Stat 532 Assignment 1

Kenny Flagg

September 2, 2015

1. The hypergeometric distribution is appropriate, so

   ```
   dhyper(x = x, m = theta, n = 12 - theta, k = 5)
   ```

   gives the appropriate probability.

2. My function:

   ```
   # theta = number of gold marbles in bucket
   # x = number of gold marbles drawn
   # N = number of marbles in bucket
   # n = number of marbles drawn
   probGold <- function(theta, x = 1, N = 12, n = 5){
     return(dhyper(x, theta, N - theta, k = n))
   }
   ```

3. The table below shows $Pr(X = x|\theta)$, with columns corresponding to θ values and rows corresponding to x values.

	0	1	2	3	4	5	6	7	8	9	10	11	12	Sum
0	1.00	0.58	0.32	0.16	0.07	0.03	0.01	0.00	0.00	0.00	0.00	0.00	0.00	2.17
1	0.00	0.42	0.53	0.48	0.35	0.22	0.11	0.04	0.01	0.00	0.00	0.00	0.00	2.17
2	0.00	0.00	0.15	0.32	0.42	0.44	0.38	0.27	0.14	0.05	0.00	0.00	0.00	2.17
3	0.00	0.00	0.00	0.05	0.14	0.27	0.38	0.44	0.42	0.32	0.15	0.00	0.00	2.17
4	0.00	0.00	0.00	0.00	0.01	0.04	0.11	0.22	0.35	0.48	0.53	0.42	0.00	2.17
5	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.03	0.07	0.16	0.32	0.58	1.00	2.17
Sum	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	13.00

4. The columns sum to 1 since each column is a probability mass function. The rows are positive but do not sum to 1; this is acceptable because the entries are not the probabilities that θ takes the given values.

5. A likelihood function describes how well a possible parameter value fits with the observed data. It has the same expression as the probability mass function $f(x|\theta)$, but the likelihood is a function of θ where x is considered fixed.

   In the marbles example, suppose we draw 5 marbles without replacement and observe $X = 1$ gold marble. We then look at the row for $x = 1$ in the above table, and we can see that $Pr(X = 1|\theta = 1) = 0.42$ and $Pr(X = 1|\theta = 2) = 0.53$. These are not probability statements about θ; they tell us that 2 is a more likely value of θ than 1 in the sense that, if the experiment were repeated many times, we would observe 1 gold marble more frequently if $\theta = 2$ than if $\theta = 1$.

