

Lecture 16: Introduction to Probability

Instructor: Jin-Yi Cai

Scribe: Aparna Das, Scott Diehl, Giordano Fusco

1 Finite Fields

We begin with a short discussion on finite fields and their representation. Finite fields are essentially finite sets on which we have defined multiplication and addition that perform exactly like the corresponding operations on reals (most importantly, we can divide).

What are examples of fields? For all primes p the set of equivalence classes of the integers modulo p forms a field, \mathbb{Z}_p . Also for all prime powers p^d there is a field of size p^d . Namely, consider the ring of single variable polynomials with coefficients in \mathbb{Z}_p , modulo an irreducible polynomial $p(x) \in \mathbb{Z}_p[x]$, written $\mathbb{Z}_p[x]/p(x)$. If $p(x)$ has degree d , then $|\mathbb{Z}_p[x]/p(x)| = p^d$. $\mathbb{Z}_p[x]/p(x)$ is a field with p^d elements. In particular, we have finite fields of size 2^n , for any $n > 0$. Therefore all binary string of length n can always be expressed in some finite field.

1.1 A Family of Useful Fields

A particularly useful polynomial when representing integers in a finite field is $p_k(x) = x^{2 \cdot 3^k} + x^{3^k} + 1$ for $k > 0$. Such a $p_k(x)$ has the property that it is always irreducible. Then consider the field $\mathbb{Z}_2[x]/p_k(x)$. All polynomials in this field have degree at most $n - 1$ where $n = 2 \cdot 3^k$ and have coefficients that are 0 or 1. Therefore we can represent each $f \in \mathbb{Z}_2[x]/p_k(x)$ uniquely by a bit string of length n .

Example:

When $k = 2$, $f = x^{17} + x^{16} + x^{10} + x^2 + 1$ can be encoded as the bit string 110000010000000101.

Operations Over $f \in \mathbb{Z}_2[x]/p_k(x)$

Arithmetic over $\mathbb{Z}_2[x]/p_k(x)$ is particularly easy to implement. The following operations can be implemented in NC^1 , and work directly on the bit-string representations:

- Addition (and since the coefficients are in \mathbb{Z}_2 , also subtraction) of $f, g \in \mathbb{Z}_2[x]/p_k(x)$ can be achieved simply by the bitwise XOR of their representations.
- Multiplication can then be implemented in the usual shift-and-add manner, with addition as above. However we might end up with a result that is of degree more than $n - 1$, in which case we need to perform a division by $p_k(x)$ and take the remainder. We describe this procedure next.
- Division by $p_k(x)$ is also simple to implement. Recall the reduction step in polynomial division: suppose $f(x) = x^{n+l} + f'(x)$ where $\deg(f'(x)) < n + l$ for $l \geq 0$, then when dividing by $p_k(x) = x^{2 \cdot 3^k} + x^{3^k} + 1$, a reduction step will remove the term x^{n+l} and subtract $x^l \cdot (x^{3^k} + 1)$

from $f'(x)$. However since our coefficients are in \mathbb{Z}_2 , this is the same as adding $x^l(x^{3^k} + 1)$. Therefore, for each bit in the representation of $f(x)$ that is in position $n+l$ for $l > 0$, we can perform the reduction step by removing each such bit and then taking the not of the bits in position $3^k + l$ and l . The entire division process can therefore be implemented by shifting the higher order bits to two different lower order locations.

2 Random Functions in \mathbb{Z}_p

Fix $a, b \in \mathbb{Z}_p$ and consider the function $f_{a,b} : \mathbb{Z}_p \rightarrow \mathbb{Z}_p$ given by $f_{a,b}(x) = ax + b \pmod{p}$. If we choose the pair a and b randomly from \mathbb{Z}_p , $f_{a,b}$ is a random function.

