

HW1

1 A Exponential Case

We write $Y \sim \text{Exp}(\theta)$ to indicate that Y has the Exponential distribution, that is, its p.d.f. is

$p(y|\theta) = \text{Exp}(y|\theta) = \theta e^{-\theta y} 1_{\{y>0\}}$

The Exponential distribution has some special properties that make it a good model for certain applications. It has been used to model the time between events (such as neuron spikes, website hits, neutrinos captured in a detector), extreme values such as maximum daily rainfall over a period of one year, or the amount of time until a product fails (lightbulbs are a standard example). Suppose you have data y_1, \dots, y_n which you are modeling as iid observations from an Exponential distribution, and suppose that your prior is $\theta \sim \text{Gamma}(a, b)$, defined as in the previous question.

(a). Derive the formula for the posterior density, $p(\theta|y_{1:n})$. Give the form of the posterior in terms of one of the following distributions: Bernoulli, Beta, Exponential, and Gamma

$$\Pr(Y_{1:n} = y_{1:n}|\theta) = p(y_{1:n}|\theta) = \prod_{i=1}^n p(y_i|\theta) = \prod_{i=1}^n \theta e^{-\theta y_i} 1_{\{y_i>0\}} = \theta^n e^{-\theta \sum y_i} \left(\prod_{i=1}^n 1_{\{y_i>0\}} \right)$$

$$p(\theta) = \text{Gamma}(\theta|a, b) = \frac{b^a}{\Gamma(a)} \theta^{a-1} e^{-b\theta} 1_{\{\theta>0\}}$$

$$p(\theta|y_{1:n}) \propto p(y_{1:n}|\theta)p(\theta) = \left(\theta^n e^{-\theta \sum y_i} \right) \left(\frac{b^a}{\Gamma(a)} \theta^{a-1} e^{-b\theta} 1_{\{\theta>0\}} \right) = \theta^{a+n-1} e^{-(b+\sum y_i)\theta} 1_{\{\theta>0\}} \propto \text{Gamma}(a+n, b+\sum y_i)$$

(b). Now, suppose you are measuring the number of seconds between lightning strikes during a storm, your prior is $\text{Gamma}(0.1, 1.0)$, and your data is

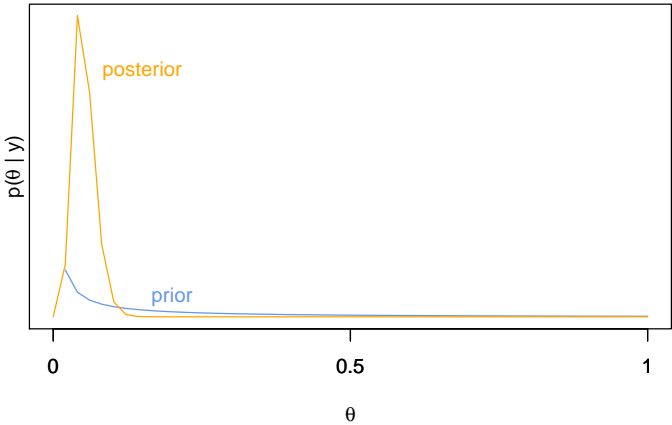
$(y_1, \dots, y_8) = (20.9, 69.7, 3.6, 21.8, 21.4, 0.4, 6.7, 10.0)$

Plot the prior and posterior pdf's. (Be sure to make your plots on a scale that allows you to clearly see the important features.)

```
y <- cumsum(c(20.9,69.7,3.6,21.8,21.4,0.4,6.7,10.0))
N <- 1:8
k <- 8
ps <- seq(0,1,length.out = 50)
prior <- dgamma(ps,shape=0.1,rate=1)
post <- dgamma(ps,shape=(0.1+N[k]),rate=(1+y[k]))

plot(x=ps,post,type="n",ylim=c(0,22),
     main="",xlab=expression(theta),ylab="",xaxt="n",yaxt="n")
axis(side = 1, at = c(0,0.5,1),labels = c(0,0.5,1), cex=0.7)
lines(x=ps,prior,type="l",ylim=c(0,22),ylab="",xaxt="n",yaxt="n",col="cornflowerblue")
text(x=0.2,1.5,labels="prior",cex=1,col="cornflowerblue")

lines(x=ps,post,type="l",ylim=c(0,22),#range(c(prior,post)),
     ylab="",xaxt="n",yaxt="n",col="orange")
axis(side = 1, at = c(0,0.5,1),labels = c(0,0.5,1), cex=0.7)
text(x=0.15,18,labels="posterior",cex=1,col="orange")
mtext(expression(paste("p(",theta," | ",y,")")),side=2,line=0)
```



(c). Give a specific example of an application where an Exponential model would be reasonable. Give an example where an Exponential model would NOT be appropriate, and explain why.

