Probability theory for cryptography*

Firas Kraïem†

16th December 2017

These notes provide an overview of the basic material in probability theory which is necessary for a good understanding of probabilistic arguments in the theory of cryptography. Although in many cases one is able to make do with an informal, intuitive understanding of probability, a formal and rigorous understanding thereof can be of great help when one encounters a more complex probabilistic argument.

These notes consist of three sections. The first section is a general introduction to probability theory, where we quickly restrict ourselves to finite probability spaces, since the probability spaces typically encountered in the theory of cryptography are finite. The second section describes probabilistic algorithms (*i.e.*, Turing machines) and how the theory developed in the first section applies to them. Finally, the third section is devoted to stating and proving several probabilistic inequalities which are used often in cryptographic arguments. We note that in the first two sections, very few statements are proved or even justified, and the reader is encouraged to prove any result which does not seem immediately obvious.

Contents

1 Probability theory **1**
- 1.1 Probability spaces . . . 2
- 1.2 Independence and conditional probabilities . . . 3
- 1.3 Random variables . . . 4
- 1.4 Expectation and variance . . . 5

2 Probability theory in cryptography **6**
- 2.1 Probabilistic algorithms . . . 6
- 2.2 Probabilistic arguments in cryptography . . . 8

3 Some useful probabilistic inequalities **10**
- 3.1 The difference lemma . . . 10
- 3.2 The averaging argument . . . 10
- 3.3 The union bound . . . 11
- 3.4 Markov's inequality . . . 12
- 3.5 Chebychev's inequality and approximation by repeated sampling . . . 12

References **15**

*The LaTeX source for this document can be found at `http://svn.fkraiem.org/listing.php?repname=crypto`.

†Information Security Laboratory, Tohoku University, Japan. `firas@isl.is.tohoku.ac.jp`

1 Probability theory

> We shall not define *set*, but shall just hope that when such expressions as "the set of all real numbers" or "the set of all members of the United States Senate" are used, people's various ideas of what is meant are sufficiently similar to make communication feasible.
>
> John B. Fraleigh, *A First Course in Abstract Algebra* [3]

In this section, all sets are assumed to be subsets of some "universal set" Ω. For a set A, we note A^c or $\overline{A}$ its *complement* $\Omega \setminus A$, which consists of all the elements of Ω which are not in A. We start by recalling the De Morgan laws, as applied to unions and intersections of sets. For completeness, we also recall the definitions of the union and intersection of a (countably) infinite number of sets.

Definition 1.1 (Union and intersection)**.** Let $A_1, A_2, \dots$ be subsets of Ω. Their *union*, noted $\bigcup_{i=1}^{\infty} A_i$ is the set of all elements ω of Ω which are in at least one of the A_i, and their *intersection*, noted $\bigcap_{i=1}^{\infty} A_i$, is the set of elements which are in all of them; formally,

$$\bigcup_{i=1}^{\infty} A_i = \{\omega \in \Omega \mid \exists i, \omega \in A_i\} \quad \text{and} \quad \bigcap_{i=1}^{\infty} A_i = \{\omega \in \Omega \mid \forall i, \omega \in A_i\}.$$

Proposition 1.2 (The De Morgan laws)**.** *For any sequence $A_1, A_2, \dots$ of subsets of Ω, we have*

$$\overline{\bigcup_{i=1}^{\infty} A_i} = \bigcap_{i=1}^{\infty} \overline{A_i} \quad \textit{and} \quad \overline{\bigcap_{i=1}^{\infty} A_i} = \bigcup_{i=1}^{\infty} \overline{A_i}.$$

We note that these definitions also cover unions and intersections of finite collections of sets: in order to construct the union (resp., intersection) of n sets $A_1, \dots, A_n$, we can set $A_i = \emptyset$ (resp., $A_i = \Omega$) for all $i > n$. Of course, the De Morgan laws still apply to such collections.

1.1 Probability spaces

We can now define the basic objects of probability theory, following [5]. We note that in general a "set of sets" is not always actually a set, because anomalies similar to Russell's paradox may occur. Thus in the general case we will speak of a *collection* of subsets of Ω, rather than of a *set* of subsets of Ω. Later, we will restrict ourselves to the *power set* (*i.e.*, the set of subsets) $\mathfrak{P}(\Omega)$ of Ω, which is indeed a set.

