Statistical Inference

Lecture 11b

ANU - RSFAS

Last Updated: Tue May 15 10:53:59 2018

Non-Parametric Methods

$y \sim Beta(a, b)$
$x \sim N(\mu, \sigma^2)$
etc.

- So far we have assumed a particular parametric model and examined the estimation of parameters in those models.
- Now we will no longer make that assumption.
- Parameters are just surrogates for some numerical characteristic of the population.
- We can think about building estimation procedures through some "function" of the underlying population distribution.
- Let's denote this quantity:

$$\theta(F)$$

where $F = F(x)$ is a is the CDF of the population distribution of numerical characteristics.

Non-Parametric Methods

- The empirical distribution function $\hat{F}$ is the CDF of a new discrete random variable, say X^*.
- It can be shown that $\hat{F}$ is a sufficient statistic for F (based on a random sample), so
- $\hat{F}$ and X^* mimics the relationship between F and the X.
- This leads to studying $(\hat{F}, X^*)$ to learn about (F, X).

$$\hat{F}(x) = \frac{1}{n} \sum_{i=1}^{n} \mathbb{I}_{(x_i \leq x)}$$
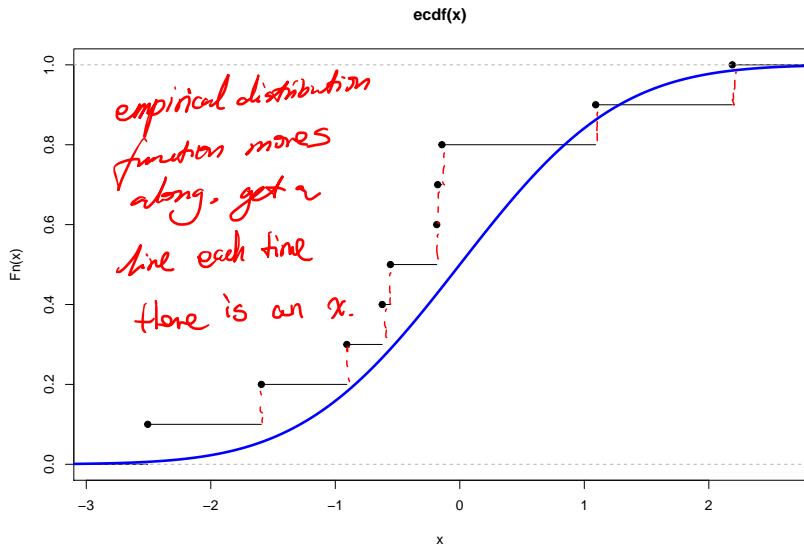
$$F(x) = P(X \leq x)$$

Non-Parametric Methods

```
set.seed(1001)
x <- rnorm(10)
sort(x)
```

```
## [1] -2.5065362 -1.5937133 -0.9074604 -0.6229437 -0.5573113
## [7] -0.1775473 -0.1435595  1.0915017  2.1886481
```
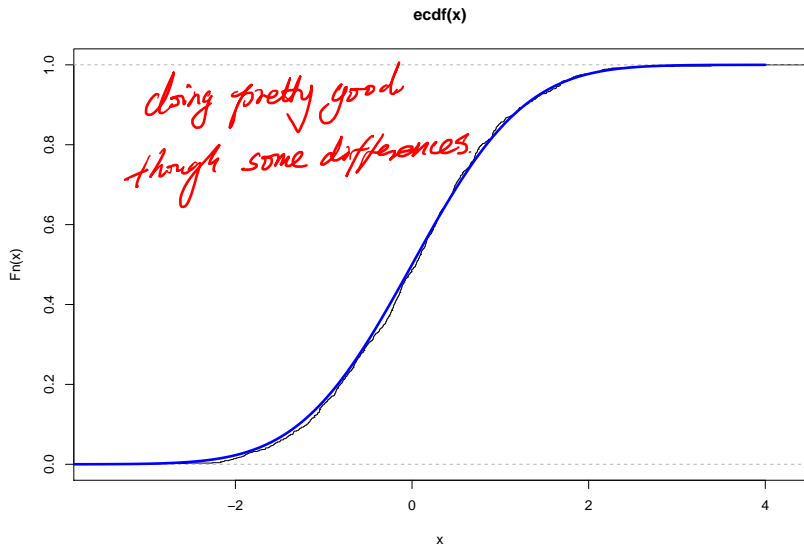
```
F <- ecdf(x)
plot(F)
z <- seq(-4,4, by=0.001)
lines(z, pnorm(z), lwd=3)
```

Non-Parametric Methods

ecdf(x)

- data (black), truth (blue)

Non-Parametric Methods $n = 1000$

ecdf(x)

closing pretty good though some differences.

