# Lecture Notes, Week 7

*Math 480A2: Mathematics of Blockchain Protocols, Fall 2022*

Lecturer: Bryan Gillespie

## Probability and the Schwartz-Zippel Lemma

**Definition 1.** Let $X$ be a finite or countably infinite set. Then a **probability function** on $X$ is a function $p : X \to [0, 1]$ satisfying

$$\sum_{x \in X} p(x) = 1$$

The value $p(x)$ for $x \in X$ is referred to as the **probability of outcome** $x$, and the set $X$ along with a probability distribution function is called a **discrete probability space**. A subset $A \subseteq X$ is called an **event** in $X$, and in general we write $p(A)$ for the total probability of all outcomes in $A$,

$$p(A) \coloneqq \sum_{x \in A} p(x)$$

**Remark 2.** If $(X, p)$ is a discrete probability space, then sometimes the notation Pr will be used to denote the probability function of $X$ without referring to its name explicitly. In this case, we may discuss a discrete probability space without giving an explicit name for a probability function, understanding that any references to probabilities on this space will be accomplished using this generic notation.

**Proposition 3.** *Let $X$ be a discrete probability space. Then*

- *(Monotonicity) If $A, B \subseteq X$ with $A \subseteq B$, then* $\Pr(A) \leq \Pr(B)$
- *(Disjoint additivity) If $A_1, A_2, \ldots \subseteq X$ are disjoint, then* $\Pr(\cup A_i) = \sum \Pr(A_i)$
- *(Subadditivity) If $A_1, A_2, \ldots \subseteq X$, then* $\Pr(\cup_i A_i) \leq \sum_i \Pr(A_i)$

*Proof.* Monotonicity follows from the definition of $\Pr(A)$ for an event $A \subseteq X$: if $A \subseteq B$, then

$$\Pr(A) = \sum_{x \in A} \Pr(x) \leq \sum_{x \in B} \Pr(x) = \Pr(B)$$

Disjoint additivity is a consequence of the fact that absolutely convergent series can be re-ordered. If $A_1, A_2, \ldots \subseteq X$ are disjoint, then

$$\Pr(\cup A_i) = \sum_{x \in \cup A_i} \Pr(x) = \sum_{i=1}^{\infty} \sum_{x \in A_i} \Pr(x) = \sum_{i=1}^{\infty} \Pr(A_i)$$

Finally, subadditivity can be seen as a consequence of monotonicity and disjoint additivity. Let $B_1 = A_1$, and for $i > 1$ let $B_i = A_i \setminus \cup_{j=1}^{i-1} A_j$. Then the sets $B_i$ are a disjoint collection satisfying $\cup_i B_i = \cup_i A_i$ and $B_i \subseteq A_i$ for each $i$, from which we conclude

$$\Pr(\cup_i A_i) = \Pr(\cup_i B_i) = \sum_i \Pr(B_i) \leq \sum_i \Pr(A_i)$$

□

The subadditivity property is sometimes also referred to as the *union bound*, as it bounds the probability of a union of events by the sum of the individual probabilities of the events.

**Example 4.** Let $X$ be a finite set, and let $\Pr(x) = 1/|X|$ for each $x \in X$. This is called the **uniform distribution** on $X$.

**Definition 5.** Let $X$ be a discrete probability space. The **indicator function** $\chi_A$ of an event $A \subseteq X$ is the function on $X$ taking value 1 if $x \in A$ and 0 if $x \notin A$. Then in particular, we can write

$$\Pr(A) = \sum_{x \in X} \chi_A(x) \Pr(x)$$

If $A$ and $B$ are events, then $\chi_{A \cap B} = \chi_A \chi_B$, and $\chi_{X \setminus A} = 1 - \chi_A$.

**Definition 6.** Let $X$ be a discrete probability space, and let $A, B$ be events with $\Pr(B) > 0$. Then the **conditional probability of** $A$ **given** $B$, written $\Pr(A \mid B)$, is defined as

$$\Pr(A \mid B) = \frac{\Pr(A \cap B)}{\Pr(B)}$$

Conditional probability describes the likelihood that an outcome in $A$ occurs when restricting the outcomes to only those in $B$. Notice in particular that a conditional probability is a probability, and takes values in $[0, 1]$.

