Probability Theory

Waltenegus Dargie

Chair for Computer Networks

• Deals with the study of random phenomena, which under repeated experiments yield different outcomes that have certain underlying patterns about them.

• When an experiment is performed under these conditions, certain elementary events $\xi_i$ occur in different but completely uncertain ways.

• One can assign nonnegative number, $P_i(\xi_i)$ as the probability of the event $\xi_i$ in various ways.
Definition

• Laplace’s Classical Definition
  – The probability of an event $A$ is defined a-priori without actual experimentation (provided all the outcomes are equally likely)

$$P(A) = \frac{\text{Number of outcomes favorable to } A}{\text{Total number of possible outcomes}}$$

• Relative Frequency Definition

$$P(A) = \lim_{n \to \infty} \frac{n_A}{n}$$

where $n_A$ is the number of occurrences of $A$ and $n$ is the total number of trials.
Probability Sets

- The totality of all $\xi_i$, known \textit{a-priori}, constitutes a set $\Omega$, the set of all experimental outcomes

$$\Omega = \{ \xi_1, \xi_2, \ldots, \xi_k, \ldots \}$$

- For any subsets $A$, $B$, $C$, $\ldots$ of $\Omega$,
  - If $\xi \in A$ then $\xi \in \Omega$
  - $A \cup B = \{ \xi | \xi \in A \text{ or } \xi \in B \}$
  - $A \cap B = \{ \xi | \xi \in A \text{ and } \xi \in B \}$
  - $\overline{A} = \{ \xi | \xi \notin A \}$
Probability Sets

\[ A \cup B \]

\[ A \cap B \]

\[ \overline{A} \]

\[ A \cap B = \phi \]

\[ A_i \cap A_j = \phi, \quad \text{and} \quad \bigcup_{i=1}^{n} A_i = \Omega. \]
De-Morgan’s Laws

\[ \overline{A \cup B} = \overline{A} \cap \overline{B}; \quad \overline{A \cap B} = \overline{A} \cup \overline{B} \]
• A collection of subsets of a nonempty set $\Omega$ forms a field $F$

(i) $\Omega \in F$

(ii) If $A \in F$, then $\bar{A} \in F$

(iii) If $A \in F$ and $B \in F$, then $A \cup B \in F$.

• Hence, the following sets, all, belong to $F$ (the term event is used only to members of $F$):

$$F = \{ \Omega, A, B, \bar{A}, \bar{B}, A \cup B, A \cap B, \bar{A} \cup B, \ldots \}$$
Axioms of Probability

1. Probability is a nonnegative number
\[ P(A) \geq 0 \]

2. The probability of a whole set is unity
\[ P(\Omega) = 1 \]

3. The probability of mutually exclusive sets is the union of the probability of each set
If \( A \cap B = \emptyset \), then \( P(A \cup B) = P(A) + P(B). \)
Axioms of Probability

- From the axioms it follows that:

\[ A \cup \overline{A} = \Omega, \quad P( A \cup \overline{A}) = P(\Omega) = 1. \]
\[ A \cap \overline{A} \in \phi, \]
\[ P( A \cup \overline{A}) = P(A) + P(\overline{A}) = 1 \quad \text{or} \quad P(\overline{A}) = 1 - P(A). \]
\[ A \cap \{\phi\} = \{\phi\}. \]
\[ P \left( A \cup \{\phi\} \right) = P(A) + P(\phi). \]
\[ A \cup \{\phi\} = A, \]
\[ P\{\phi\} = 0. \]
Conditional Dependencies

• In $N$ independent trials, suppose $N_A$, $N_B$, $N_{AB}$ denote the number of times events $A$, $B$ and $AB$ occur respectively. According to the frequency interpretation of probability, for large $N$

\[
P(A) \approx \frac{N_A}{N}, \quad P(B) \approx \frac{N_B}{N}, \quad P(AB) \approx \frac{N_{AB}}{N}.
\]
Conditional Dependencies

- Among the $N_A$ occurrences of $A$, only $N_{AB}$ of them are also found among the $N_B$ occurrences of $B$. Thus the ratio

$$\frac{N_{AB}}{N_B} = \frac{N_{AB}}{N} = \frac{P(AB)}{P(B)}$$

- The above expression is a measure of “the event $A$ given that $B$ has already occurred”. Conditional dependencies are expressed using Bayes’ Theorem:

$$P(A \mid B) = \frac{P(AB)}{P(B)}, \quad P(B) \neq 0.$$
Conditional Dependencies

