Randomized Methods in Computation*
Lecture 1: Probability and the Probabilistic Method

Notes taken by Maksim Frenkel and Safro Ilya

Spring 2001

Summary: This lecture is a general introduction to the course "Randomized methods in Computation". We review some of the basic notions of probability theory, such as the expectation and the variance of a random variable, Chebyshev's inequality and the first law of large numbers. We also introduce the Probabilistic Method and demonstrate one application of it (specifically, to the Max3SAT problem). Finally, we will develop a simple algorithm that given a 3CNF formula, finds an assignment that satisfies many of its clauses.

1 Motivation

Oded's Note: *I tried to stress three aspects regarding this course as hinted by its name: First, this course focuses on* methods *(i.e., tools and techniques) and not on concepts. Second, these methods are* randomized *and so the result is often not "(fully) explicit". This can be seen both in the form of randomized algorithms (and procedures) that do not always return the same answer (but rather have an output distribution that satisfies some desired property with high probability) as well as in arguments (e.g., in the "probabilistic method") that do not explicitly present a desired object (but rather only specifies a distribution of objects so that with high probability an object in this distribution possesses the desired property). Lastly, the third aspect of this course is that it focuses on* applications to the study of computation. *This is reflected in the emphasis put on the complexity of generating certain distributions, where complexity may mean a variety of complexity measures (such as time, space, and randomness).*

Consider the following problem: given a set of integer numbers, sort them in an ascending order. There exist a lot of different algorithms that solve this problem. These algorithms have different worst and average running times. Some of them are kno wn to perform badly on certain problem instances (for example the QuickSort algorithm) while on the average their running times are good. A possible solution is to randomly permute the given integer sequence such that it will resemble the average case inp ut. In this example our algorithm will use the sequence of outcomes of coin tosses which explicitly determine the permutation. Such an algorithm is called *randomized.*

A randomized algorithm gets as its input a sequence of n outcomes of coin toss in addition to its regular input. The algorithm makes its decisions based upon this sequence. Standard (deterministic) algorithm has one answer given one input instance. Randomized algorithm has a set of

* Lecture Notes for a course given by Oded Goldreich at the Weizmann Institute of Science, Israel.

answers (or a set of ways to reach an answer) given the problem input instance. Some of these answers (or ways to reach an answer) are good whenever other are bad. The advantage of a good randomized algorithm is its abil ity to reach the good answer (or to find a good way to this answer) fast.

> Oded's Note: *The above motivation (which is not at all what I said in class) addresses just one aspect of the course. See above note for what I did say.*

2 Basic Notions from Probability Theory

> Oded's Note: *I assume that most students know most of the contents of the current section. My goals in covering this basic material were two-fold. First, to notify all student of what they are assumed to know, and teach the few that don't know some of this material the little they don't know. Second, in a confusing business such as probabilistic analysis it is good to get reminded of the basic underlying definitions: in case of difficulty one should remember that the complex notions can and should (when in doubt) be traced back to the underlying notions. That is, when faced with a complex analysis it is a good practice to ask basic questions like "what is the underlying probability space" (which is typically only implicit in complex discussions).*

2.1 Outcomes, Events and Probability

We assume that the reader is familiar with the basic notions of Probability Theory. Therefore the purpose of this section will be to review those concepts, which will be useful for us in upcoming lectures.

