STAT 8700 Homework 1

Brian Detweiler

Friday, September 2, 2016

1. Read Chapter 1.

Done.

2. Consider an urn containing 9 balls, which can be either red or green. Let X be the number of red balls in the urn and before observing any balls we will assume that all possible values of X from 0 to 9 are equally likely. Suppose we plan to draw 3 balls from the urn, and let $Y_i = 1$ if the i^{th} ball is red, and $Y_i = 0$ if the i^{th} ball is green for $i = 1, 2, 3$. When we draw the 3 balls, we observe $Y_1 = 1, Y_2 = 1$, and $Y_3 = 0$. As per our examples in class, construct a table with columns X, Prior, Likelihood, Likelihood x Prior, and Posterior to obtain the Posterior Distribution of X.

Our likelihood is given by:

$$P(Y_1 = 1, Y_2 = 1, Y_3 = 0 | X = x)$$

We want to find the posterior distribution, such that

$$P(X = x | Y_1 = 1, Y_2 = 1, Y_3 = 0) = \frac{P(Y_1 = 1, Y_2 = 1, Y_3 = 0 | X = x) \cdot P(X = x)}{P(Y_1 = 1, Y_2 = 1, Y_3 = 0)}$$

```r
hypergeo <- function(numberRedDrawn, numberRed, numberWhite, totalDrawn) {
  return(dhyper(numberRedDrawn, numberRed, numberWhite, totalDrawn))
}

# For Y_1 = 1
totalInUrn <- 9
minNumberOfRedInUrn <- 2
# X = x for all possible x
numberOfRedInUrn <- c(minNumberOfRedInUrn:totalInUrn)
numberOfRedDrawn <- 1
numberOfWhiteInUrn <- totalInUrn - numberOfRedInUrn
totalDrawn <- 1

Y1 <- hypergeo(numberOfRedDrawn, numberOfRedInUrn, numberOfWhiteInUrn, totalDrawn)
Y1 <- append(c(0, 0), Y1)
Y1
```

```
##  [1] 0.0000000 0.0000000 0.2222222 0.3333333 0.4444444 0.5555556 0.6666667
##  [8] 0.7777778 0.8888889 1.0000000
```

```r
# For Y_2 = 1, having already drawn Y_1 = 1
totalInUrn <- 8
minNumberOfRedInUrn <- 1
# X = x for all possible x
numberOfRedInUrn <- c(minNumberOfRedInUrn:totalInUrn)
numberOfRedDrawn <- 1
numberOfWhiteInUrn <- totalInUrn - numberOfRedInUrn
totalDrawn <- 1

Y2 <- hypergeo(numberOfRedDrawn, numberOfRedInUrn, numberOfWhiteInUrn, totalDrawn)
Y2 <- append(c(0, 0), Y2)
Y2
```

```
##  [1] 0.000 0.000 0.125 0.250 0.375 0.500 0.625 0.750 0.875 1.000
```

```r
# For Y_3 = 0, having already drawn Y_1 = 1 and Y_2 = 2
totalInUrn <- 7
minNumberOfRedInUrn <- 0
# X = x for all possible x
numberOfRedInUrn <- c(minNumberOfRedInUrn:totalInUrn)
numberOfRedDrawn <- 0
numberOfWhiteInUrn <- totalInUrn - numberOfRedInUrn
totalDrawn <- 1

Y3 <- hypergeo(numberOfRedDrawn, numberOfRedInUrn, numberOfWhiteInUrn, totalDrawn)
Y3 <- append(c(0, 0), Y3)
Y3
```

```
##  [1] 0.0000000 0.0000000 1.0000000 0.8571429 0.7142857 0.5714286 0.4285714
##  [8] 0.2857143 0.1428571 0.0000000
```

```r
conditionalLikelihood <- Y1 * Y2 * Y3
conditionalLikelihood
```

```
##  [1] 0.00000000 0.00000000 0.02777778 0.07142857 0.11904762 0.15873016
##  [7] 0.17857143 0.16666667 0.11111111 0.00000000
```

```r
sum(conditionalLikelihood)
```

```
## [1] 0.8333333
```

