Ma 322: Biostatistics
Homework Assignment 4

Prof. Wickerhauser

Read Chapter 9, "An Introduction to Bayesian Data Analysis," pages 135–159 of our text.

Note: There are many errors in the relevant formulas in our textbook. Use the correct formula
$$P(A|B)P(B) = P(A \cap B) = P(B|A)P(A).$$

1. Suppose that X is a probability space with probability function P, and $A, B \subset X$ are events satisfying $P(A) = 0.6$ and $P(B) = 0.3$.

   (a) If $P(A|B) = 0.4$, what is $P(B|A)$?

   (b) If $P(A \cap B) = 0.1$, what are $P(A|B)$ and $P(B|A)$?

   (c) If $P(A \cap B) = 0.2$, what are $P(A|B^c)$ and $P(B|A^c)$? (Here A^c is the complement of A, while B^c is the complement of B.)

   Solution: (a) Use Bayes' theorem:
   $$P(B|A) = \frac{P(A|B)P(B)}{P(A)} = \frac{(0.4)(0.3))}{(0.6)} = 0.2.$$

   (b) From the definitions:
   $$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{0.1}{0.3} = \frac{1}{3} \approx 0.33; \qquad P(B|A) = \frac{P(A \cap B)}{P(A)} = \frac{0.1}{0.6} = \frac{1}{6} \approx 0.17.$$

   (c) First note that $P(A^c) = 1 - P(A) = 0.4$ and $P(B^c) = 1 - P(B) = 0.7$. Next, note that $A \cap B^c$ and $A \cap B$ are disjoint, and $(A \cap B^c) \cup (A \cap B) = A$. Thus $P(A \cap B^c) + P(A \cap B) = P(A) = 0.6$. Solving gives
   $$P(A \cap B^c) = P(A) - P(A \cap B) = 0.6 - 0.2 = 0.4.$$

Likewise,

$$P(B \cap A^c) = P(B) - P(A \cap B) = 0.3 - 0.2 = 0.1.$$

Finally, compute from the definitions:

$$P(A|B^c) = \frac{P(A \cap B^c)}{P(B^c)} = \frac{0.4}{0.7} \approx 0.57; \qquad P(B|A^c) = \frac{P(B \cap A^c)}{P(A^c)} = \frac{0.1}{0.4} = \frac{1}{4} = 0.25.$$

$\square$

2. Candidate prior pdfs from the Beta family of densities have the following form:

$$f_{\alpha,\beta}(x) = \frac{1}{\beta(\alpha,\beta)} x^{\alpha-1}(1-x)^{\beta-1}.$$

(a) Find the mean and standard deviation of $f_{\alpha,\beta}$ in terms of α and β. (Hint: see p.103 of our textbook.)

(b) For what values of α, β is $f_{\alpha,\beta}$ the uniform density on $[0,1]$? Plot this beta pdf to check your result.

(c) Professor Bayes believes that an experiment has probability $p = 0.3$ of succeeding, but concedes that his p is the mean of a beta density with standard deviation 0.1. To model these beliefs, what values of α, β should Bayes use in the prior pdf $f_{\alpha,\beta}(p)$? Plot your solution to check your work. Evaluate this with the `R` command `dbeta(x,shape1=alpha,shape2=beta)`.

Solution: (a) The mean of $f_{\alpha,\beta}$ is $\alpha/(\alpha+\beta)$. The variance is $\alpha\beta/((\alpha+\beta)^2(\alpha+\beta+1))$, so the standard deviation is $\sqrt{\alpha\beta}/(|\alpha+\beta|\sqrt{\alpha+\beta+1})$. These formulas may be found on page 103 of our text.

(b) Choosing $\alpha = \beta = 1$ gives the uniform density $f_{1,1}(x) = 1$, all $x \in [0,1]$. This is plotted in Figure HW 4, Ex.2i.

