[CSED261] Discrete Mathematics for Computer Science

Lecture 16

Discrete Probability

Suha Kwak

suha.kwak@postech.ac.kr

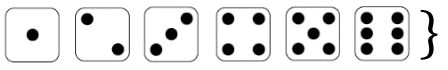
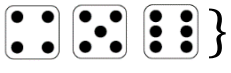
Dept. of Computer Science and Engineering

POSTECH

An Introduction to Discrete Probability

Probability of an Event

- Key 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.

- Example
  - Random experiment
    - Tossing a single six-sided dice
  - Sample space
    - $S = \{$  $\}$
  - Event
    - A subset of the sample space
    - "The number on the face > 3" $= \{$  $\}$
    - $S =$ Sure event
    - $\phi =$ Impossible event

Probability of an Event

- Definition (*by Pierre-Simon Laplace*)
  - If S is a finite sample space of equally likely outcomes, and E is an event (*i.e.*, a subset of S) then the probability of E is

$$P(E) = \frac{|E|}{|S|}.$$

  - 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 $0 \leq |E| \leq |S|$.

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

Probability of an Event

- Example 2
  - 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?
  - Answer:
    - The number of ways to choose six numbers out of 40 is $C(40,6) = 40!/(34!\,6!) = 3{,}838{,}380.$
    - Hence the probability of picking a winning combination is $\frac{1}{3{,}838{,}380} \approx 0.00000026.$

Probability of an Event

- Example 3
  - What is the probability that the numbers 11, 4, 17, 39, and 23 are drawn in that order from a bin with 50 balls labeled with the numbers 1,2, …, 50 if
    A. The ball selected is not returned to the bin.
    B. The ball selected is returned to the bin before the next ball is selected.
  - Answer:
    - Use the product rule in each case.
    A. (*Sampling without replacement*) The probability is $\frac{1}{254,251,200}$ since there are $50 \cdot 49 \cdot 48 \cdot 47 \cdot 46 = 254,251,200$ ways to choose the five balls.
    B. (*Sampling with replacement*) Since $50^5 = 312,500,000$ the probability is $\frac{1}{50^5} = \frac{1}{312,500,000}$.

Complements and Unions of Events

- Theorem: Probability of the complementary event
  - Let E be an event in sample space S.
  - The probability of the event $\bar{E} = S - E$, the complementary event of E, is given by $P(\bar{E}) = 1 - P(E)$.
  - Proof
    - Using the fact that $|\bar{E}| = |S| - |E|$,

$$P(\bar{E}) = \frac{|S| - |E|}{|S|} = 1 - \frac{|E|}{|S|} = 1 - P(E)$$

- Example
  - A sequence of 10 bits is chosen randomly.
  - What is the probability that at least one of these bits is 0?

$$P(E) = 1 - P(\bar{E}) = 1 - \frac{|\bar{E}|}{|S|} = 1 - \frac{1}{2^{10}} = \frac{1023}{1024}$$

Complements and Unions of Events

- Theorem: Probability of a union of events
  - Let E_1 and E_2 be events in the sample space S, then

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

  - Proof
    - Given the inclusion-exclusion formula,

$$|A \cup B| = |A| + |B| - |A \cap B|.$$

    - It follows that

$$P(E_1 \cup E_2) = \frac{|E_1 \cup E_2|}{|S|} = \frac{|E_1| + |E_2| - |E_1 \cap E_2|}{|S|}$$

$$= \frac{|E_1|}{|S|} + \frac{|E_2|}{|S|} - \frac{|E_1 \cap E_2|}{|S|}$$

$$= P(E_1) + P(E_2) - P(E_1 \cap E_2)$$

Complements and Unions of Events

- Example
  - Suppose we select an integer at random from the set of positive integers not exceeding 100.
  - What is the probability that the selected integer 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 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 it is divisible by 2 and 5.
  - It follows that

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

$$= \frac{50}{100} + \frac{20}{100} - \frac{10}{100} = \frac{3}{5}$$

Probability Theory

Summary

- Assigning Probabilities
- Conditional Probability
- Independence
- Bernoulli Trials and the Binomial Distribution
- Random Variables

Assigning Probabilities

- Laplace's definition of probability
  - It assumes that all outcomes are equally likely.
  - A more general definition is required to avoid the restriction.

