ESC 2022 Week2

One parameter model

학술부 김채현, 임세희

July 21, 2022

Binomial model

1. Binomial data

Ex) Happiness data

$$n = 129$$

$$Y_i = \begin{cases} 1 & \text{if happy} \\ 0 & \text{otherwise} \end{cases}$$

Conditional on θ, the Y_i's are i.i.d. binary random variables with expectation θ

$$p(y_1, \ldots, y_{129}|\theta) = \theta^{\sum_{i=1}^{129} y_i}(1-\theta)^{129-\sum_{i=1}^{129} y_i}$$

만약 $Y_i = 1$이 118명, $Y_i = 0$이 11명이면?

$$p(y_1, \ldots, y_{129}|\theta) = \theta^{118}(1-\theta)^{11}$$

Binomial model

2. A uniform prior distribution

θ is unknown number between 0 and 1

$$p(\theta) = 1 \text{ for all } \theta \in [0,1]$$

- Bayes' rule 활용하기

$$\begin{aligned} p(\theta|y_1, \dots y_{129}) &= \frac{p(y_1, \dots, y_{129}|\theta)p(\theta)}{p(y_1, \dots, y_{129})} \\ &= p(y_1, \dots, y_{129}|\theta) \times \frac{1}{p(y_1, \dots, y_{129})} \\ &\propto p(y_1, \dots, y_{129}|\theta) \end{aligned}$$

Binomial model

$$
\begin{aligned}
p(\theta|y_1, ...y_{129}) &= \frac{p(y_1, ..., y_{129}|\theta)p(\theta)}{p(y_1, ..., y_{129})} \\
&= p(y_1, ..., y_{129}|\theta) \times \frac{1}{p(y_1, ..., y_{129})} \\
&\propto p(y_1, ..., y_{129}|\theta)
\end{aligned}
$$

$p(\theta|y_1, ...y_{129})$ and $p(y_1, ..., y_{129}|\theta)$ are proportional to each other as functions of θ.

- posterior distribution is equal to $p(y_1, ..., y_{129}|\theta)$ divided by something that does not depend on θ.
- These two functions of θ have the same shape, but not necessarily the same scale.

Binomial model

More precise than this
- We need to know the scale of $p(\theta|y_1, ...y_{129})$ as well as the shape!
- We can calculate the scale or normalizing constant $\frac{1}{p(y_1,...,y_{129})}$ using

$$\int \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}\theta^{a-1}(1-\theta)^{b-1}d\theta = 1 \quad \Rightarrow \quad \int \theta^{a-1}(1-\theta)^{b-1}d\theta = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}$$

$$\Gamma(a) = (a-1)!$$

- a) $\int_0^1 p(\theta|y_1, ...y_{129})d\theta = 1$
- b) $p(\theta|y_1, ...y_{129}) = \theta^{118}(1-\theta)^{11}/p(y_1, ..., y_{129})$

$$\begin{cases} p(y_1, \cdots, y_{129}|\theta) = \theta^{118}(1-\theta)^{11} \\ p(\theta) = 1 \end{cases}$$

$1 = \int_0^1 p(\theta|y_1, \cdots, y_{129})d\theta$ using (a)

$1 = \int_0^1 \frac{\theta^{118}(1-\theta)^{11}}{p(y_1, \cdots, y_{129})} d\theta$ using (b)

$$1 = \frac{1}{p(y_1, \cdots, y_{129})} \int_0^1 \theta^{118}(1-\theta)^{11} d\theta$$

$$1 = \frac{1}{p(y_1, \cdots, y_{129})} \frac{\Gamma(119)\Gamma(12)}{\Gamma(131)} \int_0^1 \frac{\Gamma(131)}{\Gamma(119)\Gamma(12)} \theta^{118}(1-\theta)^{11} d\theta$$

$$p(y_1, \cdots, y_{129}) = \frac{\Gamma(119)\Gamma(12)}{\Gamma(131)}$$

Binomial model

$$p(y_1, ..., y_{129}) = \frac{\Gamma(119)\Gamma(12)}{\Gamma(131)}$$

- This result holds for any sequence $y_1, ..., y_{129}$ that contains 118 ones and 11 zeros.