Let us consider an experiment that results in several outcomes. Denote these outcomes by $A_1, A_2, ..., A_m$. If we will repeat this experiment n times under various conditions, A_1 will occur k_1 times, while other outcomes will occur $n - k_1$ times. We refer to a certain subset of the outcomes as *an event* and to the set of all possible outcomes as *a probability space* Ω. When we talk about *the probability* of the event X occurring over its probability space Ω we consider the ratio of the sizes of the two sets: $\frac{|X|}{|\Omega|}$. The meaning of this ratio is the answer to the question: how to measure the possibility that the event X will occur as a result of our experiment. This ratio we denote by $Prob(X)$. It is easy to see that

$$0 \leq Prob(X) \leq 1$$

Example 2.1 *Suppose we are rolling a die and we are interested in obtaining an even outcome. Here*

$$\Omega = \{1, 2, 3, 4, 5, 6\}$$

The event X will be the following subset of Ω:

$$X = \{2, 4, 6\}$$

Then

$$Prob(X) = \frac{|X|}{|\Omega|} = \frac{3}{6}$$

Remark 2.2 *In this course we will consider the probability space to be a finite set.*

Sometimes the outcome of one event may influence the outcome of another one. To measure the amount of influence we introduce the *conditional probability* of event B occurring given that the event A has occurred:

$$Prob(B|A) = \frac{Prob(A \cap B)}{Prob(A)}$$

Recall that events can be either *independent* or *dependent*. The events are independent when the trials of our experiment are carried out independently, which implies that the result of each trial does not influence the other results. If we are interested in the probability of the events A and B occurring when they are independent then

$$Prob(A \cap B) = Prob(A) \cdot Prob(B)$$

In other words, A and B are independent if the conditional probability of B given A is just equal to the probability that B occurs without paying attention to A at all, i.e.

$$Prob(B|A) = Prob(B)$$

> Oded's Note: *Indeed, probabilities are merely ratios of sizes of sets, and so probabilistic analysis is merely combinatorics. Yet, as is often the case in mathematics and science, using certain definitions and notations (rather than others) may simplify the analysis tremendously. Specifically, in many cases, the analysis is much simpler to carry out in terms of probabilities than in terms of set sizes.*

2.2 Random variables

A random variable is a function from a probability space to the reals (or complexes); that is, a random variable X is

$$X : \Omega \longrightarrow \mathbb{R} \qquad (possibly\ \mathbb{C})$$

As we remarked earlier, Ω in our case is a finite set and hence the set of values that X is mapped onto is also finite. Intuitively a random variable is a function that can be calculated as a result of an experiment. The adjective "random" reminds us that we are dealing with a sample space and are trying to describe "random events". Random variables are often defined without explicitly specifying a probability space (but such a space is always implicit in the definition). For example, the number of hits out of three shots, the number of tails in five coin tosses, etc. Once we have defined what our random variable measures we can define the probability that it will take on a certain value α:

$$Prob[X = \alpha] = \frac{|\{w \in \Omega : X(w) = \alpha\}|}{|\Omega|}$$

Let us remind also that if X and Y are random variables then

$$X + Y, \quad X - Y, \quad XY, \quad \frac{X}{Y}, \quad \alpha X + \beta Y$$

(where α and β are real numbers) are also random variables.

Discrete random variables X and Y are independent if

$$Prob(X = x\ and\ Y = y) \;=\; Prob(X = x) \cdot Prob(Y = y)$$

for all values x in the range of X and y in the range of Y.

By definition of a random variable one can not define it without specifying the corresponding probability space. However, often, when the associated probability space is obvious we will drop this convention. We will usually describe the set of values that can be associated with the random variable with their corresponding weights, instead of explicitly defining the function with its domain.

> Oded's Note: *Indeed the last paragraph refers to my main motivation in recalling all these trivialities. We will often talk of random variables without specifying the underlying probability space, and will often be able to carry out the analysis that way. However, when in doubt, one should specify explicitly the probability space that underliees the analized events or random variables.*

2.3 Expectation

For a random variable X defined on a probability space Ω, its *expectation* is the number $E(X)$ defined by the formula

$$E(X) \;=\; \sum_{\omega\in\Omega} Prob(\{\omega\})X(\omega)$$

If we think of $Prob(\{\omega\})$ as the weight attached to ω then $E(X)$ can be understood as the weighted average of the function X. The intuition of the value $E(X)$ might be obtained from its analogue in mechanics. Consider a line with n physical points distributed on it. Every point has a weight p_i. Let X_i be the location of the i-th point. Taking the sum of all the weights to be 1, the equilibrium point of the system has the coordinate:

$$X \;=\; \frac{\sum_{i=1}^n p_i X_i}{\sum_{i=1}^n p_i} \;=\; \frac{\sum_{i=1}^n p_i X_i}{1} \;=\; \sum_{i=1}^n p_i X_i$$

This is exactly the expectation of the random variable X. The following theorem introduces some properties of the expectation. These properties will be very useful in the future discussion.

