STATS 551 HW1

Minxuan Chen, Bohan Zhang, Jiaqi Zhu

2023/9/19

Problem 1

Since a population is partitioned into disjoint groups, we should have

$$P(H_1) + P(H_2) + P(H_3) = 1$$

According to Bayes theorem,

$$P(H_i|E) \propto P(E|H_i)P(H_i)$$

For $P(E|H_i)$, we have

$$P(E|H_1) = 0.1, \quad P(E|H_2) = 0.3, \quad P(E|H_3) = 0.5$$

Therefore

(a) We can set marginal probabilities to be

$$P(H_1) = 0.1, \quad P(H_2) = 0.8, \quad P(H_3) = 0.1$$

Then we have

$$0.1P(H_1) < 0.5P(H_3) < 0.3P(H_2)$$
$$\Rightarrow P(E \mid H_1)P(H_1) < P(E \mid H_3)P(H_3) < P(E \mid H_2)P(H_2)$$

Equivalently,
$$P(H_1 \mid E) < P(H_3 \mid E) < P(H_2 \mid E)$$

(b) We can set marginal probabilities to be

$$P(H_1) = 0.4, \quad P(H_2) = 0.1, \quad P(H_3) = 0.5$$

Then we have

$$0.3P(H_2) < 0.1P(H_1) < 0.5P(H_3)$$
$$\Rightarrow P(E \mid H_2)P(H_2) < P(E \mid H_1)P(H_1) < P(E \mid H_3)P(H_3)$$

Equivalently,
$$P(H_2 \mid E) < P(H_1 \mid E) < P(H_3 \mid E)$$

(c) We can set marginal probabilities to be

$$P(H_1) = 0.8, \quad P(H_2) = 0.1, \quad P(H_3) = 0.1$$

Then we have

$$0.3P(H_2) < 0.5P(H_3) < 0.1P(H_1)$$
$$\Rightarrow P(E \mid H_2)P(H_2) < P(E \mid H_3)P(H_3) < P(E \mid H_1)P(H_1)$$

Equivalently,

$$P(H_2 \mid E) < P(H_3 \mid E) < P(H_1 \mid E)$$

(d) We can set marginal probabilities to be

$$P(H_1) = 0.7, \quad P(H_2) = 0.2, \quad P(H_3) = 0.1$$

Then we have

$$0.5P(H_3) < 0.3P(H_2) < 0.1P(H_1)$$
$$\Rightarrow P(E \mid H_3)P(H_3) < P(E \mid H_2)P(H_2) < P(E \mid H_1)P(H_1)$$

Equivalently,

$$P(H_3 \mid E) < P(H_2 \mid E) < P(H_1 \mid E)$$

(e) We can set marginal probabilities to be

$$P(H_1) = 0.6, \quad P(H_2) = 0.3, \quad P(H_3) = 0.1$$

Then we have

$$0.5P(H_3) < 0.1P(H_1) < 0.3P(H_2)$$
$$\Rightarrow P(E \mid H_3)P(H_3) < P(E \mid H_1)P(H_1) < P(E \mid H_2)P(H_2)$$

Equivalently,

$$P(H_3 \mid E) < P(H_1 \mid E) < P(H_2 \mid E)$$

Check

```r
pE_H <- c(0.1,0.3,0.5)
pH <- matrix(c(0.3,0.4,0.8,0.7,0.6,
               0.5,0.1,0.1,0.2,0.3,
               0.2,0.5,0.1,0.1,0.1),ncol = 3)

pH_E <- t(pE_H*t(pH))/c(pH%*%pE_H)

#order
t(apply(pH_E, 1, order))
```

```
##      [,1] [,2] [,3]
## [1,]    1    3    2
## [2,]    2    1    3
## [3,]    2    3    1
## [4,]    3    2    1
## [5,]    3    1    2
```

Problem 2

Let L_1, L_2, L_3 denote the car is located behind Door 1,2,3, resp. Let H_1, H_2, H_3 denote the host open Door 1,2,3, resp.

Before the game,

$$P(L_1) = P(L_2) = P(L_3) = \frac{1}{3}$$

The probability of winning (if switching) is

$$P(\text{win}|H_3) = P(L_2|H_3)$$
$$= \frac{P(H_3|L_2)P(L_2)}{P(H_3|L_1)P(L_1) + P(H_3|L_2)P(L_2) + P(H_3|L_3)P(L_3)}$$
$$= \frac{P(H_3|L_2)}{P(H_3|L_1) + P(H_3|L_2) + P(H_3|L_3)}$$

For these conditional probabilities, $H_3|L_3$ is impossible, so $P(H_3|L_3) = 0$. For $H_3|L_2$, the host has not choice rather opening Door 3, so $P(H_3|L_2) = 1$.
For $H_3|L_1$, since the car is in neither Door 2 nor 3, the host can randomly choose from the two doors. Note that the prob host open Door 2 or 3 depends on his location.