$$p(\theta|y_1, ...y_{129}) = \frac{\Gamma(131)}{\Gamma(119)\Gamma(12)}\theta^{119-1}(1-\theta)^{12-1}$$

$$\sim beta(119, 12)\ (a = 119, b = 12)$$

Inference for exchangeable binary data

- Posterior inference under a uniform prior

If $Y_1, ... Y_n$ are i.i.d. binary(θ),

$$p(\theta|y_1, ...y_n) = \theta^{\sum y_i}(1-\theta)^{n-\sum y_i} \times p(\theta)/p(y_1, ...y_n)$$

$$\frac{p(\theta_a|y_1, \cdots, y_n)}{p(\theta_b|y_1, \cdots, y_n)} = \frac{\theta_a^{\sum y_i}(1-\theta_a)^{n-\sum y_i} \times p(\theta_a)/p(y_1, \cdots, y_n)}{\theta_b^{\sum y_i}(1-\theta_b)^{n-\sum y_i} \times p(\theta_b)/p(y_1, \cdots, y_n)}$$

$$= \left(\frac{\theta_a}{\theta_b}\right)^{\sum y_i} \left(\frac{1-\theta_a}{1-\theta_b}\right)^{n-\sum y_i} \frac{p(\theta_a)}{p(\theta_b)}$$

Inference for exchangeable binary data

This shows that the probability density at θ_a relative to that at θ_b depends on $y_1, ..., y_n$ only through $\sum_{i=1}^n y_i$.

$$Pr(\theta \in A | Y_1 = y_1, ..., Y_n = y_n) = Pr(\theta \in A | \sum_{i=1}^n Y_i = \sum_{i=1}^n y_i)$$

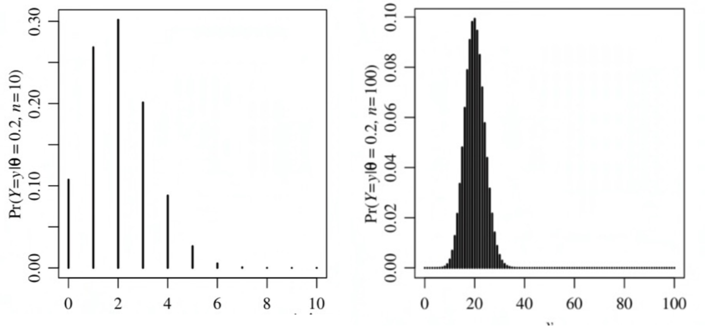
- $\sum_{i=1}^n Y_i$ contains all the information about θ from the data
- $\sum_{i=1}^n Y_i =$ "sufficient statistic" for θ and $p(y_1, ...y_n|\theta)$.
- It is sufficient to know $\sum_{i=1}^n Y_i$ in order to make inference about θ.
- The "sufficient statistic Y" $= \sum_{i=1}^n Y_i$ has a binomial distribution with parameters (n, θ) where $Y_1, ..., Y_n|\theta$ are i.i.d binary(θ) random variables.

Inference for exchangeable binary data

- The binomial distribution

A random variable $Y \in \{0, 1, ..., n\}$ has a binomial(n, θ) distribution if

$$Pr(Y = y|\theta) = \binom{n}{r} \theta^y (1-\theta)^{n-y}, \quad y \in \{0, 1, ..., n\}$$



Inference for exchangeable binary data

- Posterior inference under a uniform prior distribution

observed $Y = y$

$$p(\theta|y) = \frac{p(y|\theta)p(\theta)}{p(y)}$$

$$= \frac{\binom{n}{y}\theta^y(1-\theta)^{n-y}p(\theta)}{p(y)}$$

$$= c(y)\theta^y(1-\theta)^{n-y}p(\theta)$$

$\rightarrow$ $c(y)$ is a function of y and not of θ.

$\rightarrow$ $c(y)$를 구해야 한다!

Inference for exchangeable binary data

$c(y)$를 구하기 위해서는.. $\rightarrow$ $\mathrm{p}(\theta) = 1$

$$1 = \int_0^1 c(y)\theta^y(1-\theta)^{n-y}d\theta \quad \rightarrow \quad 1 = c(y)\int_0^1 \theta^y(1-\theta)^{n-y}d\theta$$

$$\rightarrow \quad 1 = c(y)\frac{\Gamma(y+1)\Gamma(n-y+1)}{\Gamma(n+2)} \quad \rightarrow \quad c(y) = \frac{\Gamma(n+2)}{\Gamma(y+1)\Gamma(n-y+1)}$$

