MA 677-Final Report

Xiang Li

5/12/2022

Introduction

This report is for MA677 final project. In this report, I will reproduce 4 examples in Chapter 6 in Computer Age Statistical Inference. Chapter 6 contains four examples – insurance claims (Robbins' Formula), species discovery(The Missing-Species Problem), Shakespeare's vocabulary, and lymph node counts (A Medical Example). In each example, the result of the empirical Bayes analysis is given.

Robbins' Formula

```
# Create a data frame
auto <- data.frame(Claims_x=seq(0,7),
           Counts_yx=c(7840,1317,239,42,14,4,4,1))
#auto
n <- 8
robbin1<-round(((auto$Claims+1)[1:7]*auto$Counts[2:8]/auto$Counts[1:7]),3)
```

The Missing-Species Problem

Shakespeare's vocabulary,

```
data("bardWordCount", package = "deconvolveR")
lambda <- seq(-4, 4.5, .025)
tau <- exp(lambda)
result <- deconv(tau = tau, y = bardWordCount, n = 100, c0=2)
stats <- result$stats
d <- data.frame(lambda = lambda, g = stats[, "g"], tg = stats[, "tg"],
                SE.g = stats[, "SE.g"])
indices <- seq(1, length(lambda), 5)
print(
    ggplot(data = d) +
    geom_line(mapping = aes(x = lambda, y = g)) +
    geom_errorbar(data = d[indices, ],
                  mapping = aes(x = lambda, ymin = g - SE.g, ymax = g + SE.g),
                  width = .01, color = "green") +
    labs(x = expression(log(theta)), y = expression(g(theta))) +
    ##ylim(-0.001, 0.006) +
```

```
    xlim(-4, 4) +
    geom_vline(xintercept = 0.0, linetype = "dotted", color = "blue") +
    geom_hline(yintercept = 0.0, linetype = "dotted", color = "blue") +
    geom_line(mapping = aes(x = lambda, y = tg),
              linetype = "dashed", color = "red") +
    annotate("text", x = c(-4, -3, -2, -1, 0, 1, 2, 3, 4),
             y = rep(-0.0005, 9),
             label = c("0.02", "0.05", "0.14", "0.37", "1.00", "2.72", "7.39", "20.09", "90.02"), size =
    scale_y_continuous(breaks = c(-0.0005, 0.0, 0.002, 0.004, 0.006),
                       labels = c(expression(theta), "0.000", "0.002", "0.004", "0.006"),
                       limits = c(-0.0005, 0.006)) +
    labs(caption="Figure 1")
)
```

