Discrete Probability

Probability of an Event

- Pierre-Simon Laplace's classical theory of probability:
- Definition of terms:
    - An *experiment* is a procedure that yields one of a given set of possible outcomes.
    - The *sample space* of the experiment is the set of possible outcomes.
    - An *event* is a subset of the sample space.
- The probability of an event:
  If S is a finite sample space of equally likely outcomes, and E is an event, that is, a subset of S, then the *probability* of E is $p(E) = |E|/|S|$
- For every event E, we have $0 \leq p(E) \leq 1$. This follows directly from the definition because $0 \leq p(E) = |E|/|S| \leq |S|/|S| \leq 1$, since $0 \leq |E| \leq |S|$.

Applying Laplace's Definition

- **Example:** An urn contains four blue balls and five red balls. What is the probability that a ball chosen from the urn is blue?
- **Solution:** The probability that the ball is chosen is $4/9$ since there are nine possible outcomes, and four of these produce a blue ball.
- **Example:** What is the probability that when two dice are rolled, the sum of the numbers on the dice is 7?
- **Solution:** By the product rule, there are $6^2 = 36$ possible outcomes. Six of these sum to 7. Hence, the probability of obtaining a 7 is $6/36 = 1/6$.

Applying Laplace's Definition

- **Example:** In a lottery, a player wins a large prize when they pick four digits that match, in correct order, four digits selected by a random mechanical process. What is the probability that a player wins the prize?
- **Solution:** By the product rule, there are $10^4 = 10,000$ ways to pick four digits. Since there is only one way to pick the correct digits, the probability of winning the large prize is $1/10,000 = 0.0001$.

Applying Laplace's Definition

- **Example:** In a lottery, a player wins a small prize when they pick three digits that match from four digits selected by a random mechanical process. What is the probability that a player wins the prize?
- **Solution:** If exactly three digits are matched, one of the four digits must be incorrect and the other three digits must be correct. For the digit that is incorrect, there are 9 possible choices. Hence, by the sum rule there are a total of 36 possible ways to choose four digits that match exactly three of the winning four digits. The probability of winning the small prize is $36/10,000 = 0.0036$

Applying Laplace's Definition

- **Example:** There are many lotteries that award prizes to people who correctly choose a set of six numbers out of the first n positive integers, where n is usually between 30 and 60. What is the probability that a person picks the correct six numbers out of 40?
- **Solution:** The number of ways to choose six numbers out of 40 is $C(40,6) = 3,838,380$. Hence, the probability of picking a winning combination is $1/3,838,380 \approx 0.00000026$.

The Probability of Complements and Unions of Events

- **Theorem 1:** Let E be an event in sample space S. The probability of the event $\overline{E} = S - E$, the complementary event of E, is given by
$$p(\overline{E}) = 1 - p(E)$$

- **Proof:** Using the fact that $|\overline{E}| = |S| - |E|$,
$$p(\overline{E}) = \frac{|S| - |E|}{|S|} = 1 - \frac{|E|}{|S|} = 1 - p(E)$$

The Probability of Complements and Unions of Events

- **Example:** A sequence of 10 bits is chosen randomly. What is the probability that at least one of these bits is 0?
- **Solution:** Let E be the event that at least one of the 10 bits is 0. Then $\overline{E}$ is the event that all of the bits are 1. The size of the sample space S is 2^{10}. Hence,

$$p(E) = 1 - p(\overline{E}) = 1 - \frac{|\overline{E}|}{|S|} = 1 - \frac{1}{1024} = \frac{1023}{1024}$$

The Probability of Complements and Unions of Events

- **Theorem 2:** Let E_1 and E_2 be events in sample space S. Then

$$p(E_1 \cup E_2) = p(E_1) + p(E_2) - p(E_1 \cap E_2)$$

- **Proof:** Using the inclusion-exclusion formula, it follows that

$$\begin{aligned}
p(E_1 \cup E_2) &= \frac{|E_1 \cup E_2|}{|S|} \\
&= \frac{|E_1| + |E_2| - |E_1 \cap E_2|}{|S|} \\
&= \frac{|E_1|}{|S|} + \frac{|E_2|}{|S|} - \frac{|E_1 \cap E_2|}{|S|} \\
&= p(E_1) + p(E_2) - p(E_1 \cap E_2)
\end{aligned}$$