• We have
  \[ P(A \mid B) = \frac{P(AB)}{P(B)} \geq 0, \]
  \[ P(\Omega \mid B) = \frac{P(\Omega B)}{P(B)} = \frac{P(B)}{P(B)} = 1, \]

• Given \( A \cap C = 0 \).
  \[ P(A \cup C \mid B) = \frac{P((A \cup C) \cap B)}{P(B)} = \frac{P(AB \cup CB)}{P(B)}. \]
  \[ AB \cap AC = \phi \implies P(AB \cup CB) = P(AB) + P(CB). \]
  \[ P(A \cup C \mid B) = \frac{P(AB)}{P(B)} + \frac{P(CB)}{P(B)} = P(A \mid B) + P(C \mid B). \]
• Properties of conditional probabilities

1. \( B \subseteq A, \ AB = B, \)
   \[
P(A \mid B) = \frac{P(AB)}{P(B)} = \frac{P(B)}{P(B)} = 1
   \]

2. \( A \subseteq B, \ AB = A, \)
   \[
P(A \mid B) = \frac{P(AB)}{P(B)} = \frac{P(A)}{P(B)} > P(A).\]
3. One can use the conditional probability to express the probability of a complicated event in terms of “simpler” related events.

- Let $A_1, A_2, \ldots, A_n$ are pair wise disjoint sets and their union is $\Omega$. Thus $A_i A_j = \phi$, and $\bigcup_{i=1}^{n} A_i = \Omega$.

- Thus

$$B = B(A_1 \cup A_2 \cup \cdots \cup A_n) = BA_1 \cup BA_2 \cup \cdots \cup BA_n.$$
Conditional Probability

• Since \( A_i \cap A_j = \phi \Rightarrow BA_i \cap BA_j = \phi \),

\[
P(B) = \sum_{i=1}^{n} P(BA_i) = \sum_{i=1}^{n} P(B | A_i)P(A_i).
\]

• If \( A \) and \( B \) are independent, then:

\[
P(A | B) = \frac{P(AB)}{P(B)} = \frac{P(A)P(B)}{P(B)} = P(A).
\]

• Generally,

\[
P(A_i | B) = \frac{P(B | A_i)P(A_i)}{P(B)} = \frac{P(B | A_i)P(A_i)}{\sum_{i=1}^{n} P(B | A_i)P(A_i)}.
\]
Independence

• Independence: Events $A$ and $B$ are independent if

$$P(AB) = P(A)P(B).$$

• This also implies that the following are also independent:

$$\bar{A}, B; \quad A, \bar{B}; \quad \bar{A}, \bar{B}$$
Independence

- The event of zero probability is independent of every other event. If \( P(A) = 0 \),
  \[ P(AB) \leq P(A) = 0 \Rightarrow P(AB) = 0, \quad AB \subset A \]

- Independent events cannot be mutually exclusive,
  \[ P(A) > 0, \quad P(B) > 0 \quad P(AB) > 0. \]

- More generally, a family of events \( \{A_i\} \) are said to be independent, if for every finite sub collection \( A_{i_1}, A_{i_2}, \ldots, A_{i_n} \), we have
  \[ P\left( \bigcap_{k=1}^{n} A_{i_k} \right) = \prod_{k=1}^{n} P(A_{i_k}). \]
Independence

Given a union of \( n \) independent events:

\[ A = A_1 \cup A_2 \cup A_3 \cup \cdots \cup A_n, \]

Then by De-Morgan’s law and using their independence, the following holds:

\[ \overline{A} = \overline{A_1} \overline{A_2} \cdots \overline{A_n} \]

\[
P(\overline{A}) = P(\overline{A_1} \overline{A_2} \cdots \overline{A_n}) = \prod_{i=1}^{n} P(\overline{A_i}) = \prod_{i=1}^{n} (1 - P(A_i)).
\]

\[
P(A) = 1 - P(\overline{A}) = 1 - \prod_{i=1}^{n} (1 - P(A_i)),
\]
Independence

• Example: Three switches connected in parallel operate independently. Each switch remains closed with probability $p$.

(a) Find the probability of receiving an input signal at the output.

