Probability and Statistics — MATH-235
Autumn 2025 — Mats J. Stensrud

Problem Sheet 4 [1]

Optional Revision Problems

Exercise 1 (Conditional Independence)**.** Two different diseases cause a certain weird symptom; anyone who has either or both of these diseases will experience the symptom. Let D_1 be the event of having the first disease, D_2 be the event of having the second disease, and W be the event of having the weird symptom. Suppose that D_1 and D_2 are independent with $P(D_j) = p_j$, and that a person with neither of these diseases will have the weird symptom with probability w_0. Let $q_j = 1 - p_j$, and assume that $0 < p_j < 1$.

1. Find $P(W)$.
2. Find $P(D_1|W)$, $P(D_2|W)$, and $P(D_1, D_2|W)$.
3. Determine algebraically whether or not D_1 and D_2 are conditionally independent given W.

Exercise 2 (Monty Hall Problem)**.** Consider the Monty Hall problem, except that Monty enjoys opening door 2 more than he enjoys opening door 3, and if he has a choice between opening these two doors, he opens door 2 with probability p, where $\frac{1}{2} \leq p \leq 1$.

To recap: there are three doors, behind one of which there is a car (which you want), and behind the other two of which there are goats (which you don't want). Initially, all possibilities are equally likely for where the car is. You choose a door, which for concreteness we assume is door 1. Monty Hall then opens a door to reveal a goat, and offers you the option of switching. Assume that Monty Hall knows which door has the car, will always open a goat door and offer the option of switching, and as above assume that if Monty Hall has a choice between opening door 2 and door 3, he chooses door 2 with probability p, (with $\frac{1}{2} \leq p \leq 1$).

1. Find the unconditional probability that the strategy of always switching succeeds (unconditional in the sense that we do not condition on which of doors 2 or 3 Monty opens).
2. Find the probability that the strategy of always switching succeeds, given that Monty opens door 2.
3. Find the probability that the strategy of always switching succeeds, given that Monty opens door 3.

Exercise 3 (Simpson's Paradox)**.** The book *Red State, Blue State, Rich State, Poor State* by Andrew Gelman discusses the following election phenomenon: within any U.S. state, a wealthy voter is more likely to vote for a Republican than a poor voter, yet the wealthier states tend to favor Democratic candidates!

[1] Exercises are based on the coursebook Statistics 110: Probability by Joe Blitzstein

1. Assume for simplicity that there are only 2 states (called Red and Blue), each of which has 100 people, and that each person is either rich or poor, and either a Democrat or a Republican. Make up numbers consistent with the above, showing how this phenomenon is possible, by giving a 2×2 table for each state (listing how many people in each state are rich Democrats, etc.). So within each state, a rich voter is more likely to vote for a Republican than a poor voter, but the percentage of Democrats is higher in the state with the higher percentage of rich people than in the state with the lower percentage of rich people.

2. In the setup of part 1. (not necessarily with the numbers you made up there), let D be the event that a randomly chosen person is a Democrat (with all 200 people equally likely), and B be the event that the person lives in the Blue State. Suppose that 10 people move from the Blue State to the Red State. Write P_{old} and P_{new} for probabilities before and after they move. Assume that people do not change parties, so we have $P_{\text{new}}(D) = P_{\text{old}}(D)$. Is it possible that both $P_{\text{new}}(D \mid B) > P_{\text{old}}(D \mid B)$ and $P_{\text{new}}(D \mid B^c) > P_{\text{old}}(D \mid B^c)$ are true? If so, explain how it is possible and why it does not contradict the law of total probability $P(D) = P(D \mid B)P(B) + P(D \mid B^c)P(B^c)$; if not, show that it is impossible.

Week 4 Exercises

Exercise 4. 1. Independent Bernoulli trials are performed, with probability 1/2 of success, until there has been at least one success. Find the PMF of the number of trials performed.

2. Independent Bernoulli trials are performed, with probability 1/2 of success, until there has been at least one success and at least one failure. Find the PMF of the number of trials performed.

Exercise 5. *Benford's law* states that in a very large variety of real-life data sets, the first digit approximately follows a particular distribution with about a 30% chance of a 1, an 18% chance of a 2, and in general

$$P(D = j) = \log_{10}\left(\frac{j+1}{j}\right), \text{ for } j \in \{1, 2, 3, \ldots, 9\},$$

where D is the first digit of a randomly chosen element. Check that this is a valid PMF (using properties of logs, not with a calculator).

Exercise 6. There are 100 prizes, with one worth \$1, one worth \$2, . . . , and one worth \$100. There are 100 boxes, each of which contains one of the prizes. You get 5 prizes by picking random boxes one at a time, without replacement. Find the PMF of how much your most valuable prize is worth (as a simple expression in terms of binomial coefficients).

Exercise 7. An airline overbooks a flight, selling more tickets for the flight than there are seats on the plane (figuring that it's likely that some people won't show up). The plane has 100 seats, and 110 people have booked the flight. Each person will show up for the flight with probability 0.9, independently. Find the probability that there will be enough seats for everyone who shows up for the flight (a formula is fine, you do not need to calculate an exact number).

Exercise 8. 1. In the World Series of baseball, two teams (call them A and B) play a sequence of games against each other, and the first team to win four games wins the series. Let p be the probability that A wins an individual game, and assume that the games are independent. What is the probability that team A wins the series?

2. Give a clear intuitive explanation of whether the answer to 1. depends on whether the teams always play 7 games (and whoever wins the majority wins the series), or the teams stop playing more games as soon as one team has won 4 games (as is actually the case in practice: once the match is decided, the two teams do not keep playing more games).

Exercise 9. A certain company has n + m employees, consisting of n women and m men. The company is deciding which employees to promote.

1. Suppose for this part that the company decides to promote t employees, where $1 \leq t \leq n+m$, by choosing t random employees (with equal probabilities for each set of t employees). What is the distribution of the number of women who get promoted?

2. Now suppose that instead of having a predetermined number of promotions to give, the company decides independently for each employee, promoting the employee with probability p. Find the distributions of the number of women who are promoted, the number of women who are not promoted, and the number of employees who are promoted.

3. In the set-up from part 2., find the conditional distribution of the number of women who are promoted, given that exactly t employees are promoted.