2 A Poisson Case

An ecologist records the number of eggs laid in a sample of sparrow nests of size $n = 20$. Let Y_i be the number of eggs laid in nest i for $i = 1, \dots, 20$. Based on this sample, the ecologist is interested in estimating θ , the mean number of eggs per nest in the general population of nests. Assume $Y_1, \dots, Y_n | \theta \sim \text{Poisson}(\theta)$ iid, and also for now that $\theta \in \Theta = \{0.1, 0.2, \dots, 4.9, 5.0\}$ and that $p(\theta) = 1/50$ for each $\theta \in \Theta$.

- Let $X = \sum_i Y_i$ denote the random variable that counts the total number of eggs laid in the 20 nests. Using the form of Bayes rule on page 15 of the Hoff book, write down a formula for $p(\theta|x)$ and simplify as much as possible.
- The ecologist observes that $x = 36$. Make a plot of $p(\theta|x)$ versus θ for $\theta \in \Theta$
- Find $E[\theta|x]$, the posterior mean of θ .
- Find two numbers, θ_l and θ_h such that $Pr(\theta_l \leq \theta \leq \theta_h|x) \approx 0.95$. This is an approximate 95% posterior confidence interval. Note that there is more than one way of doing this.
- Remake the plot in part (b) but with $n = 40$ and $x = 72$. Describe and explain the differences you see between this plot and the one in (b).

3 posterior predictive density

Suppose the data $y_{1:n} | \theta$ is modeled as iid $\text{Exp}(\theta)$, and the prior is

$$p(\theta) = \text{Gamma}(\theta|a, b) = \frac{b^a}{\Gamma(a)} \theta^{a-1} e^{-b\theta} 1_{\{\theta > 0\}}$$

From problem 1, we know that the posterior is $p(\theta|y_{1:n}) = \text{Gamma}(\theta|a_n, b_n)$, where $a_n = a + n$ and $b_n = b + \sum_{i=1}^n y_i$. What is the posterior predictive density $p(y_{n+1}|y_{1:n})$? Give your answer as a closed-form expression (not an integral).

$$\begin{aligned} p(y_{n+1}|y) &= p(y_{n+1}|y_{1:n}) = \int_{\theta \in \Theta} p(y_{n+1}, \theta|y_{1:n}) d\theta = \int_{\theta \in \Theta} p(y_{n+1}|\theta) p(\theta|y_{1:n}) d\theta = \int_{\theta \in \Theta} p(y_{n+1}|\theta) p(\theta|y_{1:n}) d\theta \\ &= \int_{\theta \in \Theta} \theta e^{-\theta y_{n+1}} \frac{(b + \sum y_i)^{a+n}}{\Gamma(a+n)} \theta^{a+n-1} e^{-(b + \sum y_i)\theta} d\theta \\ &= \frac{(a+n)(b + \sum y_i)^{a+n}}{(b + \sum y_i + y_{n+1})^{a+n+1}} \int_{\theta \in \Theta} \frac{(b + \sum y_i + y_{n+1})^{a+n+1}}{\Gamma(a+n+1)} \theta^{a+n} e^{-(b + \sum y_i + y_{n+1})\theta} d\theta \\ &= \frac{(a+n)(b + \sum y_i)^{a+n}}{(b + \sum y_i + y_{n+1})^{a+n+1}} \end{aligned}$$

4 (decision theory)

Suppose that in the small imaginary city of our class example, where the prevalence of a rare disease was studied (Foundations: Section 4, page 11), public health officials need to decide the amount of resources to allocate towards prevention and treatment of the disease we are concerned with, with the fraction of infected individuals θ still unknown. They will decide on the resources needed based on a fraction c of the population. If c is chosen too large, there will be wasted resources, while if it is too small, preventable cases may occur and some individuals may go untreated. After some deliberation, they tentatively adopt the following loss function:

$$\ell(\theta, c) = (\theta - c + 0.5)^2 \text{ for } c \in [0, 1]$$

- Assume that the number of people sampled is again $n = 20$, that $y = 0$, and use $\theta \sim \text{Beta}(2, 20)$ as the prior again to derive the posterior expected loss.

$$p(\theta|y) \propto p(y|\theta)p(\theta) = \binom{20}{y} \theta^y (1-\theta)^{20-y} \frac{1}{B(2, 20)} \theta^{2-1} (1-\theta)^{20-1} \propto \theta^{2+0-1} (1-\theta)^{20+(20-0)-1} \propto \text{Beta}(2, 40)$$

$$E(\theta|Y=0) = \frac{a+0}{(a+0)+(b+n-0)} = \frac{2}{42}$$

$$\begin{aligned} \text{Let } \hat{\theta} &= c - 0.5, \ell(\theta, c) = (\theta - c + 0.5)^2 = \theta^2 - 2\theta\hat{\theta} + \hat{\theta}^2, \\ \rho(\hat{\theta}, y_{1:n}) &= E(\ell(\theta, \hat{\theta})|y_{1:n}) = E(\theta^2|y_{1:n}) - 2E(\theta|y_{1:n})\hat{\theta} + \hat{\theta}^2 \\ &= E(\theta^2|y_{1:n}) - 2E(\theta|y_{1:n})(c - 0.5) + (c - 0.5)^2 \end{aligned}$$