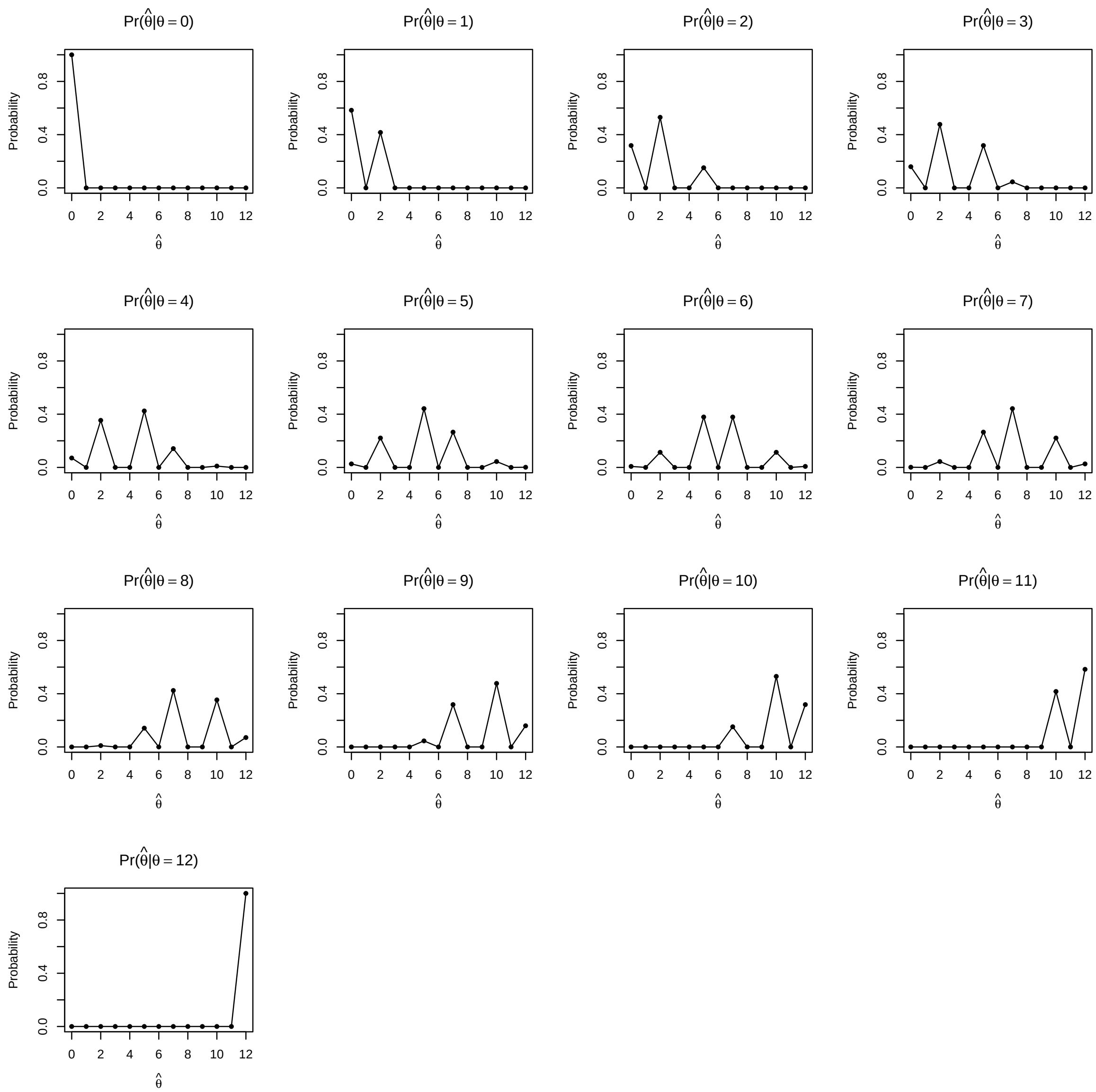
   The likelihood is denoted $L(\theta|x)$ to emphasize that that it is not a probability when considered as a function of θ. It allows for relative comparisons between different values of θ, with larger values of $L(\theta|x)$ indicating that θ is more consistent with the observation $X = x$.

6. To find a point estimate of θ, we assume that the observed result $X = x$ is not unusual and so we choose the value of θ for which the probability that $X = x$ is the largest. This is customarily denoted $\hat{\theta}$.

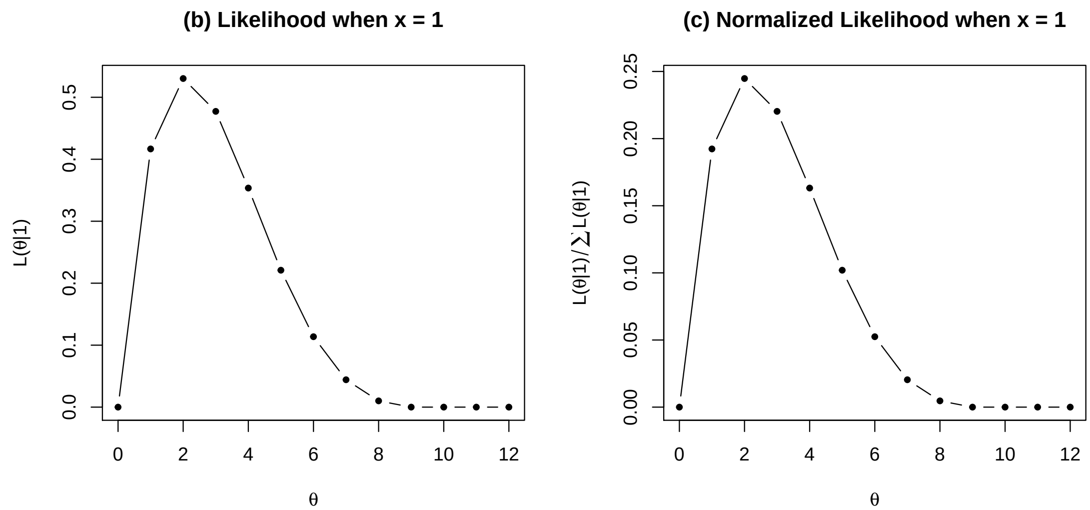
7. An estimator is a rule or procedure for computing a reasonable value of the parameter given some observed data. An estimate is the value that is computed when the rule is applied to some data. In short, an estimator is a process, and an estimate is a number.

