

Statistical Inference

Lecture 11a

ANU - RSFAS

Last Updated: Tue May 16 11:20:41 2017

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 :

$$F \rightarrow \mathbf{y} = (y_1, \dots, y_n)$$

$$G \rightarrow \mathbf{x} = (x_1, \dots, x_m)$$

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] 27 94 10 141 38 146 40
```

```
control.p
```

```
## [1] 16 30 52 51 23 99 104 197 46
```

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

```
## [1] 2.190476
```

Non-Parametric Testing

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

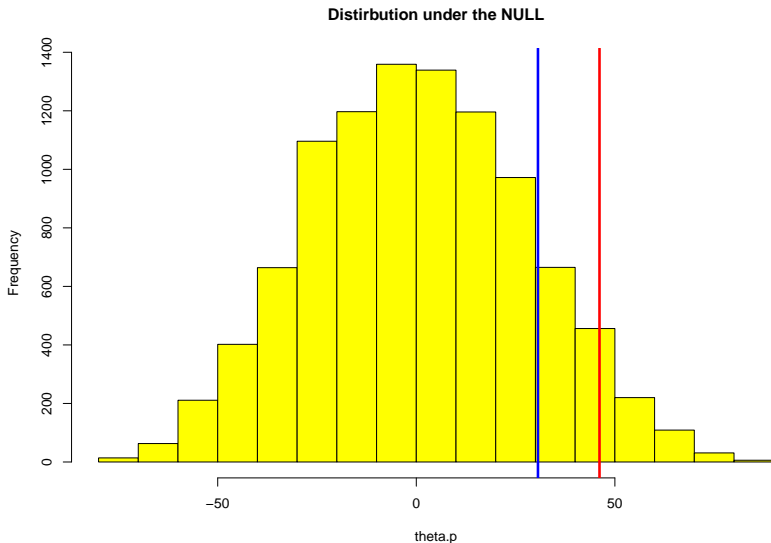
```
set.seed(1001)
B <- 10000
theta.p <- rep(0, B)

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



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:

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

```
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 $N = n + m$:

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

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

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.

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

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

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

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 $B = 10000$:

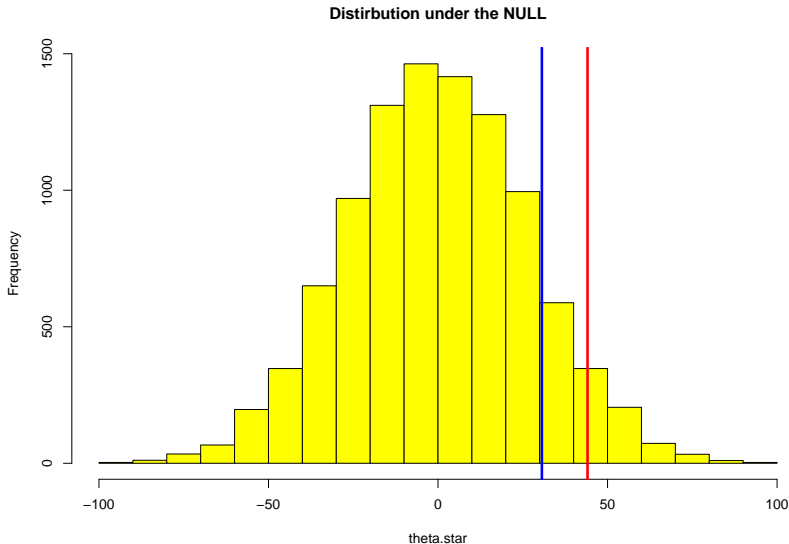
```
set.seed(1001)
B <- 10000
theta.star <- rep(0, B)
N <- n+m
theta <- mean(treat) - mean(control)

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

```
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



[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:

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

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

```
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 resample with replacement for $i = 1, \dots, 7$ to get \tilde{y}_i^* :

