1.

a. $p[R_1|I] = \frac{M}{N}$ $p[W_1|I] = \frac{N-M}{N}$

PROOF: $p[R_1|I] + p[W_1|I] = \frac{M}{N} + \frac{N-M}{N} = 1$

b.

$$p[\,R_1R_2 \mid I\,] = p[\,R_2 \mid R_1\,]p[R_1] = p[R_1]p[\,R_2 \mid R_1\,]$$
$$p[\,R_1 \mid I\,] = \frac{M}{N}$$
$$p[\,R_2 \mid R_1\,] = \frac{M-1}{N-1}$$
$$p[\,R_1R_2 \mid I\,] = \left(\frac{M}{N}\right)\left(\frac{M-1}{N-1}\right)$$

c. probability of drawing red ball on the 1st $m \leq M$ consecutive draws.

Logic: $\left(\frac{M}{N}\right)\left(\frac{M-1}{N-1}\right)\left(\frac{M-2}{M-2}\right)\dots\left(\frac{M-m}{N-m}\right)$

$$\frac{M!(N-m)!}{(M-m)!N!}$$

d) $M! = (M-1)(M-1)(M-2)\dots(M-m+1) = (M-m)!$

2. A *sample* is a set of n numbers $x = x_1, x_2, ..., x_n$. The *sample mean* is the average of the sample, $m[x] = \frac{x_1+...+x_n}{n}$, the *sample variance* is the average squared deviation of the sample values from the sample mean $s[x]^2 = \frac{(x_1-m[x])^2+...+(x_n-m[x])^2}{n}$, and the *sample standard deviation* is the square root of the sample variance.

a) Suppose we model the sample as a constant, μ. In general the likelihood that all the numbers in the sample will be to equal μ is zero. In order to allow for deviations, suppose that we assume that likelihood of the sample is proportional to $exp[-\frac{(x_1-\mu)^2+...+(x_n-\mu)^2}{n\sigma}]$, where σ is a second model parameter. Derive the expression for the posterior probability $p[\mu,\sigma|x]$, given a prior $p[\mu,\sigma]$.

2a. $p[x|\mu,\sigma] \propto e^{-\frac{(x-\mu)^2+\cdots+(x_n-\mu)^2}{n\sigma}}$

$p[\mu,\sigma|x] \propto p[\mu,\sigma]p[x|\mu,\sigma]$

$p[\mu,\sigma|x] \propto p[\mu,\sigma]e^{-\frac{(x-\mu)^2+\cdots+(x_n-\mu)^2}{n\sigma}}$

b) *Jeffreys' prior* is $p[\mu,\sigma] = d\mu\frac{d\sigma}{\sigma}$. Letting $d\mu = d\sigma = 1$ write the posterior probability with Jeffreys' prior.

2b. $p[\mu,\sigma] = d\mu\frac{d\sigma}{\sigma}$ let $d\mu = d\sigma = 1$

$\Rightarrow$

$p[\mu,\sigma] \propto \frac{1}{\sigma}$

$$p[\mu,\sigma] \propto p[x|\mu,\sigma]p[\mu,\sigma] = \frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{(x-\mu)^2+\cdots+(x_n-\mu)^2}{n\sigma}}\sigma^{-1}$$

3.

$$odds[\theta_1, \theta_2] = \frac{p[\theta_1|D]}{p[\theta_2|D]}$$

Let

$$p[\theta_1] = p[\theta_2]$$

By Bayes theorem,

$$p[\theta_1|D] = \frac{p[D|\theta_1]p[\theta_1]}{p[D]}$$

and

$$p[\theta_2|D] = \frac{p[D|\theta_2]p[\theta_2]}{p[D]}$$

$\Rightarrow$

$$\frac{p[\theta_1|D]}{p[\theta_2|D]} = \frac{\frac{p[D|\theta_1]p[\theta_1]}{p[D]}}{\frac{p[D|\theta_2]p[\theta_2]}{p[D]}} = \left(\frac{p[D|\theta_1]p[\theta_1]}{p[D]}\right)\left(\frac{p[D]}{p[D|\theta_2]p[\theta_2]}\right) = \left(\frac{p[D|\theta_1]p[\theta_1]}{p[D|\theta_2]p[\theta_2]}\right)$$

Let $p[\theta_1] = p[\theta_2]$

$\Rightarrow$

$$\frac{p[\theta_1|D]}{p[\theta_2|D]} = \left(\frac{p[D|\theta_1]}{p[D|\theta_2]}\right)$$

Exhibits that when priors = each other, likelihood = posterior….?