Proposition 2.3 *For two random variables X and Y,*

1. *$E(aX) = aE(X)$, for any real a.*
2. *$E(X+Y) = E(X) + E(Y)$, even if X and Y are dependent.*
3. *If X and Y are independent then $E(XY)=E(X)E(Y)$.*

Proof: Let X be the random variable that takes on discrete values and let a be a number. Let aX be defined on the set Ω, then

$$\begin{aligned} E(aX) &= \sum_v v\cdot Prob(aX=v) \\ &= a\cdot\sum_v \frac{v}{a}\cdot Prob(X=\frac{v}{a}) \\ &= a\cdot E(X) \end{aligned}$$

In order to prove Part 2 let us assume that Y takes on discrete values.

$$\begin{aligned}
E(X+Y) &= \sum_{(x,y)} [(x+y) \cdot Prob(X = x \text{ and } Y = y)] \\
&= \sum_x \sum_y x \cdot Prob(X = x \text{ and } Y = y) \ + \ \sum_y \sum_x y \cdot Prob(X = x \text{ and } Y = y) \\
&= \sum_x x \cdot Prob(X = x) \ + \ \sum_y y \cdot Prob(Y = y) \\
&= E(X) + E(Y)
\end{aligned}$$

And finally, for independent X and Y, we have

$$\begin{aligned}
E(XY) &= \sum_x \sum_y xy \cdot Prob(X = x \text{ and } Y = y) \\
&= \sum_x \sum_y xy \cdot Prob(X = x) \cdot Prob(Y = y) \\
&= \left(\sum_x x \cdot Prob(X = x) \right) \cdot \left(\sum_y y \cdot Prob(Y = y) \right) \\
&= E(X) \cdot E(Y)
\end{aligned}$$

■

In general, if $X_1, ... X_n$ are random variables and $X = \alpha_1 X_1 \ + \ ... \ + \ \alpha_n X_n$, then

$$E(X) = \alpha_1 E(X_1) \ + \ ... \ + \ \alpha_n E(X_n)$$

The linearity of expectation is a very useful tool in randomized methods. Its power comes from the idea that it may be possible to decompose a complex instance X as the sum of many simple instances X_i, and the linearity will hold regardless of the dependence of X_i's. The expectation of each of the X_i-s can be calculated relatively easy compared to that of X. We will see an application of this principle in Theorem 5. Another useful fact regarding the expectation is that it cannot be *always* higher than the random variable. That is:

Proposition 2.4 *For any random variable X*

$$Prob[X \geq E(X)] > 0$$

Proof Sketch: We need to show that there exists x, such that $Prob[X = x] > 0$ and $x \geq E(X)$. Let x_{max} be the maximal value that X may be equal to such that $Prob[X = x_{max}] > 0$ and $Prob[X > x_{max}] = 0$.

$$E(X) \ = \ \sum_x Prob[X = x] \cdot x \ \leq \ \sum_{x \leq x_{max}} Prob[X = x] \cdot x_{max} \ = \ x_{max}$$

Therefore $E(X) \leq x_{max}$ and since we know that $Prob[X = x_{max}] > 0$ we have shown that $Prob[X \geq E(X)] > 0$. ■

2.4 The Variance

> Oded's Note: *The following paragraph is not something I said. I would have said that the variance is a measure of the "spread" or "variability" of a random variable, and that its specific definition is adequate for various Laws of Large Numbers as well as for Chebyshev's Inequality that underlies some of them.*

