Bayesian Data Analysis: Homework 3
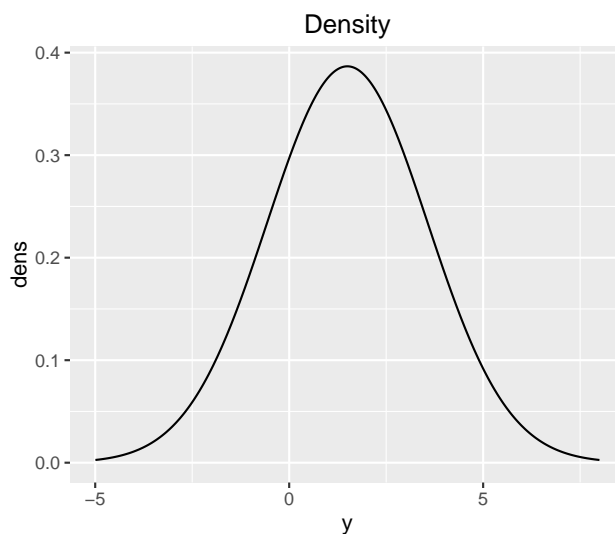
Joey Miller

February 15, 2016

Problem 1.1 Conditional probability.

a.

$$
\begin{aligned}
P(Y = y|\sigma = 2) &= P(Y = y, \theta = 1|\sigma = 2) + P(Y = y, \theta = 2|\sigma = 2) \\
&= P(\theta = 1)P(Y = y|\sigma = 2) + P(\theta = 2)P(Y = y|\sigma = 2) \\
&= \frac{1}{2}\frac{1}{2\sqrt{2\pi}}e^{\frac{-(y-1)^2}{8}} + \frac{1}{2}\frac{1}{2\sqrt{2\pi}}e^{\frac{-(y-2)^2}{8}}
\end{aligned}
$$

Density

b. Assuming $\sigma = 2$,

$$
\begin{aligned}
P(\theta = 1|y = 1) &= \frac{P(Y = 1|\theta = 1)P(\theta = 1)}{P(Y = 1)} \\
&= \frac{\frac{1}{2}\frac{1}{2\sqrt{2\pi}}e^{\frac{-(1-1)^2}{8}}}{\frac{1}{2}\frac{1}{2\sqrt{2\pi}}e^{\frac{-(1-1)^2}{8}} + \frac{1}{2}\frac{1}{2\sqrt{2\pi}}e^{\frac{-(1-2)^2}{8}}} \\
&= 0.531
\end{aligned}
$$

c. As σ increases, the distributions of $Y, \theta = 1|\sigma$ and $Y, \theta = 2|\sigma$ overlap and become less distinguishable. Thus, any observed value of y will provide less information about θ, and the probability mass function of θ will remain bernoulli with a parameter of approximately 0.5. Conversely, as σ decreases, the θs have very distinguishable implications about y, so any observed value of y will tell you with high certainty which θ it corresponds to.

Problem 1.3 Application of Conditional Probability. Note: I will use Xx to denote a heterozygote, Br to denote Xx or XX, and Bl to denote the genotype xx.

a. First, I will compute the probability that a brown-eyed child of brown-eyed parents has genotype Xx:

$$=P(\text{child=Xx}|\text{parents}=Br,\text{child=Br})$$

$$=\frac{P(\text{child=Xx},\text{parents=Br},\text{child=Br})}{P(\text{parents=Br},\text{child=Br})}=\frac{(1/2)(2p)^2(1-p)^2+(1/2)(1-p)^22p(1-p)2+0}{(3/4)(2p(1-p)^2+1(1-p)^22p(1-p)2+1(1-p)^4}$$

$$=\frac{2p}{2p+1}.$$

This was found by factoring the numerator into terms representing $P(\text{child=Xx}\mid\text{parents=Xx,Xx})P(\text{parents=Xx,Xx})$ and so on, for each permutation of brown-eyed parents. The denominator was computed likewise, but $P(\text{child=Xx}\mid\text{parents=Xx,Xx})$ becomes $P(\text{child=Br}\mid\text{parents=Xx,Xx})$ and so on.

b. In this next part, I need the probability that Judy is Xx if we know that she has n Br children and marries a Xx. I can suppress the $P(\text{husband=Xx})$ part because it will be canceled from the numerator and denominator.

