \quad \frac{n\bar{X}^2}{\chi^2_{1-\alpha/2,1}}\right]$$

$$\left[\frac{(3)2.9^2}{7.87944} \quad , \quad \frac{(3)2.9^2}{0.0000393}\right]$$

$$[3.202 \quad , \quad 642468.3]$$

Interval Estimation

- All three approaches, and everything we have considered thus far have a nice property. The distribution of the statistic does not contain parameters!

Definition: A random variable $Q(\boldsymbol{X}, \theta)$ is a pivotal quantity (or a pivot) if the distribution of $Q(\boldsymbol{X}, \theta)$ is independent of all parameters.

- If $\boldsymbol{X} \sim f(\boldsymbol{x}|\theta)$ then $Q(\boldsymbol{X}, \theta)$ has the same distribution for all values of θ.

Interval Estimation - MLEs & Asymptotics

$$\hat{\theta} \stackrel{\cdot}{\sim} \mathrm{normal}(\theta, I(\theta)^{-1})$$

$$\frac{\hat{\theta} - \theta}{1/\sqrt{I(\theta)}} \stackrel{\cdot}{\sim} \mathrm{normal}(0, 1)$$

- We have a pivotal quantity. Based on the same approach as before we can construct an asymptotic $100(1-\alpha)\%$ confidence interval as:

$$\left[\hat{\theta} - z_{\alpha/2} \frac{1}{\sqrt{I(\hat{\theta})}}\ ,\ \hat{\theta} + z_{\alpha/2} \frac{1}{\sqrt{I(\hat{\theta})}}\right]$$

Interval Estimation - MLEs & Asymptotics

- If we are interested in a function of θ, say $\tau(\theta)$, then we have:

$$\tau(\hat{\theta}) \dot{\sim} \mathrm{normal}\left(\tau(\theta), \frac{[\tau'(\theta)]^2}{I(\theta)}\right)$$

$$\frac{\tau(\hat{\theta}) - \tau(\theta)}{\sqrt{\frac{[\tau'(\theta)]^2}{I(\theta)}}} \dot{\sim} \mathrm{normal}(0, 1)$$

- We can construct an asymptotic $100(1-\alpha)\%$ confidence interval as:

$$\left[\tau(\hat{\theta}) - z_{\alpha/2}\frac{\tau'(\hat{\theta})}{\sqrt{I(\hat{\theta})}}\ ,\ \tau(\hat{\theta}) + z_{\alpha/2}\frac{\tau'(\hat{\theta})}{\sqrt{I(\hat{\theta})}}\right]$$

Interval Estimation - MLEs & Asymptotics

Example: Let $X_1, \ldots, X_n \stackrel{\text{iid}}{\sim} \text{exponential}(\theta)$:

$$f(x|\theta) = \theta \exp(-\theta x)$$

- Provide an equal tailed 95% CI for $\tau(\theta) = \theta^{-1}$.

$$\begin{aligned}
\ell &= nlog(\theta) - \theta \sum x_i \\
\ell' &= \frac{n}{\theta} - \sum x_i \\
&\quad \Rightarrow \frac{n}{\theta} - \sum x_i = 0 \\
&\quad \hat{\theta} = \frac{1}{\bar{x}} \quad \Rightarrow \widehat{\left(\frac{1}{\theta}\right)} = \frac{1}{\hat{\theta}} = \bar{x}
\end{aligned}$$

Interval Estimation - MLEs & Asymptotics

$$
\begin{aligned}
\ell'' &= -\frac{n}{\theta^2} \\
\text{Fisher Information: } I(\theta) &= -E\left[-\frac{n}{\theta^2}\right] = \frac{n}{\theta^2} \\
CRLB(\theta^{-1}) &= \frac{\left[\frac{d}{d\theta}\frac{1}{\theta}\right]^2}{\frac{n}{\theta^2}} = \frac{\left[-\frac{1}{\theta^2}\right]^2}{\frac{n}{\theta^2}} = \frac{1}{n\theta^2} \\
CRLB(\hat{\theta}^{-1}) &= \frac{1}{n\hat{\theta}^2} = \frac{\bar{x}^2}{n}
\end{aligned}
$$

- We end with the following interval for $\frac{1}{\theta}$:

$$
\left[\bar{x} - z_{\alpha/2}\frac{\bar{x}}{\sqrt{n}} \ , \ \bar{x} + z_{\alpha/2}\frac{\bar{x}}{\sqrt{n}}\right]
$$

Interval Estimation - MLEs & Asymptotics

- Note: $\tau(\hat{\theta}) = \bar{X}$, so why not use the following interval?

$$\left[\bar{x} - z_{\alpha/2}\frac{s}{\sqrt{n}} \ , \ \bar{x} + z_{\alpha/2}\frac{s}{\sqrt{n}}\right]$$

- If the data truly are exponentially distributed, then the previous interval will be more accurate.
- Of course, the this interval will be valid even in the case that the data are not truly exponentially distributed.

Interval Estimation - MLEs & Asymptotics

- Now suppose we are interested in a CI for θ:
- We constructed an interval $\tau = \frac{1}{\theta}$, so why not just take the the inverse? We can.

$$[u^{-1}, l^{-1}]$$

- So we have for θ:

$$\left[\left\{\bar{x} + z_{\alpha/2}\frac{\bar{x}}{\sqrt{n}}\right\}^{-1}, \left\{\bar{x} - z_{\alpha/2}\frac{\bar{x}}{\sqrt{n}}\right\}^{-1}\right]$$

Interval Estimation - MLEs & Asymptotics

- OK, but let's go back to the drawing-board and find the CI for θ from first principles:

$$\left[\bar{x}^{-1} - z_{\alpha/2}\frac{1}{\sqrt{n}\bar{X}}\ ,\ \bar{x}^{-1} + z_{\alpha/2}\frac{1}{\sqrt{n}\bar{x}}\right]$$

- We see that the two approaches are not the same. This is because interval construction, as we have done it, is not functionally equivalent!!

Interval Estimation

- Can we come up with an approach which does possess the equivariance property?
  - Yes, as long as the functional transformation in question is invertible.
  - Let's consider an asymptotic likelihood-based confidence interval procedure which is parameterization equivariant.
  - Specifically, this means that if we find a confidence region, C, for θ based on this new procedure and transform all of its values [which we sometimes denote as $\tau(C) = \{\tau(\theta) : \theta \in C\}$] then we will arrive at the same confidence region as if we had applied our new procedure to the parameter τ directly.

