Discrete Probability: a brief overview

CMPS 4750/6750: Computer Networks
Applications of Probability in Computer Science

• Average-case complexity
• Randomized algorithms
• Combinatorics
• Networking
• Cryptography
• Information theory
• Machine learning
• …
Sample Space

- **Experiment**: a procedure that yields one of a given set of possible outcomes
  - Ex: flip a coin, roll two dice, draw five cards from a deck, etc.

- **Sample space** $\Omega$: the set of possible outcomes
  - We focus on **countable** sample space: $\Omega$ is finite or countably infinite
  - In many applications, $\Omega$ is uncountable (e.g., a subset of $\mathbb{R}$)

- **Event**: a subset of the sample space
  - Probability is assigned to events
  - For an event $A \subseteq \Omega$, its probability is denoted by $P(A)$
    - Describes beliefs about likelihood of outcomes
Discrete uniform law

- Assume $\Omega$ consists of $n$ equally likely outcomes

- For an event $A \subseteq \Omega$, $P(A) = \frac{|A|}{n}$

- Ex. 1: An urn contains four blue balls and five red balls. What is the probability that a ball chosen at random from the urn is blue? $\frac{4}{9}$
Discrete Probability

• Sample space $\Omega$: the set of possible outcomes
  – We focus on countable $\Omega$

• Events: subsets of the sample space $\Omega$

• Discrete Probability Law
  – A function $P: \mathcal{P}(\Omega) \rightarrow [0,1]$ that assigns probability to events such that:
    • $0 \leq P\{s\} \leq 1$ for all $s \in \Omega$ (Nonnegativity)
    • $P(A) = \sum_{s \in A} P\{s\}$ for all $A \subseteq \Omega$ (Additivity)
    • $P(\Omega) = \sum_{s \in \Omega} P\{s\} = 1$ (Normalization)

• Discrete uniform probability law: $|\Omega| = n$, $P(A) = \frac{|A|}{n}$ for all $A \subseteq \Omega$
Examples

• Ex. 2: consider rolling a pair of 6-sided fair dice
  - $\Omega = \{(i,j): i,j = 1, 2, 3, 4, 5, 6\}$, each outcome has the same probability of $1/36$
  - $P(\{\text{the sum of the rolls is even}\}) = 18/36 = 1/2$

• Ex. 3: consider rolling a 6-sided biased (loaded) die
  - Assume $P(3) = \frac{2}{7}, P(1) = P(2) = P(4) = P(5) = P(6) = \frac{1}{7}$
  - $A = \{1,3,5\}, \ P(A) = \frac{1}{7} + \frac{2}{7} + \frac{1}{7} = \frac{4}{7}$
Properties of Probability Laws

• Consider a probability law, and let $A$, $B$, and $C$ be events
  
  − If $A \subseteq B$, then $P(A) \leq P(B)$
  
  − $P(\overline{A}) = 1 - P(A)$
  
  − $P(A \cup B) = P(A) + P(B) - P(A \cap B)$
  
  − $P(A \cup B) = P(A) + P(B)$ if $A$ and $B$ are disjoint, i.e., $A \cap B = \emptyset$

• Ex. 4: 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?

$$\frac{50}{100} + \frac{20}{100} - \frac{10}{100} = 0.6$$
Conditional Probability

• Conditional probability provides us with a way to reason about the outcome of an experiment, based on partial information.

