Probability

CSC 226, Spring 2022 - Lab 2

Mahdi Hajiabadi

Outline

- Sample Space/ Event/ Probability
- Random Variable/ Probability distribution/ Expected Value
- Conditional Probability/Independence
- Discussion on randomized algorithms.

Sample Space - Event

- Sample Space: is a list of all possible outcomes of the experiment. The outcomes must be mutually exclusive and exhaustive. Mutually exclusive means they are distinct and non-overlapping. Exhaustive means complete
- Examples:
  - Tossing a coin three, the sample space is: S = {HHH, HHT, HTH, THH, HTT, TTH, THT, TTT}.
  - Tossing a die twice, the sample space is: $S = \{(i, j), i, j = 1,2,\ldots,6\}$
  - Select two marbles in succession without replacement from a bag containing 1 red, 1 blue and two green marbles.
    - S = {(r,b), (r,g), (b,r), (b,g), (g,r), (g,b), (g,g)}
  - Keep tossing a coin until one gets a head, the sample space is: $S = \{H, TH, TTh, \ldots\}$
- Event: is a subset of the sample space.
- Example:
  - Event of the second marble is blue. {(r,b), (g,b)}
  - Event of the sum of tossing a die twice is 7: {(1,6) , (5,2) , (4,3) , (3,4) , (2,5) , (1,6)}
  - Tossing a die, event of the outcome is even is: {2,4,6}

Probability of an event

- Number of ways an event can occur divided by the size of sample space, we assume that all outcomes are equally likely to happen.
    - $P(E) = \dfrac{n(E)}{n(S)}$
- Probability of a sum of 7 when two dice are rolled: $P(E) = \dfrac{n(E)}{n(S)} = \dfrac{6}{36}$
- Probability of getting at least one head if we toss a coin twice: $P(E) = \dfrac{n(E)}{n(S)} = \dfrac{3}{4}$

Exercise

- Q1: 16 persons shake hands with one another in a party. How many shake hands took place?
  - Answer: Each person should shake 15 other hands in the party, so the total number of hand shakes is $\frac{15 \times 16}{2} = 120$. It should be divided by 2 because if first person shakes the hand of second person, it is equivalent that the second person shakes first person hand.
  - Answer: There is another way to approach this problem, first person should shake 15 other hands, 2nd person should shake 14 other hands and so on. The total number is $\sum_{i=1}^{15} i = \frac{15 \times 16}{2}$
- Q2: A card is drawn from a well shuffled pack of 52 cards. What is the probability of getting queen or club card?
  - Answers: Event A is getting queen and event B is getting clubs. So we need to estimate $P(A \cup B) = P(A) + P(B) - P(A \cap B)$. We know that the size of sample space is 52. So, $P(A) = \frac{4}{52}$, $P(B) = \frac{13}{52}$, $P(A \cap B) = \frac{1}{52}$, and $P(A \cup B) = \frac{16}{52}$.
- Q3: Probability of picking a number with 2 digits and the sum of it is equal to 10?
  - Answer: Number of 2 digits (size of sample space) is 90 (from 10 to 99). Event A is the set of numbers with the sum of digits is 10. So $A = \{19,28,37,46,55,64,73,82,91\}$, and $P(A) = \frac{n(A)}{n(S)} = \frac{9}{90}$

Discrete Random Variables

- A random variable X is a function $X : \Omega \to V$, where Ω is a sample space and V is some arbitrary set (the *range* of the random variable)
- Roll a dice twice, $\Omega = \{1,\ldots,6\}^2$ for every $(a, b) \in \Omega$ we have $P(a, b) = \dfrac{1}{36}$.
- Let X is defined as the random variable that associates $a + b$, then the range of X is $\{2,3,\ldots,12\}$
- Given Ω and X as a sample space and a random variable, the event $X = v$ contains all outcomes of Ω for which $X(s) = v$
- Let X is defined as the random variable that associates $a + b$, then the range of X is $\{2,3,\ldots,12\}$, the probability of $P(X = 4) = 3/36$
- Example: Toss n coins (so the the space consists of the set of all 2^n possible coin sequences) and let X be the number of heads.
- What is $P(X = k)$ in this case?
- Answer: $\dfrac{\binom{n}{k}}{2^n}, n = 10, k = 4, \dfrac{\binom{10}{4}}{2^{10}}$