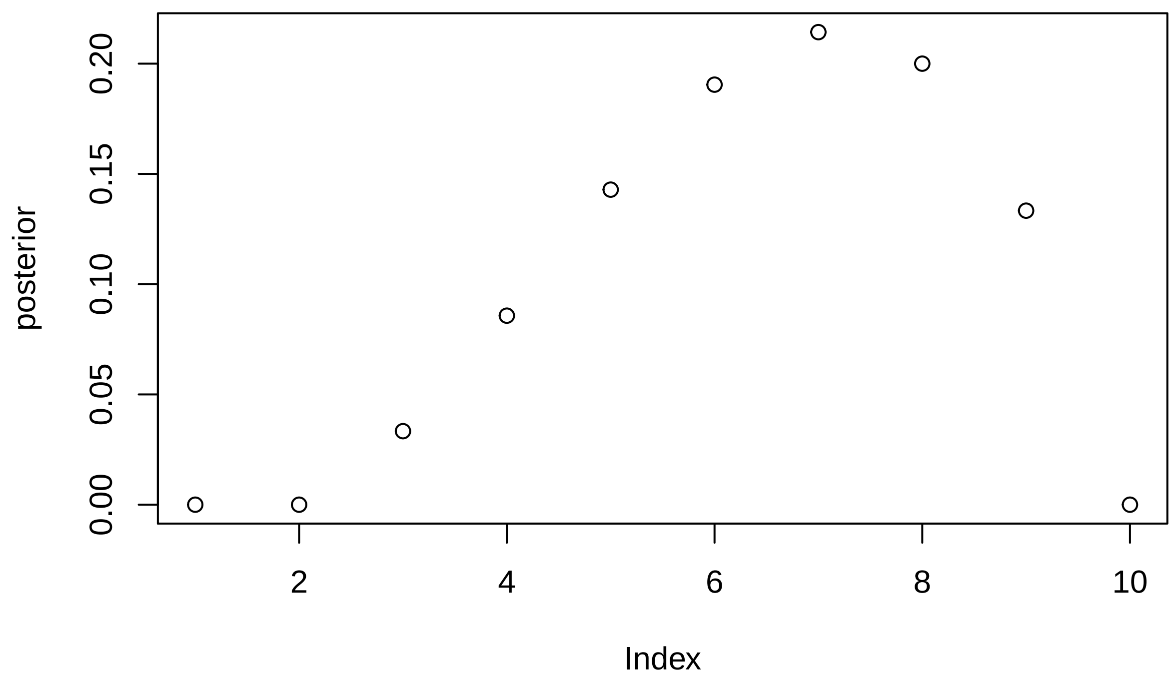
```r
prior <- 1/10

priorTimesLikelihood <- conditionalLikelihood * prior
priorTimesLikelihood
```

```
##  [1] 0.000000000 0.000000000 0.002777778 0.007142857 0.011904762
##  [6] 0.015873016 0.017857143 0.016666667 0.011111111 0.000000000
```

```r
constant <- sum(priorTimesLikelihood)
constant
```

```
## [1] 0.08333333
```

```r
posterior <- priorTimesLikelihood / constant
plot(posterior)
```




```r
posterior
```

```
##  [1] 0.00000000 0.00000000 0.03333333 0.08571429 0.14285714 0.19047619
##  [7] 0.21428571 0.20000000 0.13333333 0.00000000
```

```r
sum(posterior)
```

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

X	Prior	Likelihood	Prior $\times$ Likelihood	Prior $\times$ Likelihood / constant = Posterior
0	$\frac{1}{10}$	0	0	0
1	$\frac{1}{10}$	0	0	0
2	$\frac{1}{10}$	0.02777778	0.002777778	0.03333333
3	$\frac{1}{10}$	0.07142857	0.007142857	0.08571429
4	$\frac{1}{10}$	0.11904762	0.011904762	0.14285714
5	$\frac{1}{10}$	0.15873016	0.015873016	0.19047619
6	$\frac{1}{10}$	0.17857143	0.017857143	0.21428571
7	$\frac{1}{10}$	0.16666667	0.016666667	0.20000000
8	$\frac{1}{10}$	0.11111111	0.011111111	0.13333333
9	$\frac{1}{10}$	0	0	0
Total:	1	0.8333333	constant = 0.08333333 = $\frac{1}{12}$	1

■

3. Let Y_1 be the number of successes in $n = 10$ independent trials where each trial results in a success or failure, and θ, the probability of success in each trial is the same for each trial. Suppose we believe there are 4 possible values of θ, $\frac{1}{5}, \frac{2}{5}, \frac{3}{5}, \frac{4}{5}$, which we view as equally likely. Now suppose we observe $Y_1 = 7$, use a table similar to the previous question to find the posterior distribution of θ.

We want to find the posterior distribution

$$P(\theta|Y_1 = 7) = \frac{P(Y_1 = 7|\theta) \cdot P(\theta)}{P(Y_1 = 7)}$$

Since we have no idea what $P(\theta)$ might be, we'll start with a uniform prior distribution, $U(0, 1)$.