- 베이즈 정리 활용

$$\begin{aligned} p(\theta|y) &= \frac{\Gamma(n+2)}{\Gamma(y+1)\Gamma(n-y+1)}\theta^y(1-\theta)^{n-y} \\ &= \frac{\Gamma(n+2)}{\Gamma(y+1)\Gamma(n-y+1)}\theta^{(y+1)-1}(1-\theta)^{(n-y+1)-1} \\ &= beta(y+1, n-y+1) \end{aligned}$$

Inference for exchangeable binary data

- Happiness example

We observed that $Y \equiv \sum Y_i = 118$ (sufficient statistic)

$$n = 129, Y \equiv \sum Y_i = 118 \quad \longrightarrow \quad \theta | \{Y = 118\} \sim beta(119, 12)$$
$$p(\theta|y) = p(\theta|y_1, ..., y_n) = beta(119, 12)$$

The information contained in $\{Y_1 = y_1, ..., Y_n = y_n\}$ is the same as the information contained in $\{Y = y\}$ where $Y = \sum Y_i$ and $y = \sum y_i$.

Inference for exchangeable binary data

- Posterior distributions under beta prior distributions

The uniform prior distribution $= beta(1,1)$

$$p(\theta) = \frac{\Gamma(2)}{\Gamma(1)\Gamma(1)}\theta^{1-1}(1-\theta)^{1-1} = 1$$

$$\begin{aligned} &\text{if } \theta \sim beta(1,1), Y \sim binomial(n,\theta) \\ \rightarrow\ &\{\theta | Y = y\} \sim beta(1+y, 1+n-y) \end{aligned}$$

Inference for exchangeable binary data

Suppose $\theta \sim beta(a, b)$ and $Y|\theta \sim binomial(n, \theta)$

$$
\begin{aligned}
p(\theta|y) &= \frac{p(\theta)p(y|\theta)}{p(y)} \\
&= \frac{1}{p(y)} \times \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}\theta^{a-1}(1-\theta)^{b-1} \times \binom{n}{y}\theta^{y}(1-\theta)^{n-y} \\
&= \underline{c(n, y, a, b)} \times \theta^{a+y-1}(1-\theta)^{b+n-y-1} \\
&= beta(a+y, b+n-y)
\end{aligned}
$$

$$\frac{\Gamma(a+b+n)}{\Gamma(a+y)\Gamma(b+n-y)}$$

1. $p(\theta|y) \propto \theta^{a+y-1}(1-\theta)^{b+n-y-1}$

   $\rightarrow$ same shape

2. $p(\theta|y)$ and beta density must both integrate to 1.

   $\rightarrow$ same scale

$1 + 2$: $p(\theta|y)$ and beta density are same function.

Inference for exchangeable binary data

- Conjugacy: posterior가 prior와 같은 분포 계열에 속함.

When we use the Beta distribution as a prior, a posterior of binomial likelihood will also follow the beta distribution.

- Combining information

$\theta|\{Y=y\} \sim beta(a+y, b+n-y)$

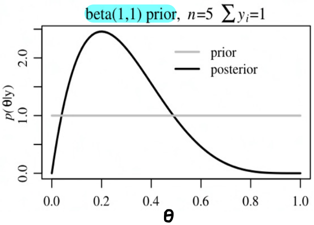
$$\begin{aligned} E[\theta|y] &= \frac{a+y}{a+b+n} \\ &= \frac{a+b}{a+b+n}\frac{a}{a+b} + \frac{n}{a+b+n}\frac{y}{n} \\ &= \frac{a+b}{a+b+n} \times \text{prior expectation} + \frac{n}{a+b+n} \times \text{data average.} \end{aligned}$$

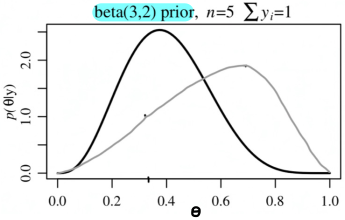
- The posterior expectation is weighted average of prior expectation and the sample mean.

