STAT 840 Chapter 1 Exercises

Vsevolod Ladtchenko 20895137

[1] 1 [1] 2 [1] 3 [1] 4 [1] 5 [1] 6 [1] 7 [1] 8 [1] 9 [1] 10 [1] 11

TRICKS TO REMEMBER

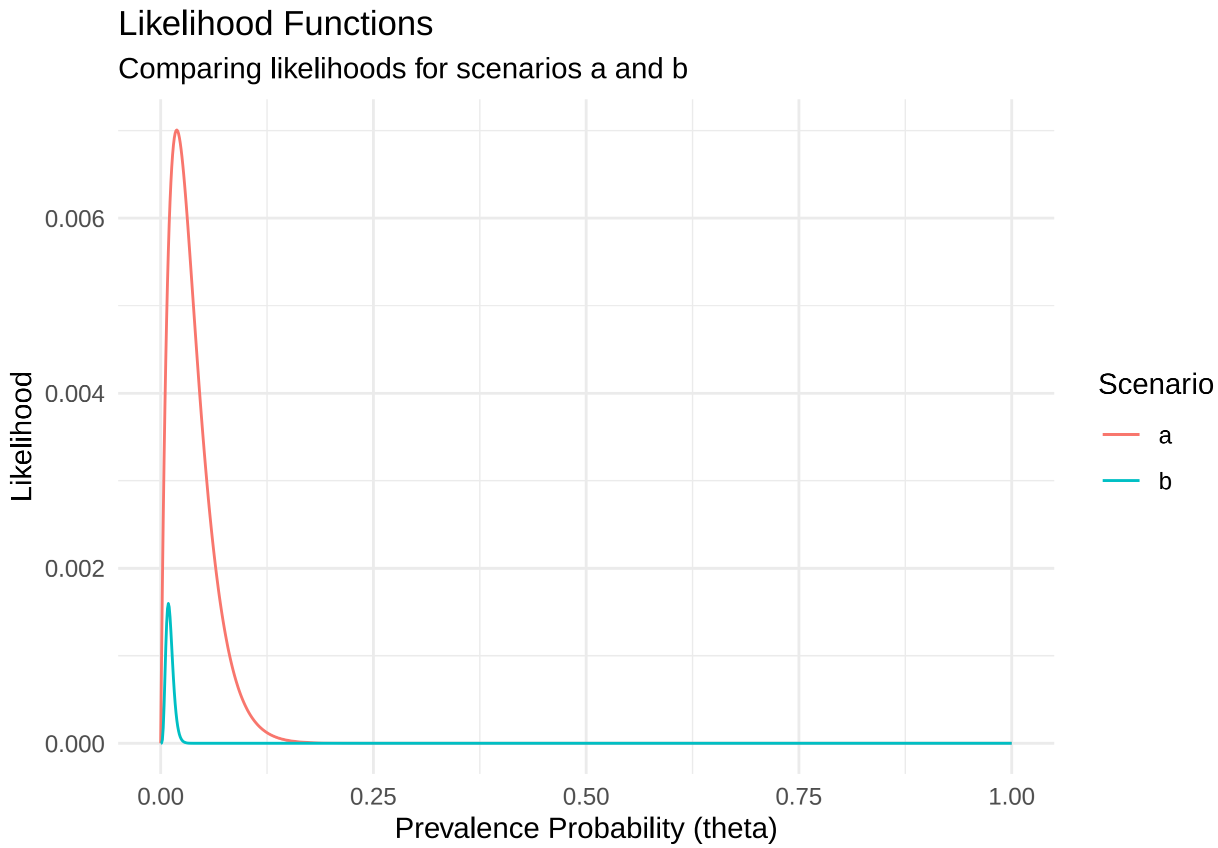
- mode = highest point of density. get it by taking derivative
- gamma/beta mean = rewrite as Gam(a+1,b) using Gam(n) = (n-1)!
- chap 1 variance decomposition: use only decomposition for xbar