4. Gamma distribution: $p[\mu|\alpha,\beta] = \frac{\beta^{\alpha}}{\Gamma[\alpha]}\mu^{\alpha-1}e^{-\beta\mu}$ $\mu,\alpha,\beta > 0$

Poisson likelihood function:

$$p[x|\mu] = \left(\prod_{i=1}^{n} x_i!\right)^{-1} e^{\log|\mu|\sum_{i=1}^{n} x_i} e^{-n\mu}$$

Kernel of gamma distribution: $gamma[\alpha,\beta] \propto \mu^{\alpha-1}e^{-\beta\mu}$

Normalizing constant of gamma distribution: $\frac{\beta^{\alpha}}{\Gamma[\alpha]}$

$$\int \mu^{\alpha-1}e^{-\beta\mu}d\mu = \frac{\beta^{\alpha}}{\Gamma[\alpha]}$$

(integral of normalizing constant * kernel = 1)

Kernel of poisson: $e^{\log|\mu|\sum_{i=1}^{n} x_i} e^{-n\mu}$

Normalizing constant of poisson: $(\prod_{i=1}^{n} x_i!)$

$$\int e^{\log|\mu|\sum_{i=1}^{n} x_i} e^{-n\mu} = \left(\prod_{i=1}^{n} x_i!\right)$$

$$\int p[x|\mu] = \left(\prod_{i=1}^{n} x_i!\right)^{-1} e^{\log|\mu|\sum_{i=1}^{n} x_i} e^{-n\mu} = 1$$

$$p[\mu] \propto \mu^{\alpha-1}e^{-\beta\mu}$$
$$p[x|\mu] \propto e^{\log|\mu|\sum_{i=1}^{n} x_i} e^{-n\mu}$$
$$p[\mu|x] \propto p[\mu]p[x|\mu]$$
$$p[\mu|x] \propto \mu^{\alpha-1}e^{-\beta\mu}e^{\log|\mu|\sum_{i=1}^{n} x_i} e^{-n\mu}$$

Observe:

$p[\mu|x] \propto \mu^{\alpha-1}e^{-\beta\mu}e^{\log|\mu|\sum_{i=1}^{n} x_i} e^{-n\mu} = \mu^{\alpha-1}e^{-\beta\mu}\mu^{\sum_{i=1}^{n} x_i}e^{-n\mu}$

$=\mu^{(\alpha-1)+\sum_{i=1}^{n} x_i}e^{(-\beta\mu)+(-n\mu)}$

$$gamma\left[\alpha + \sum_{i=1}^{n} x_i, \beta + n\right] \propto \mu^{(\alpha-1)+\sum_{i=1}^{n} x_i}e^{(-\beta\mu)+(-n\mu)} = \mu^{\alpha-1}e^{-\beta\mu}e^{\log|\mu|\sum_{i=1}^{n} x_i} e^{-n\mu}$$
$$\propto p[\mu|x]$$

$$p[\mu|x,\alpha,\beta] = gamma\left[\alpha + \sum_{i=1}^{n} x_i, \beta + n\right] = \frac{(\beta+n)^{\alpha+\sum_{i=1}^{n} x_i}}{\Gamma[\alpha\sum_{i=1}^{n} x_i]}\mu^{(\alpha-1)+\sum_{i=1}^{n} x_i}e^{-\mu(\beta+n)}$$

Integral of kernels of likelihood*prior not equal to one (without normalizing constant – which is the integral of the distribution's kernel with respect to mu.

