CS 535: Notes II

1 Independence

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.

Consider the following experiment. Toss a fair coin, if the outcome is H then output 1, otherwise output 0. Toss another fair coin, if the outcome is H then output 1, otherwise output 0. Finally output the *xor* of the first two bits. In this experiment the sample space is $\{000, 011, 101, 110\}$ and each sample point has probability 1/4. Let E_i, $1 \leq i \leq 3$, be the event that the ith bit is 1. We can see that E_1 and E_3 are independent, similarly E_2 and E_3 are also independent. Now consider the experiment where the second coin is a biased coin with probability of H being 3/4 (and the first coin is a fair coin). Are E_2 and E_3 independent? How about E_1 and E_3?

Let us return to the earlier experiment of tossing a fair coin n times. 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, $1 \leq i \leq m$, $P(E_i) = (1/2)^{\log n/2} = \frac{1}{\sqrt{n}}$. Since $\cup_i E_i \subseteq E$,

$$
\begin{aligned}
P(E) &\geq P(\cup_i E_i) \\
&= 1 - P(\overline{\cup_i E_i}) \\
&= 1 - P(\overline{E_1} \cap \overline{E_2} \cap \cdots \cap \overline{E_m})
\end{aligned}
$$

Since $\overline{E_1}, \cdots, \overline{E_m}$ are mutually independent and $P(\overline{E_i}) = 1 - \frac{1}{\sqrt{n}}$,

$$
\begin{aligned}
P(\overline{E_1} \cap \overline{E_2} \cap \cdots \cap \overline{E_m}) &= \Pi_i P(\overline{E_i}) \\
&= (1 - \frac{1}{\sqrt{n}})^{2n/\log n} \\
&= ((1 - \frac{1}{\sqrt{n}})^{\sqrt{n}})^{2\sqrt{n}/\log n} \\
&= (1/e)^{2\sqrt{n}/\log n}
\end{aligned}
$$

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 $2 \log n$.

2 Random Variables

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

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

Examples of random variables:

- Toss a coin n times. Number of heads.
- Toss a coin n times. Maximum number of consecutive heads.
- Roll a dice twice. Sum of the two outcomes.
- Toss a coin. The value of X is 1 if the outcome is head; otherwise X is 0.

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 independent if for every α and β 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 random variable from fourth experiment above. The expectation of X is calculated as follows:

$$1 \times \Pr[X = 1] + 0 \times \Pr[X = 0] = 1/2.$$

Throw a dice. Let Y be the random variable that denotes the outcome.

$$E[Y] = \sum_{i=1}^{6} i \times \Pr[Y = i] = 1/6 + 2/6 + 3/6 + 4/6 + 5/6 + 6/6 = 3.5$$

Consider the experiment with n fair coin tosses. We can define a random variable X to be the number of heads. What is the expectation of X? It can be calculated by using the definition of expectation.

$$\begin{aligned}
E[X] &= \sum_{i=0}^{n} \Pr[X = i] \times i \\
&= \sum_{i=0}^{n} \binom{n}{i} \frac{1}{2^n}
\end{aligned}$$

It is not that easy to compute the value of the expression $\sum_{i=0}^{n} \binom{n}{i} \frac{1}{2^n}$. We will compute the expectation using a different route. 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 $1 \le i \le n$, define random variable X_i as follows: $X_i = 1$, if the ith 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 $X = X_1 + X_2 + \cdots + X_n$.

$$\begin{aligned} E(X) &= E(X_1) + E(X_2) + \cdots + E(X_n) \\ &= 1/2 + 1/2 + \cdots + 1/2 \\ &= n/2. \end{aligned}$$

Let $\phi(x_1, \cdots, x_n)$ be a 3-CNF formula with n variables and m clauses. Randomly assign a value to x_i, $1 \le i \le n$, from $\{T, F\}$. Let X denote the number of clauses satisfied. We can again calculate the expectation of X using linearity. Let X_i denote a random variable whose value is 1 is ith clause of ϕ is satisfied; otherwise X_i is 0. Note that $X = X_1 + \cdots + X_m$. Thus $E[X] = E[X_1 + \cdots + X_m] = E[X_1] + \cdots + E[X_m]$. Given X_i what is the expectation of X_i?

$$E[X_i] = 1 \times \Pr[X_i = 1] + 0 \times \Pr[X_i = 0] = \Pr[X_i = 1]$$

Since a clause has three distinct variables, and the clause is a disjunction; the probability that a random assignment satisfies the clause is $7/8$. Thus $\Pr[X_i = 1] = 7/8$. Thus $E[X_i] = 7/8$, thus $E[X] = 7m/8$.

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

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

$$\begin{aligned} P(X > 0) &= P(X = 1) + P(X = 2) + P(X = 3) + \cdots \\ P(X > 1) &= + P(X = 2) + P(X = 3) + \cdots \\ P(X > 2) &= + P(X = 3) + \cdots \end{aligned}$$

Thus

$$\begin{aligned} \sum_i P(X > i) &= \sum_{i=1} i \times P(X = i) \\ &= E(X) \end{aligned}$$

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

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