Inference for exchangeable binary data

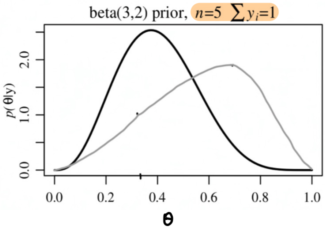
$$E[\theta|y] = \frac{a+b}{a+b+n}\frac{a}{a+b} + \frac{n}{a+b+n}\frac{y}{n} \Rightarrow \textit{If } a+b << n, \ \frac{a+b}{a+b+n} \approx 0, \ E[\theta|y] \approx \frac{y}{n}$$

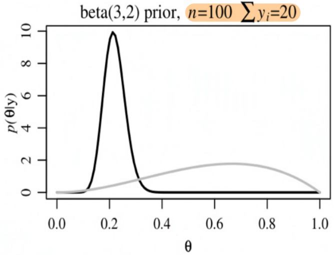
prior





sample size





Inference for exchangeable binary data

- Prediction
  - Let $y_1, .., y_n$ be the outcomes from a sample of n binary random variables.
  - Let $\tilde{Y} \in \{0, 1\}$ be an additional outcome from the same population that has yet to be observed.

Inference for exchangeable binary data

- The predictive distribution of $\tilde{Y}$

$\tilde{Y}|\{y_1, ..., y_n\}$: conditionally i.i.d. binary variables

$$Pr(\tilde{Y}=1 | y_1, \dots, y_n) = \int Pr(\tilde{Y}=1, \theta | y_1, \dots, y_n) \, d\theta$$

$$= \int Pr(\tilde{Y}=1 | \theta, y_1, \dots, y_n) P(\theta | y_1, \cdots, y_n) \, d\theta$$

$$= \int \theta P(\theta | y_1, \dots, y_n) \, d\theta = E[\underline{\theta | y_1, \dots, y_n}]$$

$\hookrightarrow$ beta$(a + \Sigma y_i , b + n - \Sigma y_i)$

$$= \frac{a + \sum_{i=1}^{n} y_i}{a + b + n}$$

$$Pr(\tilde{Y}=0 | y_1, \dots, y_n) = 1 - E[\theta | y_1, \dots, y_n] = \frac{b + \sum_{i=1}^{n}(1 - y_i)}{a + b + c}$$

- The predictive distribution depends on our observed data.

$=$ $\tilde{Y}$ is not independent of $Y_1, ..., Y_n$.

Poisson Model

Some measurements have values that are **whole numbers**.
count data일 때 Binomial model 대신 Poisson model 사용

$$p(y_i|\theta) = \frac{\theta^y e^{-\theta}}{y!} \quad \text{for } y \in \{0, 1, 2, \cdots\}$$

- $E[Y|\theta] = \theta$
- $Var[Y|\theta] = \theta$

Poisson Model

If we model $Y_1, \cdots, Y_n$ as i.i.d. Poisson with mean θ, the **joint pdf** is

$$
\begin{aligned}
p(y_1, y_2, \cdots, y_n|\theta) &= \prod_{i=1}^{n} p(y_i|\theta) \\
&= \prod_{i=1}^{n} \frac{\theta^{y_i} e^{-\theta}}{y_i!} \\
&= \frac{1}{\prod_{i=1}^{n} y_i!} \theta^{\sum y_i} e^{-n\theta}
\end{aligned}
$$

Poisson Model

1) **Prior distribution 설정**: gamma(a, b) distribution

$$p(\theta) = \frac{b^a}{\Gamma(a)}\theta^{a-1}e^{-b\theta} \quad \text{for } \theta, a, b > 0$$

2) **Posterior distribution 구하기**

$$\begin{aligned} p(\theta|y_1, \cdots, y_n) &= \frac{p(\theta)p(y_1, \cdots, y_n|\theta)}{p(y_1, \cdots, y_n)} \\ &= \frac{\frac{b^a}{\Gamma(a)}\theta^{a-1}e{-}b\theta \times \frac{1}{\prod_{i=1}^n y_i!}\theta^{\sum y_i}e^{-n\theta}}{p(y_1, \cdots, y_n)} \\ &= \text{상수} \times \theta^{a+\sum y_i - 1}e^{-(n+b)\theta} \end{aligned}$$

Posterior distribution

$$p(\theta|y_1, \cdots, y_n) = \text{상수} \times \theta^{a+\sum y_i - 1} e^{-(n+b)\theta}$$

: gamma($a + \sum y_i,\ n + b$) distribution을 따름

Conjugacy: 사전분포와 사후분포가 같음 $\rightarrow$ **conjugacy**

$$E[\theta|y_1, \cdots, y_n] = \frac{a+\sum y_i}{b+n} = \frac{b}{b+n} \times \frac{a}{b} + \frac{n}{b+n} \times \frac{\sum y_i}{n}$$

b: the number of prior observations
a: the sum of counts from b prior observations