Definition 1.3 (σ-algebras)**.** Let Ω be a set and $\mathcal{A}$ be a collection of subsets of Ω. $\mathcal{A}$ is a *σ-algebra* if the following conditions hold.

- $\Omega \in \mathcal{A}$.
- For any $A \in \mathcal{A}$, $A^c \in \mathcal{A}$.
- For any sequence $A_1, A_2, \dots$ of elements of $\mathcal{A}$, $\bigcup_{i=1}^{\infty} A_i \in \mathcal{A}$.

Remark 1.4 (More properties). Other properties can be derived from the three properties above and the De Morgan laws; for example we can see that the intersection of a sequence of elements of a σ-algebra $\mathcal{A}$ is in $\mathcal{A}$.

Example 1.5. If Ω is any set and A is any subset of Ω, the collections $\{\emptyset, \Omega\}$, $\{\emptyset, A, A^c, \Omega\}$, and $\mathfrak{P}(\Omega)$ are σ-algebras on Ω.

Definition 1.6 (Measurable spaces)**.** A pair (Ω, Σ) where Ω is a set and Σ is a σ-algebra on Ω is called a *measurable space*. The elements of Σ are called the *measurable subsets* of Ω.

Definition 1.7 (Probability measures)**.** Given a measurable space (Ω, Σ), P is a *probability measure* on (Ω, Σ) if the following conditions hold.

- For any $A \in \Sigma$, $P(A)$ is a non-negative real number, called the *probability of* A.
- $P(\Omega) = 1$.
- If $A_1, A_2, \ldots$ are pairwise disjoint elements of Σ (*i.e.*, if $A_i \cap A_j = \emptyset$ whenever $i \neq j$), then

$$P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i).$$

Remark 1.8 (More properties). Here also it is possible to derive additional properties from the three above. Most importantly, $P(A) \leq 1$ for all A, and $P(\emptyset) = 0$ if Ω is not empty. However, we note that $P(A) = 0$ does not imply $A = \emptyset$.

Example 1.9. Let Ω be a non-empty set, A be a non-empty proper subset of Ω, and $\Sigma = \{\emptyset, A, A^c, \Omega\}$. For any $p \in [0, 1]$, we can define a probability measure on (Ω, Σ) by $P(\emptyset) = 0$, $P(A) = p$, $P(A^c) = 1 - p$, and $P(\Omega) = 1$.

Definition 1.10 (Probability spaces)**.** A *probability space* is a triple (Ω, Σ, P) such that the following conditions hold.

- Ω is a set, called the *sample space*.
- Σ is a σ-algebra on Ω (*i.e.*, (Ω, Σ) is a measurable space). The elements of Σ are called *events*.
- P is a probability measure on (Ω, Σ), called the *probability distribution*.[1]

For each $\omega \in \Omega$, if $\{\omega\}$ is an event (*i.e.*, if it is in Σ), it is called an *elementary event*. A probability space is called *finite* if its sample space is finite. In the following, we assume that Ω is finite and not empty, and that $\Sigma = \mathfrak{P}(\Omega)$. It is clear that the probability distribution is then completely determined by the probabilities of all the elementary events, and that the sum of those probabilities equals 1. An important special case is when $P(\{\omega\}) = 1/|\Omega|$ for all $\omega \in \Omega$; the probability distribution P is then said to be *uniform*. In the following, we will sometimes abuse notation and write $P(\omega)$ instead of $P(\{\omega\})$ to denote the probability of the elementary event $\{\omega\}$.

Example 1.11 (Throw of a die)**.** The experiment of throwing a (fair) six-sided die can be formalised by a probability space with sample space $\{1, \ldots, 6\}$ where the probability of each elementary event is $1/6$ (*i.e.*, the probability distribution is uniform). The elementary event $\{n\}$ naturally "represents" the case when the n side comes up. We can then consider "non-elementary" events, for example the event $\{2, 4, 6\}$ represents the case where the side which comes up is even, and the probability of this event is $1/2$, as intuition dictates.

[1] In cryptography it is common to denote the probability distribution by Pr instead of P; however we will use P in these notes.