The Probability of Complements and Unions of Events

- **Example:** What is the probability that a positive integer selected at random from the set of positive integers not exceeding 100 is divisible by either 2 or 5?
- **Solution:** Let E_1 be the event that the integer is divisible by 2 and E_2 be the event that it is divisible by 5. Then the event that the integer is divisible by 2 or 5 is $E_1 \cup E_2$ and $E_1 \cap E_2$ is the event that is is divisible by 2 and 5. It follows that

$$\begin{aligned} p(E_1 \cup E_2) &= p(E_1) + p(E_2) - p(E_1 \cap E_2) \\ &= \frac{50}{100} + \frac{20}{100} - \frac{10}{100} = \frac{3}{5} \end{aligned}$$

Monty Hall Puzzle

- **Example:** You are asked to select one of the three doors to open. There is a large prize behind one of the doors and if you select that door, you win the prize. After you select a door, the game show host opens one of the non-winning doors and gives you the opportunity to switch your selection. Should you switch?
- **Solution:** You should switch. The probability that your initial pick is correct is $1/3$. This is the same whether or not you switch doors. But, the game show host always opens a door that does not have the prize, so if you switch, the probability of winning will be $2/3$ because you win if your initial pick was not the correct door and the probability that your initial pick was wrong is $2/3$.

Assigning Probabilities

- Laplace's definition assumes that outcomes are equally likely, but we can introduce a more general definition of probability that avoids this restriction.
- Let S be a sample space of an experiment with a finite number of outcomes. We assign a probablility $p(s)$ to each outcome s so that

$$\begin{aligned} i. &\quad 0 \leq p(s) \leq 1, \ \forall s \in S \\ ii. &\quad \sum_{s \in S} p(s) = 1 \end{aligned}$$

- The function p from the set of all outcomes in the sample space S is called a *probablility distribution*.

Assigning Probabilities

- **Example:** What probabilities should we assign to the outcomes H(heads) and T(tails) when a fair coin is flipped?
- **Solution:** For a fair coin we have $p(H) = p(T) = 1/2$
- **Example:** What probabilities should be assigned to these outcomes when the coin is biased so that heads comes up twice as often as tails?
- **Solution:** For a biased coin we have $p(H) = 2p(T)$. Because $p(H) + p(T) = 1$, it follows that $2p(T) + p(T) = 3p(T) = 1$. Hence, $p(T) = 1/3$ and $p(H) = 2/3$.

Uniform Distribution

- **Definition:** Suppose that S is a set with n elements. The *uniform distribution* assigns the probability $1/n$ to each element of S.
- **Example:** Consider again the coin flipping example, but with a fair coin. Now $p(H) = p(T) = 1/2$.

Probability of an Event

- **Definition:** The probability of an event E is the sum of the probabilities of the outcomes in E.

$$p(E) = \sum_{s \in E} p(s)$$

- Note that now no assumption is being made about the distribution.

Example

- **Example:** Suppose that a die is biased so that 3 appears twice as often as each other number, but that the other five outcomes are equally likely. What is the probability that an odd number appears when we roll this die?
- **Solution:** We want the probability of the event $E = \{1, 3, 5\}$. We have $p(3) = 2/7$ and $p(1) = p(2) = p(4) = p(5) = p(6) = 1/7$. Hence, $p(E) = p(1) + p(3) + p(5) = 1/7 + 2/7 + 1/7 = 4/7$.

Probabilities of Complements and Unions of Events

- **Complements:** $p(\overline{E}) = 1 - p(E)$ still holds. Since each outcome is either E or $\overline{E}$, but not both,

$$\sum_{s \in S} p(s) = 1 = p(E) + p(\overline{E}).$$

- **Unions:** $p(E_1 \cup E_2) = p(E_1) + p(E_2) - p(E_1 \cap E_2)$ also still holds under the new definition.