FACT: $(\prod_{i=1}^{n} x_i!)^{-1} = \frac{\beta^{\alpha}}{\Gamma[\alpha]}$

5.
Multinomial model:

$$p[x|\theta] = \frac{n!}{x_1! \ldots x_k!}\theta_1^{x_1} \ldots \theta_k^{x_k} = \frac{n!}{\prod_i^k x_i!}\prod_i^k \theta_i^{x_i}$$

Dirichlet distribution:

$$p[\theta|\alpha] = \frac{\Gamma[\Sigma_i \alpha_i]}{\prod_i \Gamma[\alpha_i]}\prod_{i=1}^k \theta_i^{a_i-1}$$

$$p[\theta] \propto \prod_{i=1}^k \theta_i^{a_i-1}$$

$$p[x|\theta] \propto \prod_i^k \theta_i^{x_i}$$

$$p[\theta|x] \propto p[\theta]p[x|\theta] = \prod_{i=1}^k \theta_i^{a_i-1} \prod_i^k \theta_i^{x_i} = \prod_{i=1}^k \theta_i^{x_i+a_i-1}$$

$$dirichlet[x+\alpha] \propto \prod_{i=1}^k \theta_i^{(x_i+a_i-1)} = \prod_{i=1}^k \theta_i^{a_i-1} \prod_i^k \theta_i^{x_i} \propto p[\theta|x]$$

6.

Normal distribution: $p[x|\mu,\sigma] = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$
Entropy: $s[p_i] = -\sum_{i=1}^{N} p_i Log[p_i]$

$$\Rightarrow$$

$$s[p_i] = -\int_{\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} Log\left[\frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}\right] dx = \frac{1}{2} Log[2\pi e\sigma^2]$$

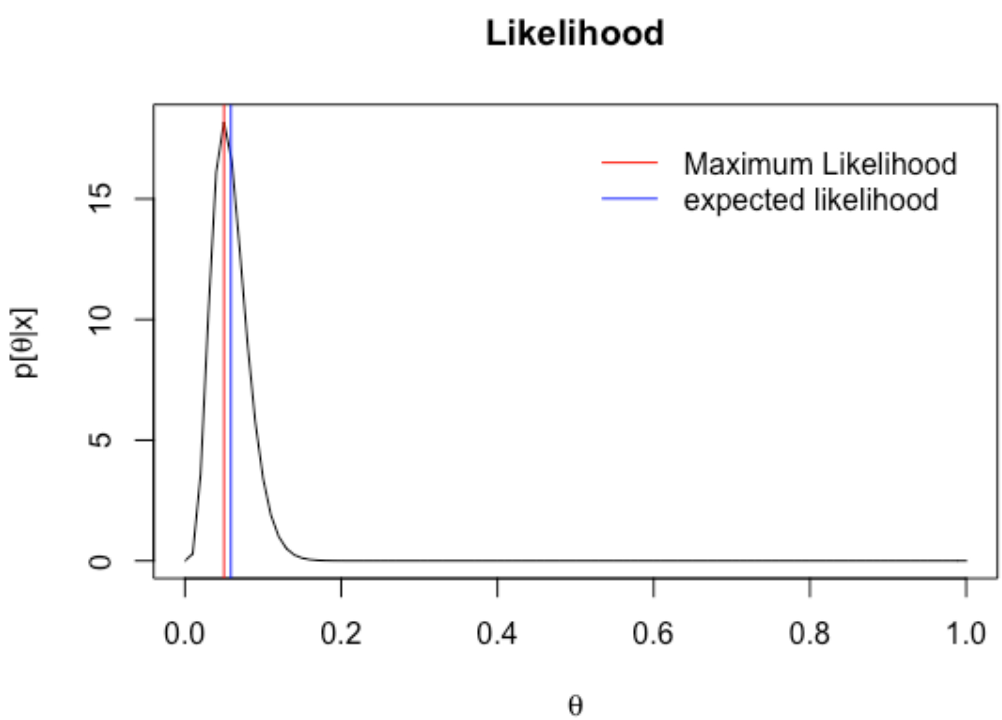
Intuitively, this makes sense since entropy, or uncertainty, should only be reliant on variance and thus the model becomes more imprecise as variance increases.

