

1. Introduction

1.1. Probability theory is a mathematical modeling of the phenomenon of chance or randomness.

2. Sample Space and Events

2.1. “Sample space” is the set of all possible outcomes of a given experiment. Notice that whether order matters, i.e. combinations versus permutations, affects the size of the sample space.

2.1.1. Examples: Dealing a five-card straight poker hand from a 52-card deck has a sample space of $\binom{52}{5} = \frac{52!}{5!(52-5)!}$, but drawing a 5-card straight, in order, has a larger sample space of $P(52, 5) = \frac{52!}{(52-5)!}$ and drawing a 5-card straight when each card is replaced in the deck each time has an even larger sample space of 52^5 .

2.2. An “event” is a subset of the sample space.

2.3. The “probability” of an event is $0 \leq \frac{n(E)}{n(S)} \leq 1$, where E is the event set, and S is the sample space.

2.3.1. $P(S) = 1$, $P(\emptyset) = 0$.

2.4. $P(A^C) = 1 - P(A)$.

2.5. $P(A \cup B)$ is the probability that either event A or event B occurs, or both.

2.6. $P(A \cap B)$ is the probability that both event A and event B occur.

2.7. Since an event is a subset, the set operators can have meaning in terms of probability experiments, but the relationship between the experiments must be clearly understood. Assume that we have sample spaces S_1 and S_2 , and events $A \subseteq S_1, B \subseteq S_2$. We must deal with two cases.

2.8. Case 1: $S_1 = S_2$. This means that the two experiments are taking place simultaneously in the same sample space.

2.8.1. This is the easiest case, and the one referred to 7.2 of the text.

2.8.2. $P(A \cup B) = \frac{n(A) + n(B) - n(A \cap B)}{n(S_1)} = P(A) + P(B) - P(A \cap B)$ by the sum rule and inclusion/exclusion principle.

2.8.2.1. Example 1: What is the probability of drawing one card from a deck, and have it be either a heart or an ace from a deck, or both? Let A be the set of cards that are hearts, and let B be the set of cards that are aces.

$$P(A \cup B) = \frac{n(A) + n(B) - n(A \cap B)}{n(S_1)} = \frac{13 + 4 - 1}{52} \cong .31.$$

2.8.3. $P(A \cap B) = \frac{n(A \cap B)}{n(S_1)}$.

2.8.3.1. Example 2: What is the probability of drawing one card from a deck, and have it be both a heart and a face card (K, Q, J) from a deck? Let A be the set of cards that are hearts, and let B be the set of cards that are face cards.

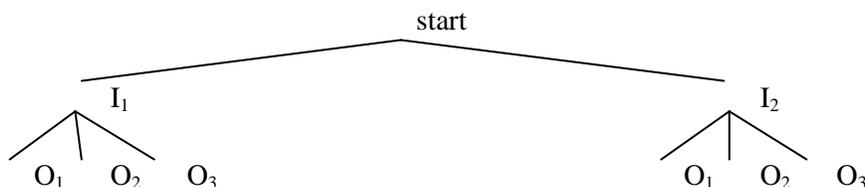
$$P(A \cap B) = \frac{n(A \cap B)}{n(S_1)} = \frac{3}{52} \cong 0.06 = \frac{13 \cdot 12}{52^2}?$$

2.8.3.1.1. Note that the example makes it look like the Multiplication Rule can always be applied. The rule can only be applied to independent events. In this case, whether a card is a heart is independent of whether it is a face card. However if we use the same property in both events, then Multiplication Rule will not work because they are not independent events. For example, let A be the set of hearts, and B be the set of spades, then $P(A \cap B) = \frac{n(A \cap B)}{n(S_1)} = \frac{0}{52} = 0 \neq \frac{13 \cdot 13}{52^2}$

2.9. Case 2: $S_1 \cap S_2 = \emptyset$. Therefore $A \cap B = \emptyset$, i.e. they are disjoint sets, and the two experiments must be independent of each other.