Combinations of Events

- **Theorem:** If $E_1, E_2, \ldots$ is a sequence of pairwise disjoint events in a sample space S, then

$$p\left(\bigcup_i E_i\right) = \sum_i p(E_i)$$

Conditional Probability

- **Definition:** Let E and F be events with $p(F) > 0$. The conditional probability of E given F, denoted by $p(E \mid F)$,

$$p(E \mid F) = \frac{p(E \cap F)}{p(F)}$$

Conditional Probability Example

- **Example:** A bit string of length four is generated at random so that each of the 16 bit strings of length 4 is equally likely. What is the probablility that it contains at least two consecutive 0s, given that its first bit is 0?
- **Solution:** Let E be the event that the bit string contains at least two consecutive 0s, and F be the event that the first bit is 0. Since $E \cap F = \{0000, 0001, 0010, 0011, 0100\}$, $p(E \cap F) = 5/16$. Also because 8 bit strings of length eight start with a 0, $p(F) = 8/16 = 1/2$. Hence,

$$p(E \mid F) = \frac{p(E \cap F)}{p(F)} = \frac{5/16}{1/2} = \frac{5}{8}.$$

Conditional Probability Example

- **Example:** What is the conditional probability that a family with two children has two boys, given that they have at least one boy? Assume that each of the possibilities BB, BG, GB, and GG is equally likely where B represents a boy and G represents a girl.

- **Solution:** Let E be the event that the family has two boys and let F be the event that the family has at least one boy. Then $E = \{BB\}$, $F = \{BB, BG, GB\}$, and $E \cap F = \{BB\}$. It follows that $p(F) = 3/4$ and $p(E \cap F) = 1/4$. Hence,

$$p(E \mid F) = \frac{p(E \cap F)}{p(F)} = \frac{1/4}{3/4} = \frac{1}{3}.$$

Independence

- **Definition:** The events E and F are independent if and only if $p(E \cap F) = p(E)p(F)$.

Independence Example

- **Example:** Suppose E is the event that a randomly generated bit string of length four begins with a 1 and F is the event that this bit string contains an even number of 1s. Are E and F independent if the 16 bit strings of length four are equally likely?
- **Solution:** There are eight bit strings of length four that begin with a 1, and eight bit strings of length four that contain an even number of 1s. Since the number of bit strings of length 4 is 16, $p(E) = p(F) = 8/16 = 1/2$. Also since $E \cap F = \{1111, 1100, 1010, 1001\}$, $p(E \cap F) = 4/16 = 1/4$. We conclude that E and F are independent because $p(E \cap F) = 1/4 = (1/2)(1/2) = p(E)p(F)$.

Independence Example

- **Example:** Assume (as in the previous example) that each of the four ways a family can have two children (BB, GG, BG, GB) is equally likely. Are the events E, that a family with two children has two boys, and F, that a family with two children has at least one boy, independent?
- **Solution:** Because $E = \{BB\}, p(E) = 1/4$. We saw previously that $p(F) = 3/4$ and $p(E \cap F) = 1/4$. The events E and F are not independent since $p(E)p(F) = 3/16 \neq 1/4 = p(E \cap F)$.

Pairwise and Mutual Independence

- **Definition:** The events $E_1, E_2, \ldots, E_n$ are *pairwise independent* if and only if $p(E_i \cap E_j) = p(E_i)p(E_j)$ for all pairs i and j with $i \leq j \leq n$.
  The events are *mutually independent* if $p(E_{i_1} \cap E_{i_2} \cap \cdots \cap E_{i_m}) = p(E_{i_1})p(E_{i_1}) \cdots p(E_{i_m})$ whenever $i_j, j = 1, 2, \ldots, m$, are integers with $1 \leq i_1 < i_2 < \cdots < i_m \leq n$ and $m \geq 2$.

Bernoulli Trials

- **Definition:** Suppose an experiment can have only two possible outcomes, for example, the flipping of a coin.
  - Each performance of the experiment is called a *Bernoulli trial*.
  - One outcome is called a *success* and the other a *failure*.
  - If p is the probability of success and q the probability of failure, then $p + q = 1$.
- Many problems involve determining the probability of k successes when an experiment consists of n mutually independent Bernoulli trials.

Bernoulli Trials

- **Example:** A coin is biased so that the probability of heads is 2/3. What is the probability that exactly fours heads occur when the coin is flipped seven times?
- **Solution:** There are $2^7 = 128$ possible outcomes. The number of ways four of the seven flips can be heads is $C(7,4)$. The probability of each of the outcomes is $(2/3)^4(1/3)^3$ since the seven flips are independent. Hence, the probability that exactly four heads occur is
$C(7,4)(2/3)^4(1/3)^3 = (35 \cdot 16)/2^7 = 560/2187$.

