BM20A9301 Statistics – Exercise set 3

To be done by 22.–26.1.2024

Text in blue or red is not part of the problem or its solution. It's there as extra information to help you learn.

Exercise 1 (Bayes' rule)**.** 2% of men have awesome manicured nails. 20% of women have awesome manicured nails. You see a random adult with awesome nails. What is the probability that it's a man? **Hint:** unless otherwise stated, in this kind of problems you can assume there's just as many men as women.

Solution. Consider the following events:

- $N =$"awesome nails"
- $M =$"man"
- $W =$"woman"

We have $P(N|M) = 0.02$ and $P(N|W) = 0.2$. By Bayes' rule

$$P(M|N) = \frac{P(N|M) \cdot P(M)}{P(N)}.$$

By the law of total probability

$$P(N) = P(N|W) \cdot P(W) + P(N|M) \cdot P(M) = 0.2 \cdot 0.5 + 0.02 \cdot 0.5 = 0.11.$$

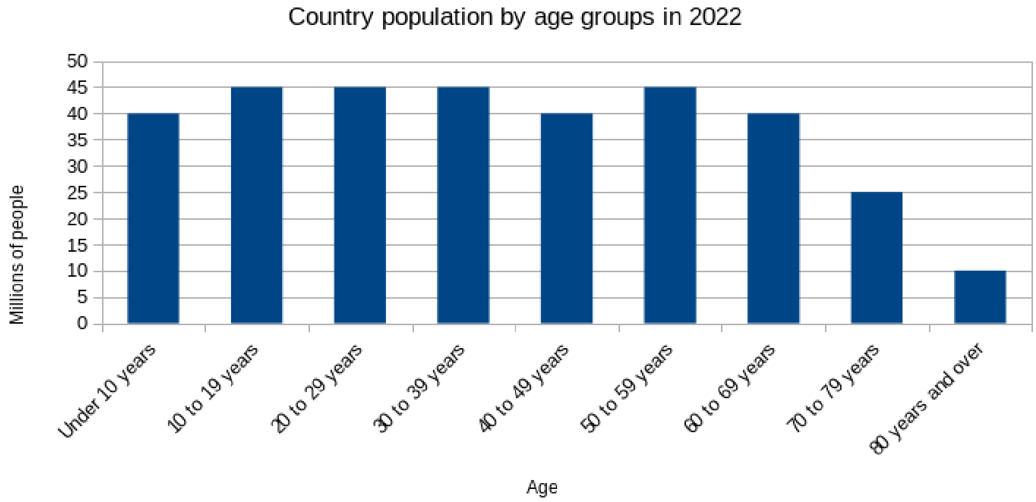
Hence

$$P(M|N) = \frac{0.02 \cdot 0.5}{0.11} \approx 0.09$$

Exercise 2 (Total probability)**.** In a Never Have I Ever television game show, the host tells some statistics about 2022:

- 1% of people under 30 years old have ever shoplifted,
- 4% of people at least 30 but under 60 years old have ever shoplifted,
- 11% of people at least 60 years old have ever shoplifted.

The next contestant is going to be a randomly selected person from the county's population. See the chart below. What is the probability that they've **never** shoplifted?



Solution. In this exercise we'll use the law of total probability. It requires a partition of the probability space. Consider the random experiment of "what is the age of the next contestant and has the next contestant ever shoplifted". Let

- $Y =$"the person is under 30 years old"
- $M =$"the person is at least 30 years but under 60 years old"
- $O =$"the person is at least 60 years old"

Denote by S the event that the random person has shoplifted. The problem statement thells us that

- $P(S|Y) = 0.01$
- $P(S|M) = 0.04$
- $P(S|O) = 0.11$

To use the law of total probability we need to find $P(Y)$, $P(M)$ and $P(O)$. These are just the number of people in the young, middle aged and old categories divided by the total population. Reading from the histogram we see that

- $40 + 45 + 45 = 130$ million people are under 30.
- $45 + 40 + 45 = 130$ million people are at least 30 but under 60.
- $40 + 25 + 10 = 75$ million people are over 60.