- data (black), truth (blue)

Non-Parametric Methods

- Let's start with a basic question:

$$
\begin{aligned}
E\{\hat{F}(x)\} &= E\left\{\frac{1}{n}\sum_{i=1}^{n}\mathbb{I}_{(x_i \leq x)}\right\} \\
&= \frac{1}{n}\sum_{i=1}^{n}E\left\{\mathbb{I}_{(x_i \leq x)}\right\} \\
&= \frac{1}{n}\sum_{i=1}^{n}Pr(X_i \leq x) = \frac{1}{n}\sum_{i=1}^{n}F(x) = F(x)
\end{aligned}
$$

Non-Parametric Methods

- Consider again the empirical distribution function:

$$\hat{F} = \frac{1}{n} \sum_{i=1}^{n} \mathbb{I}_{(x_i \leq x)} = \frac{n_x}{n}$$

where n_x is defined as the number of observed data values which are less than or equal to the value x.

Non-Parametric Methods

- Another way to see this — n_x is binomially distributed with n trials and a "success" probability of $p = Pr(X_i \leq x) = F(x)$.

$$E\left(\hat{F}\right) = E\left(\frac{n_x}{n}\right) = \frac{E(n_x)}{n} = \frac{np}{n} = \frac{nF(x)}{n} = F(x)$$

- So we also have:

$$
\begin{aligned}
Var\{\hat{F}\} &= Var\left(\frac{n_x}{n}\right) = \frac{1}{n^2} Var(n_x) \\
&= \frac{1}{n^2} np(1-p) = \frac{1}{n} F(x)(1 - F(x))
\end{aligned}
$$

Non-Parametric Methods

- Typically we are interested in a function of F. Let's consider the mean of X which has distribution F.

- A standard function of interest is:

$$\theta(F) = \int_{-\infty}^{\infty} x f(x) dx$$

- So we are interested in $\theta(F) = E_F(X)$:

$$\hat{\theta} = \theta(\hat{F}) = E_{\hat{F}}(X) = \sum_{x \in \mathcal{X}} x p_{\hat{F}}(x) = \sum_{i=1}^{n} x_i \frac{1}{n} = \bar{x}$$

- Note: $p_{\hat{F}}(x) = \frac{1}{n} \quad \forall x \in \mathcal{X}$.

Non-Parametric Methods - Bootstrap

- We are interested in the behavior of $\hat{\theta}$ under F.
- Let's examine $\hat{\theta}^* = \theta(\hat{F}^*)$ associated with $X_1^*, \ldots, X_n^*$ having distribution $\hat{F}$.
- To examine $\hat{\theta}$ under F, imagine that our observed data forms its own "population" from which we randomly sample according to the "true" distribution $\hat{F}$.
- Construct an estimate of the "true" population parameter $\theta(\hat{F})$ using the "re-sampled" data, arriving at $\theta(\hat{F}^*)$ as the estimator for $\theta(\hat{F})$.

Non-Parametric Methods - Bootstrap

normally, $x_1, \cdots, x_n \overset{iid}{\sim} Poisson(\theta)$

& then $\hat{\theta} = \bar{x}$. say its MLE. we say about E&V :-

But here, what we do $\downarrow$ *$\hat{F} \leftarrow$ "population"*   *$\leftarrow$*   *$F \leftarrow$ true population*

- Since we know the "true" distribution $\hat{F}$, we can determine exactly the bias and variance of $\theta(\hat{F}^*)$.

- So we use the bias and variance of $\theta(\hat{F}^*)$ under $\hat{F}$ as estimators for the bias and variance of $\theta(\hat{F})$ under F.

- This approach is generally referred to as the bootstrap, since we are using the data itself to estimate its behavior under F, effectively "pulling ourselves up by our own bootstraps".