Probability of k Successes in n Independent Bernoulli Trials

- **Theorem:** The probability of exactly k successes in n independent Bernoulli trials, with probability of success p and probability of failure $q = 1 - p$, is $C(n, k)p^k q^{n-k}$.

Bayes' Theorem

- **Bayes' Theorem:** Suppose that E and F are events from a sample space S such that $p(E) \neq 0$ and $p(F) \neq 0$. Then:

$$p(F \mid E) = \frac{p(E|F)p(F)}{p(E|F)p(F) + p(E \mid \overline{F})p(\overline{F})}$$

Bayes' Theorem

- **Example:** We have two boxes. The first box contains two green balls and seven red balls. The second contains four green balls and three red balls. Bob selects one of the boxes at random. Then he selects a ball from that box at random. If he has a red ball, what is the probability that he selected a ball from the first box?
- **Solution:** Let E be the event that Bob has chosen a red ball and F be the event that Bob has chosen from the first box. By Bayes' theorem, the probability that Bob has picked the first box is

$$p(F \mid E) = \frac{(7/9)(1/2)}{(7/9)(1/2) + (3/7)(1/2)} = \frac{7/18}{38/63} = \frac{49}{76}$$

Applying Bayes' Theorem

- **Example:** Suppose that one person in 100,000 has a particular disease. There is a test for the disease that gives a positive result 99% of the time when given to someone with the disease. When given to someone without the disease, 99.5% of the time it gives a negative result Find:
  - the probability that a person who tests positive has the disease.
  - the probability that a person who tests negative does not have the disease.

Applying Bayes' Theorem

- **Solution:** Let D be the event that the person has the disease, and E be the event that this person tests positive. We need to compute $p(D \mid E)$ from $p(D)$, $p(E \mid \overline{D})$ and $p(\overline{D})$.

$$p(D) = 1/100,000 = 0.000001$$
$$p(\overline{D}) = 1 - 0.00001 = 0.99999$$
$$p(E \mid D) = 0.99$$
$$p(E \mid \overline{D}) = 0.005$$
$$p(D \mid E) = \frac{(0.99)(0.000001)}{(0.99)(0.000001) + (0.005)(0.99999)} \approx 0.002$$

Applying Bayes' Theorem

- **Solution:** If the test result is negative:

$$p(D) = 1/100,000 = 0.000001$$

$$p(\overline{D}) = 1 - 0.00001 = 0.99999$$

$$p(E \mid D) = 0.99$$

$$p(\overline{E} \mid \overline{D}) = 0.995$$

$$p(\overline{D} \mid \overline{E}) = \frac{(0.995)(0.99999)}{(0.995)(0.99999) + (0.01)(0.00001)} \approx 0.99999999$$

Generalized Bayes' Theorem

- **Generalized Bayes' Theorem:** Suppose that E is an event from a sample space S and that $F_1, F_2, \ldots, F_n$ are mututally exclusive events such that $\bigcup_{i}^{n} F_i = S$. Assume that $p(E) \neq 0$ for $i = 1, 2, \ldots, n$. Then

$$p(F_j \mid E) = \frac{p(E \mid F_j)p(F_j)}{\sum_{i=1}^{n} p(E \mid F_i)p(F_i)}$$

Bayesian Spam Filters

- How do we develop a tool for determining whether an email is likely to be spam?
- If we have an initial set B of spam messages and a set G of non-spam messages, we can use this information along with Bayes' theorem to predict the probability that a new email message is spam.
- We look at a particular word w, and count the number of times that it occurs in B and in G, $n_B(w)$ and $n_G(w)$
  - The estimated probability that a spam message contains w is $p(w) = n_B(w)/|B|$
  - The estimated probability that a non-spam message contains w is $q(w) = n_G(w)/|G|$