The likelihood is given by the binomial distribution,

$$P(Y_1 = 7|\theta) = \binom{10}{7}\theta^7(1-\theta)^3$$

```
binomProb <- function(n, y, theta) {
  return(choose(n, y) * theta^y * (1 - theta)^(n - y))
}

prior <- c(1/4, 1/4, 1/4, 1/4)

n <- 10
successes <- 7

theta1 <- binomProb(n, successes, 1/5)
theta2 <- binomProb(n, successes, 2/5)
theta3 <- binomProb(n, successes, 3/5)
theta4 <- binomProb(n, successes, 4/5)

likelihood <- c(theta1, theta2, theta3, theta4)
likelihood
```

```
## [1] 0.000786432 0.042467328 0.214990848 0.201326592
```

```
likelihoodTotal <- sum(likelihood)
likelihoodTotal
```

```
## [1] 0.4595712
```

```
likelihoodTimesPrior <- likelihood * prior
likelihoodTimesPrior
```

```
## [1] 0.000196608 0.010616832 0.053747712 0.050331648
```

```r
constant <- sum(likelihoodTimesPrior)
constant
```

```
## [1] 0.1148928
```

```r
posterior <- likelihoodTimesPrior / constant
posterior
```

```
## [1] 0.00171123 0.09240642 0.46780749 0.43807487
```

```r
sum(posterior)
```

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

θ	Prior	Likelihood	Prior $\times$ Likelihood	Prior $\times$ Likelihood / constant = Posterior
$\frac{1}{5}$	$\frac{1}{4}$	0.000786432	0.000196608	0.00171123
$\frac{2}{5}$	$\frac{1}{4}$	0.04246733	0.010616832	0.09240642
$\frac{3}{5}$	$\frac{1}{4}$	0.2149908	0.053747712	0.46780749
$\frac{4}{5}$	$\frac{1}{4}$	0.2013266	0.050331648	0.43807487
Total:	1	0.4595712	constant = 0.1148928	1

■

4. Following on from the previous question, suppose we observe another 5 independent trials and $Y_2 = 2$ successes are observed in those 5 trials. Use the posterior distribution for θ from the previous question as the new prior distribution of θ and use a table to find the new posterior distribution of θ based on the added trials.

We want to find the posterior distribution

$$P(\theta|Y_2 = 2) = \frac{P(Y_2 = 2|\theta) \cdot P(\theta)}{P(Y_2 = 2)}$$

Since we have some information about $P(\theta)$, we'll use the posterior, $P(\theta|Y_2 = 2)$ for our prior, $P(\theta)$.

The likelihood is still given by the binomial distribution,

$$P(Y_2 = 2|\theta) = \binom{5}{2}\theta^2(1-\theta)^3$$

```
binomProb <- function(n, y, theta) {
  return(choose(n, y) * theta^y * (1 - theta)^(n - y))
}

prior <- posterior
prior
```

```
## [1] 0.00171123 0.09240642 0.46780749 0.43807487
```

```
n <- 5
successes <- 2

theta1 <- binomProb(n, successes, 1/5)
theta2 <- binomProb(n, successes, 2/5)
theta3 <- binomProb(n, successes, 3/5)
theta4 <- binomProb(n, successes, 4/5)

likelihood <- c(theta1, theta2, theta3, theta4)
likelihood
```

```
## [1] 0.2048 0.3456 0.2304 0.0512
```

```
likelihoodTotal <- sum(likelihood)
likelihoodTotal
```

```
## [1] 0.832
```

```
likelihoodTimesPrior <- likelihood * prior
likelihoodTimesPrior
```

```
## [1] 0.0003504599 0.0319356578 0.1077828449 0.0224294332
```

```r
constant <- sum(likelihoodTimesPrior)
constant
```

```
## [1] 0.1624984
```

```r
posterior <- likelihoodTimesPrior / constant
posterior
```

```
## [1] 0.002156698 0.196529065 0.663285594 0.138028644
```

```r
sum(posterior)
```

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

θ	Prior	Likelihood	Prior $\times$ Likelihood	Prior $\times$ Likelihood / constant = Posterior
$\frac{1}{5}$	0.00171123	0.2048	0.0003504599	0.002156698
$\frac{2}{5}$	0.09240642	0.3456	0.0319356578	0.196529065
$\frac{3}{5}$	0.46780749	0.2304	0.1077828449	0.663285594
$\frac{4}{5}$	0.43807487	0.0512	0.0224294332	0.138028644
Total:	1	0.4336128	constant = 0.1624984	1

■

5. Suppose we combine all 15 trials from questions 3 and 4 together and think of them as a single set of data in which we observed 9 successes. Starting with our initial uniform prior, use a table to nd the posterior distribution of θ. Compare your answer to your answer at the end of question 4.

