1 Probability

Each possible outcome of an experiment is called a sample point. The set of all sample points is called sample space, denoted by \( \Omega \). For example, in the experiment of 100 coin tosses, the sample space is \( \{0, 1\}^{100} \). An event \( E \) is a subset of \( \Omega \). For example in the above experiment, the set of coins tosses with even number of heads is an event. Two events \( E_1 \) and \( E_2 \) are disjoint if \( E_1 \cap E_2 = \emptyset \). A collection of events \( E_1, E_2 \cdots \) are mutually disjoint if for every \( i \neq j \), \( E_i \cap E_j = \emptyset \). A function \( P \) from \( \Omega \) to \([0, 1]\) is called a probability function if the following hold: 1) \( \sum_{x \in \Omega} P(x) = 1 \), 2) for every event \( E \), \( 0 \leq P(E) \leq 1 \), for every collection of mutually disjoint events \( \{E_i\}_i \) (possibly countably infinite), \( P(\bigcup_i E_i) = \sum_i P(E_i) \). Here \( P(E) \) denotes \( \sum_{x \in E} P(x) \). Probability of an event \( E \) is \( \sum_{x \in E} P(x) \). Given an event \( E \), \( E = \Omega - E \). Observe that \( P(E) = 1 - P(\overline{E}) \). Two events \( A \) and \( B \) are independent \( \Pr(A \cap B) = \Pr(A) \times \Pr(B) \). Observe that if \( A \) and \( B \) are independent, then \( \overline{A} \) and \( \overline{B} \) are also independent.

Union Bound. If \( E_1, E_2, \cdots E_m \) are events, then

\[
P(\bigcup_i E_i) \leq \sum_i P(E_i).
\]

Suppose we toss a fair coin \( n \) times, what is the probability that we see \( 2 \log n \) successive heads? Let us denote this event with \( E \). Let us consider the following events \( E_1, E_2, \cdots E_{n-2\log n+1} \), where \( E_i \) is the event that \( i^{th}, (i + 1)^{st}, \cdots, (i + 2\log n - 1) \) tosses are all heads. Observe that for each \( E_i, P(E_i) = (1/2)^{2\log n} = 1/n^2 \). Note that \( E = \bigcup_i E_i \). By Union bound

\[
P(E) \leq P(E_1) + P(E_2) + \cdots + P(E_{n-2\log n+1}) \leq \frac{n - 2\log n + 1}{n^2} \leq \frac{1}{n}
\]

Thus the probability the we see \( 2 \log n \) consecutive heads is very small.

Now let us consider the probability that we see \( \log n/2 \) consecutive heads. Let us denote this event with \( E \). Divide the \( n \) coin tosses into \( m = 2n/\log n \) groups of size \( \log n/2 \). Let us call these groups \( G_1, \cdots G_m \). Let \( E_i \) denote the probability that all coin tosses in group \( G_i \) are heads. Note that the events \( E_1, E_2 \cdots E_m \) are all mutually independent, and for every \( i, 1 \leq i \leq m \), \( P(E_i) = (1/2)^{\log n/2} = 1/\sqrt{n} \). Since \( \bigcup_i E_i \subseteq E \),

\[
P(E) \geq P(\bigcup_i E_i) = 1 - P(\bigcap_i E_i) = 1 - P(\overline{E_1} \cap \overline{E_2} \cap \cdots \cap \overline{E_m})
\]
Since $E_1, \ldots, E_m$ are mutually independent and $P(E_i) = 1 - \frac{1}{\sqrt{n}}$,

\[
P(E_1 \cap E_2 \cap \cdots \cap E_m) = \Pi_i P(E_i) = (1 - \frac{1}{\sqrt{n}})^{2n/\log n} = ((1 - \frac{1}{\sqrt{n}})^{\sqrt{n}})^{2\sqrt{n}/\log n} = (1/e)^{2\sqrt{n}/\log n}
\]

Thus the probability that we see at least $\log n/2$ consecutive heads is very close to 1. Thus with very high probability that numbers if consecutive heads lie between $\log n/2$ and $2\log n$.

A random variable $X$ is a function from sample space to the real numbers, $X: \Omega \rightarrow \mathbb{R}$. Given a random variable $X$ and a value $\alpha$,

\[
\begin{align*}
\Pr[X = \alpha] &= \Pr\{v \mid X(v) = \alpha\} \\
\Pr[X \geq \alpha] &= \Pr\{v \mid X(v) \geq \alpha\} \\
\Pr[X < \alpha] &= \Pr\{v \mid X(v) < \alpha\}
\end{align*}
\]

The expectation of a random variable is defined as

\[
E(X) = \sum_{\alpha} \Pr(X = \alpha) \times \alpha.
\]

$E(X)$ is average value of $X$.

Two random variable $X$ and $Y$ are indepnet if for every $\alpha$ and $\beta$ the events $\{v \mid X(v) = \alpha\}$ and $\{v \mid Y(v) = \beta\}$ are independent. This definition can be extended to more than two random variables.

Consider the experiment with $n$ fair coin tosses. We can define a random variable $X$ to be the number of heads. A very useful property of expectation is that it is linear, i.e., If $X$ and $Y$ are two random variable and $a$ is a real number, then

\[
E(aX + Y) = aE(X) + E(Y).
\]

This property can be used to compute the expectation of some random variables easily. Let $X$ be the number of heads in $n$ fair coin tosses. For $1 \leq i \leq n$, define random variable $X_i$ as follows: $X_i = 1$, if the $i$th coin toss is a head, else $X_i$ is zero.

\[
E(X_i) = \Pr(X_i = 1) \times 1 + \Pr(X_i = 0) \times 0 = 1/2.
\]

It is clear that $X = X_1 + X_2 + \cdots + X_n$.

\[
E(X) = E(X_1) + E(X_2) + \cdots + E(X_n) = 1/2 + 1/2 + \cdots + 1/2 = n/2.
\]

Let $\phi$ be a 3-CNF formula with $m$ clauses. Suppose if we randomly pick an assignment to the variables of the $\phi$. Let $X$ denote the number of classes satisfied. What is the expectation of $X$?
Let $X_i$ be a random variable whose value is 1 of $i$th clause satisfied. Otherwise $X_i = 0$. Observe that $X = X_1 + \cdots + X_m$. Since $E(X_i) = 7/8$, we have that $E(X) = 7m/8$.

