Lecture 1 – Events and Probability

A **discrete probability space** is a model to capture the set of discrete outcomes of a random process. E.g., picking a prime from the first 100 numbers. More formally, it is a pair (Ω, p), where Ω is a countable set (possibly finite) and $p : \Omega \to \mathbb{R}_{\geq 0}$ is the probability function such that $\sum_{\omega \in \Omega} p_\omega = 1$. A subset $A \subseteq \Omega$ is called an **event** and its probability is naturally defined as $\Pr(A) := \sum_{\omega \in A} p_\omega$. More examples:

1. Place r balls in n bins uniformly at random. Each assignment of balls corresponds to a map from $[r] \to [n]$. Therefore, Ω is the set of all such functions, and because of uniformity each mapping is assigned the probability $1/n^r$.
2. Randomly shuffle a deck of n cards. Ω is the set of all permutations with probability $1/n!$.
3. Throw a fair die n times. $\Omega = [6]^n$, that is, the set of n-tuples with entries from $1, \ldots, 6$ and probability $1/6^n$.

Claim: If $A_1, A_2, \ldots,$ are a countable set of pairwise mutually disjoint events then

$$\Pr(\cup_i A_i) = \sum_i \Pr(A_i).$$

Claim: For any two events $A, B \subseteq \Omega$, we have

$$\Pr(A \cup B) = \Pr(A) + \Pr(B) - \Pr(A \cap B).$$

As a consequence, by induction we have the **union bound**: For any countable sequence of events $A_1, A_2, \ldots,$

$$\Pr(\bigcup_{i \geq 1} A_i) \leq \sum_{i \geq 1} \Pr(A_i). \tag{1}$$

Two events $A, B \subseteq \Omega$ are said to be **independent** if

$$\Pr(A \cap B) = \Pr(A) \Pr(B). \tag{2}$$

A finite sequence of events $A_1, \ldots, A_k$ are said to be **mutually independent** iff for all $I \subseteq [k]$

$$\Pr(\bigcap_{i \in I} A_i) = \prod_{i \in I} \Pr(A_i).$$

What is the probability that non of the mutually independent events occur? It is $1 - \Pr(\cup_i A_i)$. By inclusion-exclusion we get

$$1 - \Pr(\bigcup_i A_i) = 1 - \sum_{\ell \geq 1} (-1)^{\ell - 1} \sum_{I \subseteq [k] : |I| = 1} \Pr(A_I) = \prod_{i=1}^{k} (1 - \Pr(A_i)).$$

¶1. Example : A set of events that are pairwise independent but not mutually independent. Toss three fair coins independently. The probability space is the set of all 3-tuples with entries {head, tail}. Let A_{ij} be the event that the ith and jth coin have the same value. Then $\Pr(A_{ij}) = 1/2$. If $i \neq j \neq k$ then

$$\Pr(A_{ij} \cap A_{jk}) = \Pr(\text{all coins are same}) = \frac{1}{4}.$$

So the events are pairwise independent. However, the events are not mutually independent since

$$\Pr(A_{12} \cap A_{23} \cap A_{13}) = \Pr(\text{all coins are same}) = \frac{1}{4}$$

which is larger than $1/8$, the product of the individual probabilities; note that $A_{12} \cap A_{23}$ implies A_{13} and so intuitively the three events are not mutually independent.

The **conditional probability** that event B occurs given A is defined as

$$\Pr(B|A) := \frac{\Pr(B \cap A)}{\Pr(A)}.$$

Note the normalization by the probability of event A.

1 Applications

Testing if two univariate polynomials $f(x)$ and $g(x)$ of degree at most d are the same. This is interesting when the polynomials are not given in their coefficient representation and are in some other form, for instance, an algebraic circuit. A simple probabilistic algorithm does the following: For a constant $c > 1$, pick a number a uniformly at random from the set $\{1, \ldots, cd\}$. If $f(a) \neq g(a)$, then output not equivalent; otherwise, output equivalent. Clearly, the algorithm can go wrong if the two polynomials are not equivalent but $f(a) = g(a)$ and it outputs equivalent. What is the probability that the algorithms errs? The probability space is the set of numbers $1, \ldots, cd$ each picked with probability $1/(cd)$. The event A that we go wrong is if a is a root of the polynomial $f(x) - g(x)$. There are at most d such roots and probability that a is one of them is at most $1/c$. Therefore, $\Pr(\text{Algo is wrong}) \leq 1/c$.

To boost this probability, we can run the experiment a couple of times say k times. In each run, we have a choice: either pick a number with replacement or without. We would expect the probability of making error in the latter format to be smaller than the former, and indeed that is the case. Let A_i be the even that we fail with replacement, where each of the runs are independent. Therefore, the probability that the algorithm fails to detect non-equivalence all the k times is

$$\Pr(\cap_{i=1}^{k} A_i) = \prod_{i=1}^{k} \Pr(A_i) \leq c^{-k}.$$

To consider probabilities without replacement, we need conditional probabilities. The probability that we fail at ith step given we have failed in the previous steps is if we pick a root of the polynomial $f - g$ that is different from the $(i-1)$ roots picked in the earlier rounds (as we also failed in the earlier rounds); this can be done in $(d - (i-1))$ ways out of the $(cd - (i-1))$ remaining numbers, that is,

$$\Pr(A_i | A_1, \ldots, A_{i-1}) \leq \frac{d - (i-1)}{cd - (i-1)}.$$

By induction,

$$\Pr(\bigcap_i A_i) = \prod_{i=1}^{k} \Pr(A_i | A_1, \ldots, A_{i-1}) \leq \prod_{i=1}^{k} \frac{d - (i-1)}{cd - (i-1)} \leq c^{-k}$$

since for $i \geq 1$, $(d - (i-1)/(cd - (i-1)) \leq 1/c$.