The variance of a random variable X measures the unpredictability of a random variable or in other words this is a measure of its variation from the mean $E(X)$. Suppose that two random variables X and Y have the same mean μ, but the values of X are distributed more closely to $E(X)$ than are the values of Y to $E(Y)$. That is, the probability is higher that X will be closer to μ than Y will be. We would like to define some measure that will reflect that fact. Let us shift both X and Y to 0 and consider the values $X-\mu$ and $Y-\mu$. Since Y was spread farther around μ than X was, $Y-\mu$ will be spread farther around 0 than $X-\mu$. Let's square the above differences to make sure that they are positive. The *expectation* of $(X-\mu)^2$ will be closer to 0 than $(Y-\mu)^2$. Therefore it makes sense to define the variance of a random variable X as follows:

$$VAR(X) \;=\; E((X-\mu)^2),\ where\ \mu = E(X)$$

Using the linearity of expectation, we see that

$$\begin{aligned}
VAR(X) &= E((X-\mu)^2) \\
&= E(X^2 - 2\mu X + \mu^2) \\
&= E(X^2) - 2\mu E(X) + E(\mu^2) \\
&= E(X^2) - 2\mu\mu + \mu^2 \\
&= E(X^2) - E(X)^2
\end{aligned}$$

The above formula provides a *computationally easier* method of evaluating the variance.

The variance is not always linear. An exceptional case when the variance is linear is when applied to a sum of paiwise-independent random variables. Following is a special case (which can be easily generalized to a sequence of pairwise independent random variables):

Proposition 2.5 *If X and Y are independent then*

$$VAR(X+Y) \;=\; VAR(X) + VAR(Y)$$

Proof: Since X and Y are independent, $E(XY) = E(X)\cdot E(Y) = \mu_X\mu_Y$, where μ_X and μ_Y equal to $E(X)$ and $E(Y)$ respectively. Then

$$\begin{aligned}
VAR(X+Y) &= E((X+Y)^2) - (\mu_X + \mu_Y)^2 \\
&= E(X^2 + 2XY + Y^2) - (\mu_X^2 + 2\mu_X\mu_Y + \mu_X^2) \\
&= E(X^2) - \mu_X^2 + E(Y^2) - \mu_Y^2 + 2E(XY) - 2\mu_X\mu_Y \\
&= VAR(X) + VAR(Y)
\end{aligned}$$

since the last two terms cancel. ■

On the other hand, in case X and Y are dependent $VAR(X+Y)$ may be different from $VAR(X)+VAR(Y)$. In particular, if $X = Y$ (i.e., $Prob(X=Y)=1$) then $VAR(X+Y) = 2VAR(X) + 2VAR(Y)$, whereas if $X = -Y$ than $VAR(X+Y) = 0$. Both claims follows from the next proposition:

Proposition 2.6 *Let a and b be constants. Then*

$$VAR(aX + b) = a^2 \cdot VAR(X)$$

Proof: Let $\mu = E(X)$ and observe that

$$E(aX + b) = aE(X) + b = a\mu + b$$

Now

$$\begin{aligned} VAR(aX + b) &= E((aX + b)^2) - (E(aX + b))^2 \\ &= E(a^2X^2 + 2abX + b^2) - (a\mu + b)^2 \\ &= a^2 \cdot E(X^2) + 2ab \cdot E(X) + b^2 - (a^2\mu^2 + 2ab\mu + b^2) \\ &= a^2 \cdot (E(X^2) - \mu^2) \\ &= a^2 \cdot VAR(X) \end{aligned}$$

■

Note that the above result is intuitive, since the values of the random variable $aX + b$ vary from its expectation $E(aX + b)$ as do the values of aX from its expectation $E(aX)$. On the other hand, simply shifting the distribution by b does not change its *spread*. Another intuitive property of the variance follows:

Proposition 2.7 *The variance of a random variable X is zero if and only if X is a constant.*

Proof: The backward direction ($\Leftarrow$) is trivial and follows from Proposition 2.6, when $a = 0$. The forward direction follows from the definition of the variance. Specifically, we have

$$VAR(X) = E((X - \mu)^2) = 0$$

where $\mu = E(X)$. Since $(X - \mu)^2$ is a nonnegative value, the fact that its expectation is zero implies that $X = \mu$. ■

2.5 Markov's and Chebyshev's Inequalities and a Law of Large Numbers

In analyzing the performance of a randomized algorithm we will often need to show that the algorithm behaves well almost all the time. While the concept of the expectation of a random variable provides information regarding the *average case scenario* (for instance, the algorithm's average running time), lots of times we will interested in the "typical" (rather than average) behavior; that is, we want to know an interval of values that is guaranteed to contain the value of the random variable, almost always. In doing so we will need to analyze the probability that a random variable deviates from its expectation by a given amount.

The first inequality we introduce provides a bound on the probability that a random variable deviates from its expectation.

Theorem 1 (Markov's Inequality): *If X is a random variable that assumes only non-negative values, then for all real and positive v*

$$Prob[X \geq v] \leq \frac{E(X)}{v}$$

Proof:

$$\begin{aligned} E(X) &= \sum_x Prob[X=x] \cdot x \\ &= \sum_{x \geq v} Prob[X=x] \cdot x \; + \; \sum_{x<v} Prob[X=x] \cdot x \\ &\geq \sum_{x \geq v} Prob[X=x] \cdot v \; + \; \sum_{x<v} Prob[X=x] \cdot 0 \\ &= Prob[X \geq v] \cdot v \end{aligned}$$

Thus

$$Prob[X \geq v] \leq \frac{E(X)}{v}$$

■

An equivalent formulation of Markov inequality is :

$$Prob[X \geq \alpha \cdot E(X)] \leq \frac{1}{\alpha}$$

As we can see the upper bound of the probability that a random variable deviates from its expectation decreases as α increases. As noted before the probability space of X is not directly mentioned in the above theorem but it is obviously implied. Usually we use the Markov Inequality when our knowledge about the distribution is limited to its expectation and at least one bound on the distribution of a random variable. If the variance of a given random variable X is known, the Markov inequality can be used to achieve a better bound on the probability that X will deviate from its expectation, as seen in the next theorem.

Theorem 2 (Chebyshev's Inequality): *For any random variable Y and any $\delta > 0$,*

$$Prob[|Y - E(Y)| \geq \delta] \leq \frac{VAR(Y)}{\delta^2}$$

Proof: Let $X = (Y - E(Y))^2$ and note that X is non-negative and that

$$E(X) = E((Y - E(Y))^2) = VAR(Y)$$

Next, observe that

$$Prob[|Y - E(Y)| \geq \delta] = Prob[X \geq \delta^2]$$

Finally, invoking Markov's inequality on X and δ^2 yields

$$Prob[X \geq \delta^2] \leq \frac{E(X)}{\delta^2} = \frac{VAR(Y)}{\delta^2}$$

which completes the proof. ■

Indeed, as expected for *any* distribution, the probability that the values of X are more than a certain *fixed* distance away from its mean gets smaller as $VAR(X)$ gets smaller. A special case where this is useful is for the sum of pairwise independent random variables.

Definition 3 *The random variables $X_1, ..., X_n$ are called* pairwise independent *if for every $i \neq j$ and all a and b,*

$$Prob[X_i = a \wedge X_j = b] = Prob[X_i = a] \cdot Prob[X_j = b]$$

Corollary 2.8 *Let $X_1, ..., X_n$ be pairwise independent random variables with common mean μ and variance σ^2. Denote $S_n = \sum_{i=1}^{n} X_i$, and let $\overline{X} = S_n/n$. Then*

$$Prob[|\overline{X} - \mu| \geq \epsilon] \leq \frac{\sigma^2}{n\epsilon^2}$$

Proof: First, observe that $VAR(S_n) = n\sigma^2$, by (a generalization of) Proposition 2.5.