Asymptotic Maximimum LRT Interval Estimation

- Let's consider the following based on the maximum likelihood ratio test, where $X_1, \ldots, X_n \overset{\text{iid}}{\sim} \text{exponential}(\theta)$; $f(x|\theta) = \theta \exp(-\theta x)$:

$$-2log\left(\frac{L(\theta|\boldsymbol{x})}{L(\hat{\theta}|\boldsymbol{x})}\right) \dot{\sim} \chi^2_1$$

$$\begin{aligned}
-2[\ell(\theta|\boldsymbol{x}) - \ell(\hat{\theta}|\boldsymbol{x})] &= 2[\ell(\hat{\theta}|\boldsymbol{x}) - \ell(\theta|\boldsymbol{x})] \\
&= 2[nlog(\hat{\theta}) - \hat{\theta}\sum x_i - nlog(\theta) + \theta \sum x_i] \\
&= 2[nlog(\frac{1}{\bar{x}}) - \frac{1}{\bar{x}}\sum x_i - nlog(\theta) + \theta \sum x_i] \\
&= -2n\ log(\bar{x}) - 2n\frac{\bar{x}}{\bar{x}} - 2nlog(\theta) + 2\theta n\bar{x} \\
&= -2n\ log(\bar{x}\theta) + 2n(\theta\bar{x} - 1)
\end{aligned}$$

Asymptotic Maximimum LRT Interval Estimation

- We reject if:

$$-2n\ log(\bar{x}\theta) + 2n(\theta\bar{x} - 1) > \chi^2_{\alpha,1}$$

- We accept if:

$$-2n\ log(\bar{x}\theta) + 2n(\theta\bar{x} - 1) \leq \chi^2_{\alpha,1}$$

- So our confidence set is:

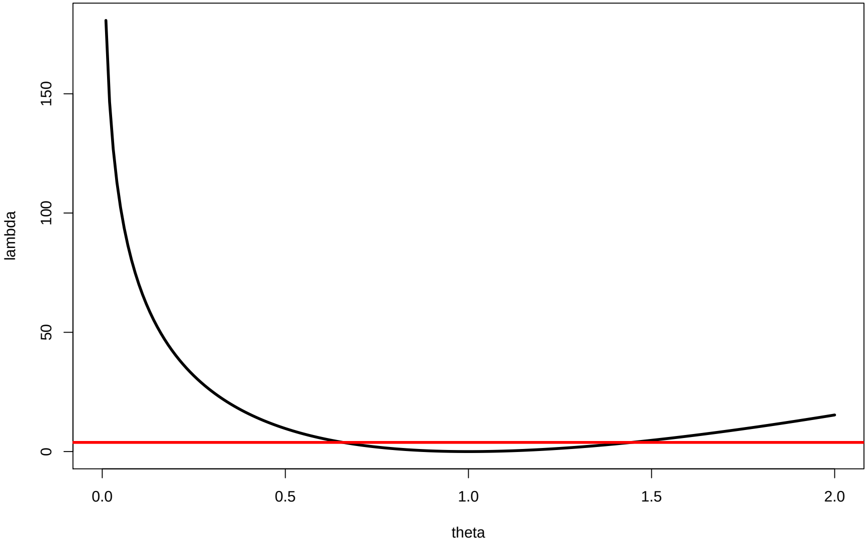
$$\begin{aligned} C &= \left\{\theta \in \Theta : -2[\ell(\theta) - \ell(\hat{\theta})] \leq \chi^2_{\alpha,1}\right\} \\ &= \left\{-2n\ log(\bar{x}\theta) + 2n(\theta\bar{x} - 1) \leq \chi^2_{\alpha,1}\right\} \end{aligned}$$

Asymptotic Maximimum LRT Interval Estimation

- We can't solve this analytically, but let's graph it:
- Suppose $\bar{X} = 1$, $n = 25$, and $\alpha = 0.05$:

```
x.bar <- 1
n <- 25

theta <- seq(0,2, by =0.01)
lambda <- -2*n*log(x.bar*theta) + 2*n*(theta*x.bar - 1)
plot(theta, lambda, lwd=3, type="l")
abline(h=qchisq(1-0.05, 1), lwd=3, col="red")
```

Asymptotic Maximimum LRT Interval Estimation



Asymptotic Maximimum LRT Interval Estimation

```
min( theta[lambda <= qchisq(1-0.05, 1)])
```

```
## [1] 0.66
```

```
max( theta[lambda <= qchisq(1-0.05, 1)])
```

```
## [1] 1.44
```

- So a 95% confidence interval for θ is:

$$[0.66\ ,\ 1.44]$$

Asymptotic Maximimum LRT Interval Estimation

- Now suppose we want the interval for $\tau = \frac{1}{\theta}$.
- Let's reparametrize the log likelihood:

$$\begin{aligned} \ell(\tau) &= \ell(\theta = \tau^{-1}) = nlog(\tau^{-1}) - \tau^{-1} \sum x_i \\ &= -nlog(\tau) - n\frac{\bar{x}}{\tau} \end{aligned}$$