2.9.1. A possibility tree can help keep track the possible outcomes by ordering them.

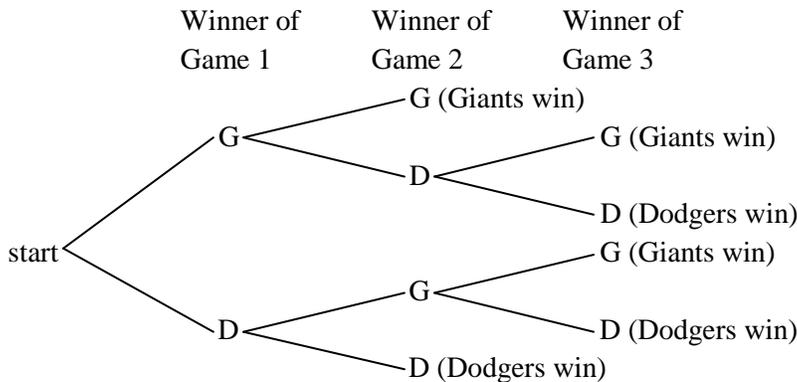
2.9.1.1. Example 3: A network router has two input lines, and three output lines how many pairs of input and output lines are there?



2.9.2. $P(A \cap B) = P(A) * P(B) = \frac{n(A)}{n(S_1)} * \frac{n(B)}{n(S_2)} = \frac{n(A)*n(B)}{n(S_1)*n(S_2)}$ which relies on the Product Rule.

2.9.2.1. Assuming each outcome is equally likely in Example 3, let A be the event that the message arrives on I_2 , and let B be the event that the message leaves on O_2 or O_3 . $P(A \text{ and } B) = P(A) * P(B) = \frac{n(A)}{n(S_1)} * \frac{n(B)}{n(S_2)} = \frac{1}{2} * \frac{2}{3} = \frac{1}{3} \cong 0.33$. Which matches the possibility tree.

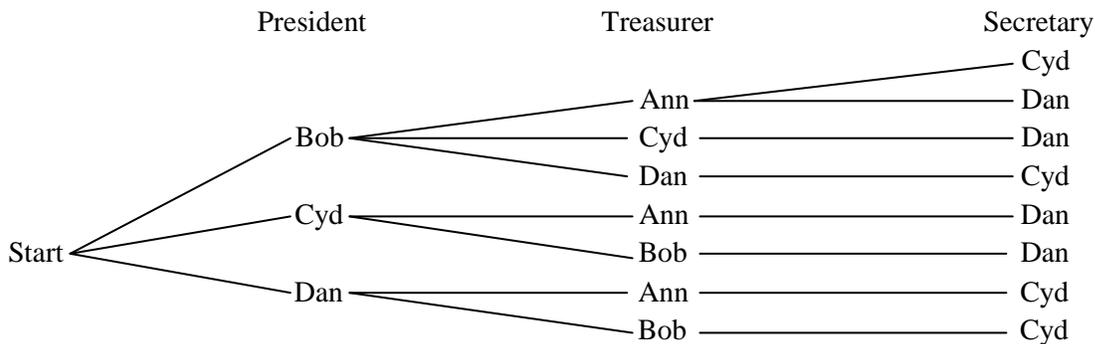
Example of problem not susceptible to the Product Rule: Two teams, the (G)iants, and the (D)odgers are in a best of 3 playoff. Such a playoff ends when one of the teams wins two games. How many ways can the playoff be played? Assuming all outcomes are equally likely, what is the probability that all three games are needed?