We want to find the posterior distribution

$$P(\theta|Y_3 = 9) = \frac{P(Y_3 = 9|\theta) \cdot P(\theta)}{P(Y_3 = 9)}$$

As stated, we'll start with a uniform prior distribution, $U(0, 1)$.

The likelihood is given by the binomial distribution,

$$P(Y_3 = 9|\theta) = \binom{15}{9}\theta^9(1-\theta)^6$$

```
binomProb <- function(n, y, theta) {
  return(choose(n, y) * theta^y * (1 - theta)^(n - y))
}

prior <- c(1/4, 1/4, 1/4, 1/4)

n <- 15
successes <- 9

theta1 <- binomProb(n, successes, 1/5)
theta2 <- binomProb(n, successes, 2/5)
theta3 <- binomProb(n, successes, 3/5)
theta4 <- binomProb(n, successes, 4/5)

likelihood <- c(theta1, theta2, theta3, theta4)
likelihood
```

```
## [1] 0.0006717597 0.0612141053 0.2065976053 0.0429926226
```

```
likelihoodTotal <- sum(likelihood)
likelihoodTotal
```

```
## [1] 0.3114761
```

```
likelihoodTimesPrior <- likelihood * prior
likelihoodTimesPrior
```

```
## [1] 0.0001679399 0.0153035263 0.0516494013 0.0107481557
```

```r
constant <- sum(likelihoodTimesPrior)
constant
```

```
## [1] 0.07786902
```

```r
posterior <- likelihoodTimesPrior / constant
posterior
```

```
## [1] 0.002156698 0.196529065 0.663285594 0.138028644
```

```r
sum(posterior)
```

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

θ	Prior	Likelihood	Prior $\times$ Likelihood	Prior $\times$ Likelihood / constant = Posterior
$\frac{1}{5}$	$\frac{1}{4}$	0.0006717597	0.0001679399	0.002156698
$\frac{2}{5}$	$\frac{1}{4}$	0.0612141053	0.0153035263	0.196529065
$\frac{3}{5}$	$\frac{1}{4}$	0.2065976053	0.0516494013	0.663285594
$\frac{4}{5}$	$\frac{1}{4}$	0.0429926226	0.0107481557	0.138028644
Total:	1	0.3114761	constant = 0.07786902	1

Interestingly, we get the exact same posterior as in **4**.

■

6. In R, install a package called Bolstad. This package includes a function called binodp, which stands for "Binomial Data, Discrete Prior", exactly like the situation described in questions 3, 4, and 5. The function requires 4 inputs, the number of successes, the number of trials, a vector containing the possible values of θ, and a vector containing the corresponding prior probabilities. Note: In R, a vector is specied by c(), so in this example, the vector containing the possible values of θ would be c(1/5,2/5,3/5,4/5). Use this binodp function to calculate the posterior distribution based on the data in question 5. The function will generate several output tables and one graph. Copy/Paste (do not manually copy) the last output table (the posterior distribution) and the graph into your assignment.