(b) Find the probability that switch $S_1$ is open given that an input signal is received at the output.
Independence

• To answer (a), let $A_i = \text{“Switch } S_i \text{ is closed”}$ and $P(A_i) = p$, for $i = 1, 2, 3$. Since the switches operate independently, we have

$$P(A_i A_j) = P(A_i) P(A_j); \quad P(A_1 A_2 A_3) = P(A_1) P(A_2) P(A_3).$$

• Let $R = \text{“input signal is received at the output”}$. Hence,

$$R = A_1 \cup A_2 \cup A_3.$$

$$P(R) = P(A_1 \cup A_2 \cup A_3) = 1 - (1 - p)^3 = 3p - 3p^2 + p^3.$$
Independence

- Alternatively

\[ P(R) = P(R \mid A_1)P(A_1) + P(R \mid \overline{A_1})P(\overline{A_1}). \]

\[ P(R \mid A_1) = 1 \]

\[ P(R \mid \overline{A_1}) = P(A_2 \cup A_3) = 2p - p^2 \]

\[ P(R) = p + (2p - p^2)(1 - p) = 3p - 3p^2 + p^3, \]
Independence

• Note that the events $A_1$, $A_2$, $A_3$ do not form a partition, since they are not mutually exclusive. Obviously any two or all three switches can be closed (or open) simultaneously. Therefore,

$$P(A_1) + P(A_2) + P(A_3) \neq 1.$$
Independence

• To answer (b), we apply Bayes theorem

\[
P(\overline{A}_1 \mid R) = \frac{P(R \mid \overline{A}_1)P(\overline{A}_1)}{P(R)} = \frac{(2p - p^2)(1-p)}{3p - 3p^2 + p^3} = \frac{2 - 2p + p^2}{3p - 3p^2 + p^3}.
\]

• Because of the symmetry of the switches, we also have

\[
P(\overline{A}_1 \mid R) = P(\overline{A}_2 \mid R) = P(\overline{A}_3 \mid R).
\]
Bernoulli Trial

• Knowledge of the success or failure of an Event $A$ in an $n$ independent trial is fundamental for many interesting problems in communication and computer networks.

• Often our interest is to find out the probability that Event $A$ occurs exactly $k$ times, $k \leq n$, in $n$ trials ($P_\omega(\omega)$), given the probability of $A$ occurring in a single trial is $p$. 
Bernoulli Trial

\[ P_0(\omega) = P(\{\xi_{i_1}, \xi_{i_2}, \cdots, \xi_{i_k}, \cdots, \xi_{i_n} \}) = P(\{\xi_{i_1} \})P(\{\xi_{i_2} \}) \cdots P(\{\xi_{i_k} \}) \cdots P(\{\xi_{i_n} \}) \]

\[ = \underbrace{P(A)P(A) \cdots P(A)}_{k} \underbrace{P(\overline{A})P(\overline{A}) \cdots P(\overline{A})}_{n-k} = p^k q^{n-k}. \]

- However the \( k \) occurrences of \( A \) can occur in any particular location inside \( \omega \).
- Let \( \omega_1, \omega_2, \cdots, \omega_N \) represent all such events in which \( A \) occurs exactly \( k \) times. Then:

"A occurs exactly \( k \) times in \( n \) trials" = \( \omega_1 \cup \omega_2 \cup \cdots \cup \omega_N \).

\( \omega_i \) are mutually exclusive and equi-probable
• Thus:

\[ P("A\ occurs\ exactly\ k\ times\ in\ n\ trials") = \sum_{i=1}^{N} P_0(\omega_i) = NP_0(\omega) = Np^k q^{n-k}, \]

• Where:

\[ N = n(n-1)\cdots(n-k+1) = \frac{n!}{k!(n-k)!} = \binom{n}{k} \]

• Generally:

\[ P_n(k) = P("A\ occurs\ exactly\ k\ times\ in\ n\ trials") = \binom{n}{k} p^k q^{n-k}, \quad k = 0,1,2,\cdots,n, \]
Similarly, we may be interested to find out the probability of at least \( k \) occurrences in an \( n \) trial, in which case:

\[
P(X_0 \cup X_1 \cup \cdots \cup X_n) = \sum_{k=0}^{n} P(X_k) = \sum_{k=0}^{n} \binom{n}{k} p^k q^{n-k}.
\]

Note that Bernoulli trial consists of repeated independent and identical experiments each of which has only two outcomes \((A\) or its complement), with \( P(A) = p \), and \( P(\overline{A}) = q \).
Bernoulli Trial