Let us look at this in a different way. For each $x \in \mathbb{Z}_p$ we define the random variable $Z_x : \mathbb{Z}_p \times \mathbb{Z}_p \rightarrow \mathbb{Z}_p$, given by $Z_x(a, b) = ax + b \pmod{p}$. Choose two elements at random from \mathbb{Z}_p so that the probability of getting any particular pair $\langle a, b \rangle$ is $1/p^2$ and apply Z_x to the pair. Fix an x . We want to determine the distribution of Z_x . For example, what is the probability that $Z_x = 0$? Let us count the number of pairs $\langle a, b \rangle$ make $ax + b = 0$. In other words, let us find the size of the set

$$\{\langle a, b \rangle \mid ax + b = 0 \pmod{p}\}$$

Since x is fixed, we know that for any a there is an unique value of b such that $ax + b = 0$, namely $b = -ax$. Thus out of p^2 possible choices for $\langle a, b \rangle$, exactly p of them satisfy $ax + b = 0$. Therefore the $\Pr_{a,b}[Z_x = 0] = 1/p$. The analysis will work the same way if we asked the same question for any α , instead of 0: how many pairs $\langle a, b \rangle$ make $ax + b = \alpha$? This shows that every α has the same weight under the distribution of Z_x . So Z_x is uniformly distributed in \mathbb{Z}_p .

Now suppose we fix x and y , such that $y \neq x$, and ask what is the $\Pr[Z_x = \alpha \wedge Z_y = \beta]$. This is the same as asking the question: $\forall \alpha, \beta \in \mathbb{Z}_p$, what is the size of the set $\{\langle a, b \rangle \mid ax + b = \alpha \text{ and } ay + b = \beta\}$. This can be found by finding the number of solutions to the linear system of equations

$$\begin{pmatrix} x & 1 \\ y & 1 \end{pmatrix} \begin{pmatrix} a \\ b \end{pmatrix} = \begin{pmatrix} \alpha \\ \beta \end{pmatrix}$$

The $\det \begin{pmatrix} x & 1 \\ y & 1 \end{pmatrix} = x - y \neq 0$, because we choose $x \neq y$. Thus we know there is an unique solution for this linear system. Therefore, $\Pr[Z_x = \alpha \wedge Z_y = \beta] = 1/p^2 = \Pr[Z_x = \alpha] \cdot \Pr[Z_y = \beta]$. This implies that Z_x and Z_y are independent random variables. Therefore $\{Z_x \mid x \in \mathbb{Z}_p\}$, is a family of pairwise independent uniformly distributed random variables.

Note that if we fixed another variable z such that x, y and z are distinct, the linear system:

$$\begin{pmatrix} x & 1 \\ y & 1 \\ z & 1 \end{pmatrix} \begin{pmatrix} a \\ b \end{pmatrix} = \begin{pmatrix} \alpha \\ \beta \\ \gamma \end{pmatrix}$$

has no solution. Therefore this family is not 3-wise independent.

To summarize, we now can think of the function $h_s(x) = ax + b$ where s is chosen uniformly from $S = \{\langle a, b \rangle \mid a, b \in \mathbb{Z}_p\}$ as a random function that will uniformly map \mathbb{Z}_p to \mathbb{Z}_p , for which we only need $2 \log p$ bits to specify (namely, to specify a and b).

3 Useful inequalities

We are now going to show some inequalities that are useful in probability theory.

3.1 Markov's inequality

THEOREM 3.1 (MARKOV'S INEQUALITY) *If $X \geq 0$ is a random variable, then for any real number $\alpha > 0$*

$$\Pr[X \geq \alpha] \leq \frac{\mathbb{E}[X]}{\alpha}$$

Proof. To prove the inequality we start from the definition of probability as a weight function

$$\Pr[X \geq \alpha] = \mu\{w \mid X(w) \geq \alpha\}$$

By the definition of measure of a set we obtain that

$$\mu\{w \mid X(w) \geq \alpha\} = \int 1_{\{X(w) \geq \alpha\}} = \int_{\{X(w) \geq \alpha\}} d\mu$$