$$P(\text{Judy=Xx}|n\text{ children=Br})=\frac{P(\text{Judy=Xx},n\text{ children=Br})}{n\text{ children=Br}}$$

$$=\frac{\frac{2p}{2p+1}(3/4)^n}{\frac{2p}{2p+1}(3/4)^n+\left(1-\frac{2p}{2p+1}\right)1^n}$$

$$=\frac{2p(3/4)^n}{2p(3/4)^n+1}$$

c. The probability that Judy's grandchild has blue eyes is computed as follows:

$$P(\text{Judy=Xx}|\text{previous information})=\frac{2p(3/4)^n}{2p(3/4)^n+1}$$

$$P(\text{Judy=XX}|\text{previous information})=\frac{1}{2p(3/4)^n+1}$$

$$P(\text{child=Xx, Judy=Xx or Judy=XX})=\frac{2p(3/4)^n}{2p(3/4)^n+1}\times1/2+\frac{1}{2p(3/4)^n+1}\times1/2=1/2$$

$$P(\text{grandchild=xx})=P(\text{child=Xx, Judy=Xx or Judy=XX})P(\text{child marries Xx})\times1/4$$
$$+P(\text{child=Xx, Judy=Xx or Judy=XX})P(\text{child marries xx})\times1/2$$
$$=1/2\times2p(1-p)\times1/4+1/2\times p^2\times1/2$$
$$=p(1-p)/4+p^2/4$$

Problem 1.6 Computation about Elvis' twin brother. Let I mean identical and F mean fraternal; $I\cup F$ therefore means any twin.

$$P(I|\text{twin brother})=\frac{P(I,\text{twin brother})}{P(\text{twin brother})}=\frac{(1/300)(1)}{(1/300)(1)+(1/125)(1/2)}$$

Problem 2.1 Calculation of the posterior density for $\theta=P(H)$ supposing a Beta(4,4) prior and a coin that is flipped 10 times and lands on heads fewer than 3 times:

$$P(\theta|y)=\frac{P(y|\theta)P(\theta)}{\int_\theta P(y|\theta)P(\theta)}\text{ where}$$
$$P(y|\theta)=P(y=0|\theta)+P(y=1|\theta)+P(y=2|\theta)$$
$$=(1-\theta)^{10}+10\theta(1-\theta)^9+45\theta^2(1-\theta)^8$$
and
$$P(\theta)\propto\theta^3(1-\theta)^3$$
thus
$$P(\theta|y)\propto\theta^3(1-\theta)^{13}+10\theta^4(1-\theta)^12+45\theta^5(1-\theta)^{11}$$

Problem 2.2 GIven two coins, C_1 and C_2 with probabilities $\theta_1 = 0.6$ and $\theta_2 = 0.4$ which were flipped at random yielding two heads, the expected number of additional spins until a heads shows up can be computed as follows:

$$P(C_1|y) = \frac{P(y|C_1)P(C_1)}{P(y)} = \frac{(0.6)^2(1/2)}{(0.6)^2(1/2) + (0.4)^2(1/2)} = \frac{36}{52}$$

$$P(C_2|y) = 1 - P(C_1|y) = \frac{16}{52}$$

Because the expected number of flips until a H occurs is simply the expected value $1/\theta$ of a geometric distribution with the same parameter,

$$E[1/\theta] = (1/\theta)P(\theta) = (1/0.4)(36/52) + (1/0.6)(16/52)$$

Problem 2.5 Let y be the number of heads of a coin with parameter θ out of n flips.

a. Assuming a uniform $[0,1]$ prior on θ, the prior predictive distribution is computed as follows:

$$P(y = k) = \int_\theta P(\theta)P(Y = k|\theta)d\theta$$

$$= \int_\theta 1 \times \binom{n}{k}\theta^k(1 - \theta)^{n-k}d\theta$$

$$= \binom{n}{k}\text{Beta}(k + 1, n - k + 1)$$

$$= \frac{n!}{(n - k)!k!}\frac{\Gamma(k + 1)\Gamma(n - k + 1)}{\Gamma(n + 2)}$$

$$= \frac{n!}{(n - k)!k!}\frac{k!(n - k)!}{n + 1)!}$$

$$= \frac{1}{n + 1}$$

b. Consider $\frac{y}{n} > \frac{x}{m}$ without loss of generality. Then,

$$xn < ym$$
$$xnm < ym^2$$
$$xnm + ymn < ym^2 + ymn$$
$$mn(x + y) < ym(m + n)$$
$$\frac{y + x}{m + n} < \frac{y}{n}$$

and

$$xn < ym$$
$$xn^2 < ymn$$
$$xmn + xn^2 < ymn + xmn$$
$$xn(m + n) < (y + x)mn$$
$$\frac{x}{m} < \frac{x + y}{m + n}.$$