Non-Parametric Methods - Bootstrap

- Define the bootstrap estimators of bias and variance as:

$$\hat{B}_B = E_{\hat{F}}\{\theta(\hat{F}^*)\} - \theta(\hat{F})$$

$$\hat{Var}_B\{\theta(\hat{F}^*)\} = E_{\hat{F}}\{\theta(\hat{F}^*)^2\} - \left[E_{\hat{F}}\{\theta(\hat{F}^*)\}\right]^2$$

- These calculations are occasionally possible exactly in the case of simple functionals $\theta(\cdot)$.
- However, generally they are estimated through computational means!
- Note: we have the observed values $x_1, \ldots, x_n$ in our possession, we can easily create realisations of the random sample $X_1^*, \ldots, X_n^*$ simply randomly drawing n values with replacement from the collection $\mathcal{X} = \{x_1, \ldots, x_n\}$.

Non-Parametric Methods - Bootstrap

- We create B "bootstrap" (re-sampled) data sets:

$$\{X_{1,1}^*, \ldots, X_{n,1}^*\}, \ldots, \{X_{1,b}^*, \ldots, X_{n,b}^*\}, \ldots, \{X_{1,B}^*, \ldots, X_{n,B}^*\}$$

- With each sample we estimate $\hat{\theta}^b = \theta(\hat{F}_b^*)$, then we can compute:

$$\hat{B}_B = E_{\hat{F}}\{\theta(\hat{F}^*)\} - \theta(\hat{F}) \approx \frac{1}{B}\sum_{b=1}^{B} \hat{\theta}_b^* - \hat{\theta}$$

$$\hat{Var}_B\{\theta(\hat{F}^*)\} \approx \frac{1}{B-1}\sum_{b=1}^{B}\left(\hat{\theta}_b^* - \frac{1}{B}\sum_{b=1}^{B}\hat{\theta}_b^*\right)^2$$

Non-Parametric Methods - Bootstrap

- Note: Casella and Berger suggest that all non unique re-samples (with replacement) should be determined.
- Eg. Suppose we have the following data:

$$2, 3, 9, 12$$

- For the first spot we have 4 possibilities. The second spot we have 4 possibilities, and so on:

$$B = n^n = 4^4 = 256$$

- Typically this is not done! And B is just set to be a large number.

Non-Parametric Methods - Bootstrap

Eg. Law School - Correlation between LSAT and GPA

Suppose that we have observed the following data pairs, which represent the average LSAT (Legal Scholastic Aptitude Test, a common entrance exam for prospective law school students in the United States) and undergraduate GPA (grade point average) scores for the 1973 entering class at a random sample of 15 U.S. Law Schools. We are interested in the an estimate of the correlation between LSAT (Y) and GPA (Z), along with an estimate of it's bias and variance:

$$\theta(F) = \rho_F = \frac{Cov_F(Y,Z)}{\sqrt{Var_F(Y)Var_F(Z)}}$$
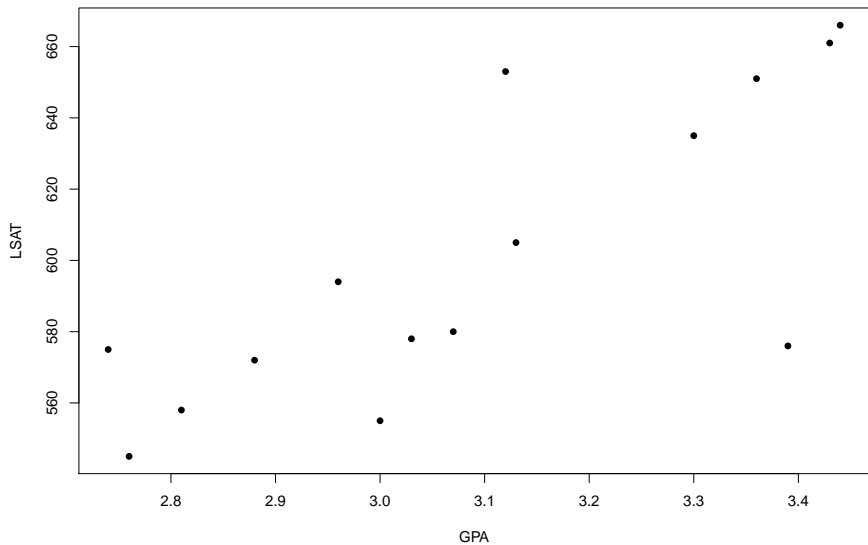
$$\hat{\rho} = \frac{\sum_{i=1}^{n}(y_i - \bar{y})(z_i - \bar{z})}{\sqrt{\sum_{i=1}^{n}(y_i - \bar{y})^2 \sum_{i=1}^{n}(z_i - \bar{z})^2}}$$

