
Algorithmen und Wahrscheinlichkeit

Peer Grading Exercise 2

SUBMIT BY MOODLE () UNTIL 16:00 ON 27.03.2025.

Exercise 1 – *Conditional probability*

- (a) Assume that $\Pr[A|X] > \Pr[B|X]$ and $\Pr[A|\bar{X}] > \Pr[B|\bar{X}]$. Does it necessarily implies that $\Pr[A] > \Pr[B]$? Prove or give a counterexample.
- (b) Assume that $\Pr[A|X \cap Y] > \Pr[A|X \cap \bar{Y}]$ and $\Pr[A|\bar{X} \cap Y] > \Pr[A|\bar{X} \cap \bar{Y}]$. Does it necessarily implies that $\Pr[A|Y] > \Pr[A|\bar{Y}]$? Prove or give a counterexample.
- (c) For two soccer teams, U and V , assume that U won $p_{U,2023} \in [0, 1]$ fraction of the games it played (not necessarily against each other) in the year 2023, and V won $p_{V,2023} \in [0, 1]$ fraction of the game it played. The $p_{U,2024}$ and $p_{V,2024}$ are defined similarly for the year 2024. Assume moreover that $p_{U,2023} > p_{V,2023}$ and $p_{U,2024} > p_{V,2024}$. (In each year each team played at least one game, so that the fractions above are well-defined.)

If p_U (respectively p_V) is the fraction of games U (resp. V) won in years 2023-2024, can we conclude that $p_U > p_V$?

Solution 1

- (a) Yes, the assumptions imply that $\Pr[A] > \Pr[B]$. Using the law of total probability we have

$$\Pr[A] = \Pr[A|X]\Pr[X] + \Pr[A|\bar{X}]\Pr[\bar{X}] \geq \Pr[B|X]\Pr[X] + \Pr[B|\bar{X}]\Pr[\bar{X}] = \Pr[B].$$

- (b) No, this statement is not true. To see this, consider a scenario where we first flip a fair coin for Y , so that $\Pr[Y] = 1/2$. Then depending on the outcome of Y , the X is a random event with respective conditional probabilities $\Pr[X|Y] = 1 - \varepsilon$ and $\Pr[X|\bar{Y}] = \varepsilon$. Finally, for each possible combination of X, \bar{X} and Y, \bar{Y} conditional probabilities of A as given as follows.

- $\Pr[A|X \cap Y] = 2\varepsilon$,
- $\Pr[A|X \cap \bar{Y}] = \varepsilon$,
- $\Pr[A|\bar{X} \cap Y] = 1 - \varepsilon$,
- $\Pr[A|\bar{X} \cap \bar{Y}] = 1 - 2\varepsilon$.

Then we will have

$$\Pr[A|Y] = \Pr[A|X \cap Y]\Pr[X|Y] + \Pr[A|\bar{X} \cap Y]\Pr[\bar{X}|Y] \leq 3\varepsilon,$$

and

$$\Pr[A|\bar{Y}] \geq \Pr[A|\bar{X} \cap \bar{Y}]\Pr[\bar{X}|\bar{Y}] = (1 - 2\varepsilon)(1 - \varepsilon) \geq 1 - 3\varepsilon.$$

Therefore, whenever $\varepsilon < 1/6$, we would have $\Pr[A|Y] < \Pr[A|\bar{Y}]$.

We can explicitly construct a probability space satisfying those conditions. Let $\Omega = \{0, 1\}^3$, where $X = \{\omega \in \Omega : \omega_1 = 1\}$, $Y = \{\omega \in \Omega : \omega_2 = 1\}$ and $A = \{\omega \in \Omega : \omega_3 = 1\}$. The probabilities will now be

- $\Pr[(0, 0, 0)] = \Pr[\bar{A}|\bar{X} \cap \bar{Y}]\Pr[\bar{X}|\bar{Y}]\Pr[\bar{Y}] = (1 - 2\varepsilon)(1 - \varepsilon)/2$.
- $\Pr[(0, 0, 1)] = \Pr[A|\bar{X} \cap \bar{Y}]\Pr[\bar{X}|\bar{Y}]\Pr[\bar{Y}] = 2\varepsilon(1 - \varepsilon)/2$.

- $\Pr[(0, 1, 0)] = \Pr[\bar{A}|\bar{X} \cap Y]\Pr[\bar{X}|Y]\Pr[Y] = \varepsilon^2/2.$
- $\Pr[(0, 1, 1)] = \Pr[A|\bar{X} \cap Y]\Pr[\bar{X}|Y]\Pr[Y] = (1 - \varepsilon)\varepsilon/2.$
- $\Pr[(1, 0, 0)] = \Pr[\bar{A}|X \cap \bar{Y}]\Pr[X|\bar{Y}]\Pr[\bar{Y}] = (1 - \varepsilon)\varepsilon/2.$
- $\Pr[(1, 0, 1)] = \Pr[A|X \cap \bar{Y}]\Pr[X|\bar{Y}]\Pr[\bar{Y}] = \varepsilon^2/2.$
- $\Pr[(1, 1, 0)] = \Pr[\bar{A}|X \cap Y]\Pr[X|Y]\Pr[Y] = (1 - 2\varepsilon)(1 - \varepsilon)/2.$
- $\Pr[(1, 1, 1)] = \Pr[A|X \cap Y]\Pr[X|Y]\Pr[Y] = 2\varepsilon(1 - \varepsilon)/2.$

(c) This is just rephrasing of the subproblem (b). Here A is “probability of a game win”, Y is “team U is playing” and X corresponds to “the year is 2024”. The same construction as in subproblem (b) can be used to show a counterexample to the statement in this problem.

Note: This, rather counterintuitive statistical phenomenon is called “Simpson’s paradox”, and has been observed in real-world data numerous times — where from the same dataset careless analyst can draw contradicting conclusion (“team U is better” or “team V is better”), depending on whether they partition the data set in a particular way, or treat it as a whole.