**Definition 7.** Let $X$ be a discrete probability space. Then a **random variable** on $X$ is a function with domain $X$.

A random variable should be thought of as an outcome associated with value randomly sampled from $X$ with its underlying probability function.

**Definition 8.** Let $X$ be a discrete probability space. If $S$ is a statement depending on the values of one or more random variables, then we write $[S]$ to denote the set

$$[S] = \{x \in X \,:\, S \text{ holds for } x\}$$

Sometimes, the notation $\Pr[S]$ will be used to denote the probability $\Pr([S])$ of this set.

**Example 9.** Let $X = \{1, 2, 3, 4, 5, 6\}$ be the set of possible rolls of a 6-sided die with uniform probability distribution, and let $P : X \to \{0, 1\}$ be the parity function, mapping even numbers in $X$ to 0, and odd numbers in $X$ to 1. Then the set of odd numbers can be written as $[P = 1]$, and their probability can be written as $\Pr[P = 1] = 1/2$.

**Definition 10.** Let $X$ be a discrete probability space, and let $R : X \to \mathbb{R}$ be a random variable. Then the **expected value** of $R$ is the sum

$$\mathbb{E}(R) = \sum_{x \in X} R(x) \Pr(x)$$

**Proposition 11** (Markov's Inequality)**.** *Let $X$ be a discrete probability space, and let $R : X \to \mathbb{R}_+$ be a nonnegative random variable on $X$. Then for any $a \in \mathbb{R}_+$,*

$$\Pr[R \geq a] \leq \frac{\mathbb{E}(R)}{a}$$

*Proof.* Let $A = [R \geq a]$. Then we have

$$\mathbb{E}(R) = \sum_{x \in X} R(x) \Pr(x) \geq \sum_{x \in X} \chi_A R(x) \Pr(x) \geq \sum_{x \in X} \chi_A a \Pr(x) = a \sum_{x \in X} \chi_A \Pr(x) = a \Pr[R \geq a]$$

Dividing through by $a$ yields the desired inequality. □

**Definition 12.** Let $X$ be a discrete probability space, and let $F_i : X \to E_i$ be random variables on $X$, $i = 1, \ldots, n$. Then the $F_i$ are said to be **independent** random variables if for each set of outputs $e_i \in E_i$, $i = 1, \ldots, n$, we have

$$\Pr[F_1 = e_1, F_2 = e_2, \ldots, F_n = e_n] = \Pr[F_1 = e_1] \Pr[F_2 = e_2] \cdots \Pr[F_n = e_n]$$

**Proposition 13.** *Let $X$ be a discrete probability space, let $F : X \to E$ be a random variable on $X$, and let $n \in \mathbb{N}$. Then it is possible to construct a discrete probability space $Y$ and independent random variables $F_1, \ldots, F_n : Y \to E$ on $Y$ such that* $\Pr[F_i = e] = \Pr[F = e]$.

The following lemma will be a fundamental tool for results in verifiable computation.

**Lemma 14** (Schwartz-Zippel lemma)**.** *Let $K$ be a field, and let $f \in K[x_1, \ldots, x_n]$ be a nonzero polynomial of total degree $d \geq 0$ over $K$. Let $S$ be a finite subset of $K$, and let $r_1, \ldots, r_n$ be chosen uniformly and independently from $S$. Then*

$$\Pr[f(r_1, \ldots, r_n) = 0] \leq \frac{d}{|S|}$$

*Proof.* We prove this fact by induction on the number of variables $n$. For one variable, the polynomial $f$ has at most $d$ roots in $K$, and so at most $d$ elements out of the finite set $S$ satisfy the condition $f(r_1) = 0$. Since the value of $r_1$ was chosen uniformly from $S$, we have that