1.2 Independence and conditional probabilities

The notion of *independence* is central in probability theory, and indeed probabilistic arguments in the theory of cryptography are no exception. It formalises the perception that past events do not influence the outcome of future ones or provide any information about them. Put another way, if two events are independent, the order in which they occur is of no consequence, and it may sometimes help to think of one occurring after the other, even if it actually occurs before.

Definition 1.12 (Independent events)**.** Let $A_1, \ldots, A_n$ be events on (Ω, Σ, P). Those events are said to be *independent* if for all subsets I of $\{1, \ldots, n\}$ we have

$$P\left(\bigcap_{i \in I} A_i\right) = \prod_{i \in I} P\left(A_i\right).$$

They are said to be *pairwise independent* if any two of them are independent, *i.e.*, if

$$P(A_i \cap A_j) = P(A_i) \cdot P(A_j)$$

whenever $i \neq j$.

Remark 1.13. The definition of independence can be generalised to a (countably) infinite sequence of events: the events $A_1, A_2, \ldots$ are independent if the events $A_1, \ldots, A_n$ are independent for all n.

Example 1.14. Considering again the case of a fair six-sided die, the events $\mathsf{odd} = \{1, 3, 5\}$ and $\mathsf{lowerhalf} = \{1, 2, 3\}$ are not independent, because $P(\mathsf{odd}) = P(\mathsf{lowerhalf}) = 1/2$, whereas $P(\mathsf{odd} \cap \mathsf{lowerhalf}) = 1/3 \neq 1/2 \cdot 1/2$. On the other hand the events odd and $\mathsf{lowerthird} = \{1, 2\}$ are independent.

The definition of independence naturally leads to conditional probabilities. Intuitively, independence of two events A and B means that the probability that B occurs is not changed if we "know" that A has occurred. Conditional probabilities formalise "the probability that B occurs if we know that A has occurred".

Definition 1.15 (Conditional probabilities)**.** Let A and B be two events, with $P(A) > 0$. Then the *conditional probability of B given A* is

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

Clearly, A and B are independent if and only if $P(B \mid A) = P(B)$.

Example 1.16. With the notation above, we have $P(\mathsf{odd} \mid \mathsf{lowerhalf}) = 2/3$. Intuitively, of the three possible "lower half" results, two are odd, so if we know that the result is in the lower half, the probability that it is odd becomes 2/3. (An equivalent experiment would be to throw an imaginary three-sided die.)

1.3 Random variables

Typically, we are interested in quantities which are defined not on the sample space Ω but on some other set. For example, say you play a game where, upon the throw of a fair six-sided die, you win three dollars if the five or the six side comes up and lose one dollar in all the other cases. Then you do not care very much about which particular side comes up, only about how much money you win or lose, which is a *function* from the sample space to the set $\mathbf{R}$ of real numbers. Random variables formalise such functions.

Definition 1.17 (Random variables)**.** A *random variable* on a probability space (Ω, Σ, P) is a function from Ω to $\mathbf{R}$.

A random variable X on a probability space (Ω, Σ, P) induces a probability distribution $\mathbf{P}$ on the measurable space $(\mathbf{R}, \mathfrak{P}(\mathbf{R}))$ by

$$\mathbf{P}(A) = P(X^{-1}(A)) = P(\{\omega \in \Omega \mid X(\omega) \in A\}) = \sum_{\substack{\omega \in \Omega \\ X(\omega) \in A}} P(\omega).$$

For a subset A of $\mathbf{R}$, $\mathbf{P}(A)$ naturally represents the probability that the random variable X takes a value in A, and in particular $\mathbf{P}(\{a\})$ for some $a \in \mathbf{R}$ represents the probability that the value of X equals a.

Example 1.18. With the example in the first paragraph of this subsection, we have as before $\Omega = \{1, \ldots, 6\}$, and the random variable X representing the amount of money gained is defined as

$$X(5) = X(6) = 3$$

and

$$X(1) = X(2) = X(3) = X(4) = -1.$$

We thus have

$$\mathbf{P}(\{3\}) = P(\{5, 6\}) = \frac{1}{3},$$

and likewise $\mathbf{P}(\{-1\}) = 2/3$, which means that the probability that you gain three dollars (resp., that you lose one dollar) is $1/3$ (resp., $2/3$).