1.1 Verifying Matrix Multiplication

$A, B, C \in \mathbb{Z}^{n \times n}$, check if $AB = C$ modulo two. Let $\boldsymbol{r} \in 0, 1^n$ be chosen uniformly at random. Check if $AB\boldsymbol{r} = C\boldsymbol{r}$; if not output not equal, otherwise yes. Simple randomized algorithm taking $\Theta(n^2)$ operations.

Claim: If $AB \neq C$, then $\Pr AB\boldsymbol{r} = C\boldsymbol{r} \leq \frac{1}{2}$. Let $D := AB - C \neq 0$. Then $D\boldsymbol{r} = 0$ gives us a n linear relations on the coordinates of $\boldsymbol{r}$. In particular, let's assume that $d_{11} \neq 0$ then we get that

$$r_1 \equiv -\sum_{j>1} r_j d_j / d_{11} \mod 2$$

That is once the r_j's, for $j = 2, \ldots, n$, are fixed then there is one choice of $r_1 \mod 2$ that works. Now we observe that choosing $\boldsymbol{r} \in \{0,1\}^n$ or choosing individual r_i's in $\{0,1\}$ give us the same probability function over $\{0,1\}^n$. Therefore, we can think of choosing the coordinates $r_2, \ldots, r_n$ and then the probability that r_1 takes the value above is at most $1/2$. To formalize this, we need the following:

Lemma 1 (Law of Total Probability) *If $A_1, \ldots,$ are mutually disjoint events in Ω that partition it then*

$$\Pr(B) = \sum_{i \geq 1} \Pr(B \cap A_i).$$

Using this we have the following chain:

$$\begin{aligned}
\Pr(AB\boldsymbol{r} \equiv C\boldsymbol{r}) &= \sum_{\boldsymbol{x} \in \{0,1\}^{n-1}} \Pr(AB\boldsymbol{r} \equiv C\boldsymbol{r} \bigcap (r_2, \ldots, r_n) = \boldsymbol{x}) \\
&\leq \sum_{\boldsymbol{x} \in \{0,1\}^{n-1}} \Pr(r_1 \equiv -\sum_{j>1} r_j d_j / d_{11} \mod 2 \bigcap (r_2, \ldots, r_n) = \boldsymbol{x}) \\
&= \sum_{\boldsymbol{x} \in \{0,1\}^{n-1}} \Pr(r_1 \equiv -\sum_{j>1} r_j d_j / d_{11} \mod 2) \cdot \Pr((r_2, \ldots, r_n) = \boldsymbol{x}) \\
&\leq \frac{1}{2} \sum_{\boldsymbol{x} \in \{0,1\}^{n-1}} \Pr((r_2, \ldots, r_n) = \boldsymbol{x}) \\
&= \frac{1}{2}.
\end{aligned}$$

1.2 Schwartz-Zippel Lemma

We now consider the following generalization of the 1-d case of polynomial identity testing.

Lemma 2 *Let $F(x_1, \ldots, x_n) \in K[x_1, \ldots, x_n]$ be a non-zero polynomial total degree d polynomial over a field K. Let $S \subseteq K$ be a finite subset and $r_1, \ldots, r_n$ be chosen independently at random from S. Then*

$$\Pr(F(\boldsymbol{r}) = 0) \leq \frac{d}{|S|}.$$

Proof. Proof is by induction on n. For $n = 1$ the result follows from the fundamental theorem of algebra.

Since $F \neq 0$, if we express as $F(x_1, \ldots, x_n) = \sum_{i=0}^{d} x_1^i F_i(x_2, \ldots, x_n)$ then there is a largest index j such that $F_j \neq 0$ and $\deg(F_j) \leq d - j$. By the law of total probability we have:

$$\begin{aligned}
\Pr(F(\boldsymbol{r}) = 0) &= \Pr(F(\boldsymbol{r}) = 0 \cap F_j(r_2, \ldots, r_n) = 0) + \Pr(F(\boldsymbol{r}) = 0 \cap F_j(r_2, \ldots, r_n) \neq 0) \\
&= \Pr(F(\boldsymbol{r}) = 0 | F_j(r_2, \ldots, r_n) = 0) \Pr(F_j(r_2, \ldots, r_n) = 0) \\
&\quad + \Pr(F(\boldsymbol{r}) = 0 \cap F_j(r_2, \ldots, r_n) \neq 0) \Pr(F_j(r_2, \ldots, r_n) = 0) \\
&\leq \Pr(F_j(r_2, \ldots, r_n) = 0) + \Pr(F(\boldsymbol{r}) = 0 \cap F_j(r_2, \ldots, r_n) \neq 0).
\end{aligned}$$

By induction the first term is $(d-j)/|S|$ and second term is $j/|S|$, which gives us the desired result. **Q.E.D.**