Discrete Probability Distribution

- Uniform: Every element in sample space is equally likely
  - $a \in \Omega, p(a) = \dfrac{1}{|\Omega|}$
    - Probability of the sum of rolling two dice is 4
    - If we pick one card from a deck of cards and its color is red.
- Binomial: When there are *exactly* two mutually exclusive outcomes, success/failure in each trial. The binomial distribution is used of observing k success in N trials and it is defined based on:
  - $P(k; p, N) = \binom{N}{k} p^k (1-p)^{N-k}$
- If we toss a fair coin ($p = \dfrac{1}{2}$) 10 times (N), what is the probability of getting 5 heads (k), $P(5, p = \dfrac{1}{2}, 10) = \binom{10}{5} \dfrac{1}{2}^{10}$
- What is the probability if the coin was bias with the probability of head is $p = \dfrac{2}{5}$
- Geometric: The probability of getting the first success after having a consecutive number of failures.
  - $P(X = k) = (1-p)^{k-1} p$
- A bias coin with $p = 0.4$ is repeated flipped until H shows up, $p(X=1) = 0.4, p(X=2) = 0.6 \times 0.4, p(X=3) = 0.6^2 \times 0.4 \ldots$
- A dice is repeatedly rolled until "3" shows up, so, $p(X=1) = \dfrac{1}{6}, p(X=2) = \dfrac{5}{6}\dfrac{1}{6}, p(X=3) = (\dfrac{5}{6})^2 \dfrac{1}{6}, \ldots$

Exercise

- Q1: Shuffle n distinct cards and let X be the position of jth card, what is the probability of $P(X = k)$ in this case?
  - Answer: There are n distinct cards which means there are n different positions to place them. The total number of shuffling is $n!$ (Why?). For the first position we have n different cards to put there , for the 2nd position we have $n - 1$ cards to put there and for the last position we just have 1 card.
  - Now, we put card k at jth position and in this case we have n-1 cards for the first position, $n - 2$ cards for the second position and so on, so the number of cases is $(n - 1)!$, and $P(X = k) = \frac{n!}{(n-1)!} = \frac{1}{n}$.
- Q2: Roll a dice twice, and let X be the maximum value of the two rolls, what is the probability of $P(X = 5)$?
  - Answer: Please note that X is the maximum values of the two rolls nut SUM. For example if first roll shows 3 and the 2nd roll shows 4, then the maximum is 4. Size of the sample space is 36 and there are 9 elements in the sample space which holds the condition : $\{(1,5), (5,1), (2,5), (5,2), (3,5), (5,3), (4,5), (5,4), (5,5)\}$, so $P(X = 5) = \frac{9}{36}$
- Q3: A random variable is defined by a modulo of the sum of a 2-digits number divided by 5, what is the probability of $P(X = 0)$?
  - Sum of two digits should be either 5, 10, or 15. Sum of 2 digits is 5: $\{14,23,32,41,50\}$, sum of 2 digits is 10: $\{19,28,37,46,55,64,73,82,91\}$, sum of 2 digits is 15: $\{69,78,87,96\}$. So the $P(X = 0) = \frac{5 + 9 + 4}{90} = \frac{18}{90}$

Expected value

- Given a random variable X with a finite range V, the expected value of X is defined as:
  - $E[X] = \sum_{v \in V} vP[X = v]$
- What is the expected value of tossing a dice?
- $E[X] = 1.\frac{1}{6} + 2.\frac{1}{6} + 3.\frac{1}{6} + 4.\frac{1}{6} + 5.\frac{1}{6} + 6.\frac{1}{6} = \frac{21}{6} = 3.5$
- It can be more understood in terms of betting: the winning probability (W) is $\frac{2}{3}, \frac{1}{6}$ is a probability of losing (L) and $\frac{1}{6}$ is a probability of a draw (D), and $X(L) = -2, X(D) = 0, X(W) = 1$, the expected value is:
- $E(X) = \frac{1}{6}(-2) + \frac{1}{6}(0) + \frac{2}{3}.1 = \frac{1}{3}$