In total there are $130 + 130 + 75 = 335$ people. Hence $P(Y) = 130/335$, $P(M) = 130/335$ and $P(O) = 75/335$.

By the law of total probability

$$\begin{aligned} P(S) &= P(S|Y)P(Y) + P(S|M)P(M) + P(S|O)P(O) \\ &= 0.01 \cdot \frac{130}{335} + 0.04 \cdot \frac{130}{335} + 0.11 \cdot \frac{75}{335} \\ &= \frac{0.01 \cdot 130 + 0.04 \cdot 130 + 0.11 \cdot 75}{335} = \frac{14.75}{335} \approx 0.044 \end{aligned}$$

so there's a 4.4% chance that the next contestant will have shoplifted. Thus the probability that they've never shoplifted is $100\% - 4.4\% = 95.6\%$.

Exercise 3 (Prosecutor problem)**.** A hundred taxis operate in a city. One of them is orange and the rest are yellow. A pedestrian gets ran over by a taxi one a dark and foggy night. Based on earlier studies, people see an orange car as yellow with 90% probability, and a yellow one as orange with 8% probability in those circumstances.

(a) If the criminal taxi was yellow, what is the probability that a witness would see an orange taxi?

(b) If the criminal taxi was orange, what is the probability that a witness would see an orange taxi?

(c) A-priori (before interviewing witnesses) we suspect that the taxi that ran over the pedestrian is orange with 1% probability and yellow with 99% probability. If we pick a witness at random, what is our subjective probability that the witness would claim to have seen an orange taxi?

(d) A witness claims the taxi was orange, so the police arrest the driver. What is the probability that the taxi that ran over the person is actually orange considering the witness account?

(e) Looking at the answer to (a) the prosecutor called for the conviction of the orange taxi's driver. Do you agree with that? Explain why.

Solution. Write Y_r and O_r for the events that the taxi that ran over the pedestrian is yellow and orange, respectively. Write also Y_w and O_w for the events that the witness thinks they say a yellow or orange car, respectively. The problem statement gives $P(Y_w|O_r) = 0.90$ and $P(O_w|Y_r) = 0.08$.

(a) This is just $P(O_w|Y_r)$ which is 8%.

(b) This is $P(O_w|O_r) = 1 - P(Y_w|O_r) = 1 - 0.90 = 10\%$ because there are only two colours of taxis.

(c) The statement gives the a-priori probabilities $P(O_r) = 0.01$ and $P(Y_r) = 0.99$. Because there's only one car that ran the pedestrian over these two events form a partition of the probability space. Thus by the law of total probability

$$\begin{aligned} P(O_w) &= P(O_w|O_r)P(O_r) + P(O_w|Y_r)P(Y_r) \\ &= 0.10 \cdot 0.01 + 0.08 \cdot 0.99 = 0.0802 = 8.02\% \end{aligned}$$

(d) Let's calculate the probability of O_r conditioned on the event O_w. Let's use Bayes' rule:

$$P(O_r|O_w) = \frac{P(O_w|O_r)P(O_r)}{P(O_w)} = \frac{0.10 \cdot 0.01}{0.0802} \approx 0.0125$$

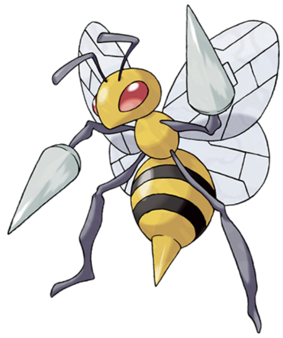
So conditioned on the witness account the probability that the orange taxi did run over the pedestrian is 1.25%.

(e) The prosecutor might have thought: "if a yellow taxi had ran over there's only an 8% chance for the witness account. Therefor the orange taxi is the culprit with 92% chance!" This is of course wrong, because $P(O_r|O_w) \neq 1 - P(O_w|Y_r)$. Instead we calculated that there's only a 1.25% probability that the orange taxi's driver is the culprit, a bit more than our guess without the witness account, but much less than what's required for conviction.

Exercise 4 (Expected value). You can solve this by simulating (Python, R, Excel, whatever) or by calculating. Up to you.