(c) Bayes sets the mean $0.3 = \alpha/(\alpha+\beta)$ to conclude that $\beta = 7\alpha/3$. Bayes then sets the standard deviation

$$0.1 = \frac{\sqrt{\alpha\beta}}{(\alpha+\beta)\sqrt{\alpha+\beta+1}} = \frac{\alpha\sqrt{\frac{7}{3}}}{(\frac{10}{3}\alpha)\sqrt{\frac{10}{\alpha}+1}} = \frac{\sqrt{\frac{7}{3}}}{\frac{10}{3}\sqrt{\frac{10}{3}\alpha+1}},$$

which is solved by $\alpha = 6$, so therefore $\beta = 14$. The resulting posterior density is plotted in Figure HW 4, Ex.2ii.

Note: the pdf $f_{\alpha,\beta}$ has mean μ and standard deviation σ if and only if

$$\alpha = \frac{\mu}{\sigma^2}\left[\mu(1-\mu) - \sigma^2\right]; \qquad \beta = \frac{1-\mu}{\sigma^2}\left[\mu(1-\mu) - \sigma^2\right].$$

The following `R` commands produced the requested beta density plots:

```
x<-seq(0,1,by=0.01);
pdf("beta0.pdf"); plot(x,dbeta(x,1,1)); dev.off();
pdf("beta2.pdf"); plot(x,dbeta(x,6,14)); dev.off();
```

□

3. Individuals from two subpopulations, A and B, of a certain species of passerines (perching birds) are mixed to form the population under study. They carry a gene mutation with respective incidences $p_A = 0.1$ and $p_B = 0.8$, but are otherwise indistinguishable. The gene expresses a detectable plasma protein marker in individuals that have it. Individuals without the gene mutation do not express the marker protein.

   (a) One individual is tested and found to have the plasma marker. Assuming that A and B are present in equal numbers, compute the probability that the individual comes from population A. Then compute the probability that the individual comes from population B.

   (b) A random sample of 93 individuals is tested and 61 of them have the plasma marker. Use Bayes' rule with the uniform prior $f_{\alpha,\beta}(a) = 1$ on the proportion a of A's; include this information and compute the posterior distribution of the proportion of A's and B's.

   Solution: (a) Write $P(A|pp)$ and $P(B|pp)$ for the respective probabilities that the individual comes from A and B, given that they express the plasma protein marker indicative of the gene mutation. Then $P(A|pp) + P(B|pp) = 1$ since the two subpopulations are exhaustive.

   Write $P(gm|A) = 0.1$ and $P(gm|B) = 0.8$ for the respective probabilities that an individual from A or B has the gene mutation. Then $P(pp|A) = 0.1$ and $P(pp|B) = 0.8$ for the respective probabilities of finding the plasma protein marker in an individual from A and B, since the marker is expressed if and only if the gene mutation is present.

   By Bayes' rule and the law of total probability, compute

$$
\begin{aligned}
P(A|pp) &= \frac{P(pp|A)P(A)}{P(pp)} = \frac{P(pp|A)P(A)}{P(pp|A)P(A) + P(pp|B)P(B)} \\
&= \frac{(0.1)(0.5)}{(0.1)(0.5) + (0.8)(0.5)} = \frac{1}{9} \approx .111
\end{aligned}
$$

   Then $P(B|pp) = 1 - P(A|pp) = 8/9 \approx .889$.

   (b) Let $P(a) = f_{1,1}(a)$ be the uniform prior pdf on the proportion a of A's, as computed during the solution of the previous exercise. Then $b = 1 - a$ is the proportion of B's in the population.

The probability of finding the plasma protein marker is given by the following combination of subpopulation probabilities:

$$p = a\,p_A + b\,p_B = 0.1a + 0.8(1-a) = 0.8 - 0.7\,a,$$

The probability of finding k individuals with the marker out of a sample of n individuals is given by the binomial distribution

$$P(pp|a) = \binom{n}{k} p^k (1-p)^{n-k} = \binom{93}{61} (0.8-0.7\,a)^{61}(1-[0.8-0.7\,a])^{93-61},$$

where we have substituted the results of the experiment: 61 individuals of 93 sampled had the marker. Using Bayes' theorem gives the posterior pdf