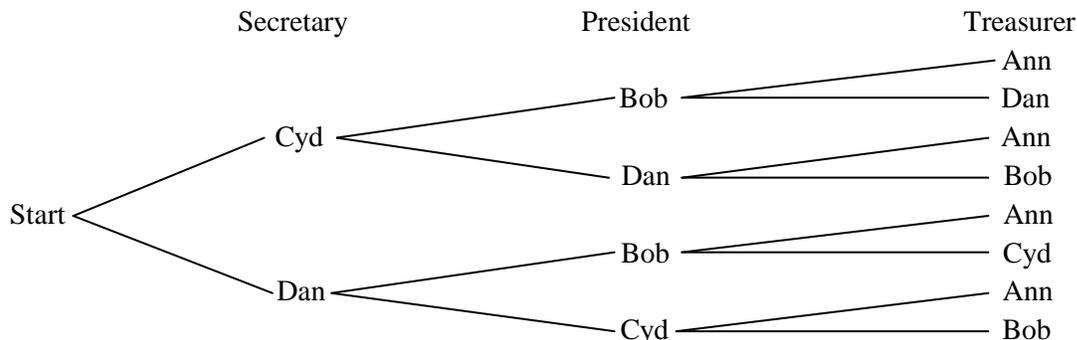
How many ways can the playoff be played? **6**

The probability that all three games are needed? $\frac{4}{6} \cong 67\%$

2.9.2.2. Example of problem that can be made susceptible to the Product Rule: Three officers—a president, a treasurer, and a secretary—are to be chosen from among four people: Ann, Bob, Cyd, and Dan. suppose that, for various reasons, Ann cannot be president, and either Cyd or Dan must be secretary. How many ways can the officers be chosen? Product Rule would say $3 * 4 * 2 = 18$, but this is wrong. If all possible outcomes are equally likely, what is the probability that Cyd will be president?



Product Rule not possible to represent this possibility tree. Reorder the steps.



Product Rule here is $2 * 2 * 2 = 8$. Cyd as president is $2/8 = .25$

2.9.3. $P(A \cup B) = 1 - P(A^C \text{ and } B^C) = 1 - P(A^C) * P(B^C) = 1 - (1 - P(A)) * (1 - P(B))$

2.9.3.1. Note that you cannot use the Addition Rule because these are not the same experiment. For example, let $P(A)$ = probability of throwing a head on first flip, 0.5, and $P(B)$ = probability of throwing a head on the second flip, 0.5, then the Addition Rule would say that $P(A \cup B) = 0.5 + 0.5 = 1$, which is clearly not true. $P(A \cup B) = 1 - (0.5 * 0.5) = .75$.

3. Finite Probability Spaces

3.1. Definition: Let S be a finite sample space, say $S = \{a_1, a_2, \dots, a_n\}$. A “finite probability space” is obtained by assigning to each point a_i in S a real number p_i , called the “probability of a_i ”, satisfying the following two properties:

3.1.1. For every p_i , $0 \leq p_i \leq 1$.

3.1.2. $\sum_{i=1}^n p_i = P(S) = 1$

3.2. If $A \subseteq B$, then $P(A) \leq P(B)$

3.3. For an event A , $P(A) = \sum p_i$ for all i where $a_i \in A$.

3.4. Example: What is the probability that the total on two dice will be 7 or 11?

3.4.1. Solution: $S = \{2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12\}$ with the corresponding $p_i =$

$\{\frac{1}{36}, \frac{2}{36}, \frac{3}{36}, \frac{4}{36}, \frac{5}{36}, \frac{6}{36}, \frac{5}{36}, \frac{4}{36}, \frac{3}{36}, \frac{2}{36}, \frac{1}{36}\}$. So $P(7 \text{ or } 11) = P(7) + P(11) = \frac{6}{36} + \frac{2}{36} = \frac{8}{36} \cong 0.22$

3.5. Definition: A finite probability space where each sample point has the same probability is called an “equiprobable space.” The expression “at random” usually indicates an equiprobable space.

4. Conditional Probability

4.1. Definition: Let $P(B) > 0$, then the probability that A occurs once B has occurred, called the “conditional probability of A given B ,” is written as $P(A|B)$ is defined as: $P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{n(A \cap B)}{n(B)}$

4.2. $P(A \cap B) = P(A | B) * P(B) = P(B \cap A)$.