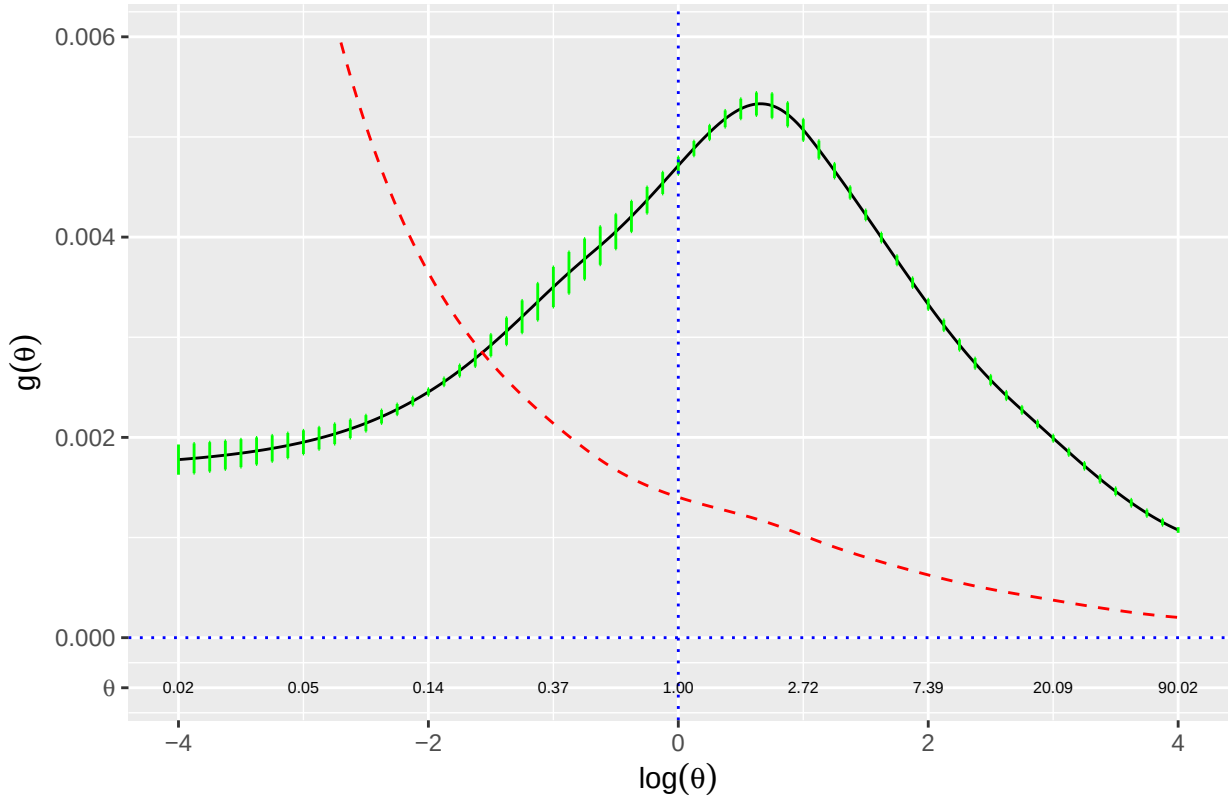