Linearity of Expectation

- Let X be a random variable and a be real then:
  - $E[aX] = aE[X]$
- Let $X_1, X_2, \ldots, X_n$ be random variables over the same sample space, then;
  - $E[X_1 + X_2 + \ldots + X_n] = E[X_1] + E[X_2] + \ldots + E[X_n]$
- In general: $E[aX_1 + bX_2] = aE[X_1] + bE[X_2]$
- Example: If we flip a fair coin n times what is the expected value of number of heads?
  - Each X_i is defined as a random variable with $E[X_i] = \frac{1}{2}$
  - $E[X_1 + X_2 + \ldots + X_n] = E[X_1] + E[X_2] + \ldots + E[X_n] = \frac{n}{2}$
- Example: If we toss a dice twice what is the expected value of the sum?
  - $E[X_1] + E[X_2] = 3.5 + 3.5 = 7$

Exercise

- Q1: Find the expected value of the max of two dice?
  - Answer: The range of the random variable is from 1 to 6 and in the following the probability of each value in the range is estimated:
  - $X = 1 : \{(1,1)\}, P(X = 1) = \frac{1}{36}, X = 2 : \{(1,2), (2,1), (2,2)\}, P(X = 2) = \frac{3}{36}, X = 3 : \{(1,3), (3,1), (3,2), (2,3), (3,3)\}, P(X = 3) = \frac{5}{36}$
  - $X = 4\{(1,4), (4,1), (2,4), (4,2), (3,4), (4,3), (4,4)\}, P(X = 4) = \frac{7}{36},$
  - $X = 5 : \{(1,5), (5,1), (2,5), (5,2), (3,5), (5,3), (4,5), (5,4), (5,5)\}, P(X = 5) = \frac{9}{36}$
  - $X = 6 : \{(1,6), (6,1), (2,6), (6,2), (3,6), (6,3), (4,6), (6,4), (5,6), (6,5), (6,6)\} = \frac{11}{36},$
  - $E[X] = \sum_{v \in \{1,\ldots,6\}} vP[X = v] = \frac{1 \times 1}{36} + \frac{2 \times 3}{36} + \frac{3 \times 5}{36} + \frac{4 \times 7}{36} + \frac{5 \times 9}{36} + \frac{6 \times 11}{36} = \frac{161}{36}$
- Q2: In the Arizona lottery called Pick 3, a player pays \$1 and then picks a three-digit number. If those three numbers are picked in that specific order the person wins \$500. What is the expected value in this game?
  - If you lose you have to pay \$1 and if you win you will receive \$500. Without the loss of generality we assume that the 3 digit number should be a valid number (first digit should not start with zero). So, the number of 3 digits numbers is 900 and the winning probability is $\frac{1}{900}$ and not winning probability is $1 - \frac{1}{900} = \frac{899}{900}$. So the expected value is $-1 \times \frac{899}{900} + 500 \times \frac{1}{900} = \frac{-399}{900}$
- Q3: Your grade = # of correct answers - (1/5)(# of incorrect answers). Suppose you guess the answer to 100 questions. What is the expected grade of test? The probability of guessing right is $\frac{1}{5}$ (X(R) = 1) and the probability of guessing wrong is $\frac{4}{5}$, X(W) = $\frac{1}{5}$.
  - $E[X] = 1 \times \frac{1}{5} - \frac{1}{5} \times \frac{4}{5} = \frac{1}{25}$. Please note that this is the expected value of one question. If there are 200 questions and the total expected value would be $E[X_1 + X_2 + \ldots + X_{200}] = 200 \times \frac{4}{25} = 32$. The expected total score from the exam.