- Assigning probabilities
  - Let S be a sample space of an experiment with a finite number of outcomes.
  - Assign a probability $P(s)$ to each outcome s, so that:
    - $0 \leq P(s) \leq 1$ for each $s \in S$
    - $\sum_{s \in S} P(s) = 1$
  - The function P from the set of all outcomes of the sample space S is called a probability distribution.

Assigning Probabilities

- Example
  - What probabilities should we assign to the outcomes *H* (heads) and *T* (tails) when a fair coin is flipped?
  - 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 fair coin, the probabilities for *H* and *T* are both $\frac{1}{2}$.
    - For the biased coin: $P(H) = 2 \cdot P(T)$, and since $P(H) + P(T) = 1$, $P(H) = \frac{2}{3}$ and $P(T) = \frac{1}{3}$.

Uniform Distribution

- Definition
  - Suppose that S is a set with n elements.
  - The uniform distribution assigns the probability $\frac{1}{n}$ to each element of S.
  - Note that we could have used Laplace's definition here.

- Example
  - For a fair coin: $P(H) = P(T) = \frac{1}{2}$
  - For a fair six-faced dice: $P(1) = P(2) = \cdots = P(6) = \frac{1}{6}$

Probability of an Event

- The probability of the event E is the sum of the probabilities of the outcomes in E:

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

  - Now no assumption is being made about the distribution.

- Example
  - Suppose that a die is biased so that ⚂ appears twice as often as each other, but the other five faces appear equally.
  - What is the probability that an odd number appears when we roll this die?

$P(⚂) = \frac{2}{7}$, and $P(⚀) = P(⚁) = P(⚃) = P(⚄) = P(⚅) = \frac{1}{7}$

$P(E) = P(⚀⚂⚄) = P(⚀) + P(⚂) + P(⚄) = \frac{4}{7}$

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 $P(E|F)$, is defined as:

$$P(E|F) = \frac{P(E \cap F)}{P(F)}$$

- Example
  - A bit string of length 4 is sampled from a uniform distribution.
  - What is the probability that it contains at least two consecutive 0s, *given* that its first bit is a 0?
    - 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 a 0.
    - $E \cap F = \{0000, 0001, 0010, 0011, 0100\}$, so $P(E \cap F) = 5/16$.
    - $P(F) = 1/2$ since there are 8 bit strings starting with a 0.
    - Hence, $P(E|F) = 5/16 \cdot 2 = 5/8$

Independence

- Definition
  - The events E and F are independent if and only if
$$P(E \cap F) = P(E) \cdot P(F)$$
- Example
  - Suppose E is the event that a randomly generated bit string of length 4 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 4 are equally likely?
  - They are independent; why?
    - There are eight bit strings of length 4 beginning with a 1, and eight bit strings of length 4 that contain an even number of 1s.
    - Since the number of bit strings of length 4 is 16, $P(E) = P(F) = \frac{1}{2}$
    - Since $E \cap F = \{1111,1100,1010,1001\}$, $P(E \cap F) = \frac{1}{4}$.

Pairwise and Mutual Independence

- Definition (*Pairwise independence)*
  - The events $E_1, E_2, \ldots, E_n$ are pairwise independent if and only if

$$P(E_i \cap E_j) = P(E_i) \cdot P(E_j)$$

for all pairs i and j with $1 \leq i < j \leq n$.

- Definition (*Mutual independence*)
  - The events $E_1, E_2, \ldots, E_n$ are mutually independent if and only if

$$P(E_{i_1} \cap E_{i_2} \cap \cdots \cap E_{i_m}) = P(E_{i_1}) \cdot P(E_{i_2}) \cdots P(E_{i_m})$$

whenever i_j $(j = 1,2, \ldots, m)$ are integers with

$1 \leq i_1 < i_2 < \cdots < i_m \leq n$ and $m \geq 2$

Pairwise and Mutual Independence

- Example
  - A fair 6-sided dice is thrown twice.
  - Three events of our interest:
    - E_1: The first roll is a 3.
    - E_2: The second roll is 4.
    - E_3: The sum of the two rolls is 7.
  - The three events are pairwise independent since
    - $P(E_1) = P(E_2) = P(E_3) = \frac{1}{6}$
    - $P(E_1 \cap E_2) = P(E_2 \cap E_3) = P(E_1 \cap E_3) = \frac{1}{6} \cdot \frac{1}{6} = \frac{1}{36}$
  - However, the events are not mutually independent since
    - $P(E_1 \cap E_2 \cap E_3) = \frac{1}{36} \neq \left(\frac{1}{6}\right)^3$