So finally,

$$\frac{x}{m} < \frac{x + y}{m + n} < \frac{y}{n}$$

and letting $x = \alpha$ and $m = \alpha + \beta$ yields the result.

c. Considering that the prior variance on a uniform $[0,1]$ distribution is $1/12$, the posterior given any positive dataset is

$$\frac{(1 + y)(1 + n - y)}{(1 + n)^2(3 + n)}.$$

3

Noting that the numerator, where y exclusively appears, maximizes when $y = n/2$, we can simply make this substitution yielding

$$\frac{(1 + n/2)(1 + n/2)}{(1 + n)^2(3 + n)}.$$

Multiplying the terms out, differentiating, and ignoring the denominator,

$$\frac{d}{dn}\frac{(1 + n/2)(1 + n/2)}{(1 + n)^2(3 + n)} \Rightarrow (n/2 + 1)(n^3 + 7n^2 + 16n + 12) - (3n^2 + 14n + 16)(n^2/4 + n + 1)$$

$$= (n^3 + 7n^2 + 16n + 12) - (3n^2 + 14n + 16)(n/2 + 1)$$

$$= n^3 + 7n^2 + 16n_12 - (3n^2/2 + 10n^2 + 22n + 16)$$

$$= -n^2/2 - 3n^2 - 6n - 4 < 0, \forall n > 0$$

d. Beginning with a Beta(1,3) prior with variance 0.0375 and observing one success yields a Beta(2,3) posterior with variance 0.04.

Problem 2.7 Noninformative prior densities.

a. For the binomial likelihood, $p(\theta) \propto \theta^{-1}(1 - \theta)^{-1}$ is the uniform prior distribution for the natural parameter of the exponential family:

$$p(y|\theta) = \binom{n}{y}(1 - \theta)^{n-y}\theta^y$$

$$= \binom{n}{y}(1 - \theta)^n\frac{\theta^y}{(1 - \theta)^y}$$

$$= \binom{n}{y}(1 - \theta)^n e^{y\log\frac{\theta}{1-\theta}}$$

implying that the natural parameter is $\phi = g(\theta) = \log\frac{\theta}{1-\theta}$ which, under a uniform distribution, implies that θ has distribution

$$p(\theta) = p(\phi) \times \frac{d}{d\theta}\log\frac{\theta}{1 - \theta}$$

$$= 1 \times \frac{d}{d\theta}\log\frac{\theta}{1 - \theta}$$

$$\propto \frac{1}{\theta(1 - \theta)}$$

b. If $y = 0$, we get

$$p(\theta|y) \propto \frac{1}{\theta(1 - \theta)} \times (1 - \theta)^n\theta^0$$

$$= \frac{(1 - \theta)^{n-1}}{\theta}$$

and if $y = n$, we get

$$p(\theta|y) \propto \frac{1}{\theta(1 - \theta)} \times (1 - \theta)^0\theta^n$$

$$= \frac{\theta^{n-1}}{1 - \theta}$$

both of which have infinite integral on (0,1).

Problem 2.8

a. As we derived in class and using the notation from the textbook,

$$\theta|y_1 \ldots y_n \sim N(\theta, \tau_n^2)$$

where $\theta = \frac{180/40^2 + 150n/20^2}{1/40^2 + n/20^2}$ and $\frac{1}{\tau_n^2} = \frac{1}{40^2} + \frac{n}{20^2}$.

b. Likewise, the posterior predictive distribution is

$$\tilde{y}|y_1 \ldots y_n \sim N(\theta, \tau_n^2 + 20^2)$$

using the same conventions as before. This result is arrived at by applying the laws of total variance and iterated expectations to the posterior predictive distribution.

c. If $n = 10$, we have a confidence interval for $\theta|y_1 \ldots y_n$ of [138.4879, 162.9755] and for $\tilde{y}|y_1 \ldots y_n$ of [109.6648, 191.7987] (see R code for calculations).

d. If $n = 100$, the confidence interval for $\theta|y_1 \ldots y_n$ becomes [146.1598, 153.9899] and for $\tilde{y}|y_1 \ldots y_n$ becomes [110.6805, 189.4691] (again see R code for calculations).