8. In this case, the estimator can take on the values 0, 2, 5, 7, 10, and 12. The parameter θ represents the number of gold marbles out of the 12 total marbles, so the estimator must only take values in $0, 1, 2, \ldots, 12$. Since the estimator is a function of the random variable X, which can take the values 0, 1, 2, 3, 4, and 5, the estimator can take at most six distinct values. By examining the table from problem 3, it can be seen that $\theta = 1, 3, 4, 6, 8, 9, 11$ do not maximize $Pr(X = x|\theta)$ for any $x = 0, 1, 2, 3, 4, 5$.

9. Below, I have plotted the sampling distribution of the MLE $\hat{\theta}$ for each θ in the parameter space.



10. (a) Looking in the row for $x = 1$ in the table from problem 3, the likelihood is maximized by $\hat{\theta} = 2$.



A histogram does not seem appropriate since histograms are meant to display counts or relative frequencies. The likelihood is neither. It is a function defined on the parameter space, which is a countable set. The mathematically correct graph would consist of discrete points, but I opt to connect the points with line segments to illustrate the changes in L between adjacent θ values.

11. (a) `lm()` assumes observations are independent and the response follows a normal distribution with constant variance. The estimates are from the least squares fit, which has a closed form.

    `glm()` assumes independent observations. There is no single distribution assumption, but a distribution must be specified by a `family` function. The estimates are found numerically by iteratively reweighted least squares.

    (b) For a normal linear model fit by least squares, confidence intervals are base upon a t distribution. For generalized linear models, confidence intervals are typically based on a normal distribution through an appeal to the Central Limit Theorem.

12. Bayes Theorem is

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}.$$

Using $f(\cdot)$ to represent a probability mass function, and using the likelihood function notation, this becomes

$$f(\theta|x) = \frac{f(x|\theta)f(\theta)}{f(x)} = \frac{L(\theta|x)f(\theta)}{f(x)}.$$

13. To find the posterior distribution $f(\theta|x)$, we first need to specify the probability model (likelihood) $f(x|\theta) = L(\theta|x)$ and prior distribution $f(\theta)$. Then we compute the marginal distribution as

$$f(x) = \sum_{\theta=0}^{12} f(x,\theta) = \sum_{\theta=0}^{12} f(x|\theta)f(\theta) = \sum_{\theta=0}^{12} L(\theta|x)f(\theta)$$

and get the result

$$f(\theta|x) = \frac{L(\theta|x)f(\theta)}{\sum_{\theta=0}^{12} L(\theta|x)f(\theta)}.$$

14. (a) If θ is the result of rolling 2 dice:

θ	$f(\theta)$	$L(\theta\|x)$	$f(x,\theta)$	$f(\theta\|x)$
0	0.0000	0.0000	0.0000	0.0000
1	0.0000	0.4167	0.0000	0.0000
2	0.0278	0.5303	0.0147	0.1230
3	0.0556	0.4773	0.0265	0.2213
4	0.0833	0.3535	0.0295	0.2459
5	0.1111	0.2210	0.0246	0.2049
6	0.1389	0.1136	0.0158	0.1317
7	0.1667	0.0442	0.0074	0.0615
8	0.1389	0.0101	0.0014	0.0117
9	0.1111	0.0000	0.0000	0.0000
10	0.0833	0.0000	0.0000	0.0000
11	0.0556	0.0000	0.0000	0.0000
12	0.0278	0.0000	0.0000	0.0000

(b) If θ is the result of rolling 1 die:

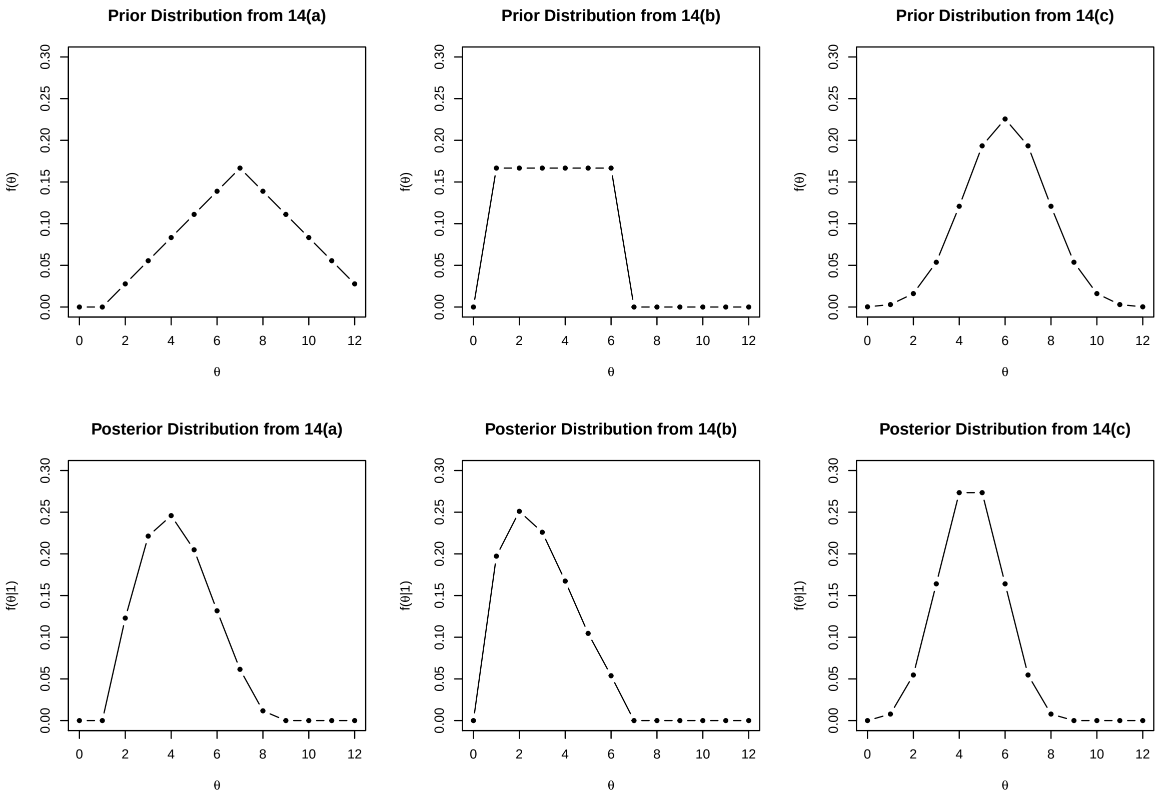
θ	$f(\theta)$	$L(\theta\|x)$	$f(x,\theta)$	$f(\theta\|x)$
0	0.0000	0.0000	0.0000	0.0000
1	0.1667	0.4167	0.0694	0.1973
2	0.1667	0.5303	0.0884	0.2510
3	0.1667	0.4773	0.0795	0.2259
4	0.1667	0.3535	0.0589	0.1674
5	0.1667	0.2210	0.0368	0.1046
6	0.1667	0.1136	0.0189	0.0538
7	0.0000	0.0442	0.0000	0.0000
8	0.0000	0.0101	0.0000	0.0000
9	0.0000	0.0000	0.0000	0.0000
10	0.0000	0.0000	0.0000	0.0000
11	0.0000	0.0000	0.0000	0.0000
12	0.0000	0.0000	0.0000	0.0000

(c) If θ is the number of heads in 12 independent coin flips:

θ	$f(\theta)$	$L(\theta\|x)$	$f(x,\theta)$	$f(\theta\|x)$
0	0.0002	0.0000	0.0000	0.0000
1	0.0029	0.4167	0.0012	0.0078
2	0.0161	0.5303	0.0085	0.0547
3	0.0537	0.4773	0.0256	0.1641
4	0.1208	0.3535	0.0427	0.2734
5	0.1934	0.2210	0.0427	0.2734
6	0.2256	0.1136	0.0256	0.1641
7	0.1934	0.0442	0.0085	0.0547
8	0.1208	0.0101	0.0012	0.0078
9	0.0537	0.0000	0.0000	0.0000
10	0.0161	0.0000	0.0000	0.0000
11	0.0029	0.0000	0.0000	0.0000
12	0.0002	0.0000	0.0000	0.0000

15. We see that θ values that cannot occur in the prior distribution, such as $\theta = 0, 1$ when rolling two dice or $\theta = 0, 7, 8, \ldots, 12$ when rolling one die, have probability 0 in the posterior distribution. This makes sense because a priori knowledge that certain values are impossible would not be changed by the observed data.