Conditional and Independence

- A conditional answers the question: "What is the chance of an event E happening, given that I have already observed some other event F?"
- I toss two coins and I tell you that at least one of those coins came up heads, what is the probability that both coins are head? (The answer is $\frac{1}{3}$)
  - Let us define $A \subset \Omega$ is an event that we already know (at least the outcome of one of those coins is head). What is the probability of $P(B|A)$?
  - $P(B|A) = \dfrac{P(B \cap A)}{P(A)}$, if $B \cap A = \phi,\ P(B|A) = 0$.
- Back to the example: $P(A) = \dfrac{n(A)}{n(S)} = \dfrac{3}{4}$, $P(B) = \dfrac{n(A)}{n(S)} = \dfrac{1}{4}$, $P(B|A) = \dfrac{1/4}{3/4} = \dfrac{1}{3}$
  - Given that you drew a red card from a deck, what's the probability that it's a four of hearts? (Answer is: $\frac{1}{26}$)

Independence

- *Two events* A and B are independent if:
  - $P(A \cap B) = P(A)\,.\,P(B)$
- If *two events* A and B are independent then:
  - $P(A \mid B) = P(A), P(B \mid A) = P(B)$
- Example: Roll a dice twice. Event A is the first dice shows 2, event B is the second dice shows 3:
  - $P(A \cap B) = \dfrac{n(A \cap B)}{n(S)} = \dfrac{1}{36}, P(A) = \dfrac{n(A)}{n(S)} = \dfrac{1}{6}, P(B) = \dfrac{1}{6}, P(A)P(B) = P(A \cap B)$
- Two random variables X, Y are independent if *for any two values* v, w the event $P(X = v), P(Y = w)$ are independent, then $E[XY] = E[X]E[Y]$.

Exercise

- Q1: A fair dice is rolled, find the probability that the number is five given that it is odd.
    - $P(B|A), P(A) = \frac{3}{6}, P(B) = \frac{1}{6}$, There are 3 odd numbers (1,3,5). Finally, $P(B|A) = \frac{\frac{1}{6}}{\frac{1}{2}} = \frac{1}{3}$
- Q2: In a sample of 902 individuals under 40 who were or had previously been married, each person was classified according to gender and age at first marriage. The results are summarized in the following two-way classification table, where the meaning of the labels is: M: male, F: female, E: a teenager when first married, W: In one's twenties when first married, H: In one's thirties when first married. Find the probability that the individual selected was a teenager at first marriage, given that the person is male?

	E	W	H	Total
M	43	293	114	450
F	82	299	71	452
Total	125	592	185	902

- $P(E|M)????\ P(M) = \frac{450}{902}, P(E) = \frac{43}{902}, P(E|M) = \frac{43}{450}$
- Q3: Are two events F: female and E: was a teenager at first marriage are independent?
    - $P(F) = \frac{452}{902}, P(E) = \frac{125}{902}, P(E \cap F) = \frac{82}{902}, P(E)P(F) \neq P(E \cap F)$, There are not independent

Discussion on Randomized Algorithms

- Advantages:
  - Avoid worst-case behavior: randomness can probabilistically guarantee the average case behaviour
  - Very practical when the size of the input is *HUGE*.
- Two types of randomized algorithm:
  - Las Vegas: (Randomized QuickSort)
    - Always gives the true answer,
    - Running time is a random variable
    - The expectation of running time is bounded (say by a polynomial)
  - Monte Carlo (Zero Polynomial problem)
    - It may produce incorrect answer
    - We are able to bound the probability of error
    - By running it many times on independent random variables, we can make the failure probability arbitrary small at the expense of running time.

Monte Carlo Example

- Suppose we want to decide whether the following equation is zero polynomial or not;
- $(x-1)(x-3)(2x+4)(x^2-5)+(2x-1)(11x-10)(5x^3+7)(3x+2)+\ldots+(x-1)(x+1)$
- It's obvious that the deterministic algorithm would expand the equation and compute the sum of it. If it is zero then it is 0 polynomial, otherwise it is not.
- n (length of the expanded equation) may be large. Suppose n is 10^{15}, if we assume 10^9 operations takes 1 second, it takes around 12 days to output result.
- Choose a uniform random numbers as the value of x, and find the final value of equation. If it is 0, then it is 0 polynomial equation, otherwise it is not.
- Error: If the random number is the root of the equation and the equation is not zero polynomial.

Las Vegas Example

- Randomized QuickSort algorithm
  - We choose a pivot randomly in each step.
  - The probability of choosing a good pivot is 1/2.
  - The expected running time is in $O(nlogn)$.

- Thank you!!! 😊
- Any questions?