R exercises:

1) A bank has made 100 mortgages of a new type (say it's 2005 and they are subprime mortgages), and all have been outstanding 5 years. Of these 100, 5 of them have defaulted. The bank would like to estimate the probability θ of default in the first five years for this type of mortgage, and get some idea of how much uncertainty there is about the probability, given the observed data. These being a new type of mortgage, the bank assigns a uniform prior over θ.

    a) What is the likelihood $p[x|\theta]$? Binomial distribution
        i) $p[x = (n,k)|\theta] = \binom{n}{k}\theta^k(1-\theta)^{n-k}$
    b) Plot the likelihood in R and indicate on the plot (e.g. use the abline() function) the location of the maximum likelihood value of θ as well as the expected value of θ.

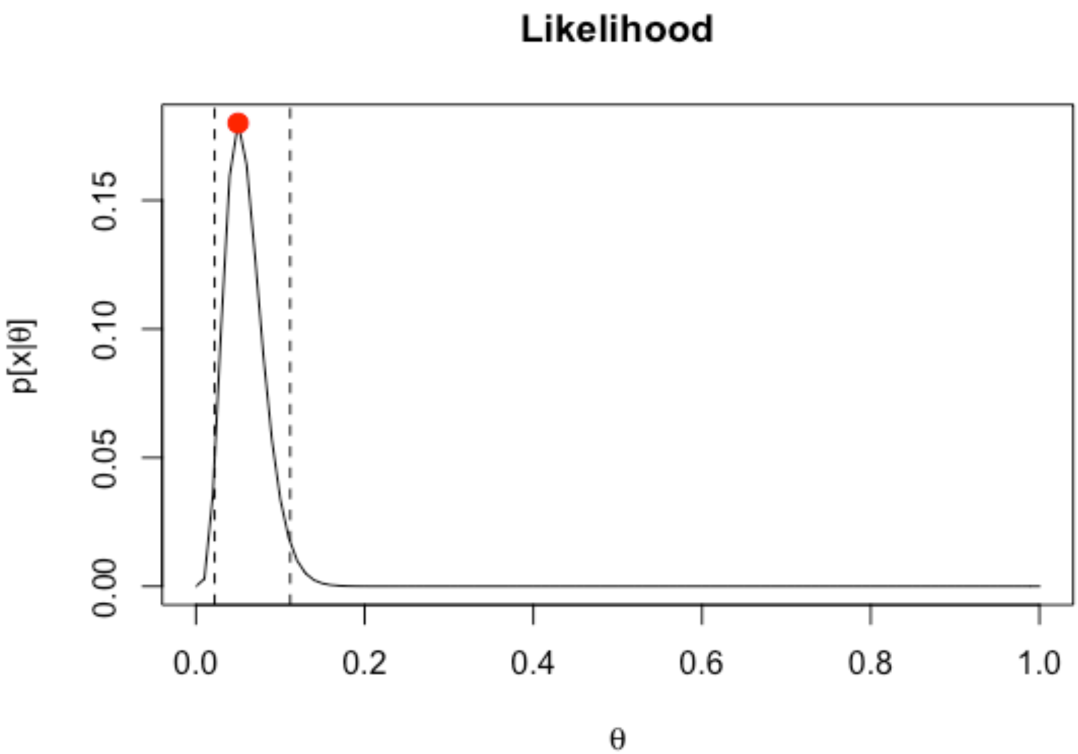
```
> optimize(log.post,interval=c(0,1),n=100,k=5,maximum=T)

$maximum                     $objective
[1] 0.05000354               [1] -1.714699
```