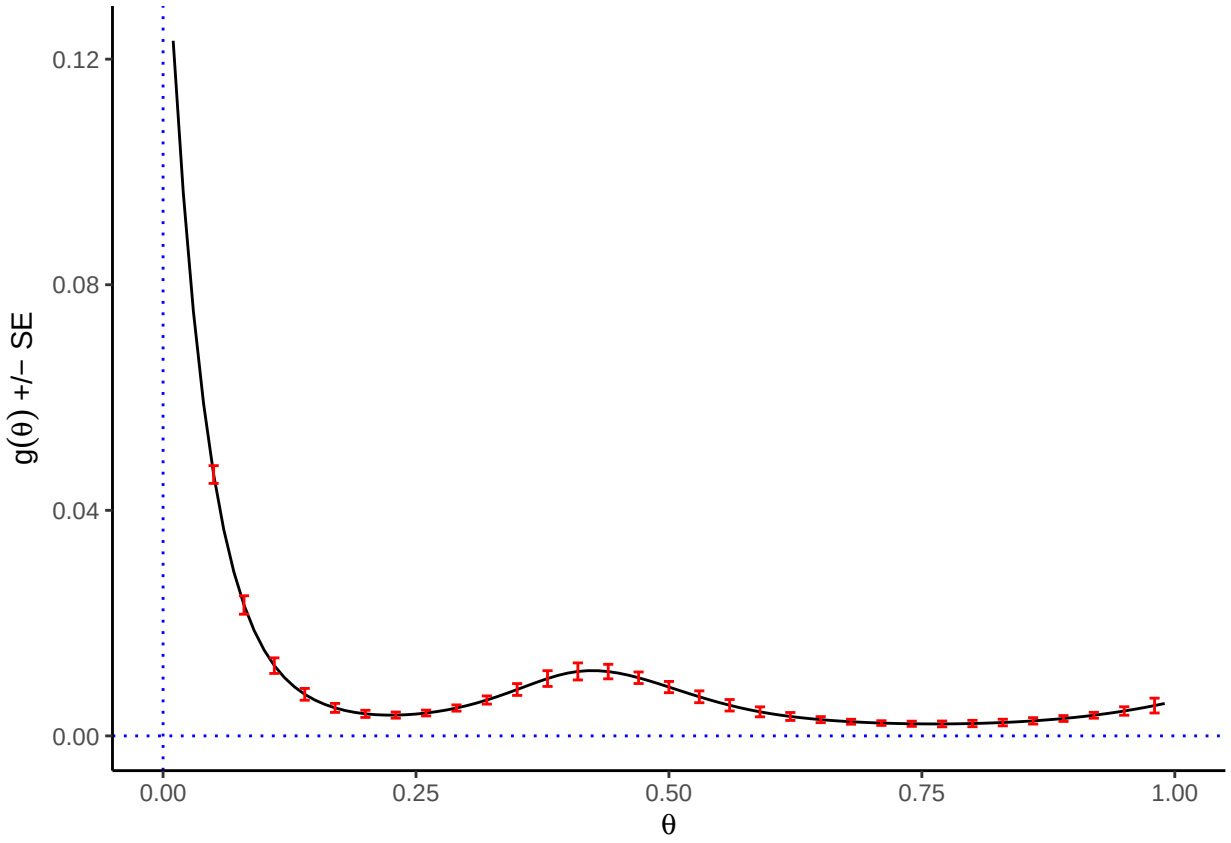


