9. Bayesian Inference for Survival Prediction on the Titanic

1. Subset of 15 Passengers

PID	Survived	Pclass	Sex
1	0	3	M
2	1	1	F
3	1	3	F
4	1	1	F
5	0	3	M
6	0	3	M
7	0	1	M
8	0	3	M
9	1	3	F
10	1	2	F
16	1	2	F
18	1	2	M
21	0	2	M
24	1	1	M
34	0	2	M

"Survived": 0 = No, 1 = Yes
"Pclass": 1 = 1st class, 2 = 2nd class, 3 = 3rd class
"Sex": male or female.

Bayesian Inference

1. Calculate Prior probabilities:

- Compute prior probability of survival, P(s) based on the overall survival rate in the sample

$$P(s) = \frac{\text{No. of survivors}}{\text{Total no. of passengers}}$$

n = 15 (sample size)

no. of survivors = 8

$$P(s) = \frac{8}{15} = 0.533$$

- Calculate Likelihoods:

- Specify a probability model

we assume a binomial distribution for binary distribution (e.g. "survived" or "not survived")

or

assume a uniform distribution.

we will choose uniform distribution on [0,1]. It does not encode strong prior belief

$$p \sim Beta[1,1]$$

---

2. Likelihood $P(D|S)$ and $P(D|\sim S)$:

$$P(D|S) = \frac{\text{No. of survivors who have feature } D}{\text{total no. of survivors}}$$

$$P(D|\sim S) = \frac{\text{No. of non-survivors who have feature } D}{\text{total no. of survivors}}$$

4 Cases

- D = Female
- D = Male
- D = Passenger is 1st Class
- D = ——"—— 2nd Class
- D = ——"—— 3rd class

(A)

$P(D = \text{Passenger is male} \mid S) = \frac{2}{8} = \frac{1}{4}$

$P(D = \text{Passenger is male} \mid \sim S) = \frac{0}{8} = \frac{7}{7} = 1$

(B)

$P(D = \text{Passenger is female} \mid S) = \frac{6}{8} = \frac{3}{4}$

$P(D = \text{Passenger is female} \mid \sim S) = \frac{0}{7} = 0$

(C)

$P(D = 1\text{st class} \mid S) = \frac{3}{8}$

$P(D = 1\text{st class} \mid \sim S) = \frac{1}{7}$

(D)

$P(D = 2\text{nd class} \mid S) = \frac{3}{8}$

$P(D = 2\text{nd class} \mid \sim S) = \frac{2}{7}$

(E)

$P(D = 3\text{rd class} \mid S) = \frac{2}{8} = \frac{1}{4}$

$P(D = 3\text{rd class} \mid \sim S) = \frac{4}{7}$

3. Calculate a Posterior probability:

(A) $P(S|D = \text{Passenger is male}) = \frac{P(D|S)\,P(S)}{P(D)}$

$P(D) = P(D|S)\,P(S) + P(D|\sim S)\,P(\sim S)$

$= \frac{\frac{1}{4} \times \frac{8}{15}}{\frac{1}{4} \times \frac{8}{15} + \frac{1}{4} \times \frac{7}{15}} = \frac{\frac{2}{15}}{\frac{2}{15} + \frac{7}{15}} = \frac{\frac{2}{15}}{\frac{9}{15}} = \frac{2}{9}$

$= 0.22$

(B) $P(S|D = \text{Passenger is Female}) = \frac{P(D|S)\,P(S)}{P(D)}$

$= \frac{\frac{3}{4} \times \frac{8}{15}}{\frac{3}{4} \times \frac{8}{15} + 0} = 1$

(C) $P(S|D = \text{1st Class}) = \frac{P(D|S)\,P(S)}{P(D)}$

$= \frac{\frac{3}{8} \times \frac{8}{15}}{\frac{3}{8} \times \frac{8}{15} + \frac{1}{8} \times \frac{7}{15}} = \frac{\frac{3}{15}}{\frac{3}{15} + \frac{1}{15}} = \frac{3}{15} \times \frac{15}{4} = \frac{3}{4}$

$= 0.75$

(D)

$$P(S|D=\text{2nd class}) = \frac{P(D|S)P(S)}{P(D)}$$

$$= \frac{\frac{3}{8}\times\frac{8}{15}}{\frac{3}{8}\times\frac{8}{15}+\frac{2}{7}\times\frac{7}{15}} = \frac{\frac{3}{15}}{\frac{5}{15}} = \frac{3}{5} = 0.60$$

(E)

$$P(S|D=\text{3rd class}) = \frac{P(D|S)P(S)}{P(D)}$$

$$= \frac{\frac{2}{8}\times\frac{8}{15}}{\frac{2}{8}\times\frac{8}{15}+\frac{4}{7}\times\frac{7}{15}} = \frac{\frac{2}{15}}{\frac{6}{15}} = \frac{1}{3} = 0.33$$

Interpretations

(A)

$P(S) = 50\%$ (prior probability)

$P(S \mid D = \text{"passenger is male"}) = 22\%$

$P(S \mid D = \text{"passenger is female"}) = 100\%$

we see that male survival rate is far less than female survivor.

female survivor rate is perfect.

∵ the sample is small, we cannot influence it for the entirity of data.

(b) $P(S) = 50\%$

$P(D = 1st\ class \mid S) = 37.5\%$

$P(D = 2nd\ class \mid S) = 37.5\%$

$P(D = 3rd\ Class \mid S) = 2.5\%$

We see that 3rd Class passenger had a lesser survival rate compared to 1st and 2nd class.

But it is less extreme.

↳ we see that being a female has the biggest impact on survival rate.

↳ second would be belonging to 1st or 2nd class it will have minimal/small effect on survival rate.