4.2.1. Generalized to $P(A_1 \cap A_2 \cap \dots \cap A_n) = P(A_1) * P(A_2 | A_1) * P(A_3 | A_1 \cap A_2) * \dots * P(A_n | A_1 \cap A_2 \cap \dots \cap A_{n-1})$

4.3. Example 5: An urn contains 5 blue and 7 gold balls. Let us say that 2 are chosen at random, one after the other, without replacement.

4.3.1. What is the probability that both balls are blue?

4.3.1.1. Solution: The probability that the first ball is blue is $\frac{5}{12}$. If the first ball is blue, then there are only 4 blue balls left, so the probability the second ball would be blue is $\frac{4}{11}$. $P(A \cap B) = P(A | B) * P(B) = \frac{4}{11} * \frac{5}{12} = \frac{20}{132} \cong 0.15$.

4.3.2. What is the probability that the first ball is gold and the second ball is blue?

4.3.2.1. Solution: The probability that the first ball is gold is $\frac{7}{12}$. If the first ball is gold, then there are only 5 blue balls left, so the probability the second ball would be blue is $\frac{5}{11}$. $P(A \cap B) = P(A | B) * P(B) = \frac{5}{11} * \frac{7}{12} = \frac{35}{132} \cong 0.27$.

4.3.3. What is the probability that the second ball is blue?

4.3.3.1. Solution: This event is composed of two disjoint sets: 1) The first ball is blue, and the second is blue (4.3.1.1), and 2) the first ball is gold, and second is blue (4.3.2.1). $P(A \cup B) = P(A) + P(B) - P(A \cap B) = \frac{20}{132} + \frac{35}{132} = \frac{55}{132} = \frac{5}{12} \cong 0.68$

5. Independent Events

5.1. Definition: Events A and B in a probability space S are said to be “independent” if the occurrence of one of them does not influence the occurrence of the other. If A and B are “independent,” then $P(A | B) = P(A)$, which from the definition of conditional probability implies that $P(A \cap B) = P(A) * P(B)$, $P(B | A) = P(B)$.

5.2. Example 6: An urn contains 5 blue and 7 gold balls. Let us say that 2 are chosen at random, one after the other, with replacement.

5.2.1. What is the probability that both balls are blue?

5.2.1.1. *Solution:* The probability that the first ball is blue is $\frac{5}{12}$. Since the first ball is placed back into the urn

before drawing the second ball, the probability the second ball would be blue is also $\frac{5}{12}$. $P(A \cap B) = P(A) * P(B) = \frac{5}{12} * \frac{5}{12} = \frac{25}{144} \cong 0.17$

6. Independent Repeated Trials, Binomial Distribution.

6.1. Definition: Let s_i be the result of the i -th trial in the finite probability space S , then the n trials are “independent repeated trials” if $P(s_1, s_2, \dots, s_n) = P(s_1) * P(s_2) * \dots * P(s_n)$.

6.2. Definition: A “binomial experiment” is a series of n independent repeated trials where there are only two outcomes, with p the probability of success, and $q = 1 - p$ the probability of failure, expressed as $B(n, p)$

6.3. The probability of exactly k successes in $B(n, p)$ is $P(k) = \binom{n}{k} p^k q^{n-k} = \frac{n! p^k q^{n-k}}{k!(n-k)!}$

6.4. Example: Suppose a die is tossed 5 times. What is the probability of getting exactly 2 fours?

6.4.1. *Solution 1:* This is a binomial experiment in which the number of trials is equal to 5, the number of successes is equal to 2, and the probability of success on a single trial is $1/6$ or about 0.167. Therefore, the binomial probability with $B(5, 0.167)$ is $P(2) = \binom{5}{2} * (0.167)^2 * (0.833)^3 = 0.161$