$$P(a|pp) \propto (0.8-0.7\,a)^{61}(0.2+0.7\,a)^{32},$$

which is a beta density $f_{\alpha,\beta}(p)$ with shape parameters $\alpha = 62$, $\beta = 33$, composed with the substitution $p \leftarrow 0.8 - 0.7\,a$. The graph of this posterior density may be obtained with the following `R` commands:

```
a<-seq(0,1,by=0.01); pA<-0.1; pB<-0.8; p<- a*pA + (1-a)*pB;
pdf("abprop.pdf"); plot(a,dbeta(p,shape1=62,shape2=33)); dev.off()
a[which.max(dbeta(p,shape1=62,shape2=33))]
```

This is a density, plotted in Figure HW 4, Ex.3(b), with a mode (maximum likelihood) near $a = 0.21$, so from the experiment we conclude that the population is about 21% A's and 79% B's. □

4. Given a gamma prior pdf

$$f_{\alpha,\beta}(\theta) = \frac{\beta^\alpha}{\Gamma(\alpha)}\theta^{\alpha-1}e^{-\beta\theta}$$

with shape parameters $\alpha = 4$ and $\beta = 2$, and an experiment that produces a count of $x = 19$ which is expected to have the Poisson density

$$P(x|\theta) = \theta^x e^{-\theta}/x!,$$

find the posterior pdf $P(\theta|x=19)$ for the parameter θ. Ignore normalization constants, but plot the result for a range of θ's that includes the likeliest.

Solution: By Bayes' theorem, compute

$$P(\theta|x=19) \propto P(x=19|\theta)\,f_{4,2}(\theta) = \frac{\theta^{19}e^{-\theta}}{19!}\,\frac{2^4}{\Gamma(4)}\theta^3 e^{-2\theta} \;\propto\; \theta^{22}e^{-3\theta},$$

which is proportional to a gamma pdf $f_{23,3}(\theta)$. Plot the results, both prior and posterior, with

```
x<-c(rgamma(100,shape=4,rate=2),rgamma(100,shape=23,rate=3));
pdf("gammpois.pdf"); plot(x,dgamma(x,shape=4,rate=2));
points(x,dgamma(x,shape=23,rate=3),pch="+"); dev.off()
```

The results are shown in Figure HW 4, Ex.4. Note how sampling the prior and posterior densities to get the abscissas (x values) produced clusters near the most likely values of θ both before and after the experiment will be shown in the plot. □

5. Given a gamma prior pdf

$$f_{\alpha,\beta}(\lambda) = \frac{\beta^\alpha}{\Gamma(\alpha)} \lambda^{\alpha-1} e^{-\beta\lambda}$$

with shape parameters $\alpha = 3$ and $\beta = 5$, and an experiment that produces a time-to-failure of $t = 2$ which is expected to have the exponential density

$$P(t|\lambda) = \lambda e^{-\lambda t},$$

find the posterior pdf $P(\lambda|t = 2)$ for the parameter λ. Ignore normalization constants, but plot the result for a range of λ's that includes the likeliest.

Solution: By Bayes' theorem, compute

$$P(\lambda|t = 2) \propto P(t = 2|\lambda)\, f_{3,5}(\lambda) = \lambda e^{-2\lambda} \frac{5^3}{\Gamma(3)} \lambda^2 e^{-5\lambda} \;\propto\; \lambda^3 e^{-7\lambda},$$

which is proportional to a gamma pdf $f_{4,7}(\lambda)$. Plot the result with

```
x<- c(rgamma(100,3,rate=5), rgamma(100,4,rate=7));
pdf("gammaexp.pdf");  plot(x,dgamma(x,4,rate=7),pch="+");
points(x,dgamma(x,3,rate=5),pch="o"); dev.off()
```

The results are shown in Figure HW 4, Ex.5. Note that the posterior gamma should be plotted first to set the y-axis range wide enough. Also, the x-values are obtained by sampling both the prior and posterior so that their most likely values are within the x range. □