Bernoulli Trials

- Definition
  - Suppose an experiment can have only two possible outcomes.
    - *E.g.*, the flipping of a coin or the random generation of a bit.
  - 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 4 heads occur when the coin is flipped 7 times?
  - Solution:
    - There are $2^7 = 128$ possible outcomes.
    - The number of ways 4 of the 7 flips can be heads is $C(7,4)$.
    - The probability of each of the outcomes is $\left(\frac{2}{3}\right)^4 \left(\frac{1}{3}\right)^3$ since the 7 flips are independent.
    - Hence, the probability that exactly 4 heads occur is

$$C(7,4) \cdot \left(\frac{2}{3}\right)^4 \left(\frac{1}{3}\right)^3 = \frac{35 \cdot 16}{3^7} = \frac{560}{2187}$$

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

$$C(n, k) \cdot p^k \cdot q^{n-k}$$

  - Proof
    - The outcome of n Bernoulli trials is an n-tuple $(t_1, t_2, \dots, t_n)$, where each t_i is either S (success) or F (failure).
    - The probability of each outcome of n trials consisting of k successes and $n - k$ failures (in any order) is $p^k q^{n-k}$.
    - Because there are a $C(n, k)$ number of n-tuple of Ss and Fs that contain exactly k Ss, the probability of k successes is $C(n, k) \cdot p^k \cdot q^{n-k}$.
  - We denote this probability by $b(k: n, p) = C(n, k) \cdot p^k \cdot q^{n-k}$, a function of k given n and p.

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.

A random variable is a function.
It is not a variable, and it is not random!

In the late 1940s W. Feller and J.L. Doob flipped a coin to see whether both would use “random variable” or the more fitting “chance variable.” Unfortunately, Feller won and the term “random variable” has been used ever since…

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.
- Example: A coin flipped 3 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 8 possible outcomes has 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 $P(X = 0) = 1/8$.

Bayes' Theorem

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:

$$P(F|E) = \frac{P(E|F)P(F)}{P(E|F)P(F) + P(E|\bar{F})P(\bar{F})}$$

  - Proof

$$P(F|E) = \frac{P(E \cap F)}{P(E)} = \frac{P(E \cap F)}{P(E \cap F) + P(E \cap \bar{F})}$$

$$= \frac{P(E|F)P(F)}{P(E|F)P(F) + P(E|\bar{F})P(\bar{F})}$$

Bayes' Theorem

- Example
  - We have two boxes.
    - The first box contains 2 green balls and 7 red balls.
    - The second contains 4 green balls and 3 red balls.
  - You select one of the boxes at random, then select a ball from the selected box at random.
  - If you have a red ball, what is the probability that you selected a ball from the first box.
  - Solution
    - E: The event that you have chosen a red ball
    - F: The event that you have chosen the first box

$$P(F|E) = \frac{P(E|F)P(F)}{P(E|F)P(F) + P(E|\bar{F})P(\bar{F})}$$

$$= \frac{(7/9)(1/2)}{(7/9)(1/2) + (3/7)(1/2)} = \frac{49}{76}$$

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 *mutually exclusive* events such that

$$\bigcup_{i=1}^{n} F_i = S$$

  - Assume that $P(E) \neq 0$ for $i = 1,2, \ldots, n$.
  - Then

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

A Little Taste of Machine Learning

- Interpreting Bayes' theorem

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

  - The event of our interest A
    - Typically latent (*i.e.*, hidden)
  - The event as an observation B
  - Prior probability $P(A)$
    - Probability of the event A with no observation
    - Based only on our *prior* knowledge about A
  - Likelihood $P(B|A)$
    - Probability of observing B when A happens
  - Posterior probability $P(A|B)$
    - Probability of A if we observed B

$$posterior \propto likelihood \times prior$$

A Little Taste of Machine Learning

- Interpreting Bayes' theorem (cont'd)

x5 x3 x2

$[prior]$

$P(\text{🍎}) = 0.5, \quad P(\text{🍐}) = 0.3, \quad P(\text{🍊}) = 0.2$

$[likelihood]$

$P(\text{👅} \mid \text{🍎}) = 0.3, \quad P(\text{👅} \mid \text{🍐}) = 0.5, \quad P(\text{👅} \mid \text{🍊}) = 0.9$