$$\tilde{y}_i^* = y_i^* + 42.1$$

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

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

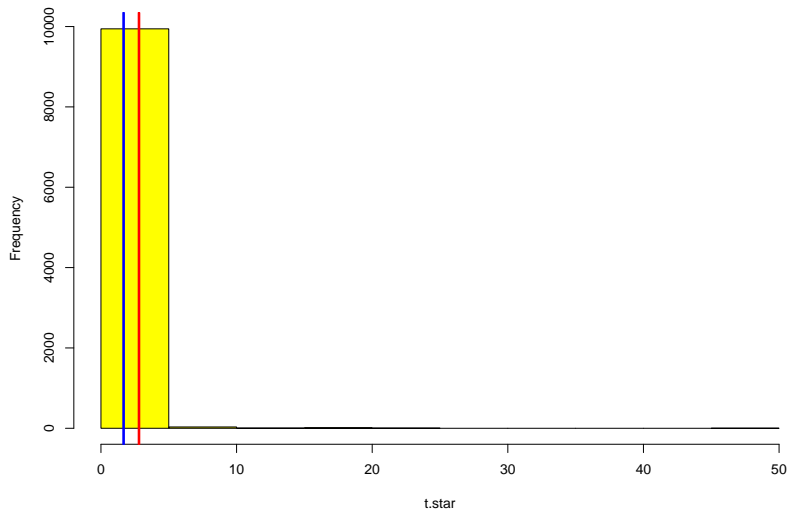
  t.star[b] <- abs( (mean(y.tilde.star) - 129)/
                    (sd(y.tilde.star)/sqrt(7)))
}
```

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 .

Bayesian Testing

- Consider testing:

$$H_0 : \theta \in \Theta_0 \quad \text{vs.} \quad H_1 : \theta \in \Theta_1$$

- The classical statistician considers θ a **fixed** unknown, thus a hypothesis test is **true** or **false**. Either θ is in Θ_0 or Θ_1 !
- Bayesians however consider θ to be random and it is quite natural to consider:

$$P(\theta \in \Theta_0 | \mathbf{x}) \quad \text{vs.} \quad P(\theta \in \Theta_1 | \mathbf{x})$$

Bayesian Testing

- One approach to Bayesian testing is to reject H_0 if:

$$P(\theta \in \Theta_1 | \mathbf{x}) > P(\theta \in \Theta_0 | \mathbf{x})$$

Bayesian Testing

Example 8.2.7: $X_1, \dots, X_n \stackrel{\text{iid}}{\sim} \text{normal}(\theta, \sigma^2)$. Let $\theta \sim \text{normal}(\mu, \tau^2)$, where σ^2, μ, τ^2 are known. Consider testing:

$$H_0 : \theta \leq \theta_0 \quad \text{vs.} \quad H_1 : \theta > \theta_0$$

$$[\theta | \mathbf{x}] \sim \text{normal} \left(\frac{\sigma^2 \mu + n\tau^2 \bar{x}}{\sigma^2 + n\tau^2}, \frac{\sigma^2 \tau^2}{\sigma^2 + n\tau^2} \right)$$

- To compare we simply examine:

$$\int_{-\infty}^{\theta_0} p(\theta | \mathbf{x}) d\theta \quad \text{vs.} \quad \int_{\theta_0}^{\infty} p(\theta | \mathbf{x}) d\theta$$

Bayesian Testing

- Based on this Bayesian approach to hypothesis testing, what if we wanted to test:

$$H_0 : \theta = \theta_0 \quad \text{vs.} \quad H_1 : \theta \neq \theta_0$$

$$[\theta|\mathbf{x}] \sim \text{normal} \left(\frac{\sigma^2 \mu + n\tau^2 \bar{x}}{\sigma^2 + n\tau^2}, \frac{\sigma^2 \tau^2}{\sigma^2 + n\tau^2} \right)$$

- $[\theta|\mathbf{x}]$ is a continuous distribution, so the probability of any single point (such as θ_0) is zero. This approach does not seem to work.

Bayesian Testing

- Notice that when we test through a Bayesian approach, we model the parameter of interest (i.e. we put a prior on it).
- If our scientific question concerns whether a parameter can be exactly θ_0 or not, then we should model that:

$$\theta \sim p\mathbf{1}_{\theta=\theta_0} + (1 - p)\text{normal}(\mu, \tau^2)$$

where $0 \leq p \leq 1$.

- In a regression or GLM setting many times we are interested in testing whether $\beta = 0$ vs. $\beta \neq 0$ and these priors or variants on them can be quite useful.

Bayesian Testing

- Another approach that does allow for consideration of $\beta = 0$ is **Bayes factors**. Let's rephrase hypothesis testing as choosing between competing models:

Model 1 (M_1): $y_i = \alpha + \epsilon_i$ $\epsilon_i \stackrel{\text{iid}}{\sim} \text{normal}(0, \sigma^2)$; $\theta_1 = \{\alpha, \sigma^2\}$

Model 2 (M_2): $y_i = \alpha + \beta x_i + \epsilon_i$ $\epsilon_i \stackrel{\text{iid}}{\sim} \text{normal}(0, \sigma^2)$; $\theta_2 = \{\alpha, \beta, \sigma^2\}$

Bayesian Testing

- Let's figure out the posterior probability for a given model i :

$$\pi(M_i|\mathbf{x}) = \frac{f(\mathbf{x}|M_i)\pi(M_i)}{m(\mathbf{x})}$$

$$f(\mathbf{x}|M_i) = \int_{\theta} f(\mathbf{x}|\theta, M_i)\pi(\theta|M_i)d\theta_i$$

$$m(\mathbf{x}) = \sum_{i=1}^2 f(\mathbf{x}|M_i)\pi(M_i)$$

Bayesian Testing

- Now consider the following ratio of the posterior model probabilities:

$$\begin{aligned}\frac{\pi(M_2|\mathbf{x})}{\pi(M_1|\mathbf{x})} &= \frac{f(\mathbf{x}|M_2)}{f(\mathbf{x}|M_1)} \times \frac{\pi(M_2)}{\pi(M_1)} \\ &= BF(M_2; M_1) \times \frac{\pi(M_2)}{\pi(M_1)}\end{aligned}$$

Where $BF(M_2; M_1)$ is called the **Bayes factor**.

- Typically $\pi(M_2) = \pi(M_1)$, so the ratio of the posterior probabilities is the Bayes factor.
- The Bayes factor looks like a likelihood ratio. However, the difference is that θ has been integrated out in both the numerator and denominator, so we have the marginal distribution of the data given the model.

Bayesian Testing

- If $f(\mathbf{x}|M_2) > f(\mathbf{x}|M_1)$ or $\frac{f(\mathbf{x}|M_2)}{f(\mathbf{x}|M_1)} > 1$ then we have support for M_2 against M_1 .
- Jeffreys, H. (1961 appendix B) suggested the following:

$BF(M_2; M_1) = B_{21}$	Evidence against model 1 (H_0)
1 to 3.2	Not worth more than a bare mention
3.2 to 10	Substantial
10 to 100	Strong
> 100	Decisive