$$\Pr[f(r_1) = 0] = \frac{|\{s \in S \,:\, f(s) = 0\}|}{|S|} = \frac{d}{|S|}$$

Now suppose that $n > 1$, and the result holds for polynomials in fewer variables. Then we can write $f$ as

$$f(x_1, \ldots, x_n) = \sum_{i=0}^{d} x_n^i f_i(x_1, \ldots, x_{n-1})$$

Since $f$ is a nonzero polynomial, there is at least one index $i$ such that $f_i$ is a nonzero polynomial. Let $j$ be the largest such index, and define events $A$ and $B$ by

$$A = [f(r_1, \ldots, r_n) = 0] \qquad B = [f_j(r_1, \ldots, r_{n-1}) = 0]$$

Since $f$ has total degree $n$, the polynomial $f_j$ has total degree at most $d - j$. Since $f_j$ doesn't depend on $x_n$, it is a polynomial in at most $n - 1$ variables, so we have by induction hypothesis that

$$\Pr(B) = \Pr[f_j(r_1, \ldots, r_{n-1}) = 0] \leq \frac{d-j}{|S|}$$

If we instead restrict our attention to the event $B^{\mathrm{c}}$, the values $r_1, \ldots, r_n$ in $B^{\mathrm{c}}$ are determined only by the first $n-1$ variables, since these are the only ones influencing when $f_j(r_1, \ldots, r_{n-1})$ is nonzero. In particular, for any fixed choices of $r_1, \ldots, r_{n-1}$ satisfying this condition, the polynomial $g(x_n) = f(r_1, \ldots, r_{n-1}, x_n)$ has degree $j$, and thus has at most $j$ roots in $K$. The probability that a random value $r_n$ is one of these roots is therefore at most $j/|S|$. This allows us to give a bound for $\Pr(A \cap B^{\mathrm{c}})$ as follows.

$$\begin{aligned}
\Pr(A \cap B^{\mathrm{c}}) &= \sum_{r_1,\ldots,r_n} \chi_{A \cap B^{\mathrm{c}}} \Pr(r_1, \ldots, r_n) \\
&= \sum_{r_1,\ldots,r_{n-1}} \chi_{B^{\mathrm{c}}} \sum_{r_n} \chi_A \frac{1}{|S|^n} \\
&\leq \sum_{r_1,\ldots,r_{n-1}} \chi_{B^{\mathrm{c}}} \frac{j}{|S|^n} \\
&= \frac{j}{|S|} \sum_{r_1,\ldots,r_{n-1}} \chi_{B^{\mathrm{c}}} \frac{1}{|S|^{n-1}} \\
&= \frac{j}{|S|} \sum_{r_1,\ldots,r_n} \chi_{B^{\mathrm{c}}} \frac{1}{|S|^n} \\
&= \frac{j}{|S|} \Pr B^{\mathrm{c}}
\end{aligned}$$

Finally, we can apply our bounds for $P(B)$ and $P(A \cap B^{\mathrm{c}})$ to find:

$$\begin{aligned}
\Pr(A) &= \Pr(A \cap B) + \Pr(A \cap B^{\mathrm{c}}) \\
&= \Pr(B) \frac{\Pr(A \cap B)}{\Pr(B)} + \Pr(B^{\mathrm{c}}) \frac{\Pr(A \cap B^{\mathrm{c}})}{\Pr(B^{\mathrm{c}})} \\
&\leq \Pr(B) + \frac{\Pr(A \cap B^{\mathrm{c}})}{\Pr(B^{\mathrm{c}})} \\
&\leq \frac{d-j}{|S|} + \frac{j}{|S|} = \frac{d}{|S|}
\end{aligned}$$

This completes the induction. □

The importance of this lemma for our purposes is in its use for a randomized algorithm for testing whether two polynomials are equal everywhere. To do this deterministically requires checking that the polynomials have equal coefficients, which can be quite inefficient in various settings. If we allow a randomized approach, then it is enough to pick random elements from a fixed set of large enough size, and compare the evaluations of the polynomials at coordinates given by the chosen random elements. If the evaluations at random coordinates are equal, then with high probability the polynomials themselves are equal as polynomials.