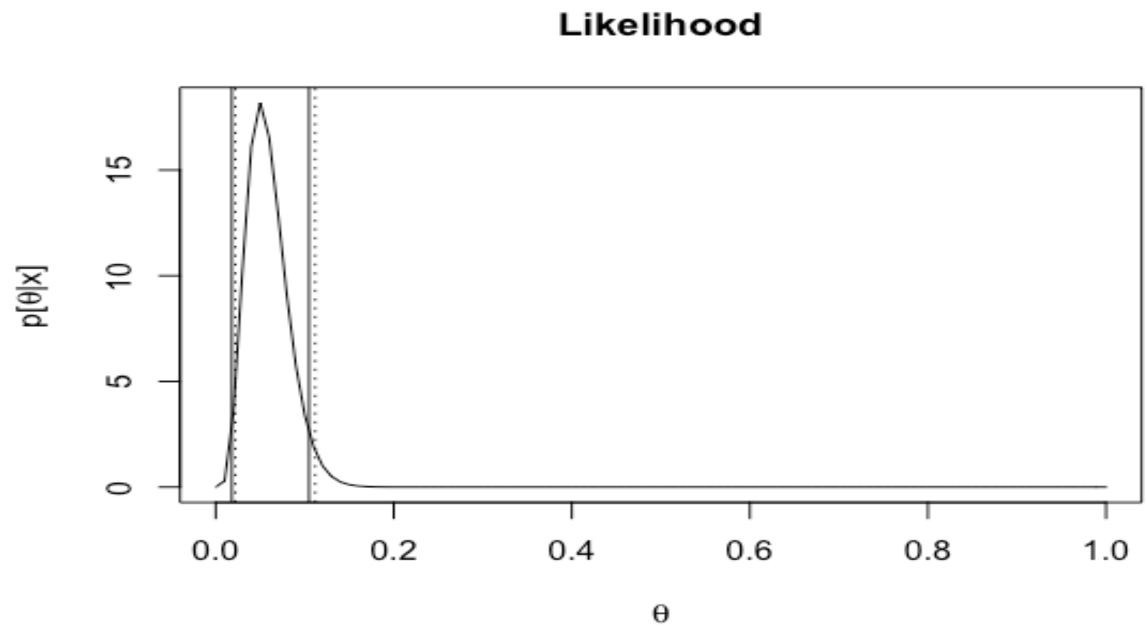
c) Using the quantile function *qbeta*() calculate and indicate asymmetric 95% confidence interval (cut off 2.5% of the left and right tail). Does this look like a reasonable confidence interval?

   i) Yes, since the highest likelihood is contained and peaks within $\theta \in (0.0221, 0.1118)$. =.897

   ii) Plotted with mle



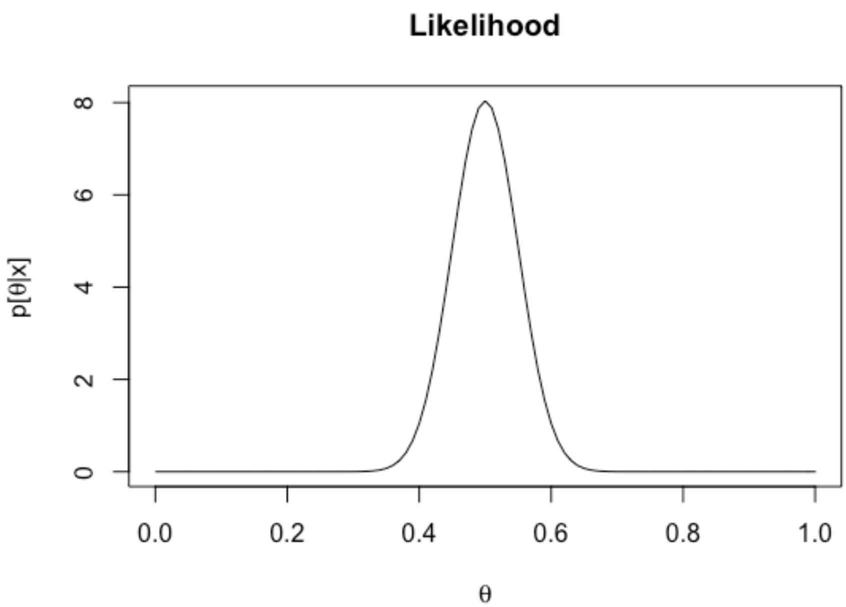
d) The shortest interval with 95% probability will have the likelihood the same height at each end. Using the package “TeachingDemos" use the HPD function to find the shortest interval.



