# CS 70 Discrete Mathematics and Probability Theory Spring 2024 Seshia, Sinclair DIS 8B

## 1 Probability Potpourri

Provide brief justification for each part.

Note 13 Note 14

- (a) For two events A and B in any probability space, show that  $\mathbb{P}[A \setminus B] \ge \mathbb{P}[A] \mathbb{P}[B]$ .
- (b) Suppose  $\mathbb{P}[D \mid C] = \mathbb{P}[D \mid \overline{C}]$ , where  $\overline{C}$  is the complement of *C*. Prove that *D* is independent of *C*.
- (c) If *A* and *B* are disjoint, does that imply they're independent?

#### **Solution:**

(a) It can be helpful to first draw out a Venn diagram:



We can see here that  $\mathbb{P}[A] = \mathbb{P}[A \cap B] + \mathbb{P}[A \setminus B]$ , and that  $\mathbb{P}[B] = \mathbb{P}[A \cap B] + \mathbb{P}[B \setminus A]$ . Looking at the RHS, we have

$$\mathbb{P}[A] - \mathbb{P}[B] = (\mathbb{P}[A \cap B] + \mathbb{P}[A \setminus B]) - (\mathbb{P}[A \cap B] + \mathbb{P}[B \setminus A])$$
$$= \mathbb{P}[A \setminus B] - \mathbb{P}[B \setminus A]$$
$$\leq \mathbb{P}[A \setminus B]$$

(b) Using the total probability rule, we have

$$\mathbb{P}[D] = \mathbb{P}[D \cap C] + \mathbb{P}[D \cap \overline{C}] = \mathbb{P}[D \mid C] \cdot \mathbb{P}[C] + \mathbb{P}[D \mid \overline{C}] \cdot \mathbb{P}[\overline{C}].$$

But we know that  $\mathbb{P}[D \mid C] = \mathbb{P}[D \mid \overline{C}]$ , so this simplifies to

$$\mathbb{P}[D] = \mathbb{P}[D \mid C] \cdot (\mathbb{P}[C] + \mathbb{P}[\overline{C}]) = \mathbb{P}[D \mid C] \cdot 1 = \mathbb{P}[D \mid C],$$

which defines independence.

CS 70, Spring 2024, DIS 8B

(c) No; if two events are disjoint, we cannot conclude they are independent. Consider a roll of a fair six-sided die. Let A be the event that we roll a 1, and let B be the event that we roll a 2. Certainly A and B are disjoint, as  $\mathbb{P}[A \cap B] = 0$ . But these events are not independent:  $\mathbb{P}[B | A] = 0$ , but  $\mathbb{P}[B] = 1/6$ .

Since disjoint events have  $\mathbb{P}[A \cap B] = 0$ , we can see that the only time when disjoint A and B are independent is when either  $\mathbb{P}[A] = 0$  or  $\mathbb{P}[B] = 0$ .

- 2 Easter Eggs
- Note 14

You made the trek to Soda for a Spring Break-themed homework party, and every attendee gets to leave with a party favor. You're given a bag with 20 chocolate eggs and 40 (empty) plastic eggs. You pick 5 eggs (uniformly) without replacement.

- (a) What is the probability that the first egg you drew was a chocolate egg?
- (b) What is the probability that the second egg you drew was a chocolate egg?
- (c) Given that the first egg you drew was an empty plastic one, what is the probability that the fifth egg you drew was also an empty plastic egg?

#### **Solution:**

- (a)  $\mathbb{P}[\text{chocolate egg}] = \frac{20}{60} = \frac{1}{3}.$
- (b) Long calculation using Total Probability Rule: let  $C_i$  denote the event that the *i*th egg is chocolate, and  $P_i$  denote the event that the *i*th egg is plastic. We have

$$\mathbb{P}[C_2] = \mathbb{P}[C_1 \cap C_2] + \mathbb{P}[P_1 \cap C_2] \\= \mathbb{P}[C_1]\mathbb{P}[C_2 \mid C_1] + \mathbb{P}[P_1]\mathbb{P}[C_2 \mid P_1] \\= \frac{1}{3} \cdot \frac{19}{59} + \frac{2}{3} \cdot \frac{20}{59} \\= \frac{1}{3}.$$

Short calculation: By symmetry, this is the same probability as part (a), 1/3. This is because we don't know what type of egg was picked on the first draw, so the distribution for the second egg is the same as that of the first. To see this rigorously observe that  $\mathbb{P}[C_2 \cap P_1] = \mathbb{P}[P_2 \cap C_1]$ and, thus:

$$\mathbb{P}[C_2] = \mathbb{P}[C_2 \cap C_1] + \mathbb{P}[C_2 \cap P_1]$$
$$= \mathbb{P}[C_2 \cap C_1] + \mathbb{P}[P_2 \cap C_1]$$
$$= \mathbb{P}[C_1]$$

(c) By symmetry, since we don't know any information about the 2nd, 3rd, or 4th eggs, we have

$$\mathbb{P}[\text{5th egg} = \text{plastic} \mid 1\text{st egg} = \text{plastic}] = \mathbb{P}[\text{2nd egg} = \text{plastic} \mid 1\text{st egg} = \text{plastic}] = \frac{39}{59}.$$