• Ex. 1: roll a six-sided fair die. Suppose we are told that the outcome is even. What is the probability that the outcome is 6?

\[
\frac{1}{3} = \frac{P(A \cap B)}{P(B)}
\]

• Let \( A \) and \( B \) be two events (of a given sample space) where \( P(B) > 0 \). The conditional probability of \( A \) given \( B \) is defined as

\[
P(A \mid B) = \frac{P(A \cap B)}{P(B)}
\]

• Given an event \( B \) with \( P(B) > 0 \), conditional probabilities \( P(A \mid B) \) form a legitimate probability law.
Independence

• We say that event $A$ is independent of event $B$ if $P(A \mid B) = P(A)$
• Two events $A$ and $B$ are independent if and only if $P(A \cap B) = P(A)P(B)$
• Ex. 3: Consider an experiment involving two successive rolls of a 4-sided die in which all 16 possible outcomes are equally likely and have probability $1/16$. Are the following pair of events independent?
  (a) $A = \{1st\; roll\; is\; 1\}$, $B = \{sum\; of\; two\; rolls\; is\; 5\}$   Yes
  (b) $A = \{1st\; roll\; is\; 4\}$, $B = \{sum\; of\; two\; rolls\; is\; 4\}$   No
• We say that the events $A_1, A_2, ... A_n$ are (mutually) independent if and only if
  $P(\cap_{i \in S} A_i) = \prod_{i \in S} P(A_i)$, for every subset $S$ of $\{1, 2, ..., n\}$
Bernoulli Trials

• Bernoulli Trial: an experiment with two possible outcomes
  – E.g., flip a coin results in two possible outcomes: head (H) and tail (T)
• Independent Bernoulli Trials: a sequence of Bernoulli trials that are mutually independent

Ex.4: Consider an experiment involving five independent tosses of a biased coin, in which the probability of heads is $p$.
  – What is the probability of the sequence $HHHTT$?
    • $A_i = \{i$–th toss is a head$\}$
    • $P(A_1 \cap A_2 \cap A_3 \cap \overline{A}_4 \cap \overline{A}_5) = P(A_1)P(A_2)P(A_3)P(\overline{A}_4)P(\overline{A}_5) = p^3(1 - p)^2$
  – What is the probability that exactly three heads come up?
    • $P($exactly three heads come up$) = \binom{5}{3} p^3(1 - p)^2$
Random Variables

• A random variable (r.v.) is a real-valued function of the experimental outcome.
• Ex. 1: Consider an experiment involving three independent tosses of a fair coin.
  − $\Omega = \{HHH, HHT, HTH, HTT, THH, THT, TTH, TTT\}$
  − $X(s) =$ the number of heads that appear for $s \in \Omega$. Then
    
    $X(HHH) = 3, X(HHT) = X(HTH) = X(THH) = 2,$
    $X(HTT) = X(THT) = X(TTH) = 1, X(TTT) = 0$

  − $P(X = 2) = P(\{s \in \Omega: X(s) = 2\}) = P(\{HHT,HTH,THH\}) = 3/8$
  − $P(X < 2) = P(\{HTT,THT,TTH,TTT\}) = 4/8 = 1/2$
Random Variables

• A random variable is a real-valued function of the outcome of the experiment.

• A function of a random variable defines another random variable.

• We can associate with each random variable certain “averages” of interest, such as the expected value and the variance.

• A discrete random variable is a real-valued function of the outcome of the experiment that can take a finite or countably infinite number of values.
Probability Mass Functions

• Let $X$ be a discrete r.v. Then the probability mass function (PMF), $p_X(\cdot)$ of $X$, is defined as:

$$p_X(x) = P(X = x) = P(s \in \Omega: X(s) = x)$$

$$- \sum_x P_X(x) = 1$$

$$- P(X \in S) = \sum_{x \in S} p_X(x)$$

The cumulative distribution function (CDF) of $X$ is defined as

$$F_X(a) = P(X \leq a) = \sum_{x \leq a} p_X(x)$$

• Ex.2: Let the experiment consist of two independent tosses of a fair coin, and let $X$ be the number of heads obtained. Find the PMF of $X$. 