Non-Parametric Methods - Bootstrap

```
## 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)



##
plot(GPA, LSAT, pch=16)
rho.hat <- cor(LSAT, GPA)
rho.hat
```
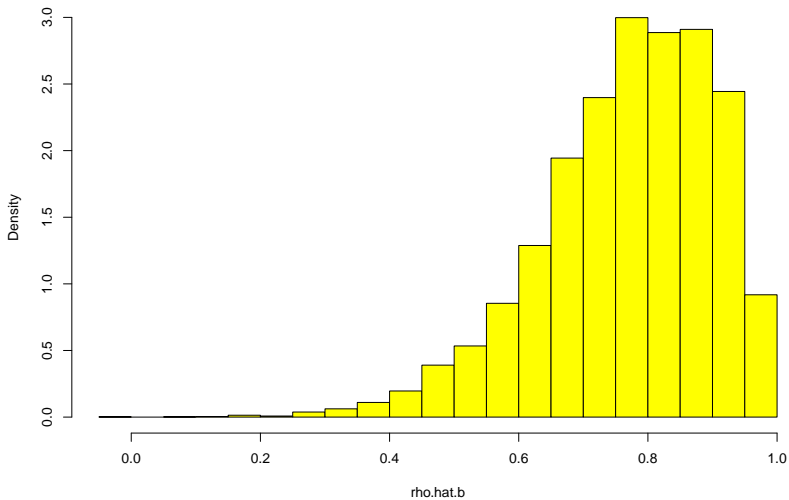
```
## [1] 0.7763745
```

```
## Bootstrap
#########################################################
## Take a single bootstrap sample
## As it is random it is different than in the notes
set.seed(2000)
S <- sample(1:n, n, replace = TRUE)
D.s <- D[S,]

### Let's do B samples
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)
    }
hist(rho.hat.b, col="yellow", prob=TRUE)
```

Histogram of rho.hat.b

- Estimate, estimate of the bias of the estimate, estimate of the variance of the estimate.

```
rho.hat
```

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

```
Bias.B.hat <- mean(rho.hat.b) - rho.hat
Bias.B.hat
```

```
## [1] -0.007520911
```

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

```
## [1] 0.01784111
```

Some Thoughts

- The idea behind the bootstrap is powerful and extremely intuitively appealing.
- Why, then, has the bootstrap not replaced parametric approaches?
    - As implemented, the bootstrap method yields a different answer every time (of course, the differences will be very small if B is large).
    - Another drawback is that if θ is complicated to calculate (perhaps because it is implicitly defined as the solution to an equation, just as the MLE was) then computing its value for each of B re-sampled data sets is computationally quite expensive and time consuming.
    - If we truly believe the parametric structure we have set up, then the parametric estimators have nice optimal properties.

Some Thoughts

- The bootstrap is a very flexible and widely applicable approach which deserves more attention than it currently gets among statistical practitioners
- The bootstrap can even be extended to circumstances beyond the *iid* setting on which we have focused here.
- But a word of warning on complicated (non iid settings):
  - We cannot always guarantee that using the bootstrap paradigm (replacing F by $\hat{F}$ and $\hat{F}$ by $\hat{F}^*$) to estimate bias and variance will yield valid estimates.

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:

  use $\hat{F}$ (original est.)   use calculation from Bootstrap

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

- 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?

  small sample.

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

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:

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

- 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 $\hat{Q}$:

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

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.

```
## 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)
```
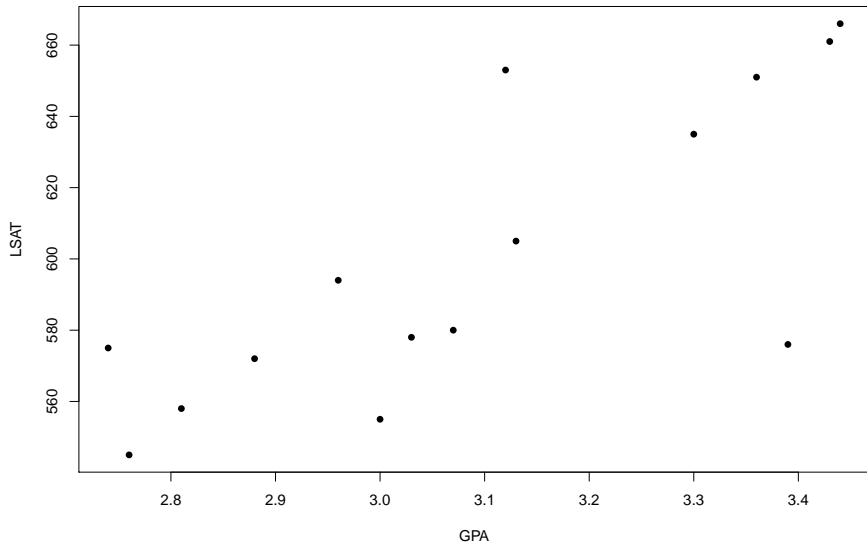
Bootstrap Interval Estimation

```
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

Histogram of rho.hat.b
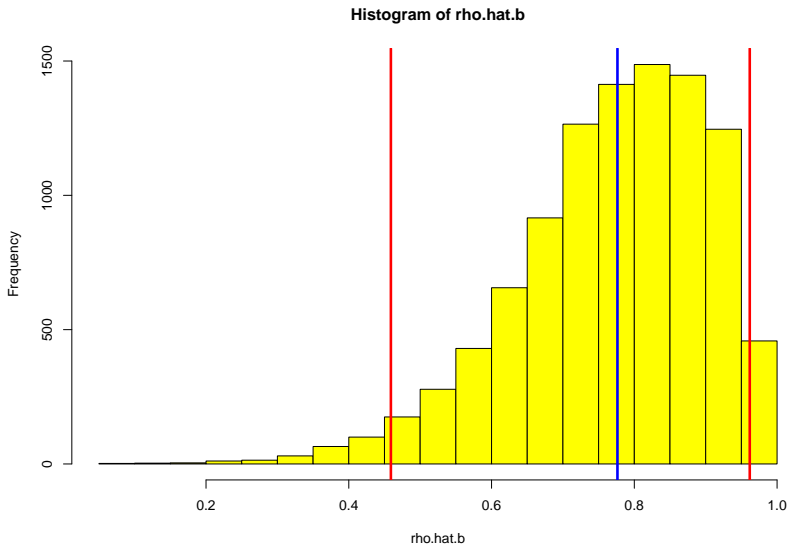
Bootstrap Interval Estimation

- 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:

$$\hat{Q}_{\alpha/2} = \hat{\rho}_{\alpha/2}^* - \hat{\rho}, \quad \hat{Q}_{1-\alpha/2} = \hat{\rho}_{1-\alpha/2}^* - \hat{\rho}$$
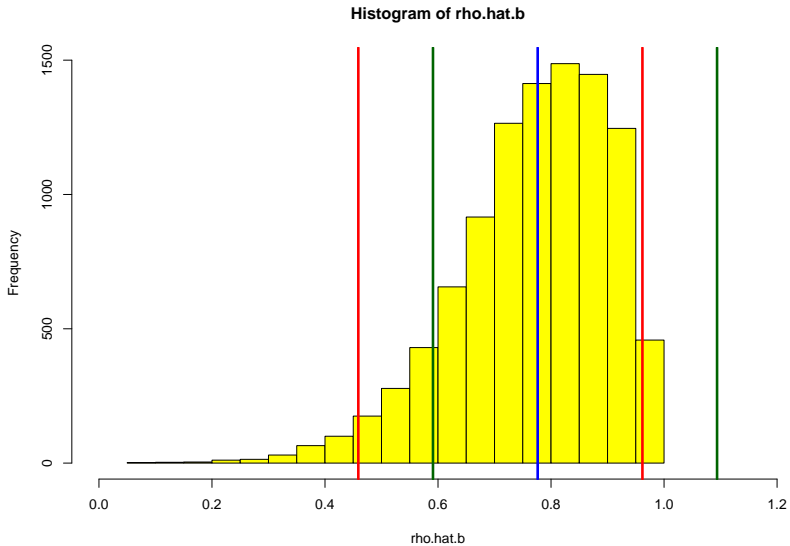
Bootstrap Interval Estimation

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

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

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

- This interval also goes outside the range of ρ.

Histogram of 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 $\gamma = \gamma(\tau) = \gamma\{\tau(\theta)\}$)

Non-Parametric Testing

- We will consider two types of tests:
  - permutation/randomization tests
  - bootstrap tests

Non-Parametric Testing

- Permutation tests are a computer-intensive approach introduced in the 1930s by R.A. Fisher.
- We will consider a two sample problem (the idea can easily be extended)
- Suppose we observe the following two data sets drawn from two potentially different probability distribution F and G:

$$
\begin{aligned}
F \rightarrow \mathbf{y} &= (y_1, \ldots, y_n) \\
G \rightarrow \mathbf{x} &= (x_1, \ldots, x_m)
\end{aligned}
$$

Non-Parametric Testing

- Data: 16 mice were randomly assigned to a treatment and control group. The survival times, in days, following a test surgery were recorded.