PROBABLY CORRECT BUT I SHOULD ASK ANYWAYS

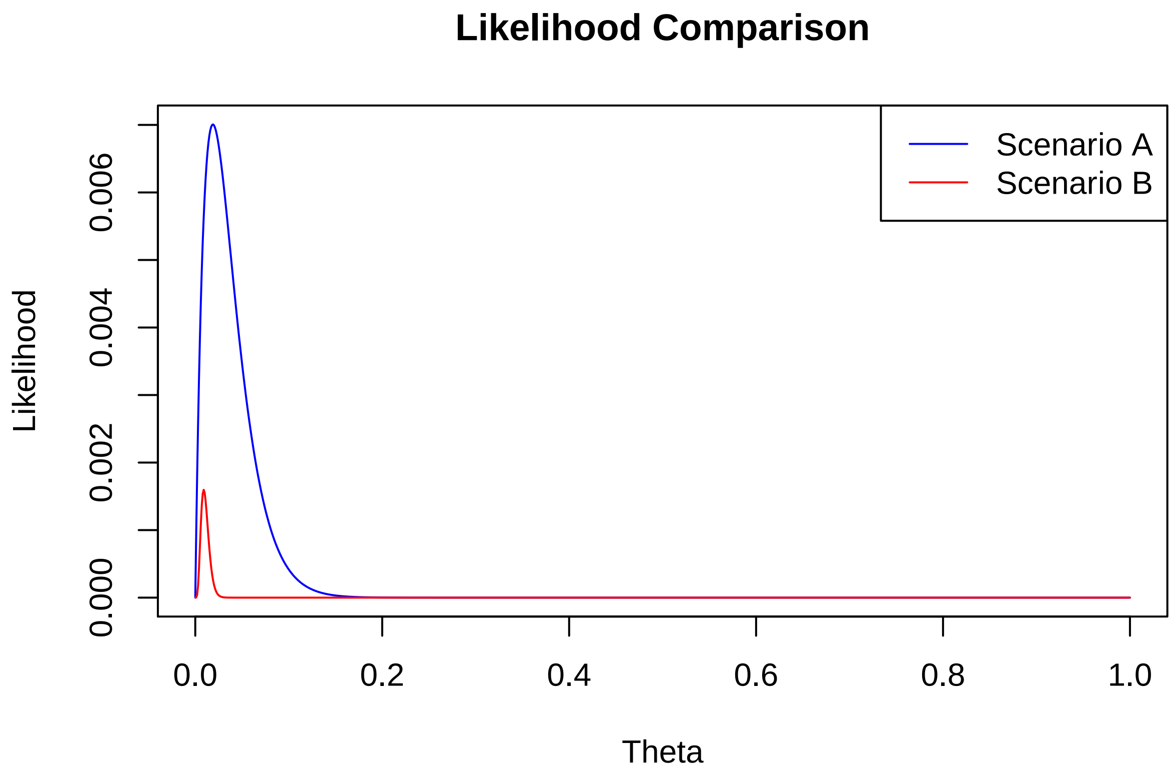
11, probably requires Rao Blackwell? 18

Problem 1.1

```r
# R script
library(ggplot2)

# Define likelihood functions
likelihood_a <- function(theta) {
  (1 - theta)^(53 - 1) * theta
}

likelihood_b <- function(theta) {
  choose(552 - 1, 5 - 1) * theta^5 * (1 - theta)^(552 - 5)
}

# Create a sequence of theta values between 0 and 1
theta_vals <- seq(0, 1, 0.001)

# Calculate likelihoods for each theta value
likelihoods_a <- sapply(theta_vals, likelihood_a)
likelihoods_b <- sapply(theta_vals, likelihood_b)

# Create a data frame for plotting
plot_data <- data.frame(
  Theta = rep(theta_vals, 2),
  Likelihood = c(likelihoods_a, likelihoods_b),
  Scenario = factor(rep(c("a", "b"), each = length(theta_vals)))
)

# Plot likelihoods
ggplot(plot_data, aes(x = Theta, y = Likelihood, color = Scenario)) +
  geom_line() +
  theme_minimal() +
  labs(title = "Likelihood Functions",
       subtitle = "Comparing likelihoods for scenarios a and b",
       x = "Prevalence Probability (theta)",
       y = "Likelihood",
       color = "Scenario")
```




```r
theta <- seq(0, 1, length.out = 1000)

likelihood_a <- (1 - theta)^(53 - 1) * theta
likelihood_b <- choose(552 - 1, 5 - 1) * theta^5 * (1 - theta)^(552 - 5)

plot(theta, likelihood_a, type = "l", col = "blue", xlab = "Theta", ylab = "Likelihood", main = "Likelil
lines(theta, likelihood_b, col = "red")

legend("topright", legend = c("Scenario A", "Scenario B"), col = c("blue", "red"), lty = 1)
```



- - - GPT-4 ANSWER START

a. If the genotype makes its first appearance on the 53rd subject analyzed, then the first 52 subjects do not have the genotype, and the 53rd subject has the genotype. Assuming the subjects are independent and have the same prevalence probability θ, the likelihood function can be modeled using a geometric distribution. The probability mass function (PMF) of a geometric distribution is:

$P(X = k) = (1 - \theta)^{(k-1)} \cdot \theta$

In this case, $k = 53$. So the likelihood function $L(\theta)$ is:

$L(\theta) = (1 - \theta)^{(53-1)} \cdot \theta$

b. If the scientists had planned to stop when they found five subjects with the genotype of interest, and they analyzed 552 subjects, we can model this using a negative binomial distribution. The PMF of a negative binomial distribution is:

$P(X = k) = C(k - 1, r - 1) \cdot \theta^r \cdot (1 - \theta)^{(k-r)}$

In this case, $r = 5$ (the number of successes or genotypes of interest), and $k = 552$ (the number of trials). So the likelihood function $L(\theta)$ is:

$L(\theta) = C(552 - 1, 5 - 1) \cdot \theta^5 \cdot (1 - \theta)^{(552-5)}$

c. We can plot both likelihood functions in R:

The plot will show the likelihood functions for both scenarios a and b. You will notice that the likelihood function in scenario a, where the genotype appears on the 53rd subject, is more spread out with a lower peak than the likelihood function in scenario b, where the scientists stop after finding five subjects with the genotype. This indicates that the data in scenario b provides more information about the prevalence probability θ, resulting in a more concentrated likelihood function around the most likely value of θ.