Prediction

Prediction about additional data can be obtained with the posterior predictive distribution.
(additional data $\tilde{y}$)

$$
\begin{aligned}
p(\tilde{y}|y_1, \cdots, y_n) &= \int_0^\infty p(\tilde{y}|\theta, y_1, \cdots, y_n)p(\theta|y_1, \cdots, y_n)d\theta \\
&= \int_0^\infty p(\tilde{y}|\theta)p(\theta|y_1, \cdots, y_n)d\theta \\
&= \int_0^\infty \{\frac{1}{\tilde{y}!}\theta^{\tilde{y}}e^{-\theta}\}\{\frac{(b+n)^{a+\sum y_i}}{\Gamma(a+\sum y_i)}\theta^{a+\sum y_i-1}e^{-(b+n)\theta}\}d\theta \\
&= \frac{(b+n)^{a+\sum y_i}}{\Gamma(\tilde{y}+1)\Gamma(a+\sum y_i)}\int_0^\infty \theta^{a+\sum y_i+\tilde{y}-1}e^{-(b+n+1)\theta}d\theta
\end{aligned}
$$

Prediction

$1 = \int_0^\infty \frac{b^a}{\Gamma(a)} \theta^{a-1} e^{-b\theta} d\theta$ for any values a, b > 0

$\int_0^\infty \theta^{a-1} e^{-b\theta} d\theta = \frac{\Gamma(a)}{b^a}$ for any values a, b > 0

substitute in $a + \sum y_i + \tilde{y}$ instead of a and $b + n + 1$ instead of b

$\int_0^\infty \theta^{a+\sum y_i + \tilde{y} - 1} e^{-(b+n+1)\theta} d\theta = \frac{\Gamma(a+\sum y_i + \tilde{y})}{(b+n+1)^{a+\sum y_i + \tilde{y}}}$

$$
\begin{aligned}
p(\tilde{y}|y_1, \cdots, y_n) &= \frac{(b+n)^{a+\sum y_i}}{\Gamma(\tilde{y}+1)\Gamma(a+\sum y_i)} \int_0^\infty \theta^{a+\sum y_i + \tilde{y} - 1} e^{-(b+n+1)\theta} d\theta \\
&= \frac{(b+n)^{a+\sum y_i}}{\Gamma(\tilde{y}+1)\Gamma(a+\sum y_i)} \frac{\Gamma(a+\sum y_i + \tilde{y})}{(b+n+1)^{a+\sum y_i + \tilde{y}}} \\
&= \frac{\Gamma(a+\sum y_i + \tilde{y})}{\Gamma(\tilde{y}+1)\Gamma(a+\sum y_i)} \left(\frac{b+n}{b+n+1}\right)^{a+\sum y_i} \left(\frac{1}{b+n+1}\right)^{\tilde{y}}
\end{aligned}
$$

negative binomial distribution with parameters $(a + \sum y_i, b + n)$

Example

$Y_{1,1}, \cdots, Y_{n_1,1}$: the numbers of children for the n_1 women without college degrees
$Y_{1,2}, \cdots, Y_{n_2,2}$: the numbers of children for the n_2 women with college degrees

Sampling model

$$Y_{1,1}, \cdots, Y_{n_1,1}|\theta_1 \sim i.i.d. Poisson(\theta_1)$$

$$Y_{1,2}, \cdots, Y_{n_2,2}|\theta_2 \sim i.i.d. Poisson(\theta_2)$$

Example

$n_1 = 111,\ \sum_{i=1}^{n_1} Y_{i,1} = 217,\ \bar{Y}_1 = 1.95$
$n_2 = 44,\ \sum_{i=1}^{n_2} Y_{i,2} = 66,\ \bar{Y}_2 = 1.50$

In the case where $\{\theta_1, \theta_2\} \sim i.i.d.\ gamma(a = 2, b = 1)$,

Posterior distribution

$$\theta_1 | \{n_1 = 111, \sum Y_{i,1} = 217\} \sim gamma(219, 112)$$

$$\theta_2 | \{n_2 = 44, \sum Y_{i,2} = 66\} \sim gamma(68, 45)$$

Exponential Families

The Binomial and Poisson models are both instances of one-parameter **exponential family models**.