> Oded's Note: *Specifically, we use $VAR(\sum_{i=1}^{n} X_i) = \sum_{i=1}^{n} VAR(X_i)$, which is proven as Proposition 2.5:*
>
> $$\begin{aligned} VAR(\sum_{i=1}^{n} X_i) &= E\left((\sum_{i=1}^{n} X_i)^2\right) - E\left(\sum_{i=1}^{n} X_i\right)^2 \\ &= E\left((\sum_{i=1}^{n} X_i^2) + (\sum_{i \neq j} X_i X_j)\right) - \left(\sum_{i=1}^{n} E(X_i)\right)^2 \\ &= \sum_{i=1}^{n} E(X_i^2) + \sum_{i \neq j} E(X_i X_j) - \sum_{i=1}^{n} E(X_i)^2 - \sum_{i \neq j} E(X_i)E(X_j) \\ &= \sum_{i=1}^{n} E(X_i^2) - \sum_{i=1}^{n} E(X_i)^2 \\ &= \sum_{i=1}^{n} VAR(X_i) \end{aligned}$$
>
> *where the key fact being used is $E(X_i X_j) = E(X_i)E(X_j)$, for $i \neq j$, which follows becuase X_i and X_j are independent.*

Now, by Proposition 2.6:

$$VAR(\overline{X}) = VAR(S_n/n) = \frac{VAR(S_n)}{n^2} = \frac{\sigma^2}{n}$$

Finally, invoking Chebyshev's inequality on $\overline{X}$, we get

$$Prob[|\overline{X} - \mu| \geq \epsilon] \leq \frac{VAR(\overline{X})}{\epsilon^2} = \frac{\sigma^2}{n\epsilon^2} \tag{1}$$

■

Remark 2.9 *The above corollary generalizes to the case where the X_i's are pairwise independent, but not necessarily with identical mean. In such a case, we let $\mu = \frac{1}{n}\sum_{i=1}^{n} E(X_i)$ and have:*

$$Prob[|\overline{X} - \mu| \geq \epsilon] \leq \frac{\max_i VAR(X_i)}{n\epsilon^2}$$

3 The Probabilistic Method

3.1 Introduction

The probabilistic method is a way to prove the existence of objects having certain properties, by proving their existence "in abundence" (with respect to some probability spcae). That is, suppose we want to prove the existence of some combinatorial object with property T. The construction of such an object may be too hard. The probabilistic method suggests to define some probability space S, and to show that the probability that an object selected uniformly in S satisfies property T with non-zero probability. This would imply the existence of such an object (i.e., satisfying property T).

3.2 An application of the probabilistic method: the 3CNF-SAT problem

Definition 4 *A 3CNF boolean formula is a boolean formula φ written in conjunctive normal form where each clause contains exactly three* different *literals:*

$$\varphi = \bigwedge_{i=1}^{m} C_i, \qquad \text{where } C_i = \bigvee_{j=1}^{3} l_{ij}$$

each literal l_{ij} can be either a boolean variable or negation of a boolean variable.

Theorem 5 *For every 3CNF boolean formula φ there exists a truth assignment that satisfies at least $\frac{7}{8}$ of clauses in φ.*

Note that the theorem is not true in case we consider arbitrary CNF formulae (or CNF formulae in which each clause may have any three literals that are not necessarily different). Consider for example, the formula

$$(x_1 \vee x_2) \wedge (\neg x_1 \vee x_2) \wedge (x_1 \vee \neg x_2) \wedge (\neg x_1 \vee \neg x_2)$$

Clearly, each possible truth assignment satisfies only 3 out of the 4 possible clauses. A similar example shows that the Theorem 5 is indeed optimal: there exists 3CNF boolean formulae for which any truth assignment satisfies only 7/8 fraction of the clauses.