Asymptotic Maximimum LRT Interval Estimation

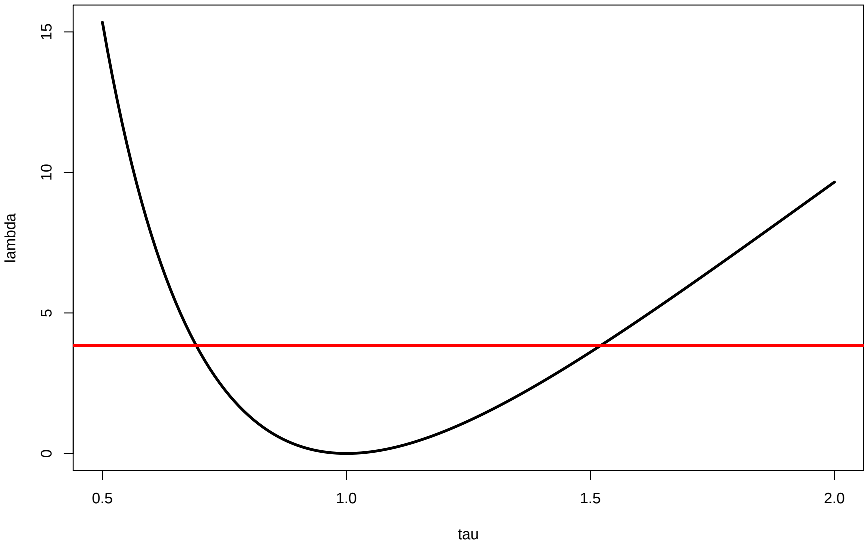
$$
\begin{aligned}
-2[\ell(\tau) - \ell(\hat{\tau})] &= 2[\ell(\hat{\tau}) - \ell(\tau)] \\
&= 2[-nlog(\hat{\tau}) - n\frac{\bar{x}}{\hat{\tau}} + nlog(\tau) + n\frac{\bar{x}}{\tau}] \\
&= 2[-nlog(\bar{x}) - n\frac{\bar{x}}{\bar{x}} + nlog(\tau) + n\frac{\bar{x}}{\tau}] \\
&= -2nlog(\bar{x}\tau^{-1}) + 2n(\tau^{-1}\bar{x} - 1)
\end{aligned}
$$

- All that was done through all the math was to replace θ with τ^{-1}!
- So our interval is:

$$[1/1.44\ ,\ 1/0.66] = [0.69\ ,\ 1.51]$$

- Let's see it in the plot
- Again, suppose $\bar{X} = 1$, $n = 25$, and $\alpha = 0.05$:

```
x.bar <- 1
n <- 25

tau <- seq(0.5, 2, by =0.01)
lambda <- -2*n*log(x.bar*(1/tau)) + 2*n*((1/tau)*x.bar - 1)
plot(tau, lambda, lwd=3, type="l")
abline(h=qchisq(1-0.05, 1), lwd=3, col="red")
```



lambda
tau
0
5
10
15
0.5
1.0
1.5
2.0

```r
min( tau[lambda <= qchisq(1-0.05, 1)])
```

```
## [1] 0.7
```

```r
max( tau[lambda <= qchisq(1-0.05, 1)])
```

```
## [1] 1.52
```

Maximimum LRT Interval Estimation

- Did we have to use the asymptotic result of the LRT for our interval. No, but it is more straightforward.

Interval Estimation - CDF Method

- Pivoting the CDF
  - A pivot Q leads to a confidence set:

$$C(\boldsymbol{x}) = \{\theta_0 : a \leq Q(\boldsymbol{x}; \theta_0) \leq b\}$$

  - If for every $\boldsymbol{x}$ the pivot is a monotone function of θ then the confidence set $C(\boldsymbol{x})$ is guaranteed to be an interval.
  - Most pivots we have considered have this property.

Interval Estimation - CDF Method

Theorem:

- Let T be a statistic with a continuous cdf $F_T(t|\theta)$ [Note: We can also work with discrete distributions - see C&B].
- Let $\alpha_1 + \alpha_2 = \alpha$ with $0 < \alpha < 1$.
- Suppose that for each $t \in T$, the functions $\theta_L(t)$ and $\theta_U(t)$ can be defined as:

1. If $F_T(t|\theta)$ is a decreasing function of θ for each t, define $\theta_L(t)$ and $\theta_U(t)$ by:

$$F_T(t|\theta_U(t)) = \alpha_1 \qquad F_T(t|\theta_L(t)) = 1 - \alpha_2$$

2. If $F_T(t|\theta)$ is an increasing function of θ for each t, define $\theta_L(t)$ and $\theta_U(t)$ by:

$$F_T(t|\theta_L(t)) = \alpha_1 \qquad F_T(t|\theta_U(t)) = 1 - \alpha_2$$

Then the interval $[\theta_L(t), \theta_U(t)]$ is a $1 - \alpha$ confidence interval for θ.

- We can prove that $F_T(t|\theta)$ is monotone in θ. See C&B.

Interval Estimation - CDF Method

Example: Consider $X_1, \ldots, X_n \stackrel{\text{iid}}{\sim} Unif(0, \theta)$.

- So we have the following CDF for X:

$$F_X(x|\theta) = \frac{x}{\theta}\mathbb{I}_{(0 \leq x \leq \theta)}$$

- We know the MLE for θ is $T = max(X_1, \ldots, X_n)$

$$\begin{aligned} F_T(t|\theta) = Pr(T \leq t) &= Pr\{max(X_1, \ldots, X_n) \leq t\} \\ &= Pr\{X_1 \leq t, \ldots, X_n \leq t\} \\ &= Pr\{X_1 \leq t\} \times \ldots \times Pr\{X_n \leq t\} \\ &= \{F_X(t|\theta)\}^n \\ &= \frac{t^n}{\theta^n}\mathbb{I}_{(0 \leq t \leq \theta)} \end{aligned}$$

Interval Estimation - CDF Method

- Note: $F_T(t|\theta)$ is a deceasing function for θ. Let $\alpha_1 = \alpha_2 = \alpha/2$. We have:

$$\begin{aligned} F_T(t|\theta_U(t)) &= \alpha/2 \\ \left(\frac{t}{\theta_U}\right)^n &= \alpha/2 \\ \theta_U &= t(\alpha/2)^{-(1/n)} \end{aligned}$$

$$\begin{aligned} F_T(t|\theta_L(t)) &= 1-\alpha/2 \\ \left(\frac{t}{\theta_L}\right)^n &= 1-\alpha/2 \\ \theta_L &= t(1-\alpha/2)^{-(1/n)} \end{aligned}$$

Interval Estimation - CDF Method

```r
##
set.seed(1001)
n <- 15
X <- runif(n, 0, 10)
t <- max(X)
alpha <- 0.05

##
theta.u <- t*(alpha/2)^(-(1/n))
theta.l <- t*(1-alpha/2)^(-(1/n))

c(theta.l, theta.u)
```

```
## [1]  9.873539 12.605028
```

Interval Estimation - CDF Method

- Interpretation: Over repeated sampling, we expect 95% of the intervals we create to contain the true value θ.
- Let's check: We set $\alpha = 0.05$, so 95% of the intervals should contain θ.

