Homework 4

Chris Cameron

9/15/2021

1. (10 points, 5 points each) Hoff, 3.10 (Change of variables).

$$p_\psi(\psi) = p_\theta(h(\psi))\left|\frac{dh}{d\psi}\right|$$

a) $\theta \sim beta(a,b)$ & $\psi = log[\theta/(1-\theta)]$. Obtain the form of p_ψ and plot it for the case that $a = b = 1$ where $\theta = h(\psi)$.

$$\psi = log\left[\frac{\theta}{1-\theta}\right]$$

$$e^\psi = \frac{\theta}{1-\theta}$$

$$\theta = e^\psi(1-\theta)$$

$$\theta = e^\psi - \theta e^\psi$$

$$e^\psi = \theta + \theta e^\psi$$

$$e^\psi = \theta(1+e^\psi)$$

$$\theta = \frac{e^\psi}{1+e^\psi}$$

$$\theta =: h(\psi) = \frac{e^\psi}{1+e^\psi}$$

$$\left|\frac{dh}{d\psi}\right| = \frac{e^\psi}{(1+e^\psi)^2}$$

$$p_\theta(h(\psi)) = p_\theta(\theta) = \frac{1}{B(a,b)}[h(\psi)]^{a-1}[1-h(\psi)]^{b-1}$$

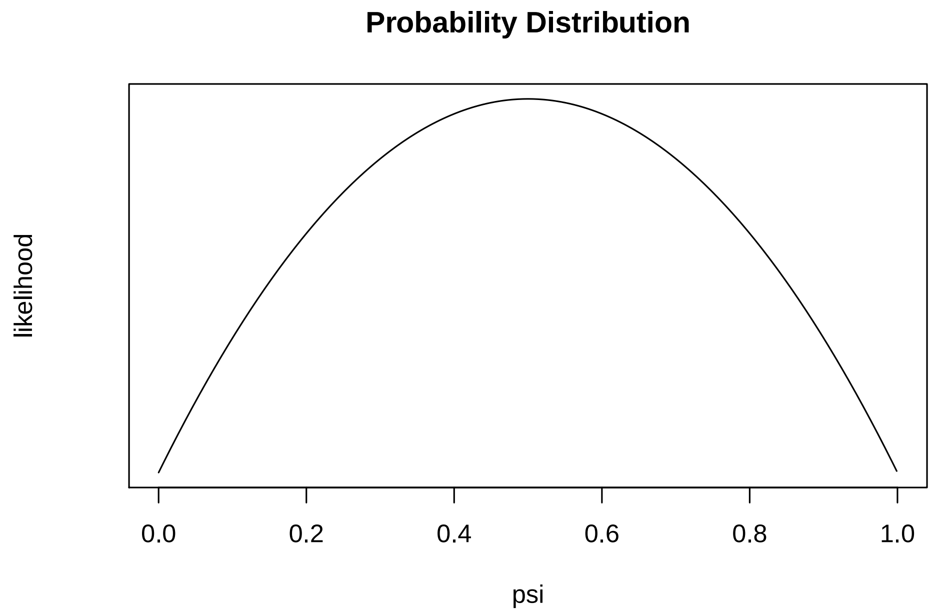
$$p_\psi(\psi) = p_\theta(h(\psi))\left|\frac{dh}{d\psi}\right| = \frac{1}{B(a,b)}[h(\psi)]^{a-1}[1-h(\psi)]^{b-1}\frac{e^\psi}{(1+e^\psi)^2}$$

$$= \frac{1}{B(a,b)} \left[\frac{e^{\psi}}{1+e^{\psi}} \right]^{a-1} \left[\frac{1}{1+e^{\psi}} \right]^{b-1} \frac{e^{\psi}}{(1+e^{\psi})^2}$$

$$= \frac{1}{B(a,b)} \left[\frac{e^{\psi}}{1+e^{\psi}} \right]^{a} \left[\frac{1}{1+e^{\psi}} \right]^{b} \frac{e^{\psi}}{(1+e^{\psi})^2} \left(\frac{e^{\psi}}{1+e^{\psi}} \right) \left(\frac{1}{1+e^{\psi}} \right)$$

$$p_{\psi}(\psi) = \frac{1}{B(a,b)} \left[\frac{e^{\psi}}{1+e^{\psi}} \right]^{a} \left[\frac{1}{1+e^{\psi}} \right]^{b}$$

```r
set.seed(1)
theta.sim <- seq(from = 0, to = 1, length.out = 1000)
a <- 1
b <- 1

y <- (gamma(a+b)/gamma(a)*gamma(b))*(((exp(log((theta.sim)/(1-theta.sim))))/(1+exp(log((thet


plot(theta.sim, y, type = "l", main = "Probability Distribution",
     xlab = "psi", yaxt = "n", ylab = "likelihood")
```



b) Let $\theta \sim gamma(a, b)$ and let $\psi = log(\theta)$. Obtain the form of p_{ψ} and plot it for the case that $a = b = 1$.