Problem 2.9

a. In terms of the mean and standard deviation,

$$\alpha = \frac{-\mu(\sigma^2 + \mu^2 - \mu)}{\sigma^2} = \frac{-0.6(0.3^2 + 0.6^2 - 0.6)}{0.3^2} = 1$$

$$\beta = \frac{(mu - 1)(\sigma^2 + \mu^2 - \mu)}{\sigma^2} = \frac{(0.6 - 1)(0.3^2 + 0.6^2 - 0.6)}{0.3^2} = 2/3$$
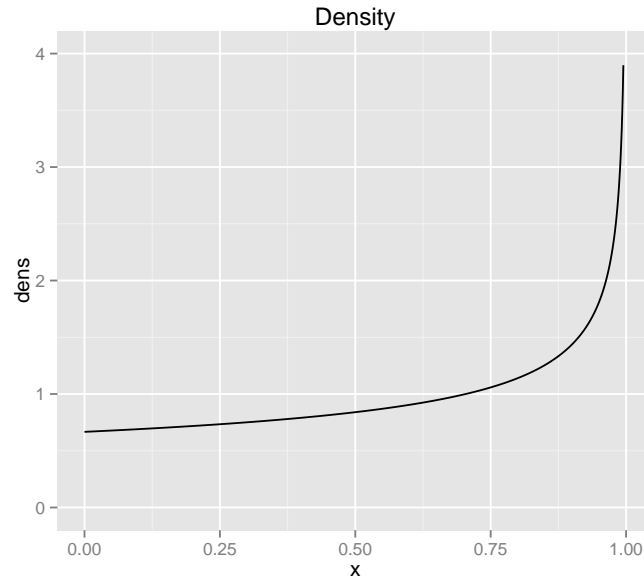
b. If out of 1000 Californians, 650 support the death penalty, then the posterior mean and variance for the population proportion is computed by
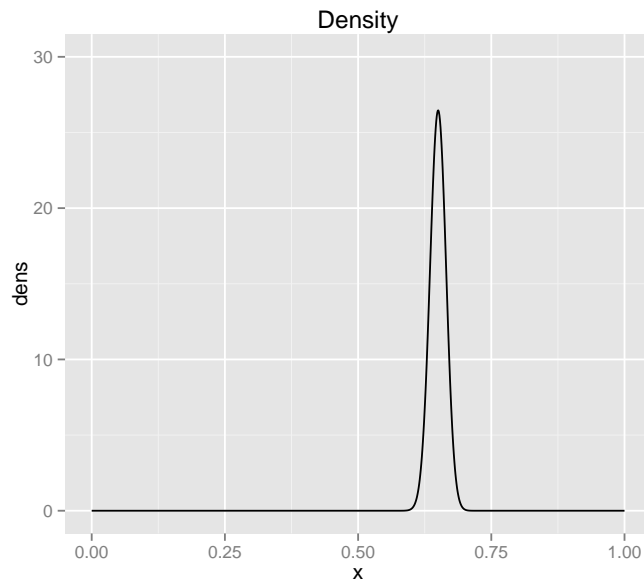
$$\alpha_{new} = 1 + 650$$
$$\beta_{new} = 2/3 + 350$$
$$E[\theta|y] = \frac{651}{651 + 350 + 2/3}$$
$$= 0.6499$$
$$\text{var}(\theta|y) = \frac{651 \times (350 + 2/3)}{(651 + 350 + 2/3)^2(651 + 350 + 2/3 + 1)}$$
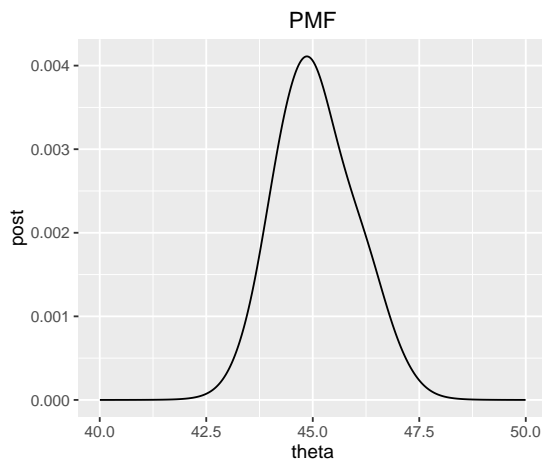$$= 0.0002269198$$

Density

c. The Jeffrey's prior for a binomial likelihood is $\alpha = \beta = 1/2$. This implies a mean of $1/2$ and variance of $1/8$. The posterior is not noticeably different from that found in part (b).