The θ value with maximum posterior probability differs between the three cases and appears to be related to the center and spread of the prior distribution. Compared to the normalized likelihood, the posterior distributions are "pulled" toward the centers of the prior distributions. The amount of the "pull" can be seen by comparing the posterior mode to the MLE, which has a value of 2. The largest shift is in 14(c), which has the prior with the smallest standard deviation.

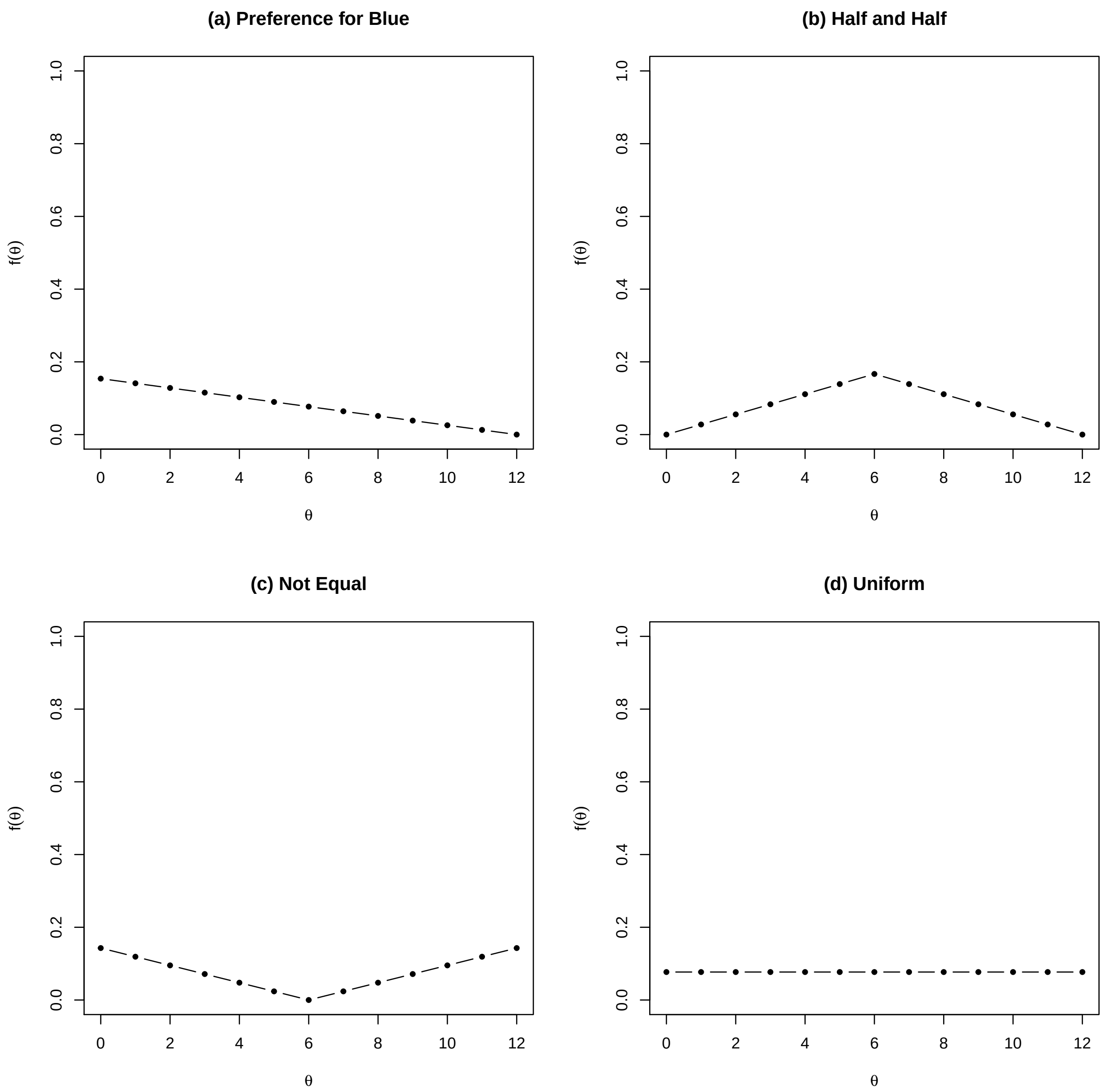


16. None of these posterior distributions match the normalized likelihood function from 10(c). Normalizing the likelihood is equivalent to choosing a uniform prior for θ, that is $f(\theta) = \frac{1}{13}$ for $\theta = 0, 1, 2, \ldots, 12$, because