- - - GPT-4 ANSWER END

Problem 1.2

- - - GPT-4 ANSWER START

- - - GPT-4 ANSWER END

Problem 1.3

- - - GPT-4 ANSWER START

- - - GPT-4 ANSWER END

Problem 1.4

- - - GPT-4 ANSWER START

- - - GPT-4 ANSWER END

Problem 1.5

- - - GPT-4 ANSWER START

- - - GPT-4 ANSWER END

Problem 1.6

- - - GPT-4 ANSWER START

- - - GPT-4 ANSWER END

Problem 1.7

- - - GPT-4 ANSWER START

- - - GPT-4 ANSWER END

Problem 1.8

- - - GPT-4 ANSWER START

- - - GPT-4 ANSWER END

Problem 1.9

- - - GPT-4 ANSWER START

- - - GPT-4 ANSWER END

Problem 1.10

- - - GPT-4 ANSWER START

- - - GPT-4 ANSWER END

Problem 1.11

- - - GPT-4 ANSWER START

Given that $X_1, ..., X_n$ are n Bernoulli(p) trials, we can model the joint distribution of X_i's and T as a binomial distribution. Recall that $T = \sum X_i$, which means T follows a binomial distribution with parameters n and p.

Conditional expectation $E[X_1|T]$:

To find the conditional expectation $E[X_1|T]$, we can use the following property of conditional expectation:

$E[X_1|T] = \sum(x * P(X_1 = x|T))$ for x in $0, 1$

We need to find the conditional probabilities $P(X_1 = 0|T)$ and $P(X_1 = 1|T)$. Using the definition of conditional probability and properties of binomial distribution, we get:

$P(X_1 = 1|T) = \frac{P(X_1=1,T=t)}{P(T=t)} = \frac{P(T-X_1=t-1)}{P(T=t)}$

Since T follows a binomial distribution with parameters n and p, $(T - X_1)$ will follow a binomial distribution with parameters $(n - 1)$ and p:

$P(T - X_1 = t - 1) = C(n - 1, t - 1) * p^{(t-1)} * (1 - p)^{(n-t)}$

$P(T = t) = C(n, t) * p^t * (1 - p)^{(n-t)}$

Therefore,

$P(X_1 = 1|T) = \frac{C(n-1,t-1)*p^{(t-1)}*(1-p)^{(n-t)}}{C(n,t)*p^t*(1-p)^{(n-t)}}$

$P(X_1 = 1|T) = \frac{t}{n}$

Since X_1 is a Bernoulli random variable, $P(X_1 = 0|T) = 1 - P(X_1 = 1|T) = \frac{(n-t)}{n}$.

Now, we can find $E[X_1|T]$:

$E[X_1|T] = \sum(x * P(X_1 = x|T)) = 0 * P(X_1 = 0|T) + 1 * P(X_1 = 1|T) = \frac{t}{n}$

Compare variance:

The variance of the estimator $E[X_1|T]$ can be calculated as:

$\text{Var}(X_1|T) = E[(X_1 - E[X_1|T])^2|T] = E[X_1^2|T] - (E[X_1|T])^2$

Since X_1 is a Bernoulli random variable, $X_1^2 = X_1$. Thus,

$E[X_1^2|T] = E[X_1|T] = \frac{t}{n}$

So,

$\text{Var}(X_1|T) = \frac{t}{n} - \left(\frac{t}{n}\right)^2$

Now let's find the variance of X_1:

$\text{Var}(X_1) = E[X_1^2] - (E[X_1])^2 = p - p^2 = p(1 - p)$

Comparing the variances, we can see that $\text{Var}(X_1|T)$ is generally different from $\text{Var}(X_1)$. However, the variance of the estimator $E[X_1|T]$ depends on the observed value of T, while the variance of X_1 depends on the unknown parameter p. In some cases, the conditional variance $\text{Var}(X_1|T)$ can be smaller than $\text{Var}(X_1)$, indicating that conditioning on T provides additional information that reduces the uncertainty about X_1.

- - - GPT-4 ANSWER END

Problem 1.12

$$
\begin{aligned}
Lik &= \prod_{i=1}^{n} \theta^{x_i}(1-\theta)^{1-x_i} \\
&= \theta^{\sum_{i=1}^{n} x_i}(1-\theta)^{n-\sum_{i=1}^{n} x_i} \\
lik &= \sum_{i=1}^{n} x_i \log(\theta) + (n - \sum_{i=1}^{n} x_i)\log(1-\theta) \\
s &= \sum_{i=1}^{n} x_i \\
\partial_\theta lik &= \frac{s}{\theta} - \frac{n-s}{1-\theta} = 0 \\
\frac{s}{\theta} &= \frac{n-s}{1-\theta} \\
s(1-\theta) &= (n-s)\theta \\
s - s\theta &= n\theta - s\theta \\
\theta_{MLE} &= s/n \\
\theta^2_{MLE} &= (s/n)^2
\end{aligned}
$$

We know that the MLE of a function is the function of the MLE, by the invariance property of the MLE.

- - - GPT-4 ANSWER START

To find the maximum likelihood estimate (MLE) of θ^2, we first need to find the likelihood function, then its derivative with respect to θ, and finally set the derivative equal to 0 and solve for θ.

Likelihood function:

Since $X_1, ..., X_n$ are independent and identically distributed (i.i.d.) random variables, the likelihood function $L(\theta)$ is given by the product of probabilities:

$L(\theta) = P(X_1 = x_1, ..., X_n = x_n|\theta) = \prod_i P(X_i = x_i|\theta)$

Given that $Pr(X_i = 1) = \theta$ and $Pr(X_i = 0) = 1 - \theta$, the likelihood function can be written as:

$L(\theta) = \theta^s \cdot (1-\theta)^{(n-s)}$

where s is the number of successes (i.e., the number of 1's) in the sample.