(1) Suppose the host is closer to Door 2, then $P(H_3|L_1) = 1 - \gamma \in (0, 1/2]$. So

$$P(\text{win}|H_3) = \frac{1}{1 - \gamma + 1} = \frac{1}{2 - \gamma} \in [\frac{2}{3}, 1)$$

Since $P(\text{win}|H_3) \geq \frac{2}{3} > \frac{1}{2}$, the probability of winning is higher by switching.

(2) Suppose the host is closer to Door 3, then $P(H_3|L_1) = \gamma \in [1/2, 1)$. So

$$P(\text{win}|H_3) = \frac{1}{\gamma + 1} \in (\frac{1}{2}, \frac{2}{3}]$$

Since $P(\text{win}|H_3) > \frac{1}{2}$, the probability of winning is higher by switching.

By (1),(2), no matter where the host is, under $\gamma \in [1/2, 1)$, the player will have a better chance of winning by always switching the door

Problem 3

First, we use a Bayesian Network to represent their relationships. Based on the significance of age (X) and country (Y), we set them as Markovian parents of most variables.

DAG:

```
Sys.setenv(LANGUAGE = "en")
library(dagitty)
library(ggdag)
```

```
##
## Attaching package: 'ggdag'

## The following object is masked from 'package:stats':
##
##     filter
```

```r
dag <- dagitty::dagitty("
      dag{
      Y->X
      Y->Z X->Z
      Y->U X->U
      X->V Y->V
      X->W Y->W
      }
       ")
tidy_dag <- tidy_dagitty(dag)
ggdag(tidy_dag)+theme_dag()
```
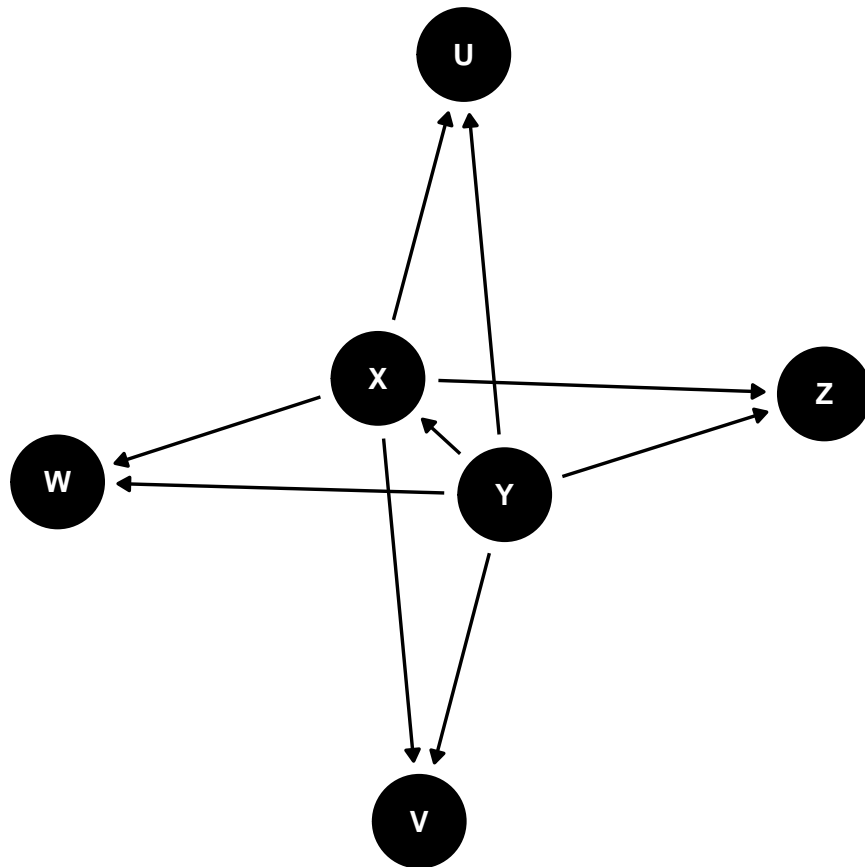
Under this assumption, the pdf of the joint distribution is:

$$f(x, y, u, v, w) = f_Y(y) f_{X|Y}(x|y) f_{Z|X,Y}(z|x,y) f_{U|X,Y}(u|x,y) f_{V|X,Y}(v|x,y) f_{W|X,Y}(w|x,y)$$

Then, we will specify the domain, distribution, and parameters of each random variable.