Figure 1

A Medical Example

```
data(surg)
p <- surg$x/surg$n
tau <- seq(from = 0.01, to = 0.99, by = 0.01)
result <- deconv(tau = tau, X = surg, family = "Binomial")
d <- data.frame(result$stats)
indices <- seq(5, 99, 3)
errorX <- tau[indices]
ggplot() +
```

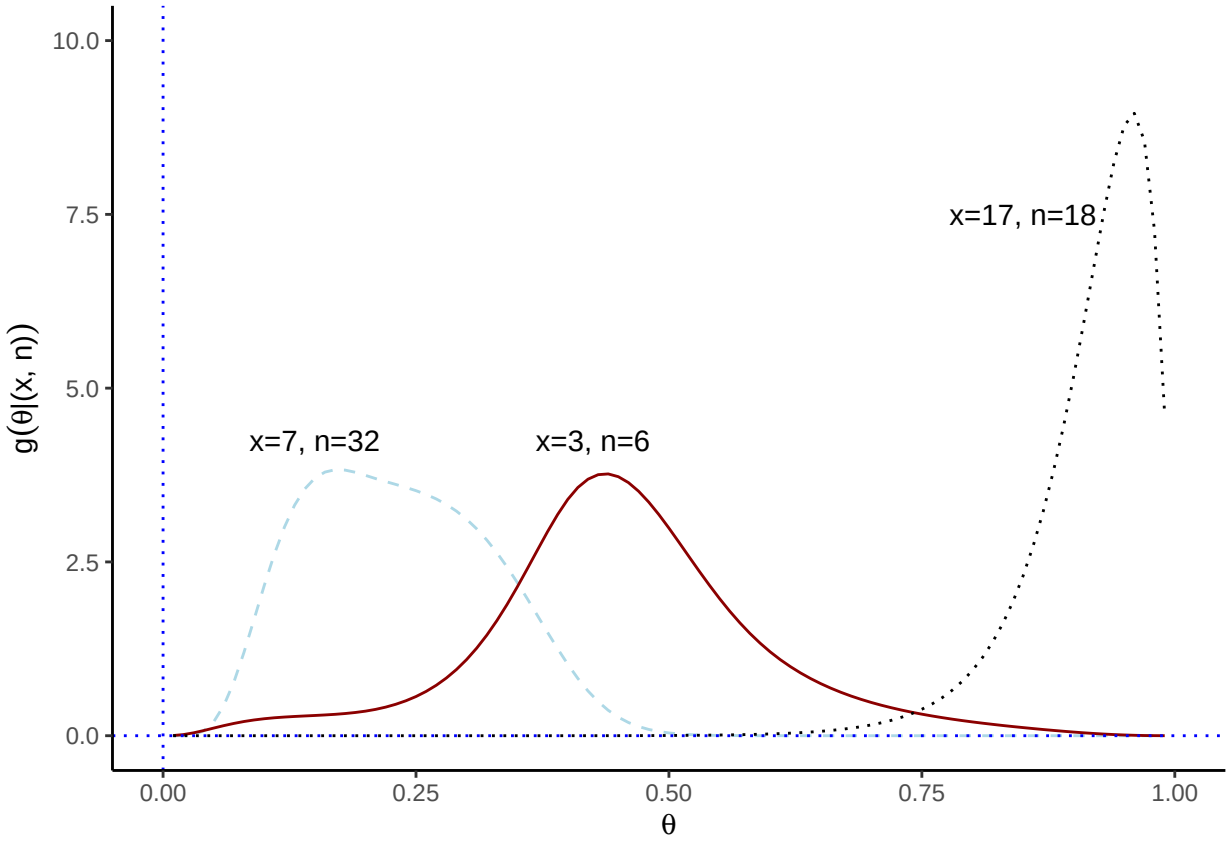
```
  geom_line(data = d, mapping = aes(x = tau, y = g)) +
  geom_errorbar(data = d[indices, ],
                mapping = aes(x = theta, ymin = g - SE.g, ymax = g + SE.g), width = .01, color = "red")
  geom_vline(xintercept = 0.0, linetype = "dotted", color = "blue") +
    geom_hline(yintercept = 0.0, linetype = "dotted", color = "blue") +
  labs(x = expression(theta), y = expression(paste(g(theta), " +/- SE")))+
  expand_limits(x=c(0,1), y=c(0, 0.12,0.02))+
  theme(panel.grid.major = element_blank(), panel.grid.minor = element_blank(),
panel.background = element_blank(), axis.line = element_line(colour = "black"))
```




```
theta <- result$stats[, 'theta']
gTheta <- result$stats[, 'g']
f_alpha <- function(n_k, x_k) {
    ## .01 is the delta_theta in the Riemann sum
    sum(dbinom(x = x_k, size = n_k, prob = theta) * gTheta) * .01
}
g_theta_hat <- function(n_k, x_k) {
    gTheta * dbinom(x = x_k, size = n_k, prob = theta) / f_alpha(n_k, x_k)
}

g1 <- g_theta_hat(x_k = 7, n_k = 32)
g2 <- g_theta_hat(x_k = 3, n_k = 6)
g3 <- g_theta_hat(x_k = 17, n_k = 18)
ggplot() +
    geom_line(mapping = aes(x = theta, y = g1), col = "lightblue",linetype = "dashed") +
    ylim(0, 10) +
```

```
    geom_line(mapping = aes(x = theta, y = g2), col = "darkred",) +
    geom_line(mapping = aes(x = theta, y = g3), col = "black", linetype ="dotted") +
    labs(x = expression(theta), y = expression(g(paste(theta, "|(x, n)")))) +
    geom_vline(xintercept = 0.0, linetype = "dotted", color = "blue") +
    geom_hline(yintercept = 0.0, linetype = "dotted", color = "blue") +
    annotate("text", x = 0.15, y = 4.25, label = "x=7, n=32") +
    annotate("text", x = 0.425, y = 4.25, label = "x=3, n=6") +
    annotate("text", x = 0.85, y = 7.5, label = "x=17, n=18") +
    expand_limits(x=c(0,1,0.2), y=c(0, 6,2))+
    theme(panel.grid.major = element_blank(), panel.grid.minor = element_blank(),
panel.background = element_blank(), axis.line = element_line(colour = "black"))
```



Reference:

https://github.com/jrfiedler/CASI_Python/blob/master/chapter06/ch06s01.ipynb

https://github.com/jrfiedler/CASI_Python/blob/master/chapter06/ch06s02.ipynb

Professor Haviland Wright's class note: "File deconvolveR hw.R"