Proof: For sake of simplicity, we assume that no clause of φ may contain two literals corresponding to the same variable (i.e., a boolean variable and its negation). Note that any clause that conatins a boolean variable and its negation is satisfied under any truth assignment. Thus, we may omitting such clauses, observing that any assignment that satisfies 7/8 of the remaining clauses also satisfies all the omitted clauses.

The idea of the proof is to select a random assignment to the variables of φ uniformly (over all possible choices) and to evaluate the expected number of clauses satisfied by it. That is, we take as a probability space the set of all possible assignments to the variables of φ. Choose a random assignment and let X be the number of clauses (of $\varphi = \bigwedge_{i=1}^{m} C_i$) satisfied by it. Let X_i be the indicator random variable such that

$$X_i = \begin{cases} 1 & \text{if the assignment satisfies } C_i \\ 0 & \text{otherwise} \end{cases} \tag{2}$$

For $\varphi = \bigwedge_{i=1}^{m} C_i$, we have $X = X_1 + ... + X_m$. These X_i-s are not necessarily independent, because some clauses may contain the same literals (or just the same variables). Nevertheless, this fact does not prevent us from considering the expectation of X and using linearity of expectation:

$$E(X) = E(\sum_{i=1}^{m} X_i) = \sum_{i=1}^{m} E(X_i)$$

Thus we have broken down the relatively complicated problem of calculating the value $E(X)$ into a number of simpler problems of calculating each of the $E(X_i)$-s. Suppose that the clause C_j contains the three literals $l_{j1},\ l_{j2},\ l_{j3}$. Since these three literals are of different variables, there are 8 possible assignments to these three literals and 7 of them satisfy C_j. Since each of these 8 possibilities is equaly likely, when an assignment to all variables (of φ) is selected uniformly, we get:

$$E(X_i) \ = \ Prob[X_i = 1] \ = \ \frac{7}{8}$$

So $E(X) \ = \ \frac{7}{8}m$. Now as we already observed (see Proposition 2.4) $Prob[X \geq E(X)] > 0$, so we can say that in our case $Prob[X \geq \frac{7}{8}m] > 0$. This implies the existence of a truth assignment for φ that satisfies at least $\frac{7}{8}$ of its clauses. ■

In the above proof we have defined the probability space Ω of all possible truth assignments to a given $3CNF$ formula, and then shown that the probability of choosing an assignment from Ω that satisfies the required property is non-zero. We then concluded that an assignment as required does exist.

3.3 Obtaining a randomized algorithm

In the previous section we have shown that there exists an assignment that satisfies at least $\frac{7}{8}$ of the clauses in a given $3CNF$ formula. Here we will present a randomized algorithm that finds such an assignment with high probability in polynomial time. The idea behind this algorithm will be to repeteadly select random assignments to the variables of the given formula, untill it finds an assignment that satisfies at least $\frac{7}{8}$ of the clauses. The expected running time of the algorithm depends on the probability that the chosen random assignment satisfies at least $\frac{7}{8}$ of the clauses. The following claim will show that this probability is quite noticeable (i.e, bigger than a polynomial fraction):

Claim 3.1 *Let φ be a 3CNF formula with m clauses. Let X be the random variable that counts the number of clauses satisfied by a uniformly random assignment to the variables of φ. Then*

$$Prob(X \ \geq \ \frac{7}{8}m) \ \geq \ \frac{1}{m+1}$$

Proof: As in the proof of Theorem 5, we assume for simplicity (and without loss of generality) that each clause in φ consists of three literals corresponding to three different variables.