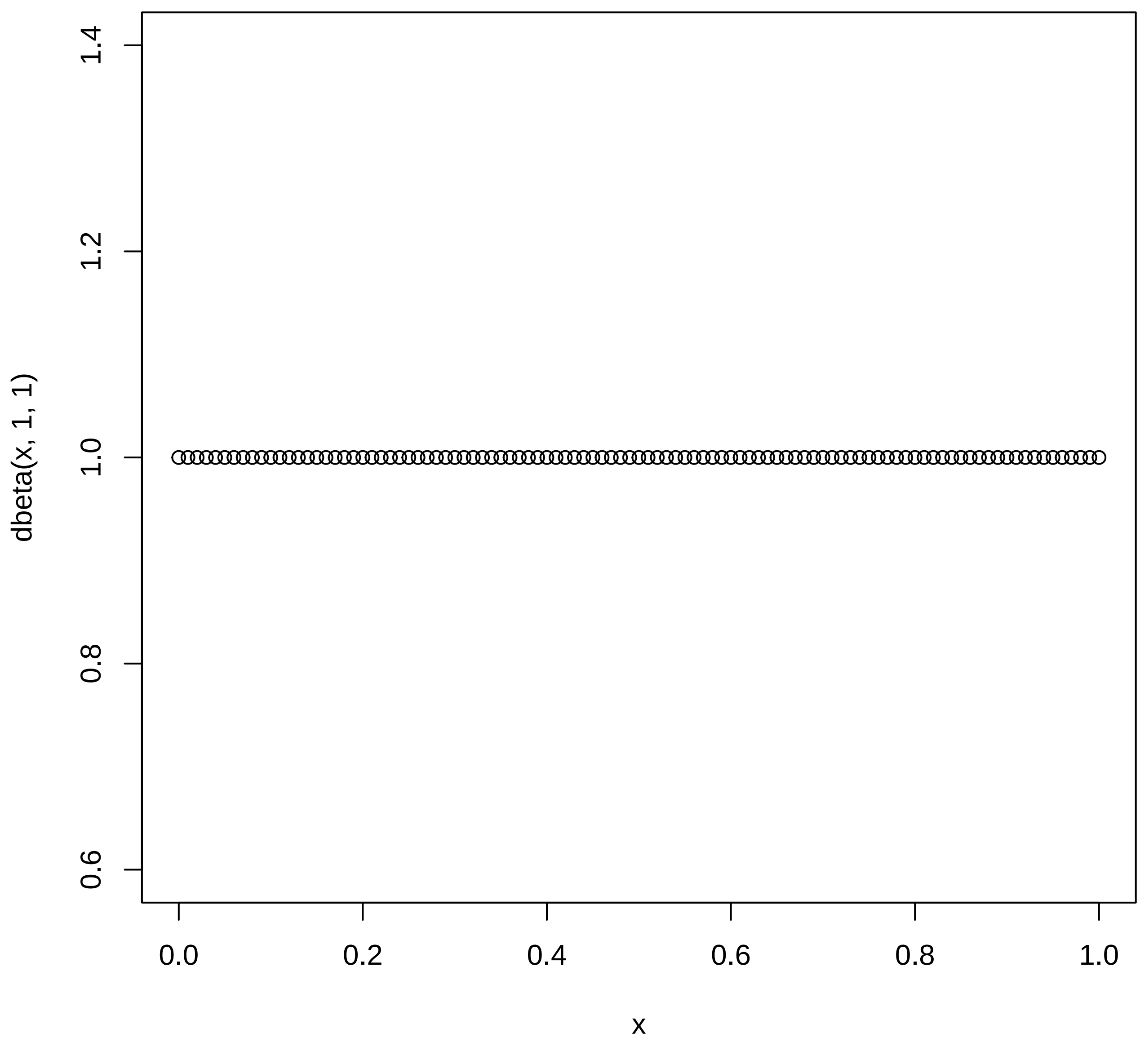

Figure 1: HW 4, Ex.2(b): Beta density, $\alpha = 1$, $\beta = 1$