Interval Estimation - CDF Method

```r
set.seed(1001)
##
S <- 10000
coverage <- rep(0, S)
theta.true <- 10

##
n <- 15
alpha <- 0.05

##
for(s in 1:S){

##
X <- runif(n, 0, theta.true)
t <- max(X)

##
theta.u <- t*(alpha/2)^(-(1/n))
theta.l <- t*(1-alpha/2)^(-(1/n))

if(theta.l < theta.true && theta.u > theta.true){coverage[s] <- 1}
}

mean(coverage)
```

```
## [1] 0.9517
```

Bayesian Interval Estimation

- Suppose we have data $X_1, \ldots, X_n$ from density $f_X(x|\theta)$ along with a prior distribution $\pi(\theta)$. As we saw we use Bayes' rule to update our 'beliefs' about θ once we observe the data:

$$\begin{aligned}\pi(\theta|\boldsymbol{x}) &= \frac{L(\theta|\boldsymbol{x})\pi(\theta)}{\int_{\theta\in\Theta} L(\theta|\boldsymbol{x})\pi(\theta)d\theta} \\ &= \frac{L(\theta|\boldsymbol{x})\pi(\theta)}{m(\boldsymbol{x})}\end{aligned}$$

- So we have the whole distribution for

$$\pi(\theta|\boldsymbol{x})$$

- This is different than the frequentist approach where find an estimator for θ, say $\hat{\theta}$ and then try to determine the distribution of $\hat{\theta}$.

Bayesian Interval Estimation

- To obtain an interval we simply consider:

$$P\pi(\theta|x)(C) = \int_C \pi(\theta|\boldsymbol{x})d\theta = 1 - \alpha$$

- Be careful. We are using α quite generically. Recall that α does have a formal definition: The probability of a Type-I error. This is based on repeated sampling. For the Bayesian case we only think about one data set an infinite number of possible data sets.
- There are quite a lot of choices for C. We will consider the 3 most common.

Bayesian Interval Estimation

- Equal tailed:

$$\int_{-\infty}^{\theta_L} \pi(\theta|\boldsymbol{x}) d\theta = \alpha/2, \quad \int_{\theta_U}^{\infty} \pi(\theta|\boldsymbol{x}) d\theta = \alpha/2$$

- Smallest length: We can choose C to minimize $\theta_U - \theta_L$.

Bayesian Interval Estimation

- Highest posterior density region (HPD): We define C to be that set with posterior probability $1-\alpha$ which satisfies the criterion:

$$\theta_1 \in C \quad \text{and} \quad \pi(\theta_2|\boldsymbol{x}) > \pi(\theta_1|\boldsymbol{x}) \Rightarrow \theta_2 \in C$$

C contains the values of θ which have the highest posterior density values, so that we can determine HPD regions as the set:

$$C = \{\theta \in \Theta : \pi(\theta|\boldsymbol{x}) > c_\alpha\}$$

- If the posterior is unimodal then this will be the smallest length interval!

Bayesian Interval Estimation

Example: $X_1, \ldots, X_n \overset{\text{iid}}{\sim} \text{exponential}(1/\theta)$ and $\pi(\theta) = \theta exp(-\theta)$.

$$
\begin{aligned}
\pi(\theta|\boldsymbol{x}) \quad &\propto \quad \left\{\prod_{i=1}^{n} \theta exp(-x_i\theta)\right\} \theta exp(-\theta) \\
&= \theta^n exp(-\sum x_i\theta)\theta exp(-\theta) \\
&= \theta^{n+1} exp(-\theta(n\bar{x}+1)) \\
&= \theta^{n+2-1} exp(-\theta(n\bar{x}+1))
\end{aligned}
$$

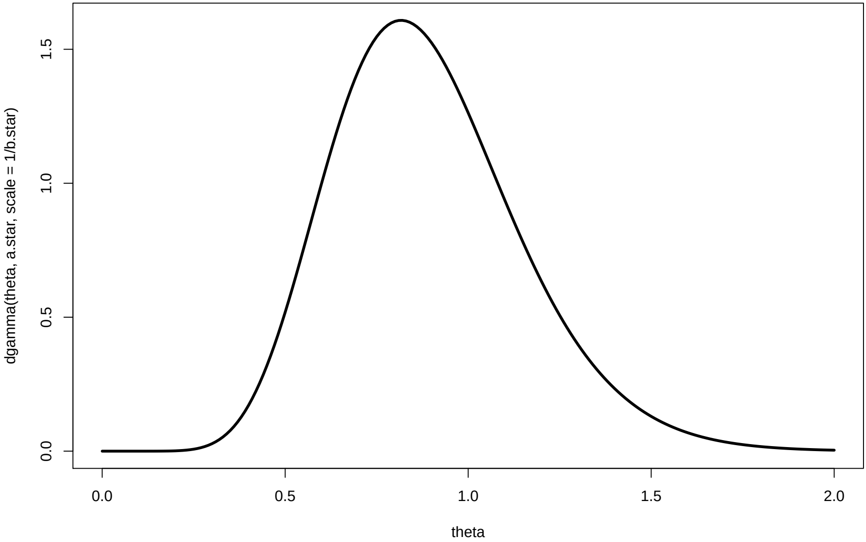
$$
[\theta|\boldsymbol{x}] \sim gamma\left(n+2, \frac{1}{n\bar{x}+1}\right)
$$

Bayesian Interval Estimation

- Let's plot the density for $n = 10$ and $\bar{x} = 1.247$.

```
n=10; x.bar <- 1.247
a.star <- n+2; b.star <- n*x.bar+1
theta <- seq(0, 2, by=0.01)
plot(theta, dgamma(theta, a.star,
                   scale=1/b.star), type="l", lwd=3)
```

Bayesian Interval Estimation



Bayesian Interval Estimation

- An equal-tailed 95% interval is given by $[\theta_l, \theta_u]$:

$$\begin{aligned} \int_0^{\theta_l} \pi(\theta|\boldsymbol{x}) &= 0.025 \\ F_{[\theta|\boldsymbol{x}]}(\theta_l) &= 0.025 \end{aligned}$$

$$\begin{aligned} \int_0^{\theta_u} \pi(\theta|\boldsymbol{x}) &= 1 - 0.025 = 0.975 \\ F_{[\theta|\boldsymbol{x}]}(\theta_u) &= 0.975 \end{aligned}$$