Let $X$ be a random variable that takes values in $\{0, 1, 2, \cdots\}$.

**Claim.** $E(X) = \sum_i P(X > i)$.

Thus

$$
\sum_i P(X > i) = \sum_{i=1}^{\infty} i \times P(X = i) = E(X)
$$

Suppose we have biased coin with probability of head being $p$. Let us consider the following experiment: Toss the coin till head appears. Let $X$ be a random variable that denotes the number of coin tosses made. What is the expectation of $X$? Since $X$ takes values in $\{0, 1, \cdots\}$, by previous Claim,

$$E(X) = \sum_i P(X > i).$$

$P(X > i)$ is the probability that the first $i$ tosses result in tails. Thus $P(X > i) = (1 - p)^i$. Thus

$$E(X) = \sum_i P(X > i) = \sum_i (1 - p)^i = \frac{1}{p}$$

Suppose we have two random variables $X$ and $Y$ which take the following values with equal probability.

$X: 47, 48, 49, 50, 51, 52, 53$

$Y: 20, 30, 40, 50, 60, 70, 80$

The expectation of both $X$ and $Y$ is 50. The values taken by $Y$ are more spread out than the values taken by $X$. We capture this fact by variance. The variance of a random variable $X$ is defined as follows.

$$Var(X) = E((X - E(X))^2).$$

Thus variance is (square of) the average distance of $X$ from its Expectation.

Given a random variable $X$, we are often interested in computing probabilities such as $Pr(X > v)$, $Pr(|X - E(X)| > a)$. The following three inequalities help us to estimate this.

**Markov’s Inequality:** Let $X$ be a nonnegative random variable.

$$Pr(X > v) \leq E(X)/v.$$
In other words,\[
\Pr(X > \alpha E(X)) \leq 1/\alpha.
\]

**Chebyshev’s Inequality:** Let \( X \) be random variable.

\[
\Pr(|X - E(X)| \geq \delta) \leq \frac{\text{Var}(X)}{\delta^2}.
\]

**Chernoff’s Bound:** Let \( X_1, X_2, \ldots, X_m \) be independent random variables that take values between 0 and 1. Let \( E(X_1) = E(X_2) = \cdots = E(X_m) = p \). Let \( X = X_1 + X_2 + \cdots + X_m \).

\[
\Pr[|X/m - p| \geq p\delta] \leq 2e^{-\delta^2mp/2}.
\]

## 2 Probabilistic Algorithms

A probabilistic algorithm is an algorithm that can toss coins during its computation. Note that the outcome of an probabilistic algorithm need not be unique. If we run the algorithm twice, we may get two different outputs, the outputs may depend on the result of coin tosses.

We say a probabilistic algorithm \( A \) computes a function \( f \), if the \( A \) outputs the correct value of \( f \) with high probability. We have to define an appropriate notion of “high”. We consider the following definition.

Let \( f : \Sigma^* \to \Sigma^* \). A probabilistic algorithm \( A \) computes \( f \), if

\[
\forall x, \Pr(A(x) = f(x)) \geq 2/3.
\]

Above definition appears weak. The error probability of the above algorithm is very high—the algorithm can go wrong 1/3rd of time. However, as the next result shows, we can design a new algorithm \( B \) whose error probability is much small, and the running time of \( B \) is a little more than the running time of \( A \).

Let \( A \) be a probabilistic algorithm that computes \( f \) in time \( t(n) \). Then there exists a probabilistic algorithm \( B \) whose running time is \( O(nt(n)) \) and

\[
\forall x, \Pr(A(x) = f(x)) \geq 1 - 1/2^n, |x| = n.
\]

The algorithm \( B \) works as follows.

1. input: \( x, |x| = n \). Let \( m = 27n \).
2. for \( i = 1 \) to \( m \) compute
3. \( a_i = A(x) \).
4. Output the majority value of \( a_1, \ldots, a_m \). If there is no majority output \( a_1 \).

We claim that \( B \) computes \( f \) with very high probability. We define few random variables. For \( 1 \leq i \leq m \) define \( X_i \) as follows: \( X_i = 1 \), if \( a_i = f(x) \), \( X_i = 0 \), if \( a_i \neq f(x) \). Note that \( E(X_i) \geq 2/3 \). Let \( X = X_1 + X_2 + \cdots + X_m \). Note that \( B \) outputs a wrong value, if majority of \( a_i \)'s are wrong.
Thus $B$ outputs a wrong value if majority of $X_i$'s are zero. We can use Chernoff’s bound to show that the probability of such event is very small.

\[
\Pr[ \text{B outputs a wrong value}] = \Pr[ \text{majority of } a_i \text{'s are wrong}] \\
= \Pr[ \text{majority of } X_i \text{'s are zeros}] \\
= \Pr[X \leq m/2] \\
= \Pr[X/m \leq 1/2] \\
\leq \Pr[|X/m - 2/3| \geq 1/6] \\
\leq 1/2^n.
\]

Thus

\[
\Pr[B(x) = f(x)] \geq 1 - 1/2^{|x|}.
\]