1. Y represents the country the person is from.

Let Y be a discrete uniform distribution over a finite set of countries $\{C_1, C_2, ..., C_N\}$ where N is the total number of countries and:

$$P(Y = C_i) = \frac{1}{N}$$

N is around 200 and Y is a real number in $(0, 1)$.

2. X represents the person's age and takes value in the set of natural numbers.

$X|Y$ can follow a Beta binomial distribution. Let n_Y be the maximum age of a person which is likely decided by the country. It's usually not over 120. Then

$$X|Y \sim BetaBin(n_Y, \alpha_Y, \beta_Y)$$

and

$$P(X = x|Y) = \binom{n}{x} \frac{B(x + \alpha_Y, n_Y - x + \beta_Y)}{B(\alpha_Y, \beta_Y)}$$

where $\alpha_Y > 0, \beta_Y > 0$ decide the shape of the distribution of ages of a specific country Y. $X|Y$ is a natural number in $[0, n_Y]$.

3. Z represents whether the person attended a picnic the first week of July, 2023.

Given it is a binary variable that can take 0 or 1, so $Z|X, Y$ can follow a Bernoulli distribution where

$$P(Z = z|X, Y) = p_{XY}^z (1 - p_{XY})^{1-z} \qquad z \in \{0, 1\}$$

where parameter $p_{XY} \in [0, 1]$ is the probability of attending a picnic decided by the person age and country custom.

4. U represents the amount of coffee the person consumed in 2023.

We assume $U|X, Y \sim Gamma(\alpha_{XY}, \beta_{XY}) \in [0, +\infty)$. The PDF is given by:

$$f(u; \alpha_{XY}, \beta_{XY}) = \frac{\beta_{XY}^{\alpha_{XY}} u^{\alpha_{XY}-1} e^{-\beta_{XY} u}}{\Gamma(\alpha_{XY})}$$

With a shape parameter $\alpha_{XY} > 0$ and a scale parameter $\beta_{XY} > 0$ decided by the person age and country custom.

5. V represents the amount of screen time the person spends in each of the 24 hours on September 1, 2023.

In each of the 24 hours, we assume the screen time (in hour unit) $V_i|X \wedge Y$ follow a beta distribution $Beta(\alpha_{XY}, \beta_{XY})$ ranging from (0,1). For simplicity, we assume the screen time of each hour is independent and follows the beta distribution with the same parameters $\alpha_{XY} > 0, \beta_{XY} > 0$ once X and Y are given. So $\mathbf{V}|X,Y$ is a random vector with the distribution

$$\mathbf{V}|X,Y = (V_1|X,Y,V_2|X,Y,\ldots,V_{24}|X,Y)$$
$$\text{where } V_i|X,Y \sim Beta(\alpha_{XY},\beta_{XY}), \quad i.i.d.$$

The pdf of the joint distribution of $\mathbf{V}|X \wedge Y$ is

$$f(\mathbf{V}|X,Y) = \prod_{i=1}^{24} \frac{v_i^{\alpha_{XY}-1}(1-v_i)^{\beta_{XY}-1}}{B(\alpha_{XY},\beta_{XY})}$$

6. W represents a passport photo of the person.

We assume a photo should be a 2-d image with $m_Y \times n_Y$ pixels, which are decided by the country. Then we assume they both follows Gamma distribution ranging from $[0,+\infty)$:

$$m_Y|Y \sim Gamma(\alpha_{mY},\beta_{mY})$$
$$n_Y|Y \sim Gamma(\alpha_{nY},\beta_{nY})$$

where $\alpha_{mY},\beta_{mY},\alpha_{nY},\beta_{nY}$ are all greater than 0.

Then for each pixel w_{ij}, it has three random variables R,G,B ranging over integer in $[0,255]$. For simplicity, we assume they all follow the same Beta binomial distribution and variables across pixels are also independent once X and Y are given:

$$w_{ijR}|X,Y \sim BetaBin(255,\alpha_{XY},\beta_{XY})$$
$$w_{ijG}|X,Y \sim BetaBin(255,\alpha_{XY},\beta_{XY})$$
$$w_{ijR}|X,Y \sim BetaBin(255,\alpha_{XY},\beta_{XY})$$

where $\alpha_{XY} > 0, \beta_{XY} > 0$.