6.4.2. *Solution 2:* Using simple counting techniques, the sample space has $6^5 = 7776$ possible ways to throw the die. There are $\binom{5}{2} = 10$ locations for the 4's, and the other three positions are free to be any of the other five numbers, so 5^3 . This gives us $10 * 5^3 = 1250$ ways to have exactly two 4's. $\frac{1250}{7776} = 0.161$. Note that we just did the binomial calculation in a different order, $(0.167)^2 * (0.833)^3 = \left(\frac{1}{6}\right)^2 * \left(\frac{5}{6}\right)^3 = \frac{5^3}{6^5}$

7. Random Variables

7.1. Definition: A “random variable” X is a function that assigns a numerical value to each outcome in a sample space. Random variables can be discrete, e.g. a count of the heads in six flipped coins, or continuous, e.g. the temperature during a day.

7.1.1. Let $s \in S$, then $P(x = a) \equiv P(\{s \mid X(s) = a\})$, and $P(a \leq X \leq b) \equiv P(\{s \mid a \leq X(s) \leq b\})$

7.2. Suppose a discrete random variable X can only take on the values x_1, x_2, \dots, x_n , then $\sum_{i=1}^n P(x = x_i) = 1$, and the “mean or expectation of X ” is $\mu = E(X) = \sum_{i=1}^n x_i P(x = x_i)$

7.3. Recording all these probabilities of output ranges of a real-valued random variable X yields the probability distribution of X , often named f for frequency. The probability distribution, “forgets” about the particular sample space used to define X and only records the probabilities of various values of X .

7.4. Let X be a random variable with mean μ and distribution f , then the “variance” of X is

$$Var(X) = \sigma^2 = (x_1 - \mu)^2 f(x_1) + (x_2 - \mu)^2 f(x_2) + \dots + (x_n - \mu)^2 f(x_n) = \sum_{i=1}^n (x_i - \mu)^2 f(x_i) = \sum_{i=1}^n x_i^2 f(x_i) - \mu^2$$

7.5. The “standard deviation”, σ , is the square root of the variance, and is very useful measure of the variation in X . For the normal distribution, and the binomial distributions with large n , ~68% of the values fall within one standard deviation of the mean, ~93% fall within two standard deviations of the mean, and more than 99% fall within three standard deviations.

7.6. The binomial distribution is the most useful distribution for discrete events. The number of k successes is a random variable, with $P(k) = \binom{n}{k} p^k q^{n-k}$. For a general binomial distribution $B(n, p)$,