```
treat <- c(94, 197, 16, 38, 99, 141, 23)
n <- length(treat)
control <- c(52, 104, 146, 10, 51, 30, 40, 27, 46)
m <- length(control)
```

- Consider testing whether:

$$H_0 : F(\text{treatment}) = G(\text{control})$$

$$H_1 : \text{The distributions are not equal}$$

- Under H_0 the assignment (label) of each data point can be permuted. Whether a data point was a "treatment" or "control" is the same!

Non-Parametric Testing

- Now we need a test statistic. Consider:

$$\theta = \bar{T} - \bar{C}$$

- Note: For our alternative we are stating the distributions are not the same. But we can consider one or two-sided tests for our alternative.
- For the data we have:

```
theta <- mean(treat) - mean(control)
theta
```

```
## [1] 30.63492
```

Non-Parametric Testing

- Under H_0 we can permute the labels of the data and calculate our test statistic:

```
data <- c(treat, control)
data.p <- sample(data)
treat.p <- data.p[1:n]
control.p <- data.p[(n+1):(n+m)]
treat.p
```

```
## [1]  16  52  94 104  46 141 197
```

```
control.p
```

```
## [1]  27  30  40  10  23  38 146  51  99
```

```
theta.p <- mean(treat.p) - mean(control.p)
theta.p
```

```
## [1] 41.30159
```

Non-Parametric Testing

- Let's repeat this process $B = 10000$:

```
set.seed(1001)
B <- 10000
theta.p <- rep(0, B)

for(b in 1:B){
  data.p <- sample(data)
  treat.p <- data.p[1:n]
  control.p <- data.p[(n+1):(n+m)]
  theta.p[b] <- mean(treat.p) - mean(control.p)
  }
```

Non-Parametric Testing

```
hist(theta.p, col="yellow",
     main="Distirbution under the NULL")

## determine the critical value for alpha = 0.05
qu <- quantile(theta.p, 0.95)
abline(v=qu, lwd=3, col="red")

## add the statistic based on the
## original data (non permuted)
abline(v=theta, lwd=3, col="blue")
```

Non-Parametric Testing

Distirbution under the NULL
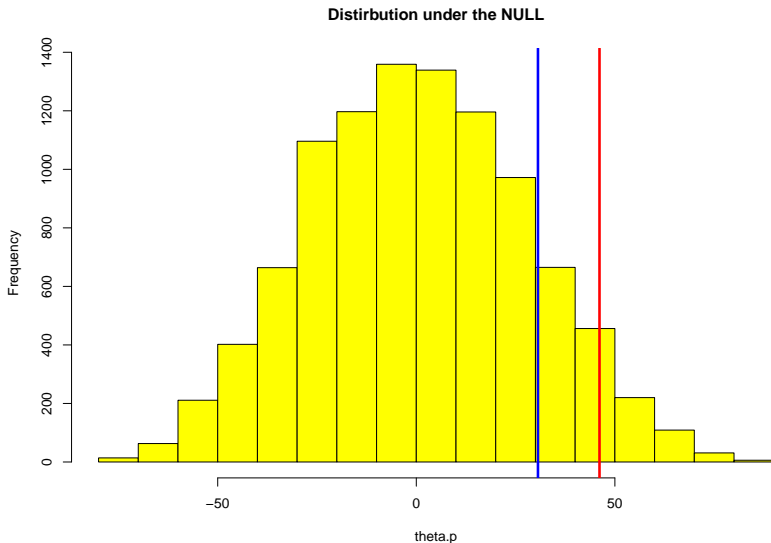
Non-Parametric Testing

- Let's consider a one sided upper tail test.
- We see that our original test statistic (based on the non permuted data) does not cross into the rejection region. Thus we do not reject H_0.
- Let's calculate the p-value:

$$P(T(\boldsymbol{X}, \boldsymbol{Y}) \geq T(\boldsymbol{x}, \boldsymbol{y}))$$