Because $X(w) \geq \alpha$ and $\alpha > 0$, we have that $\frac{X(w)}{\alpha} \geq 1$ and by the linearity of the integral we obtain:

$$\int_{\{X(w) \geq \alpha\}} d\mu = \int_{\{X(w) \geq \alpha\}} 1 d\mu \leq \int_{\{X(w) \geq \alpha\}} \frac{X(w)}{\alpha} d\mu$$

If we enlarge the domain the integral becomes bigger:

$$\int_{\{X(w) \geq \alpha\}} \frac{X(w)}{\alpha} d\mu \leq \int_{\Omega} \frac{X(w)}{\alpha} d\mu = \frac{1}{\alpha} \int_{\Omega} X(w) d\mu$$

We can recognize that the last integral is the expectation.

In a similar way we can prove the Markov's inequality in the discrete case:

$$\begin{aligned} \Pr[X \geq \alpha] &= \mu\{w \mid X(w) \geq \alpha\} = \\ &= \sum_{\{w \mid X(w) \geq \alpha\}} \mu\{w\} \\ &\leq \sum_{\{w \mid X(w) \geq \alpha\}} \frac{X(w)}{\alpha} \mu\{w\} \\ &\leq \sum_{\Omega} \frac{X(w)}{\alpha} \mu\{w\} = \\ &= \frac{\mathbb{E}[X]}{\alpha} \end{aligned}$$

To better understand the Markov's inequality we can look at the following example. Let X be the number of times a random algorithm succeeds in a certain fixed number of trials. Suppose that $\mathbb{E}[X] = 7$. We want to estimate the probability of obtaining more than 14 successes: in this case $\alpha = 14$. Using Markov's inequality we obtain:

$$\Pr[X \geq 14] \leq \frac{7}{14} = \frac{1}{2}$$

3.2 Chebychev's inequality

DEFINITION 3.2 *The variance of a random variable X is defined as*

$$\text{Var}[X] = E[(X - E[X])^2]$$

FACT 3.3

$$\text{Var}[X] = E[(X - E[X])^2] = E[X^2] - (E[X])^2$$

THEOREM 3.4 (CHEBYCHEV'S INEQUALITY) *If the random variable X has a finite expectation $E[X]$, then for any real number $\alpha > 0$*

$$\Pr[|X - E[X]| \geq \alpha] \leq \frac{\text{Var}[X]}{\alpha^2}$$

Proof. Squaring both sides of $|X - E[X]| \geq \alpha$ we obtain $(X - E[X])^2 \geq \alpha^2$. Then applying the Markov's inequality:

$$\Pr[|X - E[X]| \geq \alpha] = \Pr[(X - E[X])^2 \geq \alpha^2] \leq \frac{E[(X - E[X])^2]}{\alpha^2} = \frac{\text{Var}[X]}{\alpha^2}$$

3.3 Chernoff bound

Suppose that a process has probability p of success, then we would like to know what happens if we repeat it a lot of times. The Chernoff bound would tell us that if we repeat n times, then the probability of obtaining more than $\epsilon \cdot n$ successes away from the expected number of successes is exponentially small.

Let X_1, X_2, \dots, X_n be 0-1 variables independent identically distributed (i.i.d.) with $\Pr[X_i = 1] = p$. The sum of the variables, $Y = \sum_{i=1}^n X_i$, is the total number of successes amongst the X_i 's. Due to the linearity of the expectation:

$$E[Y] = \sum_{i=1}^n E[X_i] = \sum_{i=1}^n p = np$$

Now we are ready to introduce the Chernoff bound. For all $\delta \geq 0$,

$$\Pr[Y < (1 - \delta)E[Y]] = \Pr[Y < (1 - \delta)pn] < e^{-\frac{\delta^2 pn}{2}}$$

$$\Pr[Y > (1 + \delta)E[Y]] = \Pr[Y > (1 + \delta)pn] < e^{-\frac{\delta^2 pn}{3}}$$

We will prove the Chernoff bound during next lecture.