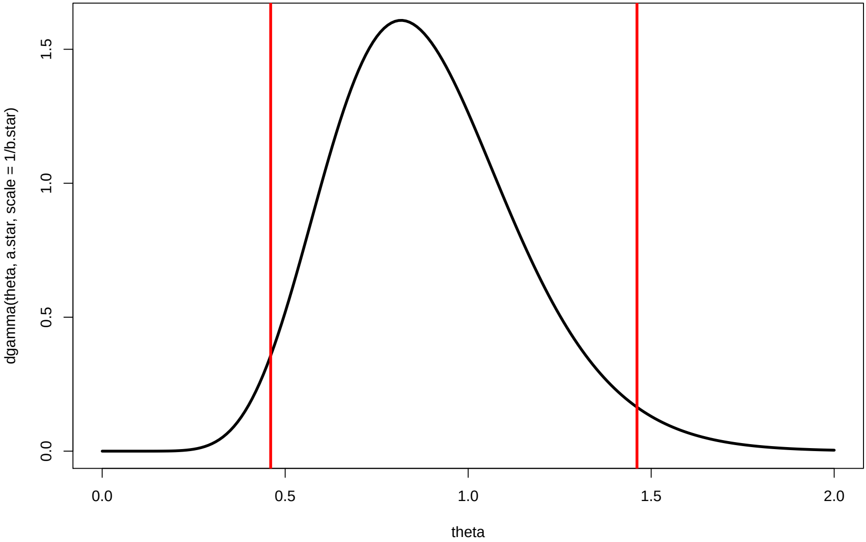
Bayesian Interval Estimation

```
theta.L <- qgamma(0.025,  a.star, scale=1/b.star)
theta.U <- qgamma(0.975,  a.star, scale=1/b.star)

c(theta.L, theta.U)
```

```
## [1] 0.4603248 1.4611758
```

```
plot(theta, dgamma(theta, a.star, scale=1/b.star), type="l", l
abline(v=c(theta.L, theta.U), lwd=3, col="red")
```

Bayesian Interval Estimation



Bayesian Interval Estimation

- If we only have tables in front of us, we can relate the gamma distribution to a χ^2 distribution as was discussed in tutorial the other week.
- If $[\theta|\boldsymbol{x}] \sim \mathrm{gamma}(a^*, b^*)$ then

$$\begin{aligned}\left[\frac{2\theta}{b^*}\middle|\boldsymbol{x}\right] &\sim \mathrm{gamma}(a^*, 2) \\ &\sim \chi^2_{p=2a^*}\end{aligned}$$

Bayesian Interval Estimation

- Using probabilities to the left. $p = 2a^* = 2n + 4$.

$$\left[\chi^2_{0.025,p} \le \left.\frac{2\theta}{b^*}\right|\boldsymbol{x} \le \chi^2_{0.975,p}\right]$$

$$\left[\chi^2_{0.025,p} \le 2\theta(n\bar{x}+1)|\boldsymbol{x} \le \chi^2_{0.975,p}\right]$$

$$\left[\frac{\chi^2_{0.025,p}}{2(n\bar{x}+1)} \le \theta|\boldsymbol{x} \le \frac{\chi^2_{0.975,p}}{2(n\bar{x}+1)}\right]$$

```
p <- 2*n + 4
theta.L <- qchisq(0.025, p)/(2*(n*x.bar+1))
theta.U <- qchisq(0.975, p)/(2*(n*x.bar+1))

c(theta.L, theta.U)
```

```
## [1] 0.4603248 1.4611758
```

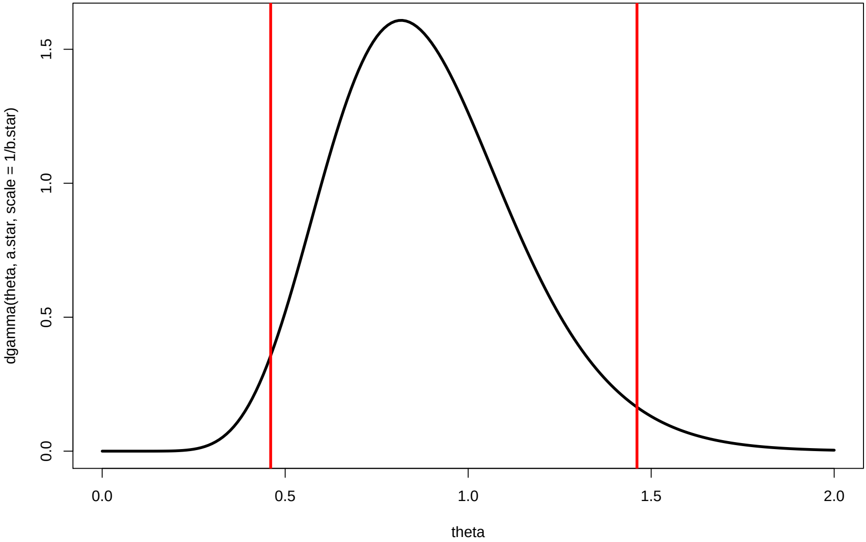
Bayesian Interval Estimation

- Is the interval $[0.4603, 1.4612]$ a HPD (highest posterior density) interval (the posterior is unimodal)?
- Recall:

$$\theta_1 \in C \quad \text{and} \quad \pi(\theta_2|\boldsymbol{x}) > \pi(\theta_1|\boldsymbol{x}) \Rightarrow \theta_2 \in C$$

- Let's see the density with the equal-tailed interval again.