Bernoulli Distribution

• Consider a Bernoulli trial with probability of success $p$. Let $X$ be a r.v. where $X = 1$ if “success” and $X = 0$ if “failure”

$$X = \begin{cases} 
1 & \text{w/prob } p \\
0 & \text{otherwise}
\end{cases}$$

We write $X \sim \text{Bernoulli}(p)$. The PMF of $X$ is defined as:

$$p_X(1) = p$$

$$p_X(0) = 1 - p$$
Binomial Distribution

• Consider an experiment of \(n\) independent Bernoulli trials, with the probability of success \(p\). Let the r.v. \(X\) be the number of successes in the \(n\) trials.

• The PMF of \(X\) is defined as:

\[
p_X(k) = P(X = k) = \binom{n}{k} p^k (1 - p)^{n-k}, \text{ where } k = 0, 1, 2, \ldots, n
\]

We write \(X \sim \text{Binomial}(p)\).
Geometric Distribution

• Consider an experiment of independent Bernoulli trials, with probability of success $p$. Let $X$ be the number of trials to get one success.

• Then the PMF of $X$ is:

$$P(X = k) = (1 - p)^{k-1}p, \text{ where } k = 1, 2, 3 \ldots$$

We write $X \sim \text{Geometric}(p)$. 
Expected Value

- The **expected value** (also called the **expectation** or the **mean**) of a random variable \(X\) on the sample space \(\Omega\) is equal to

\[
E(X) = \sum_{s \in \Omega} X(s) P\{s\} = \sum_{x} x p_X(x)
\]

**Ex. 1:** If \(X \sim \text{Bernoulli}(p)\), \(E(X) = 1 \cdot p + 0 \cdot (1 - p) = p\)

**Ex. 2:** If \(X \sim \text{Geometric}(p)\), \(E(X) = \sum_{k=1}^{\infty} k (1 - p)^{k-1}p = \frac{1}{p}\)
Linearity of Expectations

• If $X_i, i = 1, 2, \ldots, n$ are random variables on $\Omega$, and $a$ and $b$ are real numbers, then
  
  $\begin{align*}
  &-E(X_1 + X_2 + \cdots + X_n) = E(X_1) + E(X_2) + \cdots + E(X_n) \\
  &-E(aX + b) = aE(X) + b
  \end{align*}$

• Ex. 3: $X \sim \text{Binomial}(p)$
  
  $-E(X) = \sum_{k=0}^{n} \binom{n}{k} p^k (1 - p)^{n-k} = np$
Variance

- The variance of a random variable $X$ on the sample space $\Omega$ is equal to

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

$$= E \left[ (X - E(X))^2 \right]$$

- The variance provides a measure of dispersion of $X$ around its mean

- Another measure of dispersion is the standard deviation of $X$:

$$\sigma(X) = \sqrt{V(X)}$$
Variance

• Theorem: \( V(X) = E(X^2) - E(X)^2 \)

• Ex. 1: Let \( X \) be a Bernoulli random variable with parameter \( p \)
  \[
  E(X) = 1 \cdot p + 0 \cdot (1 - p) = p \quad E(X^2) = 1 \cdot p + 0 \cdot (1 - p) = p
  \]
  \[
  V(X) = E(X^2) - E(X)^2 = p - p^2
  \]

• Ex. 2: Let \( X \) be a geometric random variable with parameter \( p \)
  \[
  E(X) = \frac{1}{p}, \quad E(X^2) = \frac{2}{p^2} - \frac{1}{p}
  \]
  \[
  V(X) = E(X^2) - E(X)^2 = \frac{1-p}{p^2}
  \]
Joint Probability and Independence

- The joint probability mass function between discrete r.v.’s \( X \) and \( Y \) is defined by
  \[
  p_{X,Y}(x,y) = P\{X = x \text{ and } Y = y\}
  \]

- We say two discrete r.v.’s \( X \) and \( Y \) are independent if
  \[
  p_{X,Y}(x,y) = p_X(x) \cdot p_Y(y), \quad \forall x, y
  \]

- **Theorem** If two r.v.’s \( X \) and \( Y \) are independent, then \( E(XY) = E(X)E(Y) \)