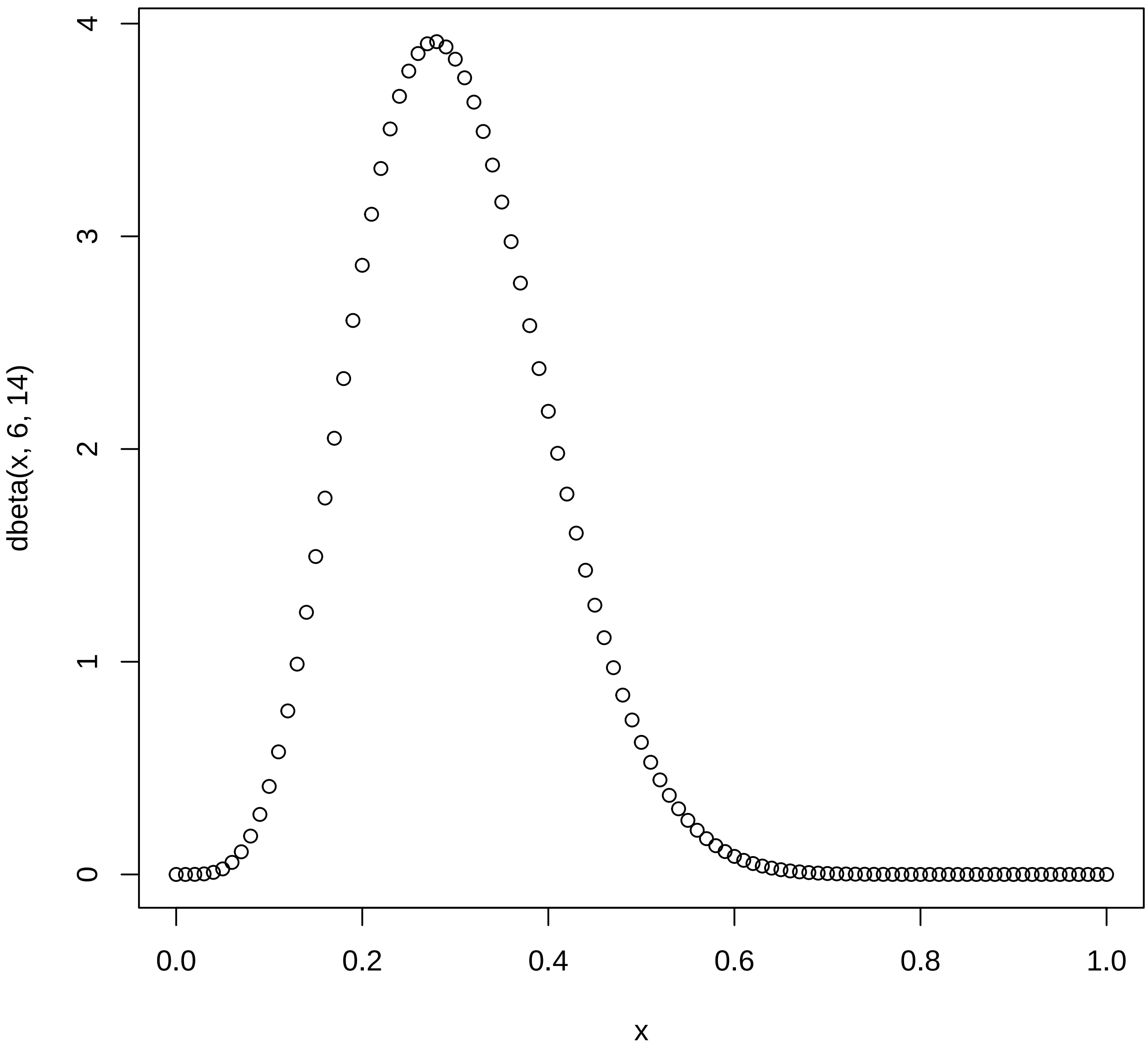

Figure 2: HW 4, Ex.2(c): Beta density, $\alpha = 6$, $\beta = 14$