$[posterior \propto likelihood \times prior]$

$P(\text{🍎} \mid \text{👅}) \propto 0.15, \quad P(\text{🍐} \mid \text{👅}) \propto 0.15, \quad P(\text{🍊} \mid \text{👅}) \propto 0.18$

A Little Taste of Machine Learning

- Bayesian spam filtering
  - Suppose that we have an initial set B of spam messages and set G of non-spam messages.
    - Datasets like B and G are called *training data* in the context of machine learning
  - 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 $n_G(w)$.
  - Let S be the event that the message is spam, and E be the event that the message contains the word w.
    - Estimated probability that a spam message contains w:
      $$P(E|S) = n_B(w)/|B|$$
    - Estimated probability that a non-spam message contains w:
      $$P(E|\bar{S}) = n_G(w)/|G|$$

A Little Taste of Machine Learning

- Bayesian spam filtering (cont'd)
  - For spam filtering, we compute the posterior probability $P(S|E)$ for the message and categorize it as spam if the probability is larger than a pre-defined threshold.
  - $P(S|E)$ indicates the probability that the message is spam when it contains the word w.

$$P(S|E) = \frac{P(E|S)P(S)}{P(E)} = \frac{P(E|S)P(S)}{P(E|S)P(S) + P(E|\bar{S})P(\bar{S})}$$

$$= \frac{P(E|S)}{P(E|S) + P(E|\bar{S})}$$

Here we assuming that the prior probabilities are uniform.

If we have data on the frequency of spam messages, we can obtain a better estimate of the priors.

A Little Taste of Machine Learning

- Bayesian spam filtering: A specific example
  - We find that the word “Rolex” occurs in 250 out of 2000 spam messages and occurs in 5 out of 1000 non-spam messages.
  - Estimate the probability that an incoming message with “Rolex” is spam.
  - Suppose our threshold for rejecting the email is 0.9.
  - Is the message a spam according to our Bayesian spam filter?

$$P(\text{"Rolex"}|S) = \frac{250}{2000} = 0.125, \qquad P(\text{"Rolex"}|\bar{S}) = \frac{5}{1000} = 0.005$$

$$P(S|\text{"Rolex"}) = \frac{P(\text{"Rolex"}|S)}{P(\text{"Rolex"}|S) + P(\text{"Rolex"}|\bar{S})} = \frac{0.125}{0.125 + 0.005} \approx 0.962$$

Since the probability $P(S|\text{"Rolex"})$ is larger than the threshold,
our spam filtering system will categorize the message as spam and reject it!

A Little Taste of Machine Learning

- Bayesian spam filtering with using multiple words
  - Accuracy can be improved by considering more than one word as evidence.
  - Consider the case where E_1 and E_2 denote the events that the message contains the words w_1 and w_2 respectively.
  - For simplicity, we make the assumption that $P(S) = 0.5$ (the uniform prior) and E_1 and E_2 are *conditionally independent* given S:

$$P(E_1 \cap E_2|S) = P(E_1|S)P(E_2|S)$$

- Then the posterior probability of S is given by

$$P(S|E_1 \cap E_2) = \frac{P(E_1|S)P(E_2|S)}{P(E_1|S)P(E_2|S) + P(E_1|\bar{S})P(E_2|\bar{S})}$$

A Little Taste of Machine Learning

- Bayesian spam filtering with using multiple words:
  A specific example
  - We have 2000 spam messages and 1000 non-spam messages.
  - The word “stock” occurs 400 times in the spam messages and 60 times in the non-spam.
  - The word “undervalued” occurs in 200 spam messages and 25 non-spam.