Log-likelihood function:

It's easier to work with the log-likelihood function, so let's take the natural logarithm of $L(\theta)$:

$\log L(\theta) = s \cdot \log(\theta) + (n-s) \cdot \log(1-\theta)$

Derivative of log-likelihood function:

To find the MLE of θ, we'll first take the derivative of the log-likelihood function with respect to θ:

$\frac{d(\log L(\theta))}{d\theta} = \frac{s}{\theta} - \frac{n-s}{1-\theta}$

Set the derivative equal to 0 and solve for θ:

$\frac{s}{\theta} - \frac{n-s}{1-\theta} = 0$

Solving this equation for θ, we get the MLE of θ:

$\hat{\theta} = \frac{s}{n}$

MLE of θ^2:

Now we have the MLE of θ, we can find the MLE of θ^2:

$(\hat{\theta})^2 = \left(\frac{s}{n}\right)^2$

So, the maximum likelihood estimate of θ^2 is $\left(\frac{s}{n}\right)^2$.

- - - GPT-4 ANSWER END

Problem 1.13

$$
\begin{aligned}
f(\theta \mid x) &\propto f(x \mid \theta) f(\theta) \\
f(\theta) &= 60\theta^2(1-\theta)^3 \\
f(x \mid \theta) &= \binom{10}{3}\theta^3(1-\theta)^7 \\
f(\theta \mid x) &\propto \theta^3(1-\theta)^7\theta^2(1-\theta)^3 \\
&\propto \theta^5(1-\theta)^{10}
\end{aligned}
$$

This is Beta 6, 11 which has mean 6/(6+11) so the Bayes estimator is 6/17.

- - - GPT-4 ANSWER START

To find the Bayes estimate of θ with respect to the squared error loss function (i.e., the posterior mean), we'll first compute the likelihood of the observed data and then use Bayes' theorem to find the posterior pdf of θ. Finally, we'll compute the posterior mean.

Prior pdf of θ:

$\pi(\theta) = 60\theta^2(1-\theta)^3$ for $0 < \theta < 1$

$\pi(\theta) = 0$ otherwise

Observed data: A random sample of 10 apples, 3 of which are bad. This follows a binomial distribution with n = 10 trials and k = 3 successes (bad apples).

Likelihood P(x|θ): Binomial distribution with parameters n = 10, k = 3, and θ.

$P(x = 3|\theta) = C(10, 3) \cdot \theta^3 \cdot (1-\theta)^{(10-3)}$

Posterior pdf P(θ|x): Use Bayes' theorem.

$P(\theta|x) \propto P(x|\theta) \cdot \pi(\theta) = C(10, 3) \cdot \theta^3 \cdot (1-\theta)^{(10-3)} \cdot 60\theta^2(1-\theta)^3$

Simplifying, we get:

$P(\theta|x) \propto \theta^5 \cdot (1-\theta)^6$ for $0 < \theta < 1$

Normalize the posterior pdf by finding the constant of proportionality (denoted by c):

To do this, we need to integrate $P(\theta|x)$ over the interval [0, 1] and set the integral equal to 1.

$c \cdot \int_0^1 \theta^5 \cdot (1-\theta)^6, d\theta = 1$

Compute the posterior mean (Bayes estimate):

$E(\theta|x) = \int_0^1 \theta \cdot P(\theta|x), d\theta = c \cdot \int_0^1 \theta^6 \cdot (1-\theta)^6, d\theta$

At this point, we need to calculate the integrals in steps 5 and 6 numerically. You can use numerical integration methods (e.g., the trapezoidal rule or Simpson's rule) or a computational tool like R or WolframAlpha to find the values.

Once you have the value of c from step 5, substitute it back into step 6 to find the Bayes estimate of θ with respect to the squared error loss function (the posterior mean).

- - - GPT-4 ANSWER END

Problem 1.14

$$
\begin{aligned}
f(\theta \mid x) &\propto f(x \mid \theta) f(\theta) \\
f(x \mid \theta) &= \binom{8}{2} \theta^2 (1-\theta)^6 \\
f(\theta = 0.1 \mid x) &\propto f(x \mid \theta = 0.1) 0.7 \\
&= 0.5418274 \\
f(\theta = 0.2 \mid x) &\propto f(x \mid \theta = 0.2) 0.3 \\
&= 0.4581726 \\
\theta_{Bayes} = \mathbb{E}[\theta \mid x] &= 0.1 f(\theta = 0.1 \mid x) + 0.2 f(\theta = 0.2 \mid x) \\
&= 0.1458173 \\
\theta_{MAP} &= \arg\max_\theta \{f(\theta = 0.1 \mid x), f(\theta = 0.2 \mid x)\} \\
&= \arg\max_\theta \{0.5418274, 0.4581726\} \\
&= 0.1
\end{aligned}
$$

theta is a discrete variable. Its expectation will be a discrete expectation (no integral) and the MAP estimate will be the larger of two possible values, because there is no density to maximize.

the posterior distribution of theta will be discrete, so there will be only two values. don't forget the normalization values, which we can get using the law of total probability.