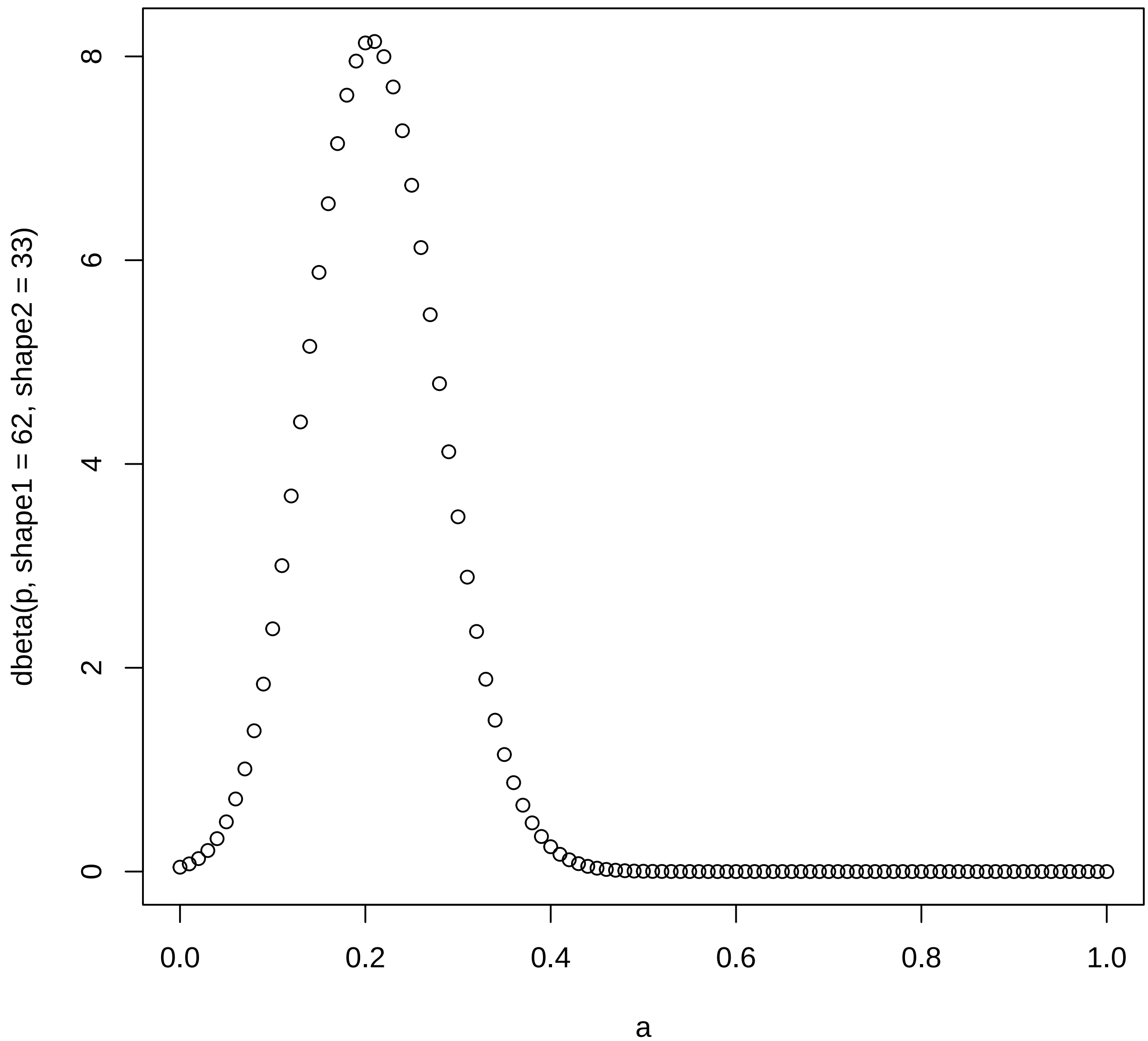

Figure 3: HW 4, Ex.3(b): Beta density, $\alpha = 62$, $\beta = 33$

