Probability Review – DSC 40A, Fall 2021

This optional worksheet provides you with probability problems from past exams. We will take this worksheet up in the Discussion section on Wednesday from 6-7PM. You do not have to submit this worksheet, but it's highly recommended that you work through it.

Problem 1

Taken from Winter 2020 Final, Problem 1 of the Midterm 2 Redemption (solutions)

The CDC is testing a vaccine for COVID-19 and has gathered a group of 44 people to experiment on. The people are classified by the state they are from, as shown in the table below.

State	#
Ohio	12
California	27
Texas	5

In each of the problems below you may leave your answers unsimplified, but they should not contain $\sum$ or $\ldots$.

a) Suppose that the CDC assigns each person in the study to one of two groups, treatment and control, by flipping 44 coins, one for each person. If the result of a person's coin is Heads, they are placed in the treatment group; if it is Tails, they are placed in the control group. How many different outcomes are possible?

Solution: 2^{44}

b) How many ways are there of choosing 22 people to be in the control group?

Solution: $\binom{44}{22}$

c) Suppose that each of the 44 people will be given the vaccine. How many different orders are there for giving the vaccine?

Solution: $44!$

d) Suppose that each of the 44 people will be given the vaccine, but the vaccine is first given to people in Ohio, then to people in Texas, and then to people in California. How many different orders are there for giving the vaccine?

> **Solution:** $12! \cdot 27! \cdot 5!$

e) How many different ways are there of choosing a group of ten people such that three are from Ohio, four are from California, and three are from Texas?

> **Solution:** $\binom{12}{3}\binom{27}{4}\binom{5}{3}$

f) Suppose that there are 13 vaccines to be tested, and each vaccine will be tested on only one person. How many ways are there of assigning each vaccine to a person? Assume that a person will not be given at most one vaccine.

> **Solution:** $\frac{44!}{(44-13)!} = \frac{44!}{31!}$

g) Now suppose that the number of vaccines has been narrowed down to four. Each vaccine will be tested on a different group of ten people (and no person will receive more than one vaccine). How many ways are there of choosing who will be given each vaccine?

> **Solution:** $\binom{44}{10}\binom{34}{10}\binom{24}{10}\binom{14}{10}$
>
> Choose 10 out of 44 for the first vaccine, 10 out of the remaining 34 for the second, and so on.

Problem 2

Taken from Winter 2021 Midterm 2, Problem 2 (exam) (solutions)

You work for a company that streams music and are developing a new feature called "k-list". In this feature, a user selects a number k, and is then given a playlist of k songs out of the n songs in that user's "Favorites" library. Assume $0 \le k \le n$ always. You are trying different implementations of this feature. You may leave terms with $\mathrm{C}(n,k)$ or $\mathrm{P}(n,k)$.

Below, Fabio has $n = 20$ songs in his library and has selected $k = 4$. Fabio's favorite song is "Hotel California". 8 of the songs in his library are from the 70's and 12 are from the 2000's.

a) [6 points] The playlist is created in the following way: k different songs are randomly selected from the n songs in the library. A playlist is a list of songs. Two lists of the same songs but in different order are considered two different playlists: "Happy birthday","Hotel California" is a different playlist from "Hotel California", "Happy Birthday". How many different playlists does this create for Fabio?

b) [6 points] Fabio apparently doesn't mind a song repeated in his playlist. The playlist is created in the following way: k songs are randomly selected from the n songs in the library, so that the same song can be selected more than once. A playlist is a list of songs. Two lists of the same songs but in different order are considered two different playlists: "Happy birthday","Hotel California" is a different playlist from "Hotel California", "Happy Birthday". What is the probability that a playlist will include "Hotel California"?

c) [6 points] The playlist is created in the following way: k different songs are randomly selected from the n songs in the library. A playlist is a list of songs. Two lists of the same songs but in different order are considered the same playlist. For example: playlists of 2 songs "Happy birthday","Hotel California" is the same playlist as "Hotel California", "Happy Birthday". What is the probability that a playlist will include "Hotel California"?

d) [6 points] The playlist is created in the following way: k different artists from the library are randomly selected and one song is selected from each artist. Fabio has 5 artists in his library with 4 songs each. What is the probability that a playlist will include "Hotel California (Artist: Eagles)"?

Solutions

a) [6 points] a. The playlist is created in the following way: k different songs are randomly selected from the n songs in the library. How many different playlists does this create for Fabio?

> **Solution:** This is a question about permutations: sequences (order matters) without repetitions (different songs). Therefore the number of playlists is $P(n,k) = n*(n-1)*\ldots*(n-k+1) = \frac{n!}{(n-k)!}$.

b) [6 points] Fabio apparently doesn't mind if a song repeated in his playlist. The playlist is created in the following way: k songs are randomly selected from the n songs in the library, so that the same song can be selected more than once. What is the probability that a playlist will include "Hotel California"?