7.6.1. Expect value $E(X) = \mu = np$.

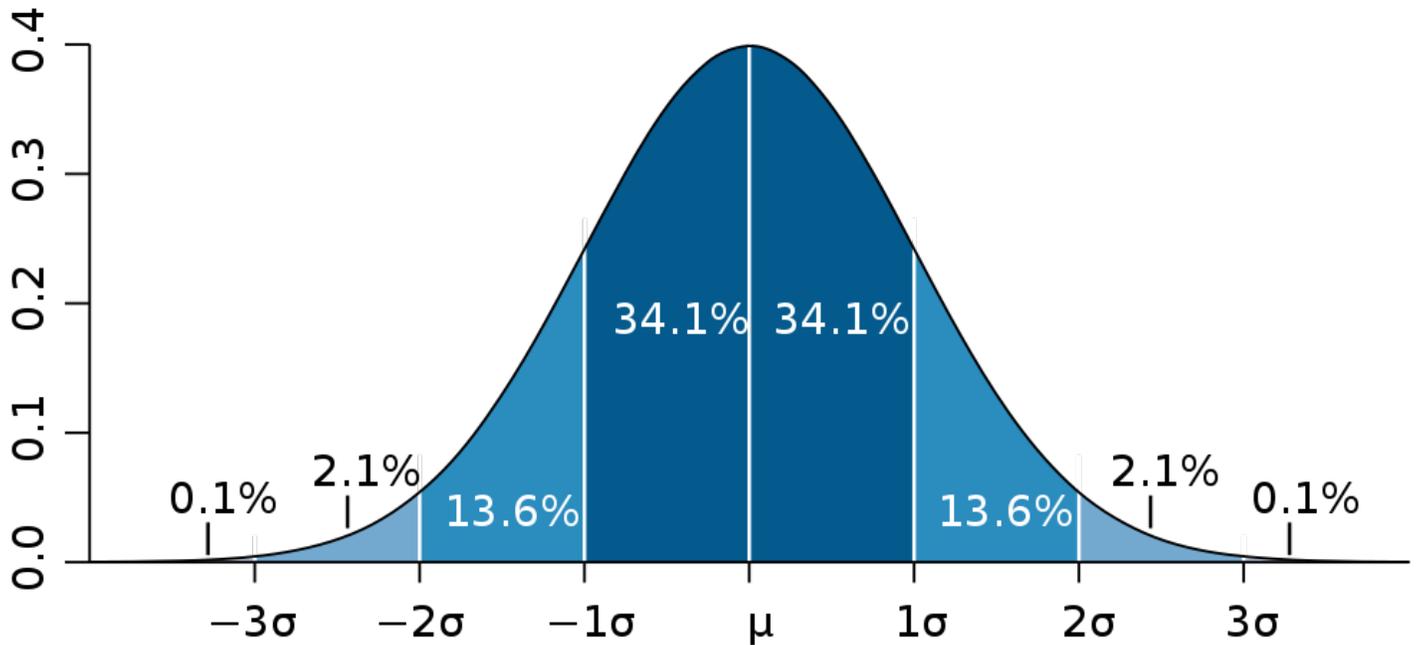
7.6.2. Variance $Var(X) = \sigma^2 = npq$

7.6.3. Standard deviation $\sigma = \sqrt{npq}$

7.6.4. Example: Assuming that the probability of having a blue eyed grandchild is $\frac{1}{4}$ for two brown eyed grandparents. What is the probability that exactly three of their twelve grandchildren will have blue eyes? What is the probability that no more than three of their grandchildren will have blue eyes?

7.6.4.1. *Solutions:* To determine the probability of an event of a single value, we should use the binomial

equation directly. $P(k) = \binom{n}{k} p^k q^{n-k}$. For this problem $p = \frac{1}{4}$, $q = \frac{3}{4}$, $n = 12$, and $k = 3$, so $P(3) =$
 $\binom{12}{3} \left(\frac{1}{4}\right)^3 \left(\frac{3}{4}\right)^9 = \frac{12 \cdot 11 \cdot 10}{3!} * \frac{3^9}{4^{12}} = \frac{11 \cdot 5 \cdot 3^9}{4^{11}} =$



7.7. The normal (or Gaussian) distribution is the most useful distribution for continuous variables. It has a very ugly formula for which you are not responsible: $f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$ This is the classic bell shaped distribution of natural phenomena.

7.7.1. Example #1: The distribution of heights of American women aged 18 to 24 is approximately normally distributed with mean 65.5 inches and standard deviation 2.5 inches. What percentage of such women are less than 5' 3"?

7.7.1.1. *Solution:* 5' 3" = 63" is 2.5 inches = 1σ less than the mean. The area to the left of $-1\sigma = 13.6\% + 2.1\% + 0.1\% = 15.8\%$

7.7.2. Example #2: When modern IQ tests are constructed, the mean (average) score within an age group is set to 100 and the standard deviation to 15. UCD is suppose to take the 10% of California's students. If a genius is defined as an IQ of 145, then what percentage of the UCD population are geniuses?

7.7.2.1. *Solution:* 145 is $3\sigma = 0.1\%$ of the general population, and thus 1% of the UCD population.

7.7.3. Example #3: the SAT has a mean of 500, and a standard deviation of 100. What percentage of test takers have at least a 400 on the SAT?

7.7.3.1. *Solution:* Since 400 is -1σ from the mean, $50\% + 13.6\% = 63.6\%$

8. Chebyshev's Inequality, Law of Large Numbers (skipped)