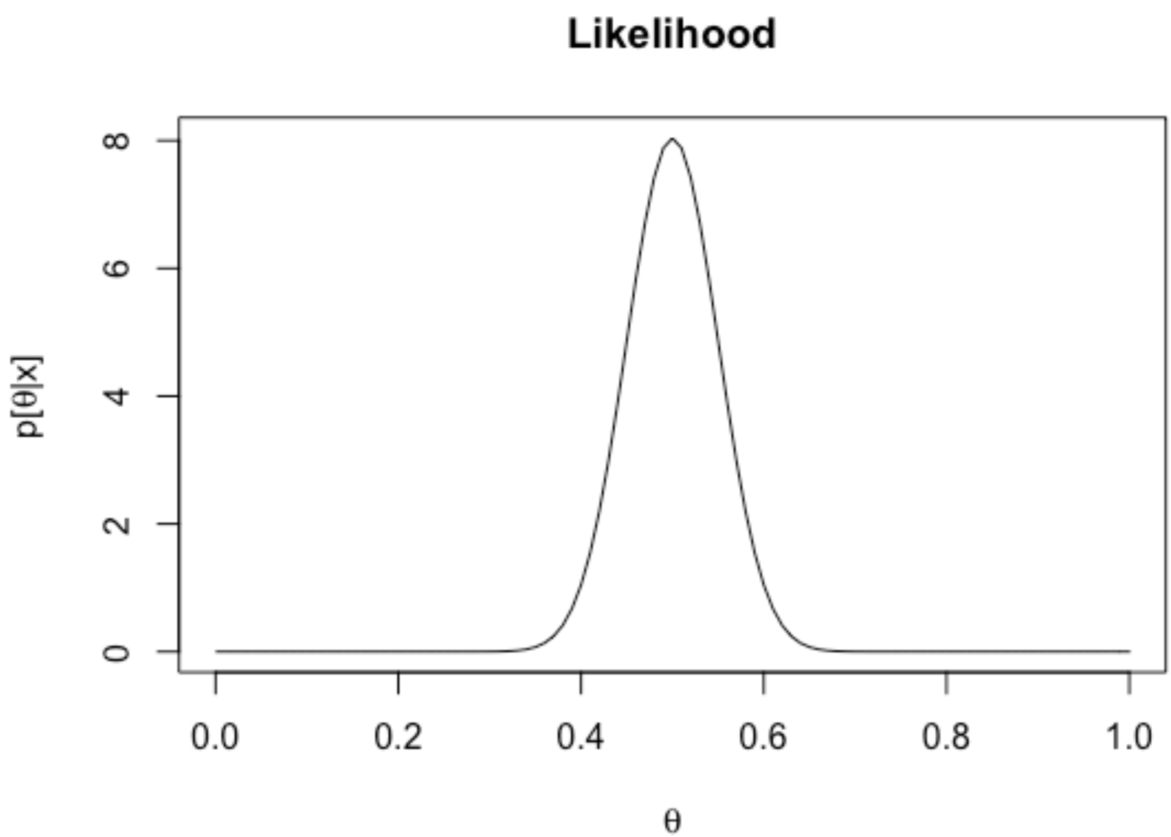
Shortest interval: $\theta \in (.0181, .1048)$

e) HPD is indeed the shortest interval

2.


Likelihood
p[θ|x]
θ
0.0
0.2
0.4
0.6
0.8
1.0
0
2
4
6
8

a)
b)


Likelihood
p[θ|x]
θ
0.0
0.2
0.4
0.6
0.8
1.0
0
2
4
6
8

c) neither since the order of the balls being drawn does not change the probability of the outcome.
d)IP
e)IP