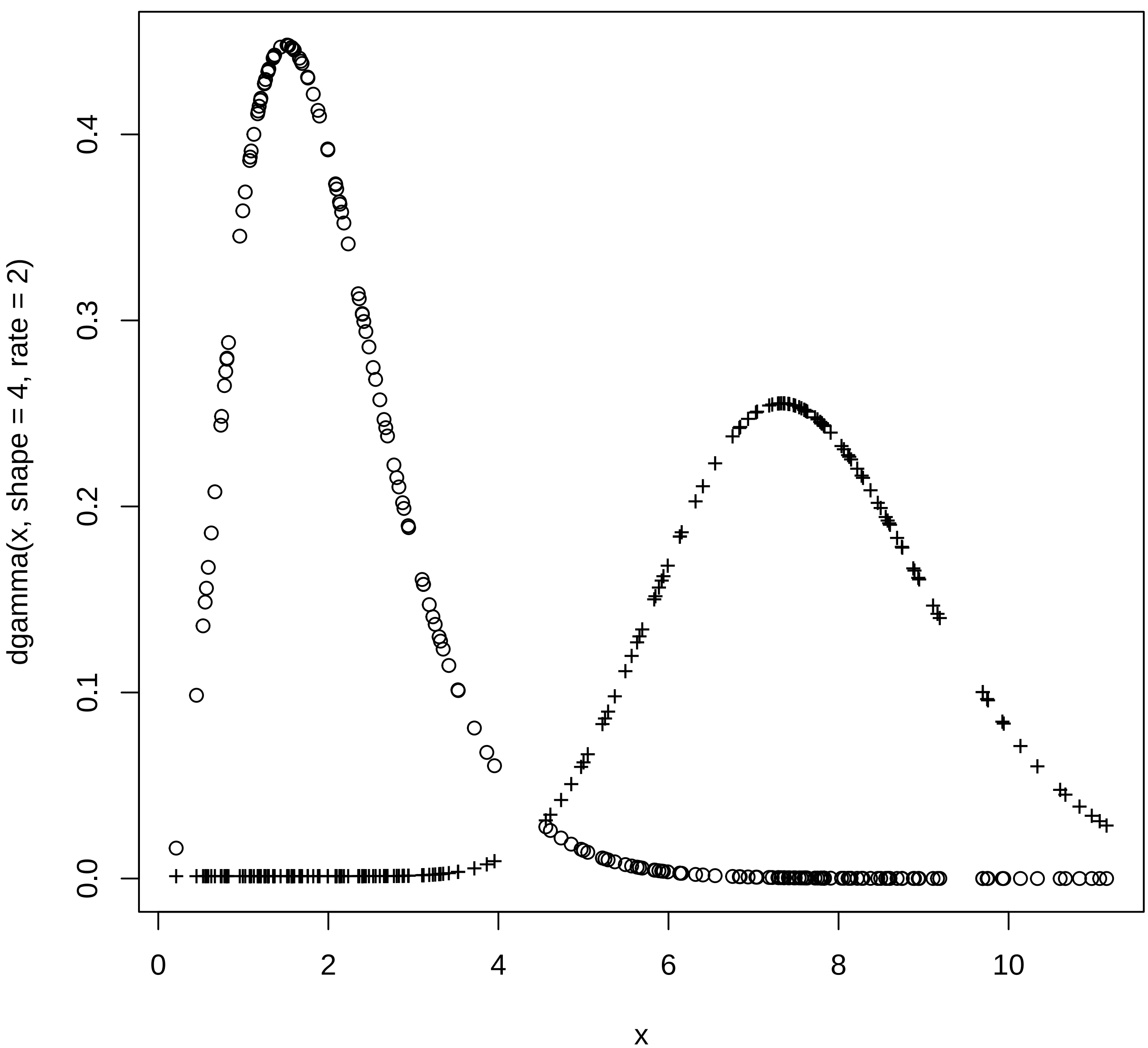

Figure 4: HW 4, Ex.4: Gamma(4,2) prior [o], Gamma(23,3) posterior [+] after Poisson evidence