```
p.value <- mean(theta.p > theta)
p.value
```

```
## [1] 0.142
```

Non-Parametric Testing

- Permutation tests: How many possible ways are there to permute? Let $N = n + m$:

$$\mathbf{z} = \{y_1, \ldots, y_n, x_1, \ldots, x_m\}$$

$$\frac{N!}{n!m!} = \binom{N}{n}$$

If we construct our test based on every possible permutation, then we have a permutation test and an exact p-value!

- If we randomly sample from the $\binom{N}{n}$ possible permutations then we have a randomization test and the p-value is approximate.

Non-Parametric Testing

- We can also consider hypothesis testing for our two sample problem via the bootstrap.

$$H_0 : F(\text{treatment}) = G(\text{control})$$

$$H_1 : \text{The distirbutions are not equal}$$

$$\mathbf{z} = \{y_1, \ldots, y_n, x_1, \ldots, x_m\}$$

Non-Parametric Testing

1. Draw B samples of size $N = n + m$ with replacement from $\mathbf{z}$. Call the first n observations $\mathbf{y}^*$ and the remaining m observations $\mathbf{x}^*$.

2. Evaluate the test statistic $t(\cdot)$ on each bootstrap sample:

$$t(\mathbf{y}^*, \mathbf{x}^*) = \bar{y}^* - \bar{x}^*$$

3. Construct an approximate p-value:

$$\text{p-value} \approx \#\{t(\mathbf{X}^*, \mathbf{Y}^*) \geq t_{obs}\}/B$$

Non-Parametric Testing

- Let's conduct a bootstrap test with $B = 10000$:

```
set.seed(1001)
B <- 10000
theta.star <- rep(0, B)
N <- n+m
theta <- mean(treat) - mean(control)

for(b in 1:B){
  data.star <- sample(data, N, replace=TRUE)
  treat.star <- data.star[1:n]
  control.star <- data.star[-c(1:n)]
  theta.star[b] <- mean(treat.star) - mean(control.star)
  }
```

Non-Parametric Testing

```r
hist(theta.star, col="yellow",
     main="Distirbution under the NULL")

## determine the critical value for alpha = 0.05
qu <- quantile(theta.star, 0.95)
abline(v=qu, lwd=3, col="red")

## add the statistic based on the original data (non permuted)
abline(v=theta, lwd=3, col="blue")

p.value <- mean(theta.star > theta)
p.value
```

Non-Parametric Testing

Distirbution under the NULL
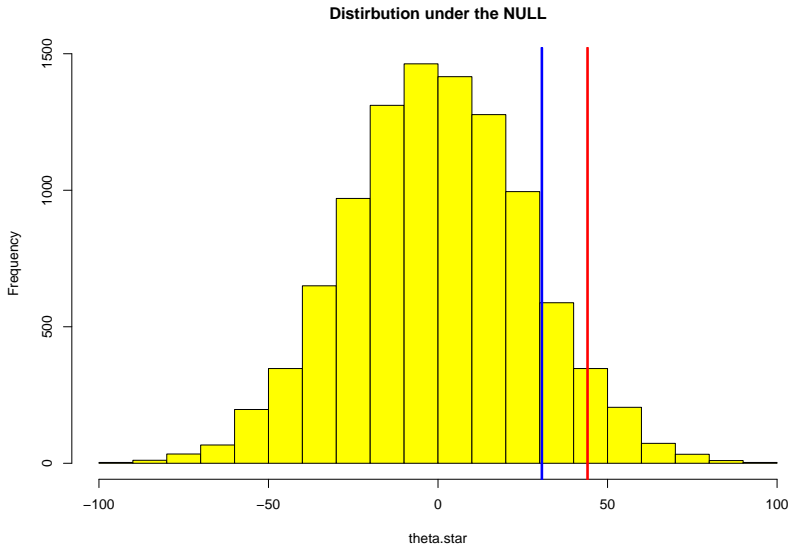
```
## [1] 0.1216
```

Non-Parametric Testing

- While permutation/randomization tests are very useful, the bootstrap approach is a little more flexible. Consider testing

$$H_0 : \mu_{treat} = 129$$

- There is no permutation/randomization test available. They require the idea of label switching and here we only have one label.

Non-Parametric Testing

- Let's use the following test statistic (t-statistic) based on a two-sided alternative:

$$H_1 : \mu_{treat} \neq 129.0$$

$$|t(\mathbf{y})| = \left| \frac{\bar{y} - \mu_0}{s/\sqrt{7}} \right| = \left| \frac{86.9 - 129.0}{66.8/\sqrt{7}} \right| = \left| -1.67 \right| = 1.67$$