When a pokémon uses Fury Attack, it hits 2–5 times according to the below probabilities. On average how many hits does the pokémon do per Fury Attack provided the attack succeeds?

number of hits	2	3	4	5
probability	3/8	3/8	1/8	1/8



©2023 Pokémon

Solution. The number of hits is a random variable H with probabilities of events $\{H = 2\}$, $\{H = 3\}$, $\{H = 4\}$ and $\{H = 5\}$ given in the table. "On average" means the number gotten as follows: repeat the experiment a lot of times and sum the total number hits, and divide this by the number of times the experiment was done. By the law of large numbers this becomes the expected value of the random variable. We have

$$E(H) = \sum_x xP(H = x) = 2 \cdot \frac{3}{8} + 3 \cdot \frac{3}{8} + 4 \cdot \frac{1}{8} + 5 \cdot \frac{1}{8} = 3.$$

Hence on average Fury Attack hits three times when it succeeds.

Exercise 5 (Valid distributions). Determine whether or not the following tables are valid probability distributions of some discrete random variable X. Explain why or why not.

(a)

x	0	1	2	3	4
$P(X = x)$	-0.25	0.5	0.35	0.1	0.3

(b)

x	home	draw	away
$P(X = x)$	0.325	0.406	0.164

(c)

x	25	26	27	28	29
$P(X = x)$	0.13	0.27	0.28	0.18	0.14

Solution. The requirements for a discrete probability distribution f is that 1) $f(x) \geq 0$ for any x, and 2) $\sum_x f(x) = 1$.

(a) This fails 1) because $f(0) < 0$.

(b) This fails 2) because $0.325 + 0.406 + 0.164 < 1$.

(c) This satisfies both 1) and 2), so it is a valid probability distribution.

Exercise 6 (Distribution of a combined random variable)**.** Suppose you have two "loaded" dice, meaning that the probability of rolling a six is twice as high as rolling any other number (the probabilities for $1, 2, 3, 4, 5, 6$ are $\frac{1}{7}, \frac{1}{7}, \frac{1}{7}, \frac{1}{7}, \frac{1}{7}, \frac{2}{7}$). You roll them both and let X denote their sum.

(a) What is the range (=possible values) of X?

(b) Write down (as a table) the probability density function (PDF) of X.

Solution.

(a) We only changed the probabilities, not the numbers on the dice. So it's just like for the sum of two ordinary dice: $\Omega = \{2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12\}$.

(b) Let X_1 be the result of the first die, and X_2 of the second die. Let's form a table of probabilities $P(X_1 = x_1, X_2 = x_2)$. The dice are independent so $P(X_1 = x_1, X_2 = x_2) = P(X_1 = x_1) \cdot P(X_2 = x_2)$. For example $P(X_1 = 2, X_2 = 6) = \frac{1}{7}\frac{2}{7} = \frac{2}{49}$. Let x_1 be on the x-axis and x_2 on the y-axis:

		X_2 1	2	3	4	5	6
X_1	1	1/49	1/49	1/49	1/49	1/49	2/49
	2	1/49	1/49	1/49	1/49	1/49	2/49
	3	1/49	1/49	1/49	1/49	1/49	2/49
	4	1/49	1/49	1/49	1/49	1/49	2/49
	5	1/49	1/49	1/49	1/49	1/49	2/49
	6	2/49	2/49	2/49	2/49	2/49	4/49

Table showing $P(X_1 = x_1, X_2 = x_2)$

Then the random variable of interest, the sum $S = X_1 + X_2$ will have the below table. It's values are calculated by summing the diagonals in the above table. For example $P(S = 4) = P(X_1 = 3, X_2 = 1) + P(X_1 = 2, x_2 = 2) + P(X_1 = 1, X_2 = 2)$.

s	2	3	4	5	6	7	8	9	10	11	12
$P(S = s)$	$\frac{1}{49}$	$\frac{2}{49}$	$\frac{3}{49}$	$\frac{4}{49}$	$\frac{5}{49}$	$\frac{8}{49}$	$\frac{7}{49}$	$\frac{6}{49}$	$\frac{5}{49}$	$\frac{4}{49}$	$\frac{4}{49}$