

# 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

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

We can calculate  $P(H_i|E)$  explicitly using the formula.

$$P(H_i|E) = \frac{P(E|H_i)P(H_i)}{\sum_{i=1}^3 P(E|H_i)P(H_i)}$$

From this, we observe that  $P(H_i|E)$  all have the same denominator, so the comparisons among them only depend on the numerator.

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 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

$$0.1P(H_1) < 0.5P(H_3) < 0.3P(H_2)$$

i.e.

$$P(E | H_1)P(H_1) < P(E | H_3)P(H_3) < P(E | H_2)P(H_2)$$

Since  $P(E) > 0$ , we can divide each of them by  $P(E)$

$$\frac{P(E | H_1)P(H_1)}{P(E)} < \frac{P(E | H_3)P(H_3)}{P(E)} < \frac{P(E | H_2)P(H_2)}{P(E)}$$

i.e.

$$\frac{P(E \cap H_1)}{P(E)} < \frac{P(E \cap H_3)}{P(E)} < \frac{P(E \cap H_2)}{P(E)}$$

Thus, we get

$$P(H_1 | E) < P(H_3 | E) < P(H_2 | E)$$

(b) We 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

$$0.3P(H_2) < 0.1P(H_1) < 0.5P(H_3)$$

i.e.

$$P(E | H_2)P(H_2) < P(E | H_1)P(H_1) < P(E | H_3)P(H_3)$$

Since  $P(E) > 0$ , we can divide each of them by  $P(E)$

$$\frac{P(E | H_2)P(H_2)}{P(E)} < \frac{P(E | H_1)P(H_1)}{P(E)} < \frac{P(E | H_3)P(H_3)}{P(E)}$$

i.e.

$$\frac{P(E \cap H_2)}{P(E)} < \frac{P(E \cap H_1)}{P(E)} < \frac{P(E \cap H_3)}{P(E)}$$

Thus, we get

$$P(H_2 | E) < P(H_1 | E) < P(H_3 | E)$$

(c) We 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

$$0.3P(H_2) < 0.5P(H_3) < 0.1P(H_1)$$

i.e.

$$P(E | H_2)P(H_2) < P(E | H_3)P(H_3) < P(E | H_1)P(H_1)$$

Since  $P(E) > 0$ , we can divide each of them by  $P(E)$

$$\frac{P(E | H_2)P(H_2)}{P(E)} < \frac{P(E | H_3)P(H_3)}{P(E)} < \frac{P(E | H_1)P(H_1)}{P(E)}$$

i.e.

$$\frac{P(E \cap H_2)}{P(E)} < \frac{P(E \cap H_3)}{P(E)} < \frac{P(E \cap H_1)}{P(E)}$$

Thus, we get

$$P(H_2 | E) < P(H_3 | E) < P(H_1 | E)$$

(d) We 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

$$0.5P(H_3) < 0.3P(H_2) < 0.1P(H_1)$$

i.e.

$$P(E | H_3)P(H_3) < P(E | H_2)P(H_2) < P(E | H_1)P(H_1)$$

Since  $P(E) > 0$ , we can divide each of them by  $P(E)$

$$\frac{P(E | H_3)P(H_3)}{P(E)} < \frac{P(E | H_2)P(H_2)}{P(E)} < \frac{P(E | H_1)P(H_1)}{P(E)}$$

i.e.

$$\frac{P(E \cap H_3)}{P(E)} < \frac{P(E \cap H_2)}{P(E)} < \frac{P(E \cap H_1)}{P(E)}$$

Thus, we get

$$P(H_3 | E) < P(H_2 | E) < P(H_1 | E)$$

(e) We 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

$$0.5P(H_3) < 0.1P(H_1) < 0.3P(H_2)$$

i.e.

$$P(E | H_3)P(H_3) < P(E | H_1)P(H_1) < P(E | H_2)P(H_2)$$

Since  $P(E) > 0$ , we can divide each of them by  $P(E)$

$$\frac{P(E | H_3)P(H_3)}{P(E)} < \frac{P(E | H_1)P(H_1)}{P(E)} < \frac{P(E | H_2)P(H_2)}{P(E)}$$

i.e.

$$\frac{P(E \cap H_3)}{P(E)} < \frac{P(E \cap H_1)}{P(E)} < \frac{P(E \cap H_2)}{P(E)}$$