$$p_{\psi}(\psi) = p_{\theta}(h(\psi)) \left| \frac{dh}{d\psi} \right|$$

$$log(\theta) = \psi$$
$$\theta = e^{\psi}$$

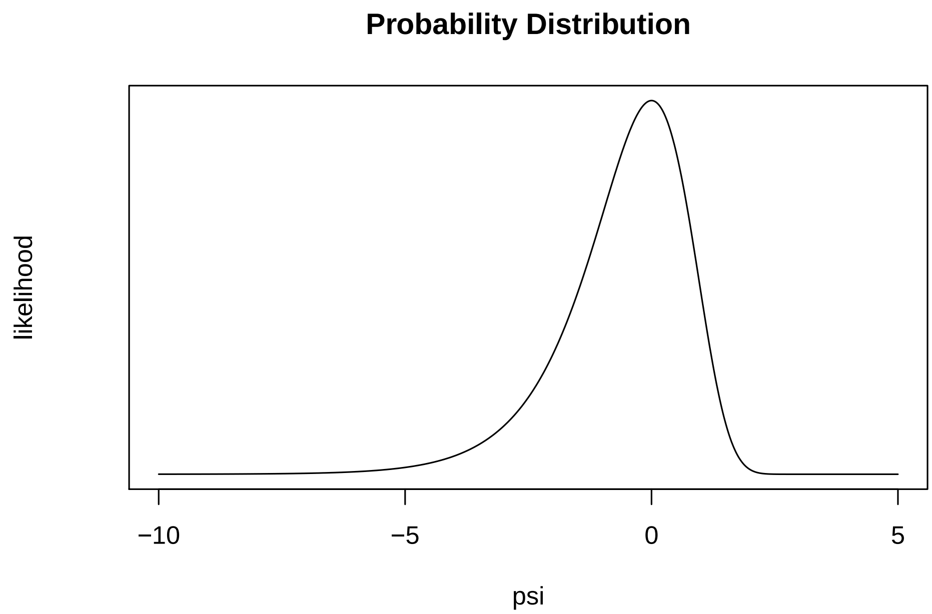
$$\theta =: h(\psi) = e^{\psi}$$

$$\left|\frac{dh}{d\psi}\right| = e^{\psi}$$

$$p_\theta(h(\psi)) = p_\theta(\theta) = \frac{b^a}{\Gamma(a)}\theta^{a-1}e^{-b\theta}$$
$$p_\theta(h(\psi)) = p_\theta(\theta) = \frac{b^a}{\Gamma(a)}(e^{\psi})^{a-1}e^{-b(e^{\psi})}$$
$$p_\theta(h(\psi)) = p_\theta(\theta) = \frac{b^a}{\Gamma(a)}e^{\psi a-\psi}e^{-be^{\psi}}$$
$$p_\theta(h(\psi)) = p_\theta(\theta) = \frac{b^a}{\Gamma(a)}e^{\psi a-\psi-be^{\psi}}$$

$$p_\psi(\psi) = p_\theta(h(\psi))\left|\frac{dh}{d\psi}\right| = \frac{b^a}{\Gamma(a)}e^{\psi a-\psi-be^{\psi}}e^{\psi}$$
$$p_\psi(\psi) = \frac{b^a}{\Gamma(a)}e^{\psi a-be^{\psi}}$$

```r
theta.sim <- seq(from = -10, to = 5, length.out = 1000)
a <- 1
b <- 1

y <- ((b^a)/(gamma(a)))*(exp((theta.sim*a)-(b*exp(theta.sim))))

y[is.nan(y)] <- 0

plot(theta.sim, y, type = "l", main = "Probability Distribution",
     xlab = "psi", yaxt = "n", ylab = "likelihood")
```



(25 points total) Please refer to lab 4 and complete tasks 4—5.

```
# input data
# spurters
x = c(18, 40, 15, 17, 20, 44, 38)
# control group
y = c(-4, 0, -19, 24, 19, 10, 5, 10,
      29, 13, -9, -8, 20, -1, 12, 21,
      -7, 14, 13, 20, 11, 16, 15, 27,
      23, 36, -33, 34, 13, 11, -19, 21,
      6, 25, 30,22, -28, 15, 26, -1, -2,
      43, 23, 22, 25, 16, 10, 29)
# store data in data frame
iqData = data.frame(Treatment = c(rep("Spurters", length(x)),
                                  rep("Controls", length(y))),
                                  Gain = c(x, y))


prior = data.frame(m = 0, c = 1, a = 0.5, b = 50)

findParam = function(prior, data){
  postParam = NULL
  c = prior$c
  m = prior$m
  a = prior$a
  b = prior$b
  n = length(data)
```

```r
  postParam = data.frame(m = (c*m + n*mean(data))/(c + n),
                 c = c + n,
                 a = a + n/2,
                 b =  b + 0.5*(sum((data - mean(data))^2)) +
                   (n*c *(mean(data)- m)^2)/(2*(c+n)))
  return(postParam)
}
postS = findParam(prior, x)
postC = findParam(prior, y)
```