Density가 $p(y|\theta) = h(y)c(\phi)e^{\phi t(y)}$ 꼴로 표현될 수 있어야 함

ϕ: unknown parameter
$t(y)$: sufficient statistic

Binomial model

Density가 $p(y|\theta) = h(y)c(\phi)e^{\phi t(y)}$ 꼴로 표현될 수 있어야 함

ϕ: unknown parameter
$t(y)$: sufficient statistic

$$
\begin{aligned}
p(y|\theta) &= \theta^y(1-\theta)^{1-y} \\
&= (\frac{\theta}{1-\theta})^y(1-\theta) \\
&= e^{\phi y}(1+e^{\phi})^{-1}
\end{aligned}
$$

$\phi = log[\theta/(1-\theta)]$

Confidence Interval

Bayesian coverage: information about the location of the true value of θ after you have observed $Y = y$.

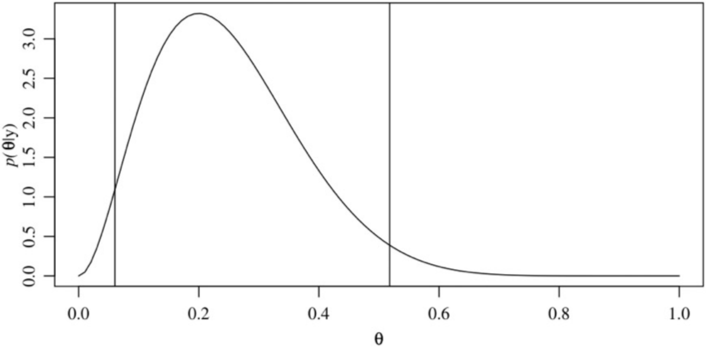
Frequentist coverage: information about the location of the true value of θ before the data are observed.

Confidence Interval

예) Beta(3, 9) posterior distribution

Quantile-based interval

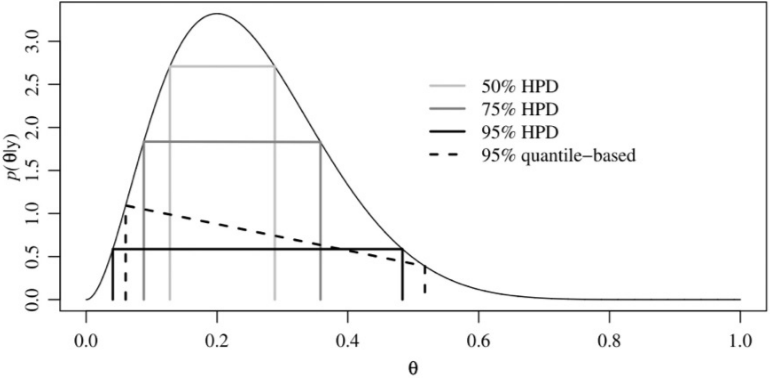
There are θ-values **outside** the quantile-based interval that have higher probability than some points **inside** the interval.



예) Beta(3, 9) posterior distribution

HPD region

All points in an HPD region have a higher posterior density than points **outside** the region.



Lab

1. Whenever you survey people about sensitive issues, you have to deal with social desirability bias, which is the tendency of people to adjust their answers to show themselves in the most positive light. One way to improve the accuracy of the results is randomized response.

As an example, suppose you want to know how many people cheat on their taxes. If you ask them directly, it is likely that some of the cheaters will lie. You can get a more accurate estimate if you ask them indirectly, like this: Ask each person to flip a coin and, without revealing the outcome,

- If they get heads, they report YES.
- If they get tails, they honestly answer the question "Do you cheat on your taxes?"

다음 슬라이드에 계속

Lab

If someone says YES, we don't know whether they actually cheat on their taxes; they might have flipped heads. Knowing this, people might be more willing to answer honestly.

Suppose you survey 100 people this way and get 80 YESes and 20 NOs. Based on this data, what is the posterior distribution for the fraction of people who cheat on their taxes?

문제 및 코드는 ESC github week2 참고하기!!

https://github.com/YonseiESC/ESC-22SUMMER/blob/main/week2/WEEK2_lab.ipynb

Homework

2. Assume home runs per game at Citizens Bank Park follow a Poisson distribution with parameter θ. Assume for θ a Gamma prior distribution with shape parameter $\alpha = 4$ and rate parameter $\beta = 2$.

1) Write an expression for the prior density $\phi(\theta)$. Find the prior mean and prior SD.

2) Suppose a single game with 1 home run is observed. Write the likelihood function.

3) Write an expression for the posterior distribution of θ given a single game with 1 home run. Find the posterior mean and posterior SD.