Problem 2.11

a. R code is copied below. Note that the true PDF of the posterior $\theta|y$ is converted to a PMF. The mass is 0 outside of the displayed interval.
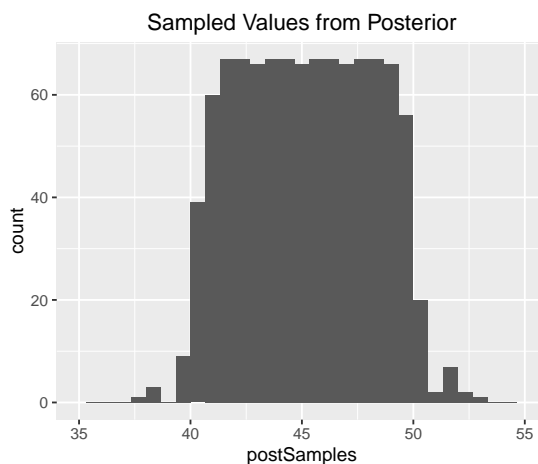

PMF

```
# Question 11a
theta <- seq(0, 100, length.out = 10000)
y <- c(43, 44, 45, 46.5, 47.5)
likelihood <- function(y, theta) {
  lprod <- 0
  for(i in 1:length(y)) {
    lprod <- log(1/(1+(y[i]-theta)^2)) + lprod
  }
  return(exp(lprod))
}

post <- likelihood(y, theta)
post <- post/sum(post)

qplot(x = theta, y = post, geom = "line", main = "PMF", xlim=c(40, 50))
```
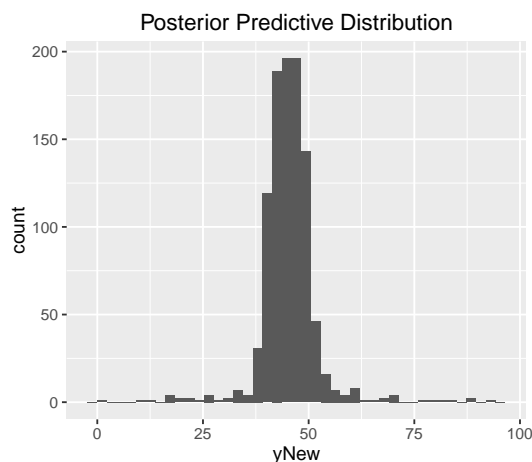
b. R code follows the histogram.

Sampled Values from Posterior

```
# Question 11b (assumes workspace of Q11a)
postSamples <- sample(x = theta, prob = post, size = 1000)
qplot(postSamples, geom = "histogram", main = "Sampled Values from Posterior", bins=30,
xlim=c(35, 55))
```

c. Using the values of θ sampled previously, the distribution $\tilde{y}|y$ can be estimated:

Posterior Predictive Distribution

```
# Question 11c (assumes workspace of Q11a and Q11b)
y <- seq(0, 110, length.out = 1000)
yNew <- rep(NA, times=1000)
for (j in 1:length(postSamples)) {
  prob <- lapply(X = y, FUN = likelihood, theta=postSamples[j])
  yNew[j] <- sample(x=y, prob = as.numeric(prob), size = 1)
}

qplot(yNew, geom = "histogram", main = "Posterior Predictive Distribution", bins=40)
```

Problem 2.20 Censored and uncensored data in an exponential model with a rate parameter distributed Gamma(α, β).