- Find an optimal value (call it \hat{c}) for c following a Bayesian decision procedure using the set up from part (a).

$$\frac{d}{d\hat{\theta}} \rho(\hat{\theta}, y_{1:n}) = -2E(\theta|y_{1:n}) + 2\hat{\theta} = 0$$

$$\hat{\theta} = \delta(y_{1:n}) = E(\theta|y_{1:n}) = \frac{2}{42} = c - 0.5 \implies \hat{c} = \frac{23}{42}$$

- Graphically compare $\ell(\theta, \hat{c})$ and $\ell(\theta, \bar{y})$ as θ ranges from 0 to 1.

```
n <- 20
a <- 2
b <- 20
y <- 0

lossfn <- function(theta,c){

  loss <- (theta-c+0.5)^2

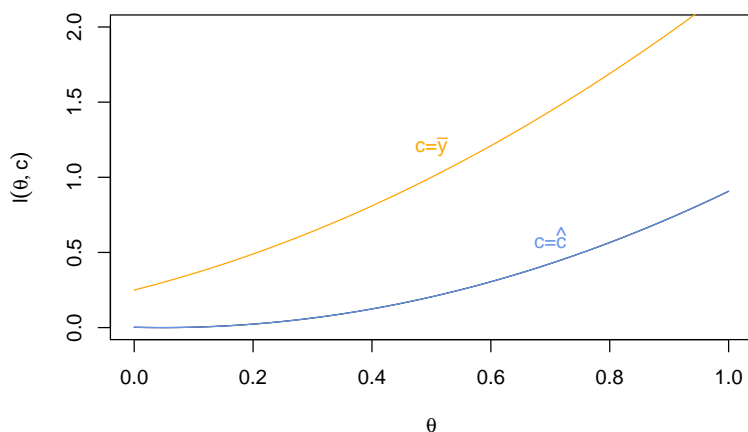
}
```

```

theta.values <- seq(0,1,by=0.005)
loss.values <- lossfn(theta.values,23/42)

plot(x=theta.values,loss.values,type="l",ylim=c(0,2),
     main="",xlab=expression(theta),ylab=expression(l(theta,c)))
lines(x=theta.values,loss.values,type="l",ylim=c(0,2),ylab="",xaxt="n",yaxt="n",col="cornflowerblue")
text(x=0.7,0.6,labels=expression(paste("c=",hat(c))),cex=1,col="cornflowerblue")
lines(x=theta.values,lossfn(theta.values,0),type="l",ylim=c(0,2),#range(c(prior,post)),
     ylab="",xaxt="n",yaxt="n",col="orange")
text(x=0.5,1.2,labels=expression(paste("c=",bar(y))),cex=1,col="orange")

```



(d). Compare the outcome from the Bayesian decision procedure to the procedures that choose (i) $c = \bar{y}$ and (ii) $c = 0.1$ constant (i.e., setting it to the prior mean). This comparison can be done by observing the optimal quantity c derived from each of the decision procedures considered while varying the value of y (the number of infected cases).

```

n <- 20
a <- 2
b <- 20
yvec <- 0:20
c.fixed <- 0.1
c.argminvec <- NA

for(y in yvec){
  rhofn <- function(c){

    ff <- function(theta){
      db <- dbeta(theta,shape1=(a+y),shape2=(b+n-y))
      (theta-c+0.5)^2 * db
    }

    res <- integrate(ff, lower=0, upper=1)
    res$value
  }

  vrhofn <- Vectorize(rhofn, vectorize.args = "c")
  c.values <- seq(0,1,by=0.005)
  rho.values <- vrhofn(c.values)
  #value of c that minimizes the posterior expected loss
  (c.argminvec[y+1] <- c.values[which(rho.values==min(rho.values))])
}

plot(x = yvec,
     y = c.argminvec,
     ylim=c(0,1),

```

```

xlab = "observed number of cases",
ylab = "optimal c",
type = "o",
lwd = 2,
pch=20,
col="orange")
lines(x = yvec, y = yvec/n,col="blue",lwd=2)
points(x = yvec, y = yvec/n,col="blue",pch=20)
lines(x = yvec, y = rep(c.fixed,(n+1)),col="darkgreen",lwd=2)
points(x = yvec, y = rep(c.fixed,(n+1)),col="darkgreen",pch=20)
#add legend to plot
legend("topleft",legend=c("Bayes",expression(bar(y)),"c=0.1"),
      col=c("orange","blue","darkgreen"),
      lty=rep(1,3),pch=rep(20,3),bty="n",ncol=1)

```