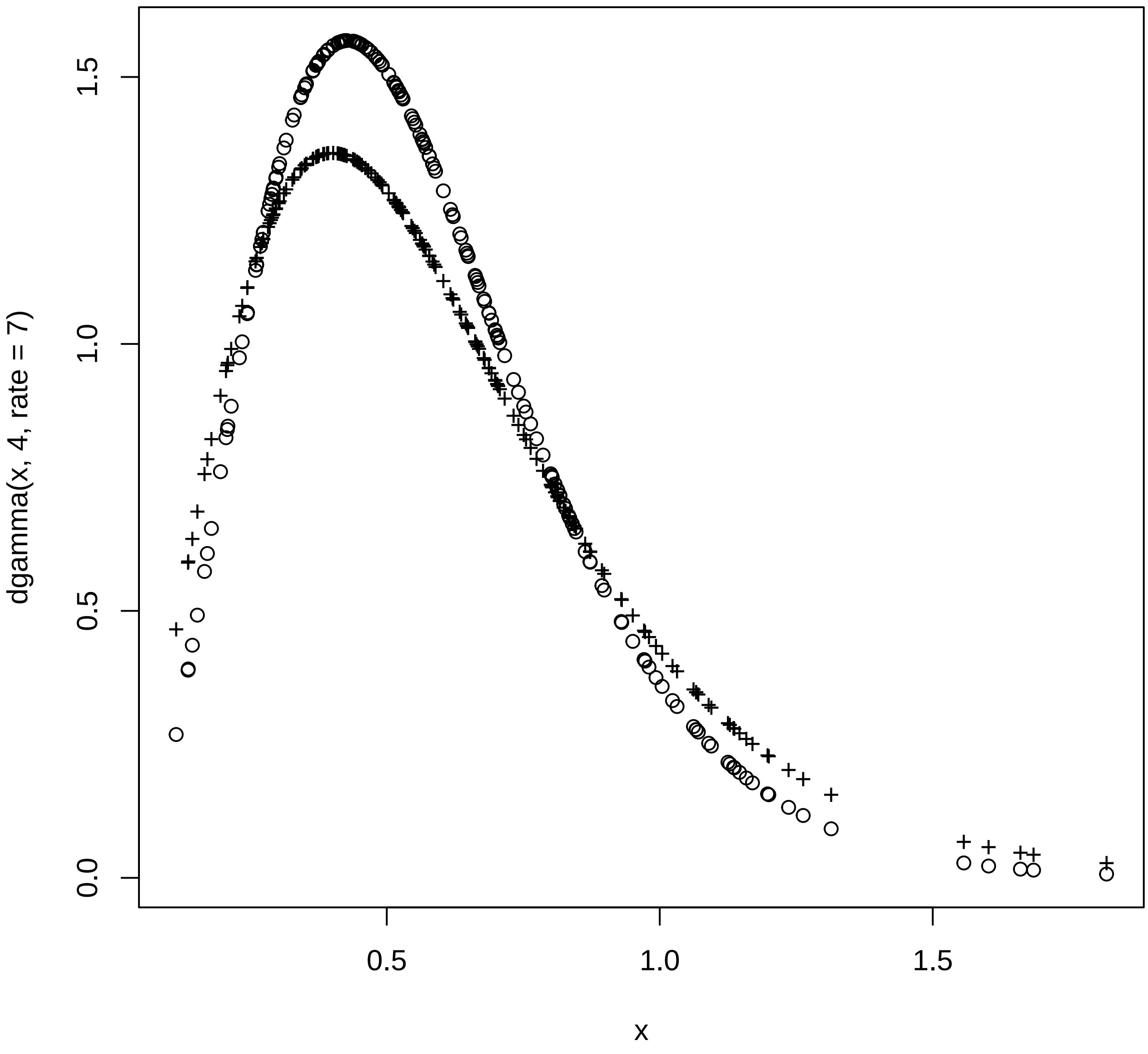

Figure 5: HW 4, Ex.5: Gamma(3,5) prior [o], Gamma(4,7) posterior [+] after Exponential evidence