2015 SISG Bayesian Statistics for Genetics R Notes: Multiple Testing

Jon Wakefield

Departments of Statistics and Biostatistics, University of Washington
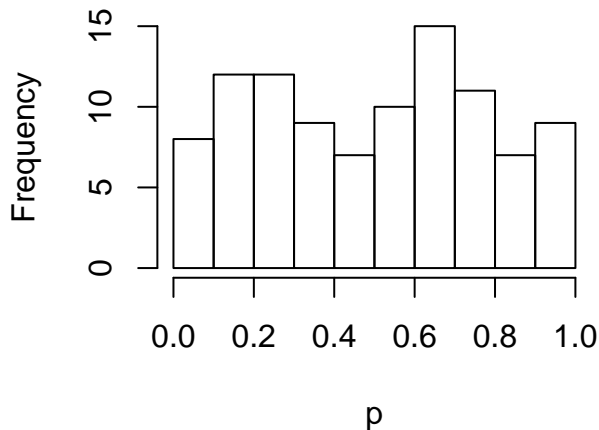
2016-07-26

Multiple testing example

Simulate $m = 100$ p-values and fix 5 to be quite small.

```
set.seed(254)
m <- 100
alpha <- 0.05
p <- runif(m)
expp <- seq(1, m)/(m + 1)
p <- sort(p)
p[1:5] <- c(2e-04, 5e-04, 0.001, 0.002, 0.01)
```

Multiple testing example

Histogram of p-values is uninformative...

```r
par(mfrow = c(1, 1))
hist(p, main = "")
```

Multiple testing example

```r
plot(expp, p, xlim = c(0, 1), ylim = c(0, 1), ylab = "Observed p",
    xlab = "Expected p")
abline(h = 0.05/m, col = "blue")
abline(h = 1/m, col = "red")
abline(a = 0, b = 1)
# points(expp,alpha*seq(1,m)/m,pch=25)
lines(expp, alpha * seq(1, m)/m, col = "green")
legend("topleft", legend = c("Bonferroni 5%", "EFD=1",
    "FDR 5%"), col = c("blue", "red", "green"), bty = "n",
    lwd = 2)
```

Multiple testing example
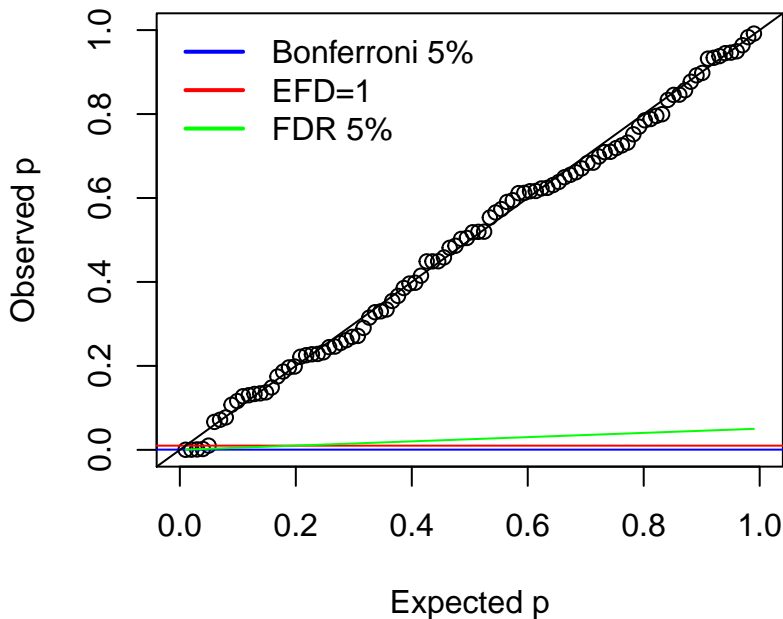
Multiple testing example

```r
plot(expp, p, xlim = c(0, 0.1), ylim = c(0, 0.1), ylab = "Observed p",
    xlab = "Expected p")
abline(h = 0.05/m, col = "blue")
abline(a = 0, b = 1)
abline(h = 1/m, col = "red")
# points(expp,alpha*seq(1,m)/m,pch=25)
lines(expp, alpha * seq(1, m)/m, col = "green")
legend("topleft", legend = c("Bonferroni 5%", "EFD=1",
    "FDR 5%"), col = c("blue", "red", "green"), bty = "n",
    lwd = 2)
```
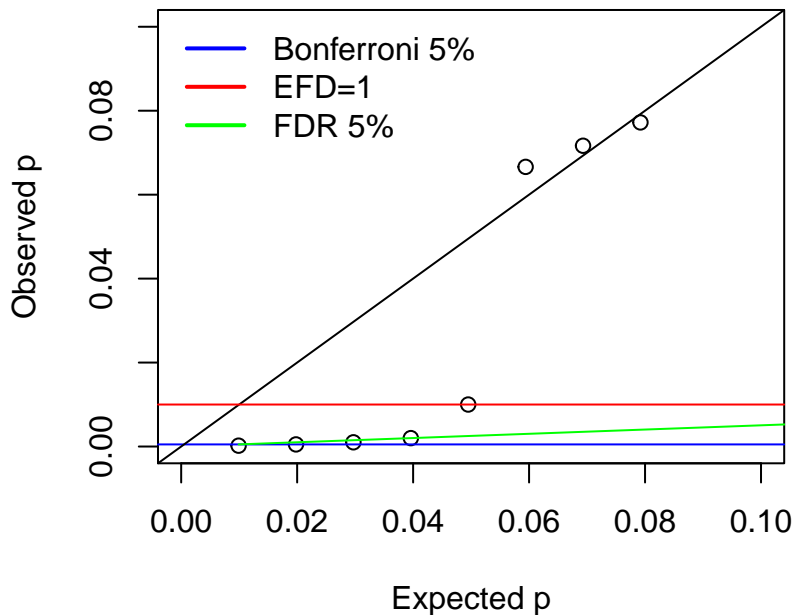
Multiple testing example

Multiple testing example

```
ymax <- max(-log10(p), -log10(0.05/m), -log10(1/m))
plot(-log10(expp), -log10(p), ylim = c(0, ymax), xlab = "-log10(expected)",
    ylab = "-log10(observed)")
abline(h = -log10(0.05/m), col = "blue")   # Bonf
abline(h = -log10(1/m), col = "red")   # EFD=1
abline(a = 0, b = 1)
# points(-log10(expp),-log10(alpha*seq(1,m)/m),pch=25)
lines(-log10(expp), -log10(alpha * seq(1, m)/m), col = "green")
legend("bottomright", legend = c("Bonferroni 5%", "EFD=1",
    "FDR 5%"), col = c("blue", "red", "green"), bty = "n",
    lwd = 2)
```

Multiple testing example

Evaluate the number passing each condition

```r
sum(-log10(p) > -log10(0.05/m))   # Bonferroni
## [1] 1
sum(-log10(p) > -log10(1/m))   # EFD=1
## [1] 4
li <- seq(1, 100) * 0.05/m
cond <- rep(0, m)
for (i in 1:m) {
    if (p[i] < li[i]) {
        cond[i] <- 1
    }
}
sum(cond)   # FDR
## [1] 3
```

Multiple testing example