- - - GPT-4 ANSWER START

To derive the posterior probability density function (pdf) of θ and find the Bayes and Maximum a Posteriori (MAP) estimators, we can use Bayes' theorem. In this problem, we have two possible values for θ: 0.1 and 0.2, with prior probabilities $\pi(0.1) = 0.7$ and $\pi(0.2) = 0.3$. We observed that 2 out of 8 items are defective.

Let X denote the number of defective items in the sample of 8 items. We want to find the posterior probabilities $P(\theta = 0.1|X = 2)$ and $P(\theta = 0.2|X = 2)$. Using Bayes' theorem:

$P(\theta|x) = \frac{P(x|\theta) \cdot \pi(\theta)}{P(x)}$

First, we need to compute the likelihood $P(x|\theta)$, which follows a binomial distribution because we have a fixed number of trials (n=8) with two possible outcomes (defective or non-defective) and a known probability of success (defective) θ.

$P(x = 2|\theta = 0.1) = C(8, 2) \cdot (0.1)^2 \cdot (1 - 0.1)^{(8-2)} = 28 \cdot 0.01 \cdot 0.9^6$

$P(x = 2|\theta = 0.2) = C(8, 2) \cdot (0.2)^2 \cdot (1 - 0.2)^{(8-2)} = 28 \cdot 0.04 \cdot 0.8^6$

Now, we can find the posterior probabilities:

$P(\theta = 0.1|X = 2) = \frac{P(X=2|\theta=0.1) \cdot \pi(0.1)}{P(X=2|\theta=0.1) \cdot \pi(0.1) + P(X=2|\theta=0.2) \cdot \pi(0.2)}$

$P(\theta = 0.2|X = 2) = \frac{P(X=2|\theta=0.2) \cdot \pi(0.2)}{P(X=2|\theta=0.1) \cdot \pi(0.1) + P(X=2|\theta=0.2) \cdot \pi(0.2)}$

For the Bayes estimator, we compute the expected value of θ given the data:

$E(\theta|X = 2) = \theta_1 \cdot P(\theta = 0.1|X = 2) + \theta_2 \cdot P(\theta = 0.2|X = 2)$

The MAP estimator is the value of θ that maximizes the posterior probability, which is the one with the highest posterior probability:

$MAP(\theta|X = 2) = \arg\max P(\theta = 0.1|X = 2), P(\theta = 0.2|X = 2)$

Compute the values using the formulas above to find the Bayes and MAP estimators of θ.

- - - GPT-4 ANSWER END

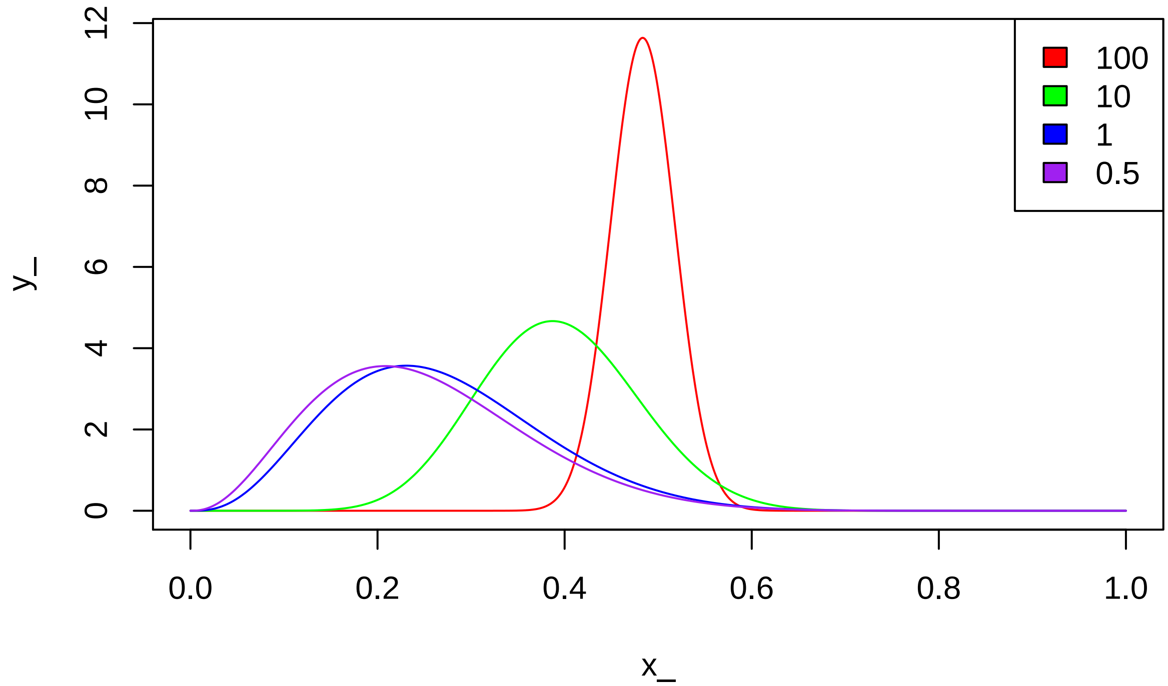
```r
problem_1_14 = function()
{
  lik = function(theta)
  {
    ret = choose(8, 2) * theta^2 * (1 - theta)^6
    return(ret)
  }

  f01_ = lik(theta = 0.1) * 0.7 # unnormalized posterior probabilities
  f02_ = lik(theta = 0.2) * 0.3
  norm = f01_ + f02_ # normalization constant is their sum in discrete case

  f01 = f01_ / norm # normalized posterior probabilities
  f02 = f02_ / norm

  print(f01)
  print(f02)

  bayes = 0.1 * f01 + 0.2 * f02
  print(bayes)
}

problem_1_14()
```

```
## [1] 0.5418274
## [1] 0.4581726
## [1] 0.1458173
```