$$
\begin{aligned}
f(\theta|x) &= \frac{L(\theta|x)f(\theta)}{\sum_{\theta=0}^{12} L(\theta|x)f(\theta)} \\
&= \frac{\frac{1}{13}L(\theta|x)}{\frac{1}{13}\sum_{\theta=0}^{12} L(\theta|x)} \\
&= \frac{L(\theta|x)}{\sum_{\theta=0}^{12} L(\theta|x)}.
\end{aligned}
$$

None of the situations described in problem 14 give equal probability to each possible value of θ.

17. Reasonable prior distributions for several situations:



R Code Appendix

Problem 3

```
# I think the matrices are an easier (though less memory-effectient) way
# than apply/sapply for this.
theta <- t(matrix(0:12, nrow = 13, ncol = 6)) # Possible theta values
x <- matrix(0:5, nrow = 6, ncol = 13) # Possible x values

probTable <- probGold(theta = theta, x = x)

# Sum down columns then across columns.
probTable <- rbind(probTable, apply(probTable, 2, sum))
```

```r
probTable <- cbind(probTable, apply(probTable, 1, sum))
colnames(probTable) <- c(0:12, 'Sum')
rownames(probTable) <- c(0:5, 'Sum')

print(xtable(probTable, align = '|c|ccccccccccccc|c|'),
      hline.after = c(-1, 0, 6, 7), floating = FALSE)
```

Problem 9

```r
# A generalizable function to compute the ML estimate when the parameter
# space is finite. I don't expect problems with nonuniqueness in this case.
mlGold <- function(x, L = probGold, theta = 0:12){
  Ls <- L(theta = theta, x = x)
  return(theta[Ls == max(Ls)])
}

par(mfrow = c(4, 4))

MLEs <- sapply(0:5, mlGold) # Get the MLE under each x in the support of X.
for(theta in 0:12){ # Loop for each actual theta.
  probX <- probGold(theta = theta, x = 0:5) # Distribution of X given theta
  probMLE <- rep(NA, 13)
  for(thetaHat in 0:12){ # Loop for each possible MLE value
    # Compute Pr(X is such that the ML estimate is thetaHat)
    probMLE[thetaHat + 1] <- sum(probX[MLEs == thetaHat])
  }
  plot(0:12, probMLE, type = 'o', pch = 20, ylim = c(0, 1),
       main = bquote(paste('Pr(', hat(theta), '|', theta==.(theta),')')),
       xlab = expression(hat(theta)), ylab = 'Probability')
}
```

Problem 10

```r
par(mfrow = c(1, 2))

plot(0:12, probGold(0:12), type = 'b', pch = 20,
     main = '(b) Likelihood when x = 1',
     xlab = expression(theta), ylab = expression(paste('L(',theta,'|1)')))

plot(0:12, probGold(0:12) / sum(probGold(0:12)), type = 'b', pch = 20,
     main = '(c) Normalized Likelihood when x = 1', xlab = expression(theta),
     ylab = expression(paste('L(',theta,'|1)')/sum(paste('L(',theta,'|1)'))))
```