Bayesian Interval Estimation



Bayesian Interval Estimation

- Note that the density seems to be higher for $\theta = 0.40$ than $\theta = 1.4612$:

```
dgamma(0.4, a.star, scale=1/b.star)
```

```
## [1] 0.1713707
```

```
dgamma(1.4612, a.star, scale=1/b.star)
```

```
## [1] 0.1641042
```

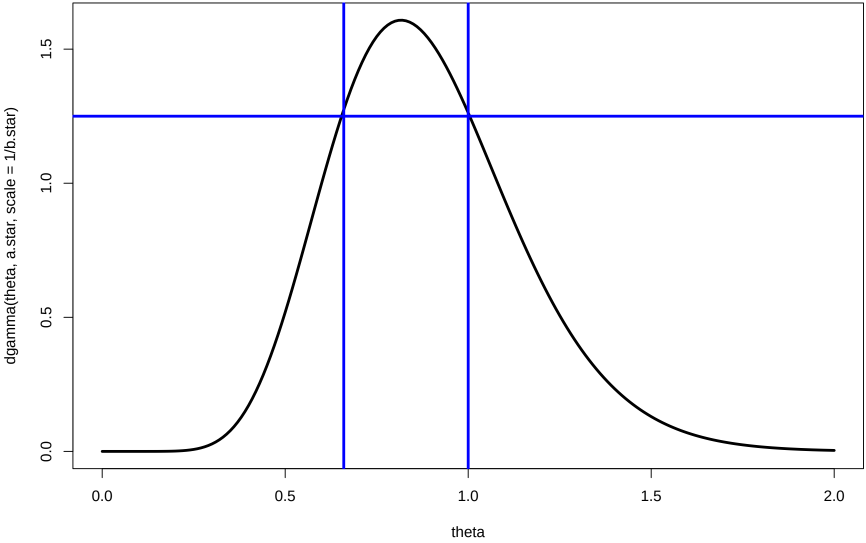
- So the equal-tailed interval is not a HPD interval!

Bayesian Interval Estimation

- To get the HPD interval we take horizontal slices across the density till we get the appropriate probability.

```
plot(theta, dgamma(theta, a.star, scale=1/b.star), type="l", lwd=3)
abline(h=1.25, lwd=3, col="blue")

##
theta <- seq(0, 2, by=0.01)
dens <- dgamma(theta, a.star, scale=1/b.star)

##
hpd.cut <- 1.25
theta.L <- min(theta[dens>=hpd.cut])
theta.U <- max(theta[dens>=hpd.cut])
abline(v=c(theta.L, theta.U), lwd=3, col="blue")

## interval probability
pgamma(theta.U, a.star, scale=1/b.star) -
  pgamma(theta.L, a.star, scale=1/b.star)
```

Bayesian Interval Estimation



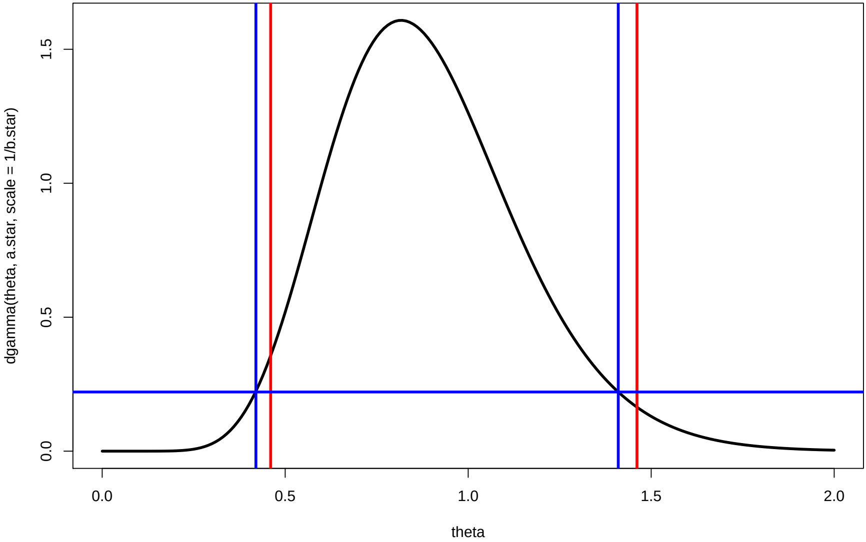

```
## [1] 0.5062717
```

Bayesian Interval Estimation

```
hpd.cut <- sort(seq(0.1, 1.25, by=0.0001), decreasing =TRUE)
c <- 1
cred.int <- 0.5063

while(cred.int<0.95){
theta.L <- min(theta[dens>=hpd.cut[c]])
theta.U <- max(theta[dens>=hpd.cut[c]])

## interval probability
cred.int <- pgamma(theta.U, a.star, scale=1/b.star) -
  pgamma(theta.L, a.star, scale=1/b.star)
c <- c+1
}

HPD <- c(theta.L,theta.U)
HPD
```

```
## [1] 0.42 1.41
```

Bayesian Interval Estimation



Bayesian Interval Estimation

- Let's check the length of each interval:
  - equal-tailed: $1.46 - 0.460 = 1.00$
  - HPD: $1.41 - 0.42 = 0.99$
- HPD is the shorter interval, but not by much.

Bootstrap Interval Estimation

- The bootstrap was used asses the bias and variability of an estimator, $\hat{\theta}$.
- The estimated standard deviation of any estimator $\hat{\theta}$ was derived by constructing some large number, B, of re-samples (with replacement) from the observed values of the sample, $X_1, \ldots, X_n$.
- The estimator was then applied to each of the B re-samples to construct:

$$\hat{\theta}^*_b, \quad b = 1, \ldots B$$

- The estimated standard deviation was:

$$\hat{\sigma}_B(\hat{\theta}) = \sqrt{\frac{1}{B-1}(\hat{\theta}^*_b - \bar{\hat{\theta}}^*)^2}$$

Bootstrap Interval Estimation

- We saw MLEs are asymptotically normal, and in fact many estimators are, we could just use that idea:

$$\left[\hat{\theta} - z_{\alpha/2}\hat{\sigma}_B(\hat{\theta}) \ \ , \ \ \hat{\theta} + z_{\alpha/2}\hat{\sigma}_B(\hat{\theta})\right]$$

- For means, we relied on the Central Limit Theorem to construct intervals when we didn't know the underlying probability distribution.
- Roughly, the bootstrap interval is a natural extension to the Central Limit Theorem for estimators which are not in the form of an average.
- For small samples we still might be in trouble so why not use a Student's t-distribution quantiles instead of the standard normal quantiles?

$$t_{n-1,\ \alpha/2}$$

Bootstrap Interval Estimation

- Recall, that $\hat{\theta}^*_b$ not only provide us with estimates of the bias and standard deviation of our estimator, $\hat{\theta}$, but also of its entire distribution.
- We can use the empirical quantiles of the "bootstrap distribution". So we have:

$$P_{\hat{F}}(\hat{\theta}^* \leq \hat{\theta}^*_L) = \alpha/2, \qquad P_{\hat{F}}(\hat{\theta}^* \leq \hat{\theta}^*_U) = 1 - \alpha/2$$