$$P(E_1|S) = \frac{400}{2000} = 0.2, \quad P(E_1|\bar{S}) = \frac{60}{1000} = 0.06$$

$$P(E_2|S) = \frac{200}{2000} = 0.1, \quad P(E_2|\bar{S}) = \frac{25}{1000} = 0.025$$

$$P(S|E_1 \cap E_2) = \frac{(0.2)(0.1)}{(0.2)(0.1) + (0.06)(0.025)} \approx 0.93$$

A Little Taste of Machine Learning

- Bayesian spam filtering in general
  - The more words we consider, the more accurate the spam filter we can obtain.
  - With the independence assumption if we consider k words:

$$P\left(S\middle|\bigcap_{i=1}^{k} E_i\right) = \frac{\prod_{i=1}^{k} P(E_i|S)}{\prod_{i=1}^{k} P(E_i|S) + \prod_{i=1}^{k} P(E_i|\bar{S})}$$

  - We can further improve the filter by considering pairs of words as a single block of certain types of strings.

Expected Value and Variance

Expected Value

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

$$E(X) = \sum_{s \in S} P(s) \cdot X(s)$$

- 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, or 6.
    - Each has probability $1/6$.
    - 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 2
  - Suppose that n mutually independent Bernoulli trials are performed, where p is the probability of success on each trial.
  - Then the expected number of successes in this case is np.
  - Proof
    - Let X be the random variable equal to the number of success in n trials.
    - Then,

$$P(X = k) = \binom{n}{k} p^k (1-p)^{n-k}$$

    - Hence,

$$E(X) = \sum_{k=1}^{n} k \cdot P(X = k)$$

$$= \sum_{k=1}^{n} k \cdot \binom{n}{k} p^k (1-p)^{n-k}$$

Expected Value

- Theorem 2
  - Proof (cont'd)

$$E(X) = \sum_{k=1}^{n} k \cdot P(X = k) = \sum_{k=1}^{n} k \cdot \binom{n}{k} p^k (1-p)^{n-k}$$

$$= \sum_{k=1}^{n} n \cdot \binom{n-1}{k-1} p^k (1-p)^{n-k}$$

$$= np \sum_{k=1}^{n} \binom{n-1}{k-1} p^{k-1} (1-p)^{n-k}$$

Shift the index of sum with $j = k - 1$.

$$= np \sum_{j=0}^{n-1} \binom{n-1}{j} p^j (1-p)^{n-1-j}$$

Recall the binomial theorem:

$$(x+y)^n = \sum_{j=0}^{n} \binom{n}{j} x^{n-j} y^j$$

$$= np \cdot (p + 1 - p)^{n-1} = np \cdot 1 = np$$

Linearity of Expectations

- Theorem
  - X_i are random variables on S where $i = 1,2, \dots, n$
  - a and b are real numbers.
  - Then

(i) $E(X_1 + X_2 + \cdots X_n) = E(X_1) + E(X_2) + \cdots + E(X_n)$

(ii) $E(aX + b) = aE(X) + b$

- Proof of (i)
  - When $n = 2$,

$$E(X_1 + X_2) = \sum_{s \in S} P(s)\big(X_1(s) + X_2(s)\big)$$

$$= \sum_{s \in S} P(s)X_1(s) + \sum_{s \in S} P(s)X_2(s) = E(X_1) + E(X_2)$$

  - We can generalize the result for $n > 2$ by math induction.

Linearity of Expectations

- Theorem
  - X_i are random variables on S where $i = 1,2, \dots, n$
  - a and b are real numbers.
  - Then

(i) $E(X_1 + X_2 + \cdots X_n) = E(X_1) + E(X_2) + \cdots + E(X_n)$

(ii) $E(aX + b) = aE(X) + b$

- Proof of (ii)

$$E(aX + b) = \sum_{s \in S} P(s)(aX(s) + b)$$

$$= a \sum_{s \in S} P(s)X(s) + b \sum_{s \in S} P(s)$$

$$= aE(X) + b \cdot 1 = aE(X) + b$$

Average-Case Computational Complexity

- The average-case complexity of an algorithm can be found by computing the expected value of a random variable.
  - Let the sample space of an experiment be the set of possible inputs a_j where $j = 1,2, \dots, n$.
  - Let the random variable X be the assignment to a_j of the number of operations used by the algorithm when given a_j as input.
  - Assign a probability $P(a_j)$ to each possible input value a_j.
  - The expected value of X is the average-case computational complexity of the algorithm.