> **Solution:** Since order matters, this is a question about sequences. The samples space is sequences of length k with entries in $\{1,..,n\}$. The probability that a playlist contains "Hotel California" is
>
> $$\frac{\text{number of sequences of length } k \text{ out of } n \text{ that include HC}}{\text{number of sequences of length } k \text{ out of } n}$$
>
> $$= 1 - \frac{\text{number of sequences of length } k \text{ out of } n \text{ that don't include HC}}{\text{number of sequences of length } k \text{ out of } n} = 1 - \frac{(n-1)^k}{n^k}.$$
>
> This question is based on Example 5 in Lecture 17.

Fabio wants a unique listening experience every time and doesn't want to hear the same k songs but in different order. A song collection is a set of k different songs.

Example for collection of 2 songs: "Happy birthday", "Hotel California" is the same song collection as "Hotel California", "Happy Birthday".

c) [6 points] The song collection is created by randomly selecting k different songs from the n songs in the library. What is the probability that a song collection will include "Hotel California"?

> **Solution:** Since order doesn't matter, this is a question about combinations (sets without replacement). The probability that a playlist contains "Hotel California" is
>
> $$\frac{\text{number of sets of length } k \text{ out of } n \text{ that include HC}}{\text{number of sets of length } k \text{ out of } n} = \frac{C(n-1,k-1)}{C(n,k)}$$
>
> This question is based on Example 6 in Lecture 18.

d) [6 points] The song collection is created in the following way: k different artists from the library are randomly selected and one song is selected from each artist. Fabio has 6 artists in his library with 4 songs each. What is the probability that a playlist will include "Hotel California (Artist: Eagles)"?

Solution: The probability of selecting the Eagles is $\frac{4}{6}$. The probability of selecting one song from an artist given that an artist has been selected is $\frac{1}{4}$. The probability that a playlist will "Hotel California (Artist: Eagles)" is

$$P(\text{playlist includes Hotel California}) = P(\text{HC}|\text{Eagles}) * P(\text{Eagles}) = \frac{1}{4} * \frac{2}{3} = \frac{1}{6}.$$

Note that the solution to part (d) is for a different version of the question. In the version of the question without solutions, Fabio has 5 artists in his library with 4 songs each, so the probability that a playlist includes Hotel California by the Eagles is

P(playlist includes Hotel California) = P(HC | Eagles) * P(Eagles) = (1 / 4) * (4 / 5) = 1 / 5

Problem 3

Taken from Winter 2021 Midterm 2, Problem 3 (exam) (solutions)

A sample space S is partitioned by three disjoint events $E_1 \cup E_2 \cup E_3 = S$. Let $A \subset S$ be an event and you are given the probabilities $P(E_1), P(E_2), P(A|E_2), P(A|E_3), P(A \cap E_1)$. $\bar{A}$ denotes the complement of A.

a) [6 points] True or false: the probability $P(E_1|\bar{A})$ can be calculated using the above terms (no probability terms are missing in order to make this calculation).

b) [5 points]

Multiple Choice. Choose the correct answer.

1. $P(\bar{E}_1|A) = P(E_2|A) + P(E_3|A)$
2. $P(\bar{E}_1|A) = 1 - P(\bar{E}_1|\bar{A})$
3. $P(\bar{A}|E_1) = 1 - P(\bar{E}_1|\bar{A})$
4. None of the above are correct.

c) [5 points] Multiple Choice. Choose the correct answer. For events E_1, E_2, E_3, which of the following properties is always true?

1. $P(\overline{E_1} \cap \overline{E_2}) = 0$
2. $P(\overline{E_1} \cup \overline{E_2}) \neq 0$
3. $E1$, E_2 and E_3 are independent events.
4. $P(E_1 \cup E_3) = 1$

d) [5 points] The event E_1 is fully contained within event A such that every outcome in E_1 is also an outcome in A. This is denoted by $E_1 \subseteq A$. Select all of the following properties that are always true.

1. If $E_1 \subseteq A$, then $P(E_1) \leq P(A)$.
2. If $P(E_1) \leq P(A)$, then $E_1 \subseteq A$.
3. If $E_1 \subseteq A$, then $P(E_1 \cap A) \leq P(E_2)$.
4. If $E_1 \subseteq A$, then $P(E_1 \cap A) = P(E_1)$.
5. If $E_1 \subseteq A$, then E_1 and A are independent events.
6. If $E_1 \subseteq A$, then E_3 and A are independent events.

Solutions

a) [6 points] True or false: the probability $P(E_1|\bar{A})$ can be calculated using the above terms (no probability terms are missing in order to make this calculation).

Solution: True.

$$P(E_1|\bar{A}) = \frac{P(\bar{A} \cap E_1)}{P(\bar{A})}$$

For the numerator

$$P(\bar{A} \cap E_1) = P(\bar{A}|E_1) * P(E_1)$$

which we can calculate as

$$P(\bar{A}|E_1) = 1 - P(A|E_1)$$

$$P(A|E_1) = \frac{P(A \cap E_1)}{P(E_1)}$$

Therefore $P(\bar{A} \cap E_1) = P(E_1) - P(A \cap E_1)$.