Problem 14

```r
theta <- 0:12
priora <- c(0, # theta = 0
            0, # theta = 1
            1/36, # theta = 2
            2/36, # theta = 3
            3/36, # theta = 4
            4/36, # theta = 5
            5/36, # theta = 6
            6/36, # theta = 7
            5/36, # theta = 8
            4/36, # theta = 9
            3/36, # theta = 10
            2/36, # theta = 11
            1/36) # theta = 12
likelihood <- probGold(theta)
jointa <- priora * likelihood
marginala <- sum(jointa) # x = 1 is fixed, sum over theta
posteriora <- jointa / marginala
```

```r
likeliTable <- data.frame(theta, priora, likelihood, jointa, posteriora)
colnames(likeliTable) <- c('\\(\\theta\\)',
                           '\\(f(\\theta)\\)',
                           '\\(L(\\theta|x)\\)',
                           '\\(f(x,\\theta)\\)',
                           '\\(f(\\theta|x)\\)')
print(xtable(likeliTable, digits = c(0, 0, 4, 4, 4, 4)),
      floating = FALSE, include.rownames = FALSE,
      sanitize.colnames.function = function(x){return(x)})
```

```r
priorb <- c(0, # theta = 0
            1/6, # theta = 1
            1/6, # theta = 2
            1/6, # theta = 3
            1/6, # theta = 4
            1/6, # theta = 5
            1/6, # theta = 6
            0, # theta = 7
            0, # theta = 8
            0, # theta = 9
            0, # theta = 10
            0, # theta = 11
            0) # theta = 12
jointb <- priorb * likelihood
marginalb <- sum(jointb) # x = 1 is fixed, sum over theta
posteriorb <- jointb / marginalb

likeliTable[,c(2, 4, 5)] <- cbind(priorb, jointb, posteriorb)

print(xtable(likeliTable, digits = c(0, 0, 4, 4, 4, 4)),
      floating = FALSE, include.rownames = FALSE,
      sanitize.colnames.function = function(x){return(x)})
```

```r
priorc <- dbinom(0:12,12,0.5)
jointc <- priorc * likelihood
marginalc <- sum(jointc) # x = 1 is fixed, sum over theta
posteriorc <- jointc / marginalc

likeliTable[,c(2, 4, 5)] <- cbind(priorc, jointc, posteriorc)

print(xtable(likeliTable, digits = c(0, 0, 4, 4, 4, 4)),
      floating = FALSE, include.rownames = FALSE,
      sanitize.colnames.function = function(x){return(x)})
```

Problem 15

```r
par(mfrow = c(2, 3))

# Priors
plot(theta, priora, type = 'b', pch = 20, ylim = c(0, 0.30),
     main = 'Prior Distribution from 14(a)', xlab = expression(theta),
     ylab = expression(paste('f(',theta,')')))

plot(theta, priorb, type = 'b', pch = 20, ylim = c(0, 0.30),
     main = 'Prior Distribution from 14(b)', xlab = expression(theta),
     ylab = expression(paste('f(',theta,')')))

plot(theta, priorc, type = 'b', pch = 20, ylim = c(0, 0.30),
     main = 'Prior Distribution from 14(c)', xlab = expression(theta),
     ylab = expression(paste('f(',theta,')')))
```

```r
# Posteriors
plot(theta, posteriora, type = 'b', pch = 20, ylim = c(0, 0.30),
     main = 'Posterior Distribution from 14(a)', xlab = expression(theta),
     ylab = expression(paste('f(',theta,'|1)')))

plot(theta, posteriorb, type = 'b', pch = 20, ylim = c(0, 0.30),
     main = 'Posterior Distribution from 14(b)', xlab = expression(theta),
     ylab = expression(paste('f(',theta,'|1)')))

plot(theta, posteriorc, type = 'b', pch = 20, ylim = c(0, 0.30),
     main = 'Posterior Distribution from 14(c)', xlab = expression(theta),
     ylab = expression(paste('f(',theta,'|1)')))
```

Problem 17

```r
par(mfrow = c(2, 2))

# I'm tired of the connected-dots plots, but I might as well keep
# the theme going...
plot(0:12, (12:0)/78, type = 'b', pch = 20, ylim = c(0, 1),
     main = '(a) Preference for Blue',
     xlab = expression(theta), ylab = expression(f(theta)))

plot(0:12, c(0:6, 5:0)/36, type = 'b', pch = 20, ylim = c(0, 1),
     main = '(b) Half and Half',
     xlab = expression(theta), ylab = expression(f(theta)))

plot(0:12, c(6:0, 1:6)/42, type = 'b', pch = 20, ylim = c(0, 1),
     main = '(c) Not Equal',
     xlab = expression(theta), ylab = expression(f(theta)))

plot(0:12, rep(1/13, 13), type = 'b', pch = 20, ylim = c(0, 1),
     main = '(d) Uniform',
     xlab = expression(theta), ylab = expression(f(theta)))
```