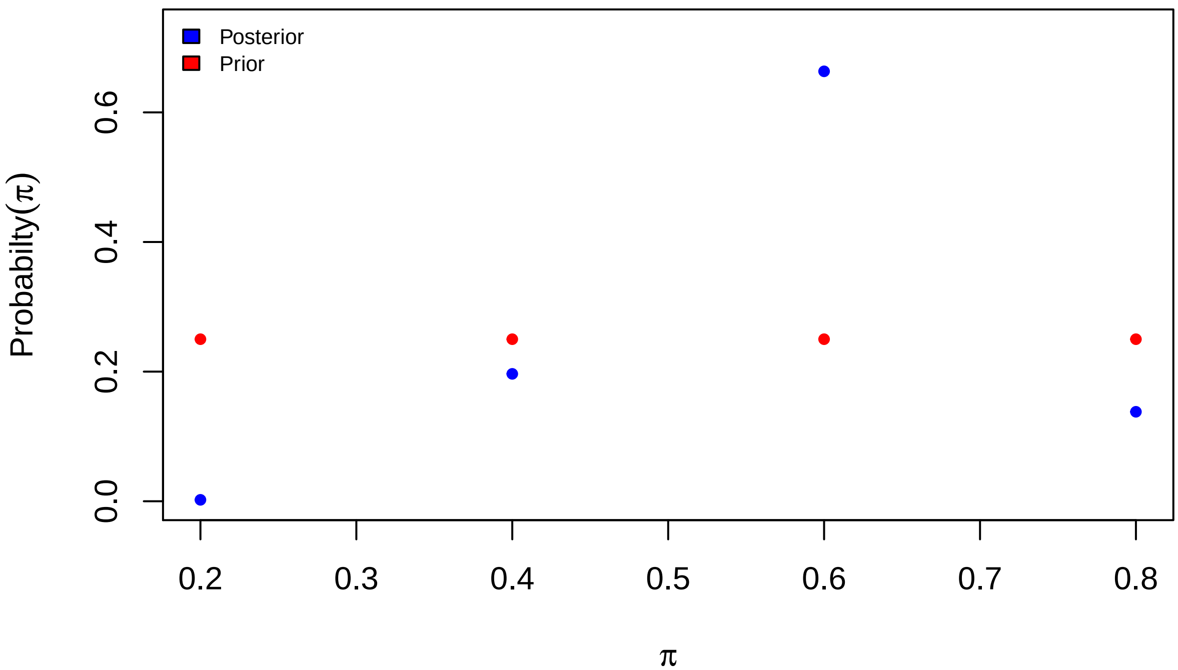
```r
library('Bolstad')
```

```
## Warning: package 'Bolstad' was built under R version 3.2.4
```

```
##
## Attaching package: 'Bolstad'
```

```
## The following objects are masked from 'package:stats':
##
##     IQR, sd, var
```

```r
binodp(x=9, n=15, pi=c(1/5, 2/5, 3/5, 4/5), pi.prior=c(1/4, 1/4, 1/4, 1/4))
```




```
## Conditional distribution of x given pi and  n:
##
##          0      1      2      3      4      5      6      7      8      9
## 0.2 0.0352 0.1319 0.2309 0.2501 0.1876 0.1032 0.0430 0.0138 0.0035 0.0007
## 0.4 0.0005 0.0047 0.0219 0.0634 0.1268 0.1859 0.2066 0.1771 0.1181 0.0612
## 0.6 0.0000 0.0000 0.0003 0.0016 0.0074 0.0245 0.0612 0.1181 0.1771 0.2066
## 0.8 0.0000 0.0000 0.0000 0.0000 0.0000 0.0001 0.0007 0.0035 0.0138 0.0430
##         10     11     12     13     14     15
## 0.2 0.0001 0.0000 0.0000 0.0000 0.0000 0.0000
## 0.4 0.0245 0.0074 0.0016 0.0003 0.0000 0.0000
## 0.6 0.1859 0.1268 0.0634 0.0219 0.0047 0.0005
## 0.8 0.1032 0.1876 0.2501 0.2309 0.1319 0.0352
##
## Joint distribution:
##
##           0      1      2      3      4      5      6      7      8      9
## [1,] 0.0088 0.0330 0.0577 0.0625 0.0469 0.0258 0.0107 0.0035 0.0009 0.0002
## [2,] 0.0001 0.0012 0.0055 0.0158 0.0317 0.0465 0.0516 0.0443 0.0295 0.0153
## [3,] 0.0000 0.0000 0.0001 0.0004 0.0019 0.0061 0.0153 0.0295 0.0443 0.0516
## [4,] 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0002 0.0009 0.0035 0.0107
##          10     11     12     13     14     15
## [1,] 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000
## [2,] 0.0061 0.0019 0.0004 0.0001 0.0000 0.0000
## [3,] 0.0465 0.0317 0.0158 0.0055 0.0012 0.0001
## [4,] 0.0258 0.0469 0.0625 0.0577 0.0330 0.0088
##
```

```
## Marginal distribution of x:
##
##          0      1      2      3      4      5      6      7      8      9
## [1,] 0.0089 0.0342 0.0633 0.0788 0.0805 0.0784 0.0779 0.0781 0.0781 0.0779
##         10     11     12     13     14     15
## [1,] 0.0784 0.0805 0.0788 0.0633 0.0342 0.0089
##
##
##     Prior   Likelihood   Posterior
## 0.2  0.25 0.0006717597 0.002156698
## 0.4  0.25 0.0612141053 0.196529065
## 0.6  0.25 0.2065976053 0.663285594
## 0.8  0.25 0.0429926226 0.138028644
```

■