Note: Each term, such as α_{XY}, n_Y, and so forth, as mentioned above, represents that the value of these parameters are related to the random variables indicated in their subscripts (i.e. some function of these random variables). In this convention, we aim to express the properties of conditional probability explicitly and comprehensibly.

Problem 4

$$\theta \sim \text{Beta}(a,b) : p(\theta) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}\theta^{a-1}(1-\theta)^{b-1}$$

Therefore

$$\text{mode}[\theta] = \max_\theta \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}\theta^{a-1}(1-\theta)^{b-1}$$

$$\frac{dp(\theta)}{d\theta} = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}\left((a-1)\theta^{a-2}(1-\theta)^{b-1} - \theta^{a-1}(b-1)(1-\theta)^{b-2}\right) = 0$$

we have
$$(a-1)\theta^{a-2}(1-\theta)^{b-1} - \theta^{a-1}(b-1)(1-\theta)^{b-2} = 0 \rightarrow \theta = \frac{a-1}{a+b-2}$$

Note that if $a > 1, b > 1$, $p(\theta = 0) = p(\theta = 1) = 0$, since $p(\theta) \geq 0 \, \forall \theta$ and $p(\theta)$ is continuous, so the solution we get above can maximize $p(\theta)$.

$$
\begin{aligned}
\mathbb{E}(\theta) &= \int_0^1 \theta \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \theta^{a-1}(1-\theta)^{b-1} d\theta \\
&= \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \int_0^1 \theta^{a+1-1}(1-\theta)^{b-1} d\theta \\
&= \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \frac{\Gamma(a+1)\Gamma(b)}{\Gamma(a+b+1)} \\
&= \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \frac{a\Gamma(a)\Gamma(b)}{(a+b)\Gamma(a+b)} \\
&= \frac{a}{a+b} \\
\mathbb{E}(\theta^2) &= \int_0^1 \theta^2 \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \theta^{a-1}(1-\theta)^{b-1} d\theta \\
&= \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \int_0^1 \theta^{a+2-1}(1-\theta)^{b-1} d\theta \\
&= \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \frac{\Gamma(a+2)\Gamma(b)}{\Gamma(a+b+2)} \\
&= \frac{a(a+1)}{(a+b)(a+b+1)}
\end{aligned}
$$

thus

$$\text{Var}[\theta] = \mathbb{E}(\theta^2) - (\mathbb{E}(\theta))^2 = \frac{a(a+1)}{(a+b)(a+b+1)} - \frac{a^2}{(a+b)^2} = \frac{ab}{(a+b)^2(a+b+1)}$$

Problem 5

Let θ be the probability of head of the chosen coin. Since we choose one coin at random,

$$P(\theta = 0.6) = P(\theta = 0.4) = \frac{1}{2}$$

Let T_i to represent the event that the i_{th} spin is tail. Then we have

$$T_i | \theta \sim \text{Bernoulli}(\theta), \text{i.i.d}$$

Posterior of θ is

$$p(\theta | T_1, T_2) = \frac{p(T_1, T_2 | \theta) p(\theta)}{p(T_1, T_2)}$$

we have
$$P(\theta = 0.6|T_1, T_2) = \frac{(1 - 0.6) \times (1 - 0.6) \times 1/2}{0.6 \times 0.6 \times 1/2 + 0.4 \times 0.4 \times 1/2} = \frac{4}{13},$$
$$P(\theta = 0.4|T_1, T_2) = 1 - P(\theta = 0.4|T_1, T_2) = \frac{9}{13}$$

Let Y denote the number of additional spins till a head show up, it follows Geometric distribution.
$$Y|\theta \sim \text{Geom}(\theta)$$
$$\mathbb{E}(Y|\theta) = \frac{1}{\theta}$$

Therefore
$$\begin{aligned}
\mathbb{E}(Y|T_1, T_2) &= \mathbb{E}(\mathbb{E}(Y|\theta, T_1, T_2)|T_1, T_2) \\
&= \sum_\theta P(\theta|T_1, T_2) \cdot \mathbb{E}(Y|\theta) \\
&= P(\theta = 0.6|T_1, T_2) \cdot \frac{1}{0.6} + P(\theta = 0.4|T_1, T_2) \cdot \frac{1}{0.4} \\
&= \frac{4}{13} \cdot \frac{1}{0.6} + \frac{9}{13} \cdot \frac{1}{0.4} \\
&= \frac{175}{78} \\
&\approx 2.2436
\end{aligned}$$