Task 4

```r
# Number of posterior simulations
sim = 100000

# initialize vectors to store samples
mus = NULL
lambdas = NULL
muc = NULL
lambdac = NULL

# Following formula from the NormalGamma with
# the update paramaters accounted accounted for below

lambdas = rgamma(sim, shape = postS$a, rate = postS$b)
lambdac = rgamma(sim, shape = postC$a, rate = postC$b)


mus = sapply(sqrt(1/(postS$c*lambdas)),rnorm, n = 1, mean = postS$m)
muc = sapply(sqrt(1/(postC$c*lambdac)),rnorm, n = 1, mean = postC$m)

# Store simulations
simDF = data.frame(lambda = c(lambdas, lambdac),
                   mu = c(mus, muc),
                   Treatment = rep(c("Spurters", "Controls"),
                                   each = sim))

simDF$lambda = simDF$lambda^{-0.5}



sgtc = 0 #Stands for "spurters greater than control"
```

```r
#Calculate proportion of spurters with greater change in IQ than the control group
for(i in 1:100000){
  if(simDF[i, "mu"] > simDF[i+100000, "mu"]){
    sgtc <- sgtc+1
  }
}

print(sgtc/100000)
```

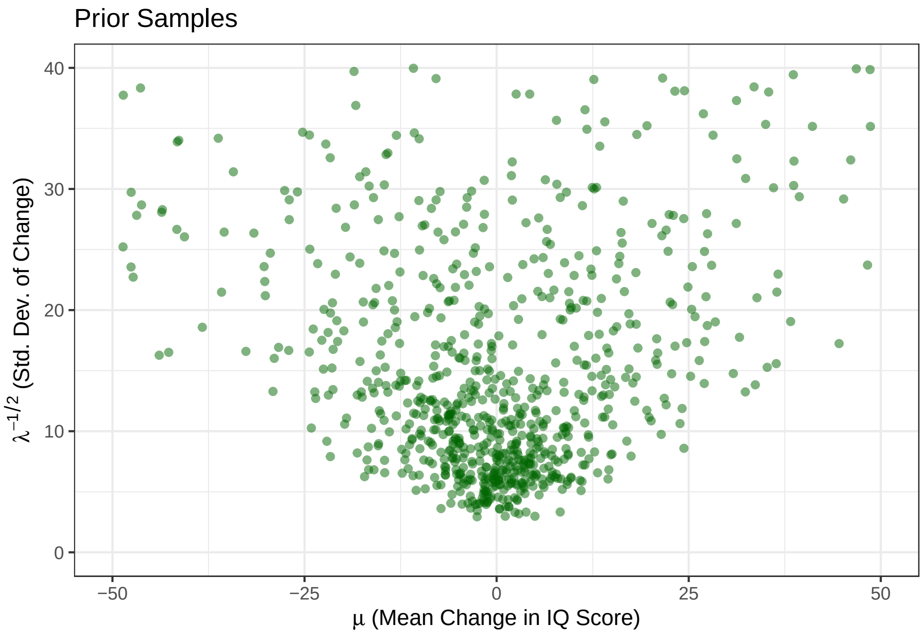
```
## [1] 0.97111
```

There is a predicted 97% chance that a randomly selected spurter will experience a greater increase in their IQ score over a year than a randomly selected child from the control group.

Task 5

```r
# Following formula from the NormalGamma

lambda_prior = rgamma(1000, shape = prior$a, rate = prior$b)

mu_prior = sapply(sqrt(1/(prior$c*lambda_prior)),rnorm, n = 1, mean = prior$m)

# Store simulations
simDF = data.frame(lambda = lambda_prior,
                   mu = mu_prior)

simDF$lambda = simDF$lambda^{-0.5}

# Plot the simulations
ggplot(data = simDF, aes(x = mu, y = lambda)) +
  geom_point(color = "#006600", alpha = 0.5) +
  xlim(-50, 50) + ylim(0, 40) +
  labs(x = expression(paste(mu, " (Mean Change in IQ Score)")),
       y = expression(paste(lambda^{-1/2}, " (Std. Dev. of Change)")))  +
  ggtitle("Prior Samples")+
  theme(plot.title = element_text(hjust = 0.5)) +
  theme_bw()
```

```
## Warning: Removed 212 rows containing missing values (geom_point).
```



The above plot does seem to correspond to our beliefs about the prior. The points are centered around a mean change of 0, which matches our initial belief that the change in IQ score is 0. Fuethermore, while the majority of the points lie below the predicted standard deviation of 10, the distribution of deviations appears to be skewed right, meaning that the mean standard deviation is likely 10.