```
y <- treat
mean(y)
```

```
## [1] 86.85714
```

```
sd(y)
```

```
## [1] 66.76683
```

Non-Parametric Testing

- Now we need a way to construct a distribution of $t(\cdot)$ under the Null hypothesis.
- We need to recenter our data:

$$\tilde{y}_i = y_i - \bar{y} + 129 = y_i - 86.9 + 129.0 = y_i + 42.1$$

- Now we sample $\tilde{y}_i$ with replacement for $i = 1, \ldots, 7$ to get $\tilde{y}_1^*, \ldots \tilde{y}_7^*$:

$$|t(\tilde{\boldsymbol{y}}^*)| = \Big|\frac{\bar{\tilde{y}}^* - \mu_0}{\tilde{s}^*/\sqrt{7}}\Big| = \Big|\frac{\bar{\tilde{y}}^* - 129.0}{\tilde{s}^*/\sqrt{7}}\Big|$$

Non-Parametric Testing

- Let's conduct a bootstrap test with $B = 10000$:

```
set.seed(1001)
B <- 10000
m <- length(y)
t.obs <- abs( (mean(y) - 129)/(sd(y)/sqrt(7)))
t.star <- rep(0, B)

for(b in 1:B){
y.star <- sample(y, m, replace=TRUE)
y.tilde.star <- y.star + 42.1

t.star[b] <- abs( (mean(y.tilde.star) - 129)/
                  (sd(y.tilde.star)/sqrt(7)))
  }
```
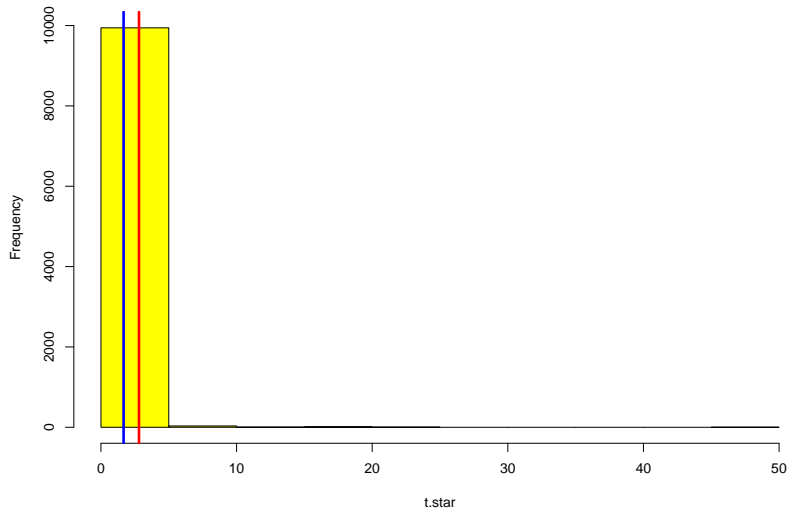
Non-Parametric Testing

```
hist(t.star, col="yellow",
     main="Distirbution under the NULL")

## determine the critical value for alpha = 0.05
qu <- quantile(t.star, 0.95)
abline(v=qu, lwd=3, col="red")

## add the statistic based on the original data
abline(v=t.obs, lwd=3, col="blue")

p.value <- mean(t.star > t.obs)
p.value
```

Non-Parametric Testing

Distribution under the NULL

[1] 0.1501

Non-Parametric Testing

- Based on either the p-value or examination based on the critical region, we see that we do not reject H_0.