$$E(X) = \sum_{j=1}^{n} P(a_j) X(a_j)$$

Average-Case Computational Complexity

- Linear search

Procedure linear_search(x: integer, $\{a_1, a_2, \ldots, a_n\}$: integers)

$i := 1$

while $i \leq n$ and $x \neq a_i$

$i := i + 1$

if $i \leq n$ **then** $\ell := i$

else $\ell := 0$

return ℓ

- Suppose that
  - The probability that x is in the list is p,
  - It is equally likely that x is any of the n elements of the list.
- Then, what is the average-case computational complexity of this algorithm?

Average-Case Computational Complexity

- Linear search (cont'd)
  - There are $n + 1$ possible types of input:
    - One type for each of the n numbers on the list
    - One additional type for the numbers not on the list.
  - Recall that we need:
    - $2i + 1$ comparisons if x equals the i^{th} element of the list.
    - $2n + 2$ comparisons if x is not on the list.
  - The probability that x equals a_j is p/n and the probability that x is not in the list is $q = 1 - p$.
  - The average-case computational complexity of the linear search algorithm is:

$$\begin{aligned} E = {} & 3p/n + 5p/n + \cdots + (2n+1)p/n + (2n+2)q \\ = {} & (p/n)\big(3 + 5 + \cdots + (2n+1)\big) + (2n+2)q \\ = {} & (p/n)\big((n+1)^2 - 1\big) + (2n+2)q \\ = {} & p(n+2) + (2n+2)q. \end{aligned}$$

For Independent Random Variables

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

$$P(X = r_1 \text{ and } Y = r_2) = P(X = r_1) \cdot P(Y = r_2)$$

- Theorem
  - If X and Y are independent variables on a sample space S, then $E(XY) = E(X) \cdot E(Y)$.
- Proof
  - The event $XY = r$ is the disjoint union of the events $X = r_1$ and $Y = r_2$ over all $r_1 \in X(S)$ and $r_2 \in Y(S)$ with $r = r_1 \cdot r_2$.

$$E(XY) = \sum_{r \in XY(S)} r \cdot P(XY = r)$$

$$= \sum_{r_1 \in X(S)} \sum_{r_2 \in Y(S)} r_1 r_2 \cdot P(X = r_1 \text{ and } Y = r_2)$$

For Independent Random Variables

- Theorem
  - Proof (cont'd)

$$E(XY) = \sum_{r_1 \in X(S)} \sum_{r_2 \in Y(S)} r_1 r_2 \cdot P(X = r_1 \text{ and } Y = r_2)$$

$$= \sum_{r_1 \in X(S)} \sum_{r_2 \in Y(S)} r_1 r_2 \cdot P(X = r_1) \cdot P(Y = r_2)$$

$$= \sum_{r_1 \in X(S)} \left(r_1 \cdot P(X = r_1) \cdot \sum_{r_2 \in Y(S)} r_2 \cdot P(Y = r_2) \right)$$

$$= \sum_{r_1 \in X(S)} r_1 \cdot P(X = r_1) \cdot E(Y)$$

$$= E(Y) \left(\sum_{r_1 \in X(S)} r_1 \cdot P(X = r_1) \right) = E(Y)E(X)$$

Variance

- Deviation
  - 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: Variance and Standard deviation
  - Let X be a random variable on the sample space S.
  - The variance of X, denoted by $V(X)$ is

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

  - That is $V(X)$ is the weighted average of the square of the deviation of X.
  - The standard deviation of X, denoted by $\sigma(X)$ is defined to be

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

Variance

- Theorem
  - If X is a random variable on a sample space S, then
$$V(X) = E(X^2) - E(X)^2$$
- Corollary
  - If X is a random variable on a sample space S and $E(X) = \mu$,
$$V(X) = E\big((X - \mu)^2\big)$$
- Example
  - For a random variable X,
    - $X(t) = 1$ if a Bernoulli trial t is a success,
    - $X(t) = 0$ if it is a failure.
  - p is the probability of success and q is that of failure.
  - What is the variance of the random variable X?

$$V(X) = E(X^2) - E(X)^2 = p - p^2 = p(1 - p) = pq$$