lab3

```r
library(opendatatoronto)
library(tidyverse)
```

```
-- Attaching packages --------------------------------------- tidyverse 1.3.2 --
v ggplot2 3.4.0      v purrr   0.3.5
v tibble  3.1.8      v dplyr   1.0.10
v tidyr   1.2.1      v stringr 1.5.0
v readr   2.1.3      v forcats 0.5.2
-- Conflicts ------------------------------------------ tidyverse_conflicts() --
x dplyr::filter() masks stats::filter()
x dplyr::lag()    masks stats::lag()
```

```r
library(stringr)
library(janitor)
```

```

Attaching package: 'janitor'

The following objects are masked from 'package:stats':

    chisq.test, fisher.test
```

```r
library(lubridate)
```

```
Loading required package: timechange

Attaching package: 'lubridate'
```

```
The following objects are masked from 'package:base':

    date, intersect, setdiff, union
```

```r
library(ggrepel)
library(dplyr)
library(bayestestR)
```

Question 1

$Y|\theta \sim Bin(n, \theta)$
likelihood function: $L(y, \theta, n) = \binom{n}{y}\theta^y(1-\theta)^{n-y}$
log-likelihood function: $l(y, \theta, n) = log(L) = log(\binom{n}{y}\theta^y(1-\theta)^{n-y}) = log\binom{n}{y} + log(\theta^y) + log((1-\theta)^{n-y}) = log\binom{n}{y} + ylog(\theta) + (n-y)log(1-\theta)$
take derivative respect to θ and set to 0:
$\frac{dl}{d\theta} = \frac{y}{\theta} - \frac{n-y}{1-\theta} = 0$
$\hat{\theta} = \frac{y}{n} = \frac{118}{129} = 0.91$
Fisher Information:
$I(\hat{\theta}) = -E[l''(\hat{\theta})] = \frac{118}{0.91^2} + \frac{11}{(1-0.91)^2} = 1500.52$
95% confidence interval:
$(\hat{\theta} - 1.96\frac{1}{\sqrt{I(\hat{\theta})}}, \hat{\theta} + 1.96\frac{1}{\sqrt{I(\hat{\theta})}})$
$(0.91 - 0.05, 0.91 + 0.05)$
$(0.86, 0.96)$

Question 2

θ ~$beta(1, 1)$
$p(\theta) = \frac{\gamma(1+1)}{\gamma(1)\gamma(1)}\theta^{1-1}(1-\theta)^{1-1} = 1$
$p(\theta|y) = \frac{p(y|\theta)p(\theta)}{p(y)} = \frac{p(y|\theta)p(\theta)}{\int p(y|\theta')d\theta'}$
$= \frac{\binom{n}{y}\theta^y(1-\theta)^{n-y}}{\int_0 1\binom{n}{y}\theta'^y(1-\theta')^{n-y}d\theta'}$
$= \frac{1}{Z}\theta^y(1-\theta)^{n-y}$, where $Z = \frac{\gamma(y+1)\gamma(n-y+1)}{\gamma(n+2)}$
the posterior distribution is :
$\theta|y$~$Beta(y+1, n-y+1)$
The posterior mean:
$E(\theta|y) = \frac{y+1}{y+1+n-y+1} = \frac{y+1}{n+2} = \frac{118+1}{129+2} = 0.91$
95% Credible interval:

```
posterior <- distribution_beta(1000, 119, 12)
ci_eti <- ci(posterior, method = "ETI")
ci_eti
```

```
95% ETI: [0.85, 0.95]
```

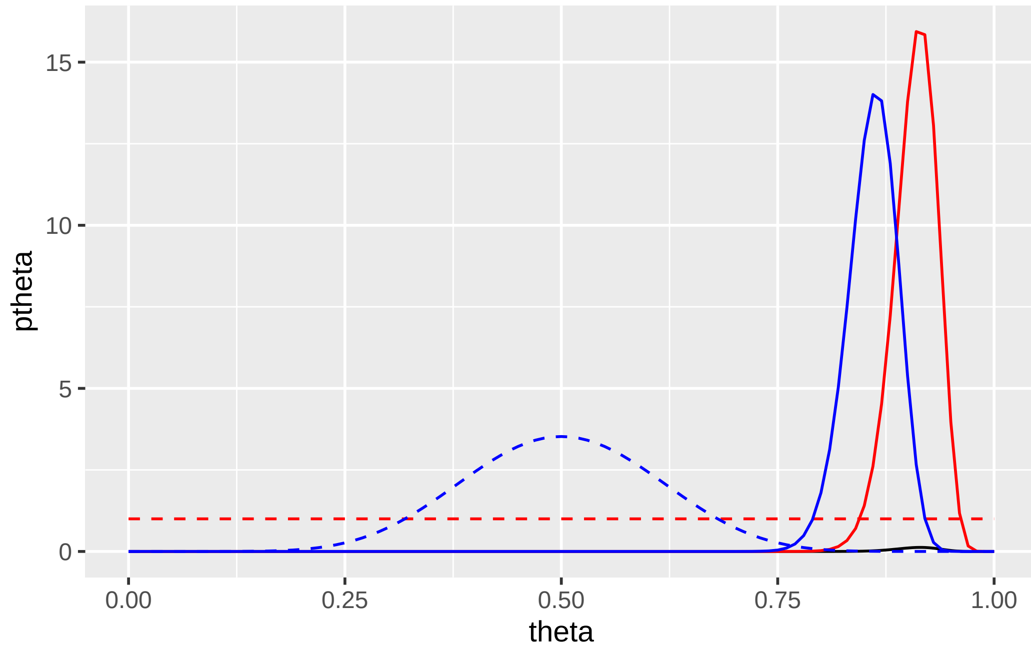
Question 3

$\theta \sim beta(10, 10)$
interpretation: my subjective beliefs about the parameter θ is distributed $beta(10, 10)$.
we assuming we know more amount of information as the prior used in Question 2, since the the prior in Question 2 is uniform, and contain no information.

Question 4

black: The likelihood
red,dashed: The prior distribution of prior beta(1,1)
red,solid: The posteriors distribution of prior beta(1,1)
blue, dashed: The prior distribution of prior beta(10,10)
blue, dahsed: The posteriors distribution of prior beta(10,10)

```
p = seq(from = 0, to =1, by = 0.001)
data = data.frame(theta=p,ptheta=dbinom(118, size = 129, p))
ggplot(data, aes(x=theta,y=ptheta))+geom_line()+stat_function(fun=dbeta, args=list(shape1=
```



Question 5

A noninformative prior: has distribution which is flat over the entire real number line and contain no information, for example, uniform distribution.

A subjective/informative prior based on your best knowledge: has a distribution which containing specific, unambiguous information about the variable, for example, based on the distribution of the average improvement in success probability.