a. The posterior given an observation $y \geq 100$:

$$P(\theta | y \geq 100) \propto P(Y \geq 100|\theta)P(\theta)$$

$$\propto \int_{100}^{\infty} \theta e^{-\theta y} dy \theta^{\alpha-1} e^{-\beta\theta}$$

$$= e^{-100\theta} \theta^{\alpha-1} e^{-\beta\theta}$$

$$= \theta^{\alpha-1} e^{-(\beta+100)\theta}$$

$$\theta | y \sim \text{Gamma}(\alpha, \beta + 100)$$

$$E[\theta|y] = \frac{\alpha}{\beta + 100}$$

$$\text{var}(\theta|y) = \frac{\alpha}{(\beta + 100)^2}$$

b. The posterior given an observation $y = 100$:

$$P(\theta | y = 100) \propto P(Y = 100|\theta)P(\theta)$$

$$= \theta e^{-100\theta} \theta^{\alpha-1} e^{-\beta\theta}$$

$$= \theta^{\alpha+1-1} e^{-(\beta+100)\theta}$$

$$\theta | y \sim \text{Gamma}(\alpha + 1, \beta + 100)$$

$$E[\theta|y] = \frac{\alpha + 1}{\beta + 100}$$

$$\text{var}(\theta|y) = \frac{\alpha + 1}{(\beta + 100)^2}$$

c. The posterior has higher variance in part (b) because a $y \geq 100$ is a very diffuse result that doesn't alter the shape parameter of the Gamma distribution, only its rate. This doesn't contradict identity 2.8 because that identity only says that the expected value of the variance of the posterior given more information is lower than the variance of the prior. In particular cases, extreme examples can raise the variance of the posterior relative to the prior.

Appendix

```
library("ggplot2")

# Question 1a
y <- seq(-10, 10, length.out=1000)
dens <- dnorm(y, mean = 1, sd = 2) + dnorm(y, mean = 2, sd = 2)
qplot(x = y, y = dens, geom = "line", main = "Density", xlim = c(-5, 8))

# Question 8c
# n=10
mu <- (180/40^2 + 150*10/20^2)/(1/40^2+10/20^2)
tau <- (1/40^2 + 10/20^2)^-1
upper <- qnorm(0.975, mean = mu, sd = sqrt(tau))
lower <- qnorm(0.025, mean = mu, sd = sqrt(tau))


mu <- (180/40^2 + 150*10/20^2)/(1/40^2+10/20^2)
tau <- (1/40^2 + 10/20^2)^-1 + 20^2
upper <- qnorm(0.975, mean = mu, sd = sqrt(tau))
lower <- qnorm(0.025, mean = mu, sd = sqrt(tau))

# Question 8d
# n=100
mu <- (180/40^2 + 150*100/20^2)/(1/40^2+100/20^2)
tau <- (1/40^2 + 100/20^2)^-1
upper <- qnorm(0.975, mean = mu, sd = sqrt(tau))
lower <- qnorm(0.025, mean = mu, sd = sqrt(tau))
```

```r
mu <- (180/40^2 + 150*100/20^2)/(1/40^2+100/20^2)
tau <- (1/40^2 + 100/20^2)^-1 + 20^2
upper <- qnorm(0.975, mean = mu, sd = sqrt(tau))
lower <- qnorm(0.025, mean = mu, sd = sqrt(tau))


# Question 9a
x <- seq(0, 1, length.out=1000)
dens <- dbeta(x, shape1 = 1, shape2 = 2/3)
qplot(x = x, y = dens, geom = "line", main = "Density", ylim=c(0,4))

# Question 9b
x <- seq(0, 1, length.out=1000)
dens <- dbeta(x, shape1 = 651, shape2 = (350+2/3))
qplot(x = x, y = dens, geom = "line", main = "Density", ylim=c(0,30))

# Question 11a
theta <- seq(0, 100, length.out = 10000)
y <- c(43, 44, 45, 46.5, 47.5)
likelihood <- function(y, theta) {
  lprod <- 0
  for(i in 1:length(y)) {
    lprod <- log(1/(1+(y[i]-theta)^2)) + lprod
  }
  return(exp(lprod))
}

post <- likelihood(y, theta)
post <- post/sum(post)

qplot(x = theta, y = post, geom = "line", main = "PMF", xlim=c(40, 50))

# Question 11b (assumes workspace of Q11a)
postSamples <- sample(x = theta, prob = post, size = 1000)
qplot(postSamples, geom = "histogram", main = "Sampled Values from Posterior", bins=30, xlim=c(35, 55))

# Question 11c (assumes workspace of Q11a and Q11b)
y <- seq(0, 110, length.out = 1000)
yNew <- rep(NA, times=1000)
for (j in 1:length(postSamples)) {
  prob <- lapply(X = y, FUN = likelihood, theta=postSamples[j])
  yNew[j] <- sample(x=y, prob = as.numeric(prob), size = 1)
}

qplot(yNew, geom = "histogram", main = "Posterior Predictive Distribution", bins=40)
```