Problem 1.15

$$
\begin{aligned}
f(x \mid \theta) &\propto \theta^x (1-\theta)^{1-x} \\
f(X \mid \theta) &\propto \prod_{i=1}^{n} \theta^{x_i} (1-\theta)^{1-x_i} \\
&\propto \theta^{\sum_{i=1}^{n} x_i} (1-\theta)^{n-\sum_{i=1}^{n} x_i} \\
f(\theta) &\propto \theta^{\alpha-1} (1-\theta)^{\beta-1} \\
f(\theta \mid X) &= f(X \mid \theta) f(\theta) \\
&\propto \theta^{\sum_{i=1}^{n} x_i} (1-\theta)^{n-\sum_{i=1}^{n} x_i} \theta^{\alpha-1} (1-\theta)^{\beta-1} \\
&\propto \theta^{\alpha-1+\sum_{i=1}^{n} x_i} (1-\theta)^{\beta-1+n-\sum_{i=1}^{n} x_i} \\
&\sim \mathrm{Beta}(\alpha_{prior} + \sum_{i=1}^{n} x_i, \beta_{prior} + n - \sum_{i=1}^{n} x_i)
\end{aligned}
$$

```r
problem_1_15 = function()
{
  x = c(0, 1, 0, 1, 0, 0, 0, 0, 1, 0, 0, 0, 0)

  bay_est = function (a,b) a/(a+b)

  map_est = function (a,b) (a-1)/(a+b-2)

  params = c(100, 10, 1, 0.5)
  colors = c('red','green','blue','purple')

  for (i in 1:4)
  {
    a_prior = params[i]
    b_prior = params[i]

    a_post = a_prior + sum(x)
    b_post = b_prior + length(x) - sum(x)

    x_ = seq(0,1,1/1000)
    y_ = dbeta(x_, a_post, b_post)

    if (i==1)
    {
      plot(x=x_,y=y_,type='l', col=colors[i])
      legend(x = "topright", legend=params, fill = colors)
    }
    else lines(x=x_,y=y_, col=colors[i])

    print(paste("for prior (", a_prior, ",",b_prior, ")"))
    print(round(c(Bay=bay_est(a_post, b_post), MAP=map_est(a_post, b_post)),4))
  }
}

problem_1_15()
```




```
## [1] "for prior ( 100 , 100 )"
##    Bay    MAP
## 0.4836 0.4834
## [1] "for prior ( 10 , 10 )"
##    Bay    MAP
## 0.3939 0.3871
## [1] "for prior ( 1 , 1 )"
##    Bay    MAP
## 0.2667 0.2308
## [1] "for prior ( 0.5 , 0.5 )"
##    Bay    MAP
## 0.2500 0.2083
```

Problem 1.16

$$
\begin{aligned}
X_i &\sim Pois(\theta) \\
f(x_i \mid \theta) &= \frac{\theta^{x_i} e^{-\theta}}{x_i!} \\
f(X \mid \theta) &= \prod_{i=1}^{n} \frac{\theta^{x_i} e^{-\theta}}{x_i!} \\
f(\theta) &= \frac{\beta^\alpha}{\Gamma(\alpha)} \theta^{\alpha-1} e^{-\beta\theta} \\
f(\theta \mid X) &= f(X \mid \theta) f(\theta) \\
&= \frac{\beta^\alpha}{\Gamma(\alpha)} \theta^{\alpha-1} e^{-\beta\theta} \prod_{i=1}^{n} \frac{\theta^{x_i} e^{-\theta}}{x_i!} \\
&\propto \theta^{\alpha-1} e^{-\beta\theta} \prod_{i=1}^{n} \theta^{x_i} e^{-\theta} \\
&\propto \theta^{\alpha-1} e^{-\beta\theta - n\theta} \theta^{\sum_{i=1}^{n} x_i} \\
&\propto \theta^{\alpha-1+\sum_{i=1}^{n} x_i} e^{-\beta\theta - n\theta} \\
&\propto \theta^{(\alpha+\sum_{i=1}^{n} x_i)-1} e^{-\theta(\beta+n)} \\
&\propto \theta^{(\alpha+\sum_{i=1}^{n} x_i)-1} e^{-(\beta+n)\theta} \\
&\sim \text{Gamma}(\alpha_{post} = \alpha_{prior} + \sum_{i=1}^{n} x_i, \beta_{post} = \beta_{prior} + n)
\end{aligned}
$$

Posterior mode:

$$
\begin{aligned}
f(\theta) &= \frac{\beta^\alpha}{\Gamma(\alpha)} \theta^{\alpha-1} e^{-\beta\theta} \\
f(\theta) &\propto \theta^{\alpha-1} e^{-\beta\theta} \\
f'(\theta) &\propto \theta^{\alpha-1}(-\beta) e^{-\beta\theta} + (\alpha-1)\theta^{\alpha-2} e^{-\beta\theta} \\
0 &= \theta^{\alpha-1}(-\beta) + (\alpha-1)\theta^{\alpha-2} \\
0 &= (-\beta\theta + \alpha - 1)\theta^{\alpha-2} \\
&\quad \text{assume } \theta \neq 0 \text{ and later check value at } 0 \\
0 &= -\beta\theta + \alpha - 1 \\
\beta\theta &= \alpha - 1 \\
\theta &= \frac{\alpha-1}{\beta} \\
&\quad \text{we know support is positive} \\
&\quad \text{this is positive for } \alpha > 1 \\
&\quad \alpha < 1 \text{ implies max at } 0
\end{aligned}
$$