Bootstrap Interval Estimation

- Let's consider a general third approach. Remember what we are doing:

$$\hat{F}^* \to \hat{F} \to F$$

- Instead of bootstrapping we might instead choose to bootstrap some other quantity $Q(F, \hat{F})$ and use its simulated quantiles to construct an interval.
- The simplest example of such an approach is to consider a quantity which we believe is (approximately) pivotal; for example:

$$Q = Q(F, \hat{F}) = \frac{\theta(\hat{F}) - \theta(F)}{\hat{\sigma}(F)}$$

Bootstrap Interval Estimation

- The trick of course is that we equate $Q(F, \hat{F})$ with $Q(\hat{F}, \hat{F}^*)$.

$$1 - \alpha = P(q_L \leq Q(F, \hat{F}) \leq q_U) = P(q_L \leq Q(\hat{F}, \hat{F}^*) \leq q_U)$$

- We can see that q_l and q_U can be estimated by generating re-samples:

$$\hat{Q}_b = Q(\hat{F}, \hat{F}^*) \quad \text{(i.e. bootstrap samples)}$$

Bootstrap Interval Estimation

1. Using B re-samples, calculate $\hat{Q}_b$ for each re-sample and approximate q_l and q_U with (where $\alpha_1 + \alpha_2 = \alpha$) using the empirical distribution of $\hat{Q}$:

$$\hat{q}_l = \hat{Q}_{\alpha_1} \qquad \hat{q}_l = \hat{Q}_{1-\alpha_2}$$

2. Construct the confidence interval using a "pivoting" argument:

$$\begin{aligned}
1-\alpha &= P(q_L \leq Q(F,\hat{F}) \leq q_U) \\
&\approx P(\hat{q}_L \leq Q(F,\hat{F}) \leq \hat{q}_U) \\
&= P\left(\hat{q}_L \leq \frac{\theta(\hat{F}) - \theta(F)}{\hat{\sigma}(F)} \leq \hat{q}_U\right) \\
&= P\left(\theta(\hat{F}) - \hat{q}_U\hat{\sigma}(F) \leq \theta(F) \leq \theta(\hat{F}) - \hat{q}_L\hat{\sigma}(F)\right)
\end{aligned}$$

$$[\theta(\hat{F}) - \hat{q}_U\hat{\sigma}(F)\ ,\ \theta(\hat{F}) + \hat{q}_L\hat{\sigma}(F)]$$

Bootstrap Interval Estimation

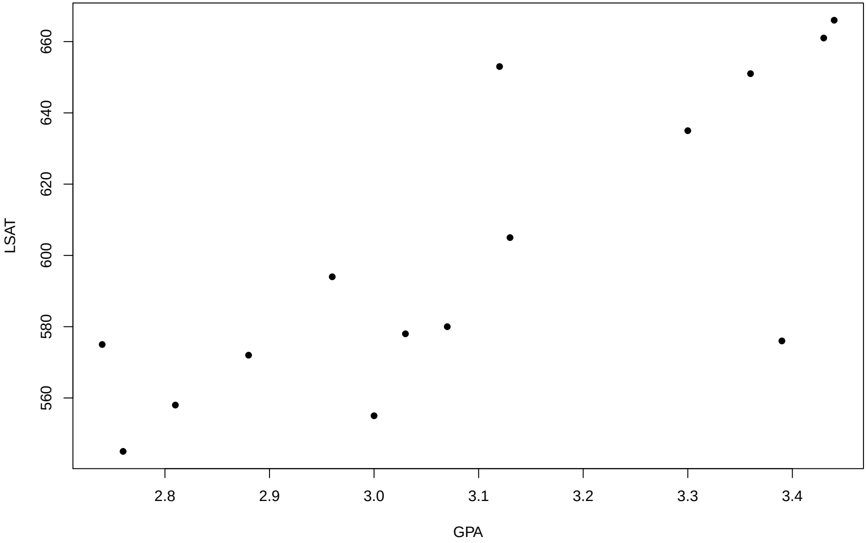
- Let's revisit our Law School example again.

```r
## Law School Example

## law school data
LSAT <- c(576, 578, 555, 605, 545, 635, 666, 661, 653,
          572, 558, 580, 651, 575, 594)
GPA <- c(3.39, 3.03, 3.00, 3.13, 2.76, 3.30, 3.44, 3.43,
         3.12, 2.88, 2.81, 3.07, 3.36, 2.74, 2.96)

D <- data.frame(LSAT, GPA)
n <- nrow(D)
```

Bootstrap Interval Estimation

```
plot(GPA, LSAT, pch=16)

##
rho.hat <- cor(LSAT, GPA)
rho.hat
```

Bootstrap Interval Estimation




```
## [1] 0.7763745
```

Bootstrap Interval Estimation

- Let's bootstrap ρ:

```
### Let's do B samples
set.seed(1001)
B <- 10000

D.B <- array(list(), B)

rho.hat.b <- rep(0, B)

for(b in 1:B){
S <- sample(1:n, n, replace = TRUE)
D.B[[b]] <- D[S,]

##
rho.hat.b[b] <- cor(D.B[[b]]$LSAT, D.B[[b]]$GPA)
    }
```

Bootstrap Interval Estimation

- Let's create intervals based on asymptotic normality:

$$\hat{\rho} = 0.7764$$

```
Var.B.hat <- var(rho.hat.b)
Var.B.hat
```

```
## [1] 0.01790244
```

```
## interval based on asymptotic normality
alpha <- 0.05

c(rho.hat - qnorm(1-alpha/2)*sqrt(Var.B.hat),
  rho.hat + qnorm(1-alpha/2)*sqrt(Var.B.hat))
```

```
## [1] 0.5141313 1.0386177
```

Bootstrap Interval Estimation

- So we have the interval:

$$[0.5141\ ,\ 1.0386]$$

But this extends beyond the range of ρ. Recall: $-1 \leq \rho \leq 1$.