• Often, it is interesting to determine the most likely value of $k$ for a given $n$ and $p$

• The most probable value of $k$ is the number which maximizes (see the figure below) $P_n(k)$.

\[
P_n(k) \quad n = 12, \quad p = 1/2.
\]
Bernoulli Trial

- To obtain this value, consider the ratio:

\[
\frac{P_n(k-1)}{P_n(k)} = \frac{n! \, p^{k-1} \, q^{n-k+1}}{(n-k+1)! \, (k-1)!} \cdot \frac{(n-k)! \, k!}{n! \, p^k \, q^{n-k}} = \frac{k \, q^{k-1} \, p}{n-k+1}.
\]

- Thus \( P_n(k) \geq P_n(k-1) \), if \( k(1-p) \leq (n-k+1)p \)

or \( k \leq (n+1)p \). Thus \( P_n(k) \) as a function of \( k \) increases until \( k = (n+1)p \)

- If \( k \) is an integer, or the largest integer \( k_{max} \) is less than \( (n+1)p \), \( P_n(k) \) represents the most likely number of successes in \( n \) trials.
Bernoulli’s Theorem

- Let $A$ denote an event whose probability of occurrence in a single trial is $p$. If $k$ denotes the number of occurrences of $A$ in $n$ independent trials, then

$$P \left( \left\{ \left| \frac{k}{n} - p \right| > \varepsilon \right\} \right) < \frac{pq}{n \varepsilon^2}.$$ 

- The above expression states that the frequency definition of probability of an event $\frac{k}{n}$ and its axiomatic definition ($p$) can be made compatible to any degree of accuracy.
Bernoulli’s Theorem

• Proof

\[
\sum_{k=0}^{n} k P_n(k) = \sum_{k=1}^{n-1} k \frac{n!}{(n-k)!k!} p^k q^{n-k} = \sum_{k=1}^{n} \frac{n!}{(n-k)!(k-1)!} p^k q^{n-k}
\]

\[
= \sum_{i=0}^{n-1} \frac{n!}{(n-i-1)!i!} p^{i+1} q^{n-i-1} = np \sum_{i=0}^{n-1} \frac{(n-1)!}{(n-1-i)!i!} p^i q^{n-1-i}
\]

\[
= np(p + q)^{n-1} = np.
\]

• Proceeding in a similar manner:

\[
\sum_{k=0}^{n} k^2 P_n(k) = \sum_{k=1}^{n} k \frac{n!}{(n-k)!(k-1)!} p^k q^{n-k} = \sum_{k=2}^{n} \frac{n!}{(n-k)!(k-2)!} p^k q^{n-k}
\]

\[
+ \sum_{k=1}^{n} \frac{n!}{(n-k)!(k-1)!} p^k q^{n-k} = n^2 p^2 + npq.
\]
Bernoulli’s Theorem

• Note that:

\[
\left| \frac{k}{n} - p \right| > \varepsilon \quad \text{is equivalent to} \quad (k - np)^2 > n^2 \varepsilon^2,
\]

\[
\sum_{k=0}^{n} (k - np)^2 P_n(k) > \sum_{k=0}^{n} n^2 \varepsilon^2 P_n(k) = n^2 \varepsilon^2.
\]

\[
\sum_{k=0}^{n} (k - np)^2 P_n(k) = \sum_{k=0}^{n} k^2 P_n(k) - 2np \sum_{k=0}^{n} k P_n(k) + n^2 p^2
\]

\[= n^2 p^2 + npq - 2np \cdot np + n^2 p^2 = npq.\]
Bernoulli’s Theorem

- The middle expression in the previous slide can be expressed as:

\[ \sum_{k=0}^{n} (k - np)^2 P_n(k) = \sum_{|k-np| \leq n\varepsilon} (k - np)^2 P_n(k) + \sum_{|k-np| > n\varepsilon} (k - np)^2 P_n(k) \]

\[ \geq \sum_{|k-np| > n\varepsilon} (k - np)^2 P_n(k) > n^2 \varepsilon^2 \sum_{|k-np| > n\varepsilon} P_n(k) \]

\[ = n^2 \varepsilon^2 P\{|k - np| > n\varepsilon\}. \]

- Rearranging and combining terms yields:

\[ P\left(\left\{\left|\frac{k}{n} - p\right| > \varepsilon\right\}\right) < \frac{pq}{n \varepsilon^2}. \]