Posterior mean:

$$
\begin{aligned}
f(\theta) &= \frac{\beta^\alpha}{\Gamma(\alpha)}\theta^{\alpha-1}e^{-\beta\theta} \\
\mathbb{E}[\theta] &= \int_0^\infty \theta\frac{\beta^\alpha}{\Gamma(\alpha)}\theta^{\alpha-1}e^{-\beta\theta}d\theta \\
&= \int_0^\infty \frac{\beta^\alpha}{\Gamma(\alpha)}\theta^{(\alpha+1)-1}e^{-\beta\theta}d\theta \\
&= \int_0^\infty \frac{1}{\beta}\frac{\beta^{\alpha+1}}{\Gamma(\alpha)}\theta^{(\alpha+1)-1}e^{-\beta\theta}d\theta \\
&\Gamma(\alpha) = (\alpha-1)! \\
&\Gamma(\alpha+1) = \alpha! \\
&\Gamma(\alpha+1) = \alpha\Gamma(\alpha) \\
&= \int_0^\infty \frac{\alpha}{\beta}\frac{\beta^{\alpha+1}}{\Gamma(\alpha+1)}\theta^{(\alpha+1)-1}e^{-\beta\theta}d\theta \\
&= \frac{\alpha}{\beta}\int_0^\infty \frac{\beta^{\alpha+1}}{\Gamma(\alpha+1)}\theta^{(\alpha+1)-1}e^{-\beta\theta}d\theta \\
&= \frac{\alpha}{\beta}
\end{aligned}
$$

Thus the posterior mean and mode of θ are:

$$
\begin{aligned}
\theta_{mode} &= \frac{\alpha_{prior} + \sum_{i=1}^n x_i - 1}{\beta_{prior} + n} \\
\theta_{mean} &= \frac{\alpha_{prior} + \sum_{i=1}^n x_i}{\beta_{prior} + n}
\end{aligned}
$$

Problem 1.17

$$
\begin{aligned}
\theta &\sim \mathrm{Beta}(\alpha = 5, \beta = 10) \\
f(\theta) &\propto \theta^{\alpha-1}(1-\theta)^{\beta-1} \\
&\propto \theta^4(1-\theta)^9 \\
f(x \mid \theta) &= \binom{n}{k}\theta^k(1-\theta)^{n-k} \\
&\propto \theta^1(1-\theta)^{19} \\
f(\theta \mid x) &= f(x \mid \theta)f(\theta) \\
&\propto \theta^1(1-\theta)^{19}\theta^4(1-\theta)^9 \\
&\propto \theta^5(1-\theta)^{28} \\
&\sim \mathrm{Beta}(6, 29)
\end{aligned}
$$

MAP estimate (mode):

$$
\begin{aligned}
f(\theta \mid x) &\propto \theta^{\alpha-1}(1-\theta)^{\beta-1} \\
f'(\theta \mid x) &\propto -(\beta-1)\theta^{\alpha-1}(1-\theta)^{\beta-2} + (\alpha-1)\theta^{\alpha-2}(1-\theta)^{\beta-1} \\
0 &= \left(-(\beta-1)\theta + (\alpha-1)(1-\theta)\right)\theta^{\alpha-2}(1-\theta)^{\beta-2} \\
&\quad \text{assume } \theta, (1-\theta) \neq 0 \text{ later check value at } 0 \\
0 &= -(\beta-1)\theta + (\alpha-1)(1-\theta) \\
0 &= (1-\beta)\theta + \alpha(1-\theta) - (1-\theta) \\
0 &= \theta - \beta\theta + \alpha - \alpha\theta - 1 + \theta \\
1-\alpha &= 2\theta - \alpha\theta - \beta\theta \\
1-\alpha &= (2-\alpha-\beta)\theta \\
\theta &= \frac{1-\alpha}{2-\alpha-\beta}\frac{-1}{-1} \\
\theta &= \frac{\alpha-1}{\alpha+\beta-2}
\end{aligned}
$$

Bayes estimate (mean):

$$
\begin{aligned}
f(\theta \mid x) &= \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\theta^{\alpha-1}(1-\theta)^{\beta-1} \\
\mathbb{E}[\theta] &= \int_0^1 \theta\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\theta^{\alpha-1}(1-\theta)^{\beta-1}d\theta \\
&= \int_0^1 \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\theta^{(\alpha+1)-1}(1-\theta)^{\beta-1}d\theta \\
&\quad \Gamma(\alpha+1) = \alpha\Gamma(\alpha) \\
&\quad \Gamma(\alpha+\beta+1) = (\alpha+\beta)\Gamma(\alpha+\beta) \\
&= \int_0^1 \frac{\alpha}{\alpha+\beta}\frac{\Gamma(\alpha+1+\beta)}{\Gamma(\alpha+1)\Gamma(\beta)}\theta^{(\alpha+1)-1}(1-\theta)^{\beta-1}d\theta \\
&= \frac{\alpha}{\alpha+\beta}\int_0^1 \frac{\Gamma(\alpha+1+\beta)}{\Gamma(\alpha+1)\Gamma(\beta)}\theta^{(\alpha+1)-1}(1-\theta)^{\beta-1}d\theta \\
&= \frac{\alpha}{\alpha+\beta}
\end{aligned}
$$