Rigorously, notice that  $\mathbb{P}[C_5 \cap P_2 | P_1] = \mathbb{P}[P_5 \cap C_2 | P_1]$  and therefore:

$$\mathbb{P}[P_5 \mid P_1] = \mathbb{P}[P_5 \cap C_2 \mid P_1] + \mathbb{P}[P_5 \cap P_2 \mid P_1]$$
$$= \mathbb{P}[C_5 \cap P_2 \mid P_1] + \mathbb{P}[P_5 \cap P_2 \mid P_1]$$
$$= \mathbb{P}[P_2 \mid P_1]$$

One could also brute force this with Total Probability Rule (like in the previous part), but the calculation is quite tedious.

### 3 Balls and Bins

Note 14 Suppose you throw *n* balls into *n* labeled bins one at a time.

- (a) What is the probability that the first bin is empty?
- (b) What is the probability that the first *k* bins are empty?
- (c) Let A be the event that at least k bins are empty. Let m be the number of subsets of k bins out of the total n bins. If we assume  $A_i$  is the event that the *i*th set of k bins is empty. Then we can write A as the union of  $A_i$ 's:

$$A = \bigcup_{i=1}^{m} A_i.$$

Compute *m* in terms of *n* and *k*, and use the union bound to give an upper bound on the probability  $\mathbb{P}[A]$ .

- (d) What is the probability that the second bin is empty given that the first one is empty?
- (e) Are the events that "the first bin is empty" and "the first two bins are empty" independent?
- (f) Are the events that "the first bin is empty" and "the second bin is empty" independent?

**Solution:** Since the balls are thrown one at a time, there is an ordering, and so we are sampling with replacement where order matters rather than where it doesn't (which would correspond to each configuration in the stars and bars setup being equally likely).

(a) Note that this is a uniform sample space, with outcomes representing all possible ways to throw each ball individually into the bins. Here,  $|\Omega| = n^n$ , as each of the *n* balls has *n* possible bins to fall into, and out of these possibilities,  $(n-1)^n$  of them leave the first bin empty—each ball would then have n-1 possible bins to fall into. This gives us an overall probability  $\left(\frac{n-1}{n}\right)^n$  that the first bin is empty.

Equivalently, we can note that each throw is independent of all of the other throws. Since the probability that ball *i* does not land in the first bin is  $\frac{n-1}{n}$ , the probability that all of the balls do not land in the first bin is  $\left(\frac{n-1}{n}\right)^n$ .

(b) Similar to the previous part, we have the same uniform sample space of size  $n^n$ . Now, there are a total of  $(n-k)^n$  possible ways to throw the balls into bins such that the first k bins are empty—each ball has n-k possible bins to fall into.

Alternatively, we can similarly make use of independence. Since the probability that ball *i* does not land in the first *k* bins is  $\frac{n-k}{n}$ , the probability that all of the balls do not land in the first *k* bins is  $\left(\frac{n-k}{m}\right)^n$ .

k bins is 
$$\left(\frac{n-n}{n}\right)$$
.

(c) We use the union bound. Then

$$\mathbb{P}[A] = \mathbb{P}\left[\bigcup_{i=1}^{m} A_i\right] \le \sum_{i=1}^{m} \mathbb{P}[A_i].$$

We know the probability of the first k bins being empty from part (b), and this is true for any set of k bins, so

$$\mathbb{P}[A_i] = \left(\frac{n-k}{n}\right)^n.$$

Then,

$$\mathbb{P}[A] \le m \cdot \left(\frac{n-k}{n}\right)^n = \binom{n}{k} \left(\frac{n-k}{n}\right)^n$$

(d) Using Bayes' Rule:

$$\mathbb{P}[2nd \text{ bin empty} \mid 1st \text{ bin empty}] = \frac{\mathbb{P}[2nd \text{ bin empty} \cap 1st \text{ bin empty}]}{\mathbb{P}[1st \text{ bin empty}]}$$
$$= \frac{(n-2)^n/n^n}{(n-1)^n/n^n}$$
$$= \left(\frac{n-2}{n-1}\right)^n$$

Alternate solution: We know bin 1 is empty, so each ball that we throw can land in one of the remaining n-1 bins. We want the probability that bin 2 is empty, which means that each ball cannot land in bin 2 either, leaving n-2 bins. Thus for each ball, the probability that bin 2 is empty given that bin 1 is empty is  $\frac{n-2}{n-1}$ . For *n* total balls, this probability is  $\left(\frac{n-2}{n-1}\right)^n$ .

- (e) They are dependent. Knowing the latter means the former happens with probability 1.
- (f) In part (c) we calculated the probability that the second bin is empty given that the first bin is empty:  $\left(\frac{n-2}{n-1}\right)^n$ . The probability that the second bin is empty (without any prior information) is  $\left(\frac{n-1}{n}\right)^n$ . Since these probabilities are not equal, the events are dependent.