Recall that $E(X) \ = \ \sum_x x \cdot Prob[X = x]$. We can decompose this sum into two components. The first component G will represent the values of X that are greater or equal to $E(X)$. The second component R will represent the rest (i.e. the other possible x-s). Thus

$$E(X) \ = \ \sum_{x \in G} x Prob[X = x] \ + \ \sum_{x \in R} x Prob[X = x]$$

The possible values that X can get are integers. Let T be the largest integer that is smaller than $E(X)$. Using the definition of T and the fact that X is an integer, we have that $X < E(X)$ implies $X \leq T$. Thus:

$$E(X) \ \leq \ m \cdot Prob(X \geq E(X)) \ + \ T \cdot Prob(X < E(X)) \tag{3}$$

On the other hand

$$Prob(X \geq E(X)) \ + \ Prob(X < E(X)) \ = \ 1$$

Denote by α the quantity $Prob(X \geq E(X))$. Then $Prob(X < E(X)) = 1 - \alpha$. Hence, plugging the appropriate values into Equation 3 above, we obtain

$$E(X) \leq m\alpha \;+\; T \cdot (1-\alpha) \tag{4}$$

Recalling the definition of T, we have

$$T \;=\; \lceil E(X) - 1 \rceil \;\leq\; E(X) - \frac{1}{8}$$

where the inequality holds because $E(X) = \frac{7}{8}m$, which equals $a + \frac{b}{8}$ for some non-negative integers a and b so that $b \leq 7$. (Extra help: if $b = 0$ then $\lceil a + \frac{b}{8} - 1 \rceil = a - 1 < (a + \frac{b}{8}) - \frac{1}{8}$, whereas if $b > 0$ then $\lceil a + \frac{b}{8} - 1 \rceil = a \leq (a + \frac{b}{8}) - \frac{1}{8}$.) Thus, Equation 4 yields:

$$E(X) \leq m\alpha \;+\; (E(X) - \frac{1}{8}) \cdot (1-\alpha) \tag{5}$$

By a sequence of manipulations, we get the following sequence of equivalent conditions:

$$\alpha E(X) \;\leq\; m\alpha \;-\; \frac{1-\alpha}{8} \tag{6}$$

$$\frac{1}{8} \;\leq\; (m - E(X) + \frac{1}{8}) \cdot \alpha \tag{7}$$

$$\alpha \;\geq\; \frac{1}{8m - 8E(X) + 1} \;=\; \frac{1}{8m - 7m + 1} \tag{8}$$

We conclude that $\alpha \geq \frac{1}{m+1}$, and the claim follows. ■

Having proved the above claim, we can now present a randomized algorithm which will get a formula φ as its input and output an assignment to the variables of φ that will satisfy $\frac{7}{8}$ of its clauses with exponentially high probability. The algorithm is as hinted above: it will selected at random truth assignments to the variables of φ until finding one as desired (which happens with probability at least $1/|\varphi|$ at each attempt). In fact, we prefer to present a version in which the number of attempts is bounded although the algorithm fails with small probability. The algorithm is as follows:

Repeat Steps 1–3 for $|\varphi|^2$ times

1. **select** a random assignment for the formula φ
2. **check** how many clauses of φ the chosen assignment satisfies
3. **if** the assignment satisfies at least $\frac{7}{8}$ of the clauses **then** output the assignment

Each iteration of the algorithm takes polynomial time, and so does the entire algorithm. Using the Claim 3.1 the probability that the algorithm will succeed in each single iteration is at least $\frac{1}{m+1}$, and this is at least $\frac{1}{|\varphi|}$ (since $|\varphi| > m$). Thus, the probability that the algorithm will fail in one iteration is less or equal to $1 - \frac{1}{|\varphi|}$. It follows that the probability that we fail in all $|\varphi|^2$ iterations is less or equal to $(1 - \frac{1}{|\varphi|})^{|\varphi|^2}$ which is approximately $e^{-|\varphi|}$. We conclude that the probability that algorithm succeeds is at least $1 - e^{-|\varphi|}$.