Thus, we get

$$P(H_3 | E) < P(H_1 | E) < P(H_2 | E)$$

Check

```
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

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

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. We use the following information to construct this DAG.

Second, we specify the domain, distribution, and parameters of each random variable resp.

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

The random variable  $X$  is a discrete random variable given that the support is a countably infinite set as

$$\{x \mid x \in (0, 1, 2, \dots, )\}$$

We assume it follows Beta-binomial distribution. The PDF is given by:

$$P(X = k) = \binom{n}{k} \frac{B(k + \alpha, n - k + \beta)}{B(\alpha, \beta)}$$

where  $B()$  is the Beta function. It is characterized by three parameters:  $\alpha$ ,  $\beta$ ,  $n$ .  $n$  is the number of Bernoulli trials,  $n$  can be any non-negative integer, i.e.

$$\{n \mid n \in (0, 1, 2, \dots, )\}$$

$\alpha$  is one of the two shape parameters of the underlying Beta distribution. It must be positive. Thus, its range is  $\alpha > 0$ .  $\beta$  is the second shape parameter of the underlying Beta distribution, it must also be positive. Thus, its range is  $\beta > 0$

2.  $Y$  represents the country the person is from.

The random variable  $Y$  is a discrete random variable given that the support is a finite set as

$$\{y \mid y \in (C_1, C_2, C_3, \dots, C_{195})\}$$

We know that the possibility that a person is from a country depends on the population. Since we know the number is known, there is no parameter. The PDF is given by:

$$P(Y = C_i) = \frac{\text{Population of country } C_i}{\text{Population of the world}}$$

3.  $Z$  represents whether the person attended a picnic the first week of July, 2023. The random variable  $Z$  is a binary random variable that can take 0 or 1.

$$\{z \mid z \in \{0, 1\}\}$$

Obviously it follows Bernoulli distribution. The PDF is given by:

$$P(Z = z) = p^z(1 - p)^{1-z} \quad z \in \{0, 1\}$$

where parameter  $p$  is the probability of attending a picnic, here we consider the range of it is  $[0, 1]$ .

4.  $U$  represents the amount of coffee the person consumed in 2023. The random variable  $U$  is a continuous random variable that takes value greater than zero.

$$\{u \mid u \in [0, \infty)\}$$

We assume it follows gamma distribution. The PDF is given by:

$$f(x; \alpha, \beta) = \frac{\beta^\alpha x^{\alpha-1} e^{-\beta x}}{\Gamma(\alpha)}$$

With a shape parameter *alpha* and a scale parameter *beta*. The range of the parameters are  $\alpha > 0$  and  $\beta > 0$

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)  $v_i$  follow a beta distribution which takes value from 0 to 1. And it's reasonable to assume each hour is independent. So  $V$  is a vector random variable with the distribution

$$\mathbf{V} = (V_1, V_2, \dots, V_{24}) \text{ where for each } i \in \{1, 2, \dots, 24\} : V_i \sim \text{Beta}(\alpha_i, \beta_i), \quad i.i.d.$$

The pdf of the joint distribution is

$$f(\mathbf{V}) = \prod_{i=1}^{24} \frac{v_i^{\alpha_i-1} (1-v_i)^{\beta_i-1}}{B(\alpha_i, \beta_i)}$$

with the support of each  $V_i$  in the following. And the parameters' range is  $\alpha_i > 0$  and  $\beta_i > 0$ .

$$\{v_i \mid v_i \in [0, 1]\}$$

6.  $W$  represents a passport photo of the person.
7. Country ( $Y$ ) influences age ( $X$ ) structure.
8. Country ( $Y$ ) and age ( $X$ ) influence whether the person attended a picnic ( $Z$ ), since country reflects cultural and different age has different opinion and tendency for a picnic.
9. Age ( $X$ ) influences coffee consumed ( $U$ ) and screen time spent ( $V$ ). For example, young age needs working, so maybe more coffee and screen time.
10. Screen time ( $V$ ) influences coffee consumed ( $U$ ).
11. Country ( $Y$ ) influences passport photo ( $W$ ).
12. All the term "influence" above means Markovian parents of the variable.
13. All distribution above also means conditional distribution.

Since there's no collider or any separate node in this DAG, we conclude that  $X, Y, Z, U, V, W$  are not pairwise independently distributed.

As for joint distribution, it's easy to read from the DAG.

$$f(x, y, z, u, v, w) = f(x|y)f(y)f(z|x, y)f(u|x, v)f(v|x)f(w|y)$$

## 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  $a > 0, b > 0$ ,  $p(\theta = 0) = p(\theta = 1) = 0$ , since  $p(\theta) \geq 0 \forall \theta$  and  $p(\theta)$  is continuous, The mode 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

The parameter we care about is  $\theta$ , which means the probability of head of the chosen coin. Our prior is  $P(\text{heads}|C_1) = 0.6, P(\text{heads}|C_2) = 0.4$ , since we choose one coin at random,

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

Use  $T_1, T_2$  to represent the first two spins are tails. This event is data and

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

Posterior of  $\theta$  is

$$p(\theta|T_1T_2) = \frac{p(T_1T_2|\theta)p(\theta)}{p(T_1T_2)}$$

we have

$$P(\theta = 0.6|T_1T_2) = \frac{4}{13}, \quad P(\theta = 0.4|T_1T_2) = \frac{9}{13}$$

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

$$\begin{aligned} \mathbb{E}(Y|T_1T_2) &= \mathbb{E}(\mathbb{E}(Y|\theta, T_1T_2)|T_1T_2) \quad (\text{Law of total expectation}) \\ &= \mathbb{E}(\mathbb{E}(Y|\theta)|T_1T_2) \\ &= \sum_{\theta} P(\theta|T_1T_2) \cdot \mathbb{E}(Y|\theta) \\ &= P(\theta = 0.6|T_1T_2) \cdot \frac{1}{0.6} + P(\theta = 0.4|T_1T_2) \cdot \frac{1}{0.4} \\ &= \frac{4}{13} \cdot \frac{1}{0.6} + \frac{9}{13} \cdot \frac{1}{0.4} \\ &= \frac{175}{78} \\ &\approx 2.2436 \end{aligned}$$