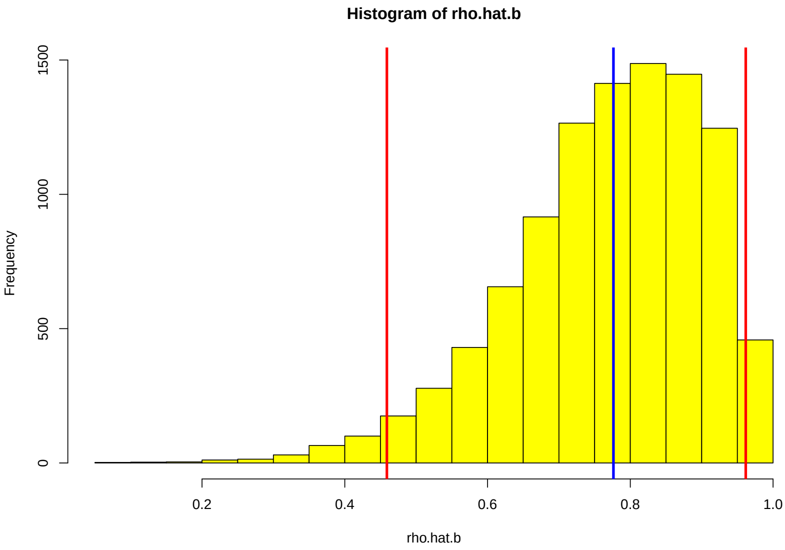
Bootstrap Interval Estimation

- Let's just look at the density of the bootstrap values and use the bootstrap percentile method (i.e. the empirical quantiles):

```
## Now let's just examine the
## bootstrapped values and use
## the empirical values
hist(rho.hat.b, col="yellow", freq=TRUE)
alpha=0.05
qu <- quantile(rho.hat.b, c(alpha/2, 1-alpha/2))
qu
abline(v=qu, col="red", lwd=3)
abline(v=rho.hat, col="blue", lwd=3)
```

```
##      2.5%     97.5%
## 0.4589734 0.9617267
```

Bootstrap Interval Estimation



Bootstrap Interval Estimation

- We see that this interval remains within the allowable range for correlation coefficients.
- Also, that this interval is not symmetric around the point estimate $\hat{\rho} = 0.7764$, which is clear from the skewness of the bootstrap histogram.

Bootstrap Interval Estimation

Example: In this question consider constructing a 95% interval ($\alpha_1 = \alpha_2 = \alpha/2$) for ρ based on:

$$Q = \hat{\rho} - \rho$$

- We use the $B = 10000$ re-sampled values:

$$\hat{Q}_b = \hat{\rho}^*_b - \hat{\rho}$$

- Notice that the only resampled part of the equation is $\hat{\rho}^*_b$ so:

$$\hat{Q}_{\alpha/2} = \hat{\rho}^*_{\alpha/2} - \hat{\rho}, \qquad \hat{Q}_{1-\alpha/2} = \hat{\rho}^*_{1-\alpha/2} - \hat{\rho}$$

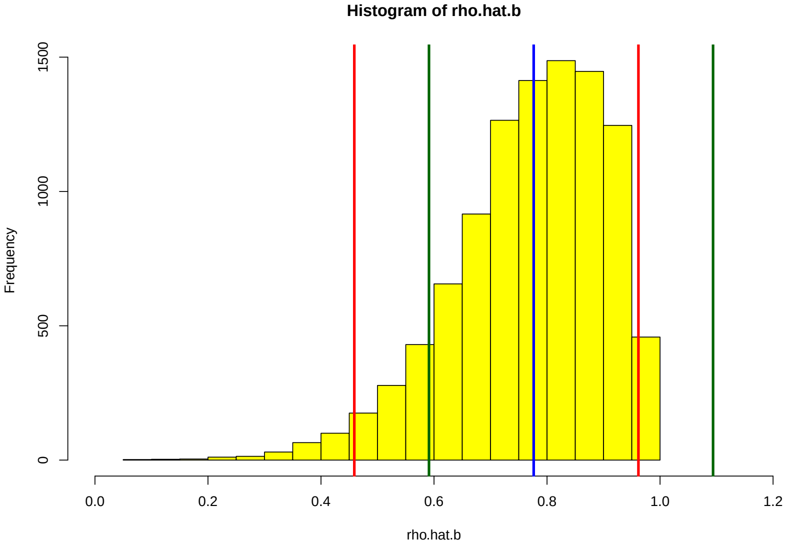
Bootstrap Interval Estimation

- Now let's form our interval and pivot ($\alpha = 0.05$):

$$
\begin{aligned}
1-\alpha = 0.95 \quad \approx \quad & P(\hat{Q}_{\alpha/2} \leq \hat{\rho} - \rho \leq \hat{Q}_{1-\alpha/2}) \\
& = P(\hat{\rho}^*_{\alpha/2} - \hat{\rho} \leq \hat{\rho} - \rho \leq \hat{\rho}^*_{1-\alpha/2} - \hat{\rho}) \\
& = P(\hat{\rho}^*_{\alpha/2} - 2\hat{\rho} \leq -\rho \leq \hat{\rho}^*_{1-\alpha/2} - 2\hat{\rho}) \\
& = P(2\hat{\rho} - \hat{\rho}^*_{1-\alpha/2} \leq \rho \leq 2\hat{\rho} - \hat{\rho}^*_{\alpha/2})
\end{aligned}
$$

$$
\begin{gathered}
[2\hat{\rho} - \hat{\rho}^*_{[1-\alpha/2]} \ , \ 2\hat{\rho} - \hat{\rho}^*_{[\alpha/2]}] \\
[2 \times 0.7764 - 0.9617 \ , \ 2 \times 0.7764 - 0.4590] \\
[0.5911 \ , \ 1.0938]
\end{gathered}
$$

- This interval also goes outside the range of ρ.


Histogram of rho.hat.b
Frequency
0
500
1000
1500
0.0
0.2
0.4
0.6
0.8
1.0
1.2
rho.hat.b

Properties of Intervals

- What we like:
  - Shortest intervals for a given confidence or credibility (eg. 95%).
  - Range respecting
  - Parameterization equivariance (We would like our interval construction procedures to transform appropriately if we change our focus from $\tau = \tau(\theta)$ to $\gamma = \gamma(\tau) = \gamma\{\tau(\theta)\}$)