So in our problem:

$$\theta_{mode} = \frac{6-1}{6+29-2}$$
$$\theta_{mean} = \frac{6}{6+29}$$

Problem 1.18

$$\begin{aligned}
P(\text{positive}) &= P(\text{positive} \cap \text{sensitive}) + P(\text{positive} \cap \text{not-sensitive}) \\
&\quad \text{law of total probability} \\
&= P(\text{positive} \mid \text{sensitive})P(\text{sensitive}) + P(\text{positive} \mid \text{not-sensitive})P(\text{not-sensitive}) \\
&\quad \text{definition of conditional probability} \\
&= P(\text{positive} \mid \text{sensitive})0.5 + P(\text{positive} \mid \text{not-sensitive})0.5 \\
&= p0.5 + 0.25 \\
0.5p &= P(\text{positive}) - 0.25 \\
p &= 2P(\text{positive}) - 0.5
\end{aligned}$$

Now we need to get the MLE of the probability of a positive response. Since this involves counting a group of binary outcomes, it is a binomial distribution.

$$\begin{aligned}
f(k) &= \binom{n}{k}\theta^k(1-\theta)^{n-k} \\
\frac{d}{d\theta}f(k) &= -(n-k)\binom{n}{k}\theta^k(1-\theta)^{n-k-1} + k\binom{n}{k}\theta^{k-1}(1-\theta)^{n-k} \\
0 &= \Bigg(-(n-k)\theta + k(1-\theta)\Bigg)\theta^{k-1}(1-\theta)^{n-k-1} \\
&\quad \text{assume } \theta, (1-\theta) \neq 0 \text{ later check value at } 0 \\
0 &= -(n-k)\theta + k(1-\theta) \\
0 &= (k-n)\theta + k - k\theta \\
0 &= k\theta - n\theta + k - k\theta \\
0 &= -n\theta + k \\
n\theta &= k \\
\theta &= k/n
\end{aligned}$$

Because of the invariance property of the MLE, we know that any function of the MLE is the MLE of that function. So in our case:

$$\begin{aligned}
P(\text{positive})_{MLE} &= X/n \\
p_{MLE} &= 2P(\text{positive})_{MLE} - 0.5 \\
p_{MLE} &= 2X/n - 0.5
\end{aligned}$$

Problem 1.19

i)

$$\begin{aligned}
\sum_{i=1}^{n}(x_i - \bar{x}_n)^2 &= \sum_{i=1}^{n} x_i^2 - 2\bar{x}_n \sum_{i=1}^{n} x_i + \sum_{i=1}^{n} \bar{x}_n^2 \\
&= \sum_{i=1}^{n} x_i^2 - 2\bar{x}_n n\bar{x}_n + n\bar{x}_n^2 \\
&\quad \text{since } n\bar{x}_n = \sum_{i=1}^{n} x_i \\
&= \sum_{i=1}^{n} x_i^2 - n\bar{x}_n^2
\end{aligned}$$

ii)

$$\begin{aligned}
\bar{x}_{n+1} &= \frac{1}{n+1}\sum_{i=1}^{n+1} x_i \\
&= \frac{1}{n+1}x_{n+1} + \frac{1}{n+1}\sum_{i=1}^{n} x_i \\
&= \frac{1}{n+1}x_{n+1} + \frac{n}{n+1}\bar{x}_n \\
&= \bar{x}_n + \frac{1}{n+1}(x_{n+1} - \bar{x}_n)
\end{aligned}$$

$$\begin{aligned}
\sum_{i=1}^{n+1}(x_i - \bar{x}_{n+1})^2 &= \sum_{i=1}^{n+1}\left(x_i - \bar{x}_n - \frac{1}{n+1}(x_{n+1} - \bar{x}_n)\right)^2 \\
&\quad \text{from previous part, plug in } \bar{x}_{n+1} \\
&= \sum_{i=1}^{n+1}(x_i - \bar{x}_n)^2 - 2\frac{1}{n+1}(x_{n+1} - \bar{x}_n)\sum_{i=1}^{n+1}(x_i - \bar{x}_n) + \sum_{i=1}^{n+1}\frac{1}{(n+1)^2}(x_{n+1} - \bar{x}_n)^2 \\
&\quad \text{note: } \sum_{i=1}^{n+1}(x_i - \bar{x}_n) = (n+1)(\bar{x}_{n+1} - \bar{x}_n) \\
&\quad \text{from previous part, } \bar{x}_{n+1} - \bar{x}_n = \frac{1}{n+1}(x_{n+1} - \bar{x}_n) \\
&= \sum_{i=1}^{n+1}(x_i - \bar{x}_n)^2 - \frac{2}{n+1}(x_{n+1} - \bar{x}_n)^2 + \frac{1}{n+1}(x_{n+1} - \bar{x}_n)^2 \\
&= \sum_{i=1}^{n}(x_i - \bar{x}_n)^2 + \frac{n}{n+1}(x_{n+1} - \bar{x}_n)^2
\end{aligned}$$