Bayesian Spam Filters

- Let S be the event that the message is spam, and E be the event that the message contains the word w.
- Using Bayes' Theorem:

$$p(S \mid E) = \frac{p(E \mid S)p(S)}{p(E \mid S)p(S) + p(E \mid \overline{S})p(\overline{S})}$$

$$p(S \mid E) = \frac{p(E \mid S)}{p(E \mid S) + p(E \mid \overline{S})}, \text{ Assume } p(S) = 1/2$$

$$r(w) = \frac{p(w)}{p(w) + q(w)}, \text{ Use estimates of } p(E \mid S) \text{ and } (E \mid \overline{S})$$

Bayesian Spam Filters

- **Example:** We find that the word "Rolex" occurs in 250 out of 2000 spam messages and occurs 5 out of 1000 non-spam messages. Estimate the probability that an incoming message is spam.
- **Solution:** $p(Rolex) = 250/2000 = 0.125$ and $q(Rolex) = 5/1000 = 0.005$.

$$r(Rolex) = \frac{p(Rolex)}{p(Rolex) + q(Rolex)} = \frac{0.125}{0.125 + 0.005} \approx 0.962$$

Random Variables

- **Definition:** A *random variable* is a function from the sample space of an experiment to the set of real numbers. That is, a random variable assigns a real number to each possible outcome.

Random Variables

- **Definition:** The *distribution* of a random variable X on a sample space S is the set of pairs $(r, p(X = r))$ for all $r \in X(S)$, where $p(X = r)$ is the probability that X takes the value r.

Random Variables

- **Example:** Suppose that a coin is flipped three times. Let $X(t)$ be the random variable that equals the number of heads that appear when t is the outcome. Then $X(t)$ takes on the following values:
  $X(HHH) = 3$
  $X(TTT) = 0$
  $X(HHT) = X(HTH) = X(THH) = 2$
  $X(TTH) = X(THT) = X(HTT) = 1.$
  Each of the eight possible outcomes has a probability $1/8$. So, the distribution of $X(t)$ is
  $p(X = 3) = 1/8, p(X = 2) = 3/8, p(X = 1) = 3/8$, and
  $p(X = 0) = 1/8.$

Expected Value

- **Definition:** The *expected value* (or *expectation* or *mean*) of the random variable $X(s)$ on the sample space S is equal to

$$E(X) = \sum_{x \in S} p(s)X(s)$$

Expected Value

- **Example:** Let X be the number that comes up when a fair die is rolled. What is the expected value of X?
- **Solution:** The random variable X takes the values 1,2,3,4,5 and 6. Each has a probability of $1/6$. It follows that:

$$E(X) = \frac{1}{6} \cdot 1 + \frac{1}{6} \cdot 2 + \cdots + \frac{1}{6} \cdot 6 = \frac{21}{6} = \frac{7}{2}$$

Expected Value

- **Theorem:** If X is a random variable and $p(X = r)$ is the probability that $X = r$, so that

$$p(X = r) = \sum_{s \in S,\, X(s) = r} p(s)$$

then

$$E(X) = \sum_{r \in X(S)} p(X = r) r$$

Independent Random Variables

- **Definition:** The random variables X and Y on a sample space S are independent if

$$p(X = r_1 \wedge Y = r_1) = p(X = r_1)p(Y = r_2)$$

- **Theorem:** If X and Y are independent variables on a sample space S, then $E(XY) = E(X)E(Y)$

Variance

- **Definition:** The *deviation* of X at $s \in S$ is $X(s) - E(X)$, the difference between the value of X and the mean of X.
- **Definition:** Let X be a random variable on the sample space S. The *variance* of X, denoted by $V(X)$ is

$$V(X) = \sum_{s \in S} (X(s) - E(X))^2 p(s)$$

- **Theorem:** If X is a random variable on a sample space S, then $V(X) = E(X^2) - E(X)^2$.

Variance

- **Example:** What is the variance of the random variable X, where X is the number that comes up when a fair die is rolled?
- **Solution:** We have $V(X) = E(X^2) - E(X)^2$. From the previous example, we know that $E(X) = 7/2$. To find $E(X^2)$ note that X^2 takes the values i^2 for $i = 1, 2, \ldots, 6$, each with a probability of $1/6$. It follows that:

$$E(X^2) = \frac{1}{6}\left(1^2 + 2^2 + 3^2 + 4^2 + 5^2 + 6^2\right) = \frac{91}{6}$$

Then

$$V(X) = \frac{91}{6} - \left(\frac{7}{2}\right)^2 = \frac{35}{12}$$