For the denominator

$$P(\bar{A}) = 1 - P(A)$$

$$P(A) = P(A|E_1)P(E_1) + P(A|E_2)P(E_2) + P(A|E_3)P(E_3)$$

which we can calculate using

$$P(E_3) = 1 - P(E_2) - P(E_1)$$

The answers in the multiple choice questions below appeared in random order in the alternate versions.

b) [5 points]

Choose the correct answer.

1. $P(\bar{E}_1|A) = P(E_2|A) + P(E_3|A)$
2. $P(\bar{E}_1|A) = 1 - P(\bar{E}_1|\bar{A})$
3. $P(\bar{A}|E_1) = 1 - P(\bar{E}_1|\bar{A})$
4. None of the above are correct.

Solution: Correct - $P(\bar{E}_1|A) = 1 - P(E_1|A) = P(E_2|A) + P(E_3|A)$.

Incorrect: $P(\bar{E}_1|A) = 1 - P(E_1|A) \neq 1 - P(\bar{E}_1|\bar{A})$

Incorrect: $P(\bar{A}|E_1) = 1 - P(A|E_1) \neq 1 - P(\bar{E}_1|\bar{A})$

c) [5 points] Choose the correct answer. For events E_1, E_2, E_3, which of the following properties is always true?

1. $P(\overline{E_1} \cap \overline{E_2}) = 0$
2. $P(\overline{E_1} \cup \overline{E_2}) \neq 0$
3. $E1$, E_2 and E_3 are independent events.
4. $P(E_1 \cup E_3) = 1$

Solution: Correct: $\overline{E_1} \cup \overline{E_2} = (E_2 \cup E_3) \cup (E_1 \cup E_3) = S \rightarrow P(\overline{E_1} \cup \overline{E_2}) = 1$.

Incorrect: $P(\overline{E_1} \cap \overline{E_2}) = P((E_2 \cup E3) \cap (E_1 \cup E_3)) = P(E_3) \neq 0$

Incorrect: $P(E_1 \cup E_3) = 1 - P(E_2) < 1$ since E_2 is non empty.

Incorrect $E1$, E_2 and E_3 are not independent since $P(E_1, E_2, E_3) = P(E_1 \cap E_2 \cap E_3) = P(\emptyset) = 0 \neq P(E_1)P(E_2)P(E_3) > 0$

d) [5 points] Choose the correct answer. The event E_1 is fully contained within event A such that every outcome in E_1 is also an outcome in A. This is denoted by $E_1 \subseteq A$.

1. If $E_1 \subseteq A$, then $P(E_1) \leq P(A)$.
2. If $P(E_1) \leq P(A)$, then $E_1 \subseteq A$.
3. If $E_1 \subseteq A$, then $P(E_1 \cap A) \leq P(E_2)$.
4. If $E_1 \subseteq A$, then $P(E_1 \cap A) = P(E_1)$.
5. If $E_1 \subseteq A$, then E_1 and A are independent events.
6. If $E_1 \subseteq A$, then E_3 and A are independent events.

Solution: Alternate versions included either 1 or 4, both are true.

Correct: $P(E_1) = \sum_{s \in E_1} p(s)$ and $P(A) = \sum_{s \in A} p(s)$. Since $\forall s \in E_1 : s \in A$, then $P(E_1) \leq P(A)$.

Correct: Since $\forall s \in E_1 : s \in A$, then $E_1 \cap A = E_1$, therefore $P(E_1 \cap A) = P(E_1)$.

Problem 4

Taken from Winter 2021 Final Part 2, Problem 2 (solutions)

A new course is being offered through DSC: DSC 178 - Data Science with Combinatorics. There is a list of $n = 172$ students who want to enroll in the course, including you and your best friend.

a) The course is capped and the professor randomly selects $k = 53$ students from the list. What is the probability you and your best friend take the course?

(Alternate version: $n = 162,\ k = 57$)

Solution: This is a questions about sets (no replacements). The number of combinations selecting k students out of n: C(n,k). The number of combinations selecting k students including the two of you out of n: C(n-2,k-2). Overall the probability is $\frac{C(n-2,k-2)}{C(n,k)} = \frac{k}{n}\frac{k-1}{n-1}$.

b) $m = 22$ of the students who want to enroll are CSE majors and the rest are DSC majors. If the professor randomly selects $k = 53$ students from the list, what is the probability that *at least* 20 CS majors are in the class?

(Alternate version: $n = 162,\ k = 57$)

Solution: S - sets of k students chosen out of n.

$$P = \frac{\#\text{ sets in S with } m-2 \text{ CS } + \#\text{ sets in S with } m-1 \text{ CS } + \#\text{ sets in S with } m \text{ CS}}{\text{number of sets in S}}$$
$$= \frac{C(m,m-2)C(n-m,k-m+2) + C(m,m-1)C(n-m,k-m+1)}{C(n,k)} +$$
$$+ \frac{C(m,m)C(n-m,k-m)}{C(n,k)}$$
$$= \frac{C(22,20)C(150,33) + C(22,21)C(150,32) + C(22,22)C(150,31)}{C(172,53)}.$$