Since random variables are functions, we can perform the usual functional operations on them. For example if X and Y are two random variables defined on the same probability space (Ω, Σ, P), we can define the random variable $X + Y$ as $(X + Y)(\omega) = X(\omega) + Y(\omega)$ for all $\omega \in \Omega$. We can likewise define the random variables $X - Y$, $X \cdot Y$, etc., and if $f : \mathbf{R} \to \mathbf{R}$ is a function, we can define the random variable $f(X)$.

In the following, P will usually be uniform, and we will only be interested in $\mathbf{P}$. Thus we identify the two, and for a subset A of $\mathbf{R}$ we write $P(X \in A)$ instead of $\mathbf{P}(A)$. Moreover, if $A = \{a\}$ for some $a \in \mathbf{R}$, we write $P(X = a)$. Thus in the end when we write $P(X \in A)$ (resp., $P(X = a)$) we mean $P(X^{-1}(A))$ (resp., $P(X^{-1}(\{a\}))$). Similarly, we will write $P(X < a)$, $P(X \leq a)$, etc. When we write $P(A)$, without any symbol, A is an element of Σ, as in the formal definition of P.

In later sections, we will abuse terminology slightly and consider random variables which are mappings from Ω to some set S other than $\mathbf{R}$. If S is finite, we then say that a random variable $X : \Omega \to S$ is *uniformly distributed* if $P(X = s) = 1/|S|$ for all $s \in S$.

1.4 Expectation and variance

Before agreeing to play the game mentioned above, you may want to consider whether it is fair. That is, whether you are more likely to win or lose. Of course, in a straightforward sense it is obvious that you are more likely to lose, since you lose with probability $2/3$ (if the result of the roll is 1, 2, 3, or 4). However, it is certainly possible that you will win, and in that case you gain more money than you would lose. The *expectation* of the random variable X can tell you exactly how much money you can expect to gain or lose in an average sense. In the following, all random variables are defined on the same probability space (Ω, Σ, P), and we recall that we assume that Ω is finite and not empty, and that $\Sigma = \mathfrak{P}(\Omega)$.

Definition 1.19 (Expectation)**.** Let X be a random variable. The *expectation* (or *expected value*) of X is

$$E(X) = \sum_{\omega \in \Omega} P(\omega) \cdot X(\omega) = \sum_{x \in \mathbf{R}} x \cdot P(X = x).$$

Example 1.20. The expectation of the variable X representing the money gained in our game is

$$3 \times \frac{1}{3} - 1 \times \frac{2}{3} = \frac{1}{3},$$

and you should definitely play since on average you will gain one third of a dollar each time.

Proposition 1.21 (Linearity of expectation)**.** *Let X and Y be two random variables, and $a \in \mathbf{R}$. Then we have*

- $E(a) = a$ *(the value a can be seen as a constant random variable,* i.e.*, $\omega \mapsto a$ for all $\omega \in \Omega$);*
- $E(a \cdot X) = a \cdot E(X)$*; and*
- $E(X + Y) = E(X) + E(Y)$.

Definition 1.22 (Independent random variables)**.** Two random variables X and Y are *independent* if for all $x, y \in \mathbf{R}$, we have

$$P(X = x \cap Y = y) = P(X = x) \cdot P(Y = y)$$

(*i.e.*, the events $X^{-1}(\{x\})$ and $Y^{-1}(\{y\})$ are independent). We similarly say that n random variables $X_1, \ldots, X_n$ are independent if for all $x_1, \ldots, x_n \in \mathbf{R}$, we have

$$P\left(\bigcap_{i=1}^{n} (X_i = x_i)\right) = \prod_{i=1}^{n} P(X_i = x_i),$$

and that they are *pairwise independent* if A_i and A_j are independent whenever $i \neq j$.

Proposition 1.23. *Let X and Y be two* independent *random variables. Then*

$$E(X \cdot Y) = E(X) \cdot E(Y).$$

The *variance* of a random variable measures in an average sense the amount by which it differs from its expectation. We note that since the variance is the expectation of a random value which takes non-negative values, it is non-negative.

Definition 1.24 (Variance)**.** The *variance* of a random variable X, noted[2] $V(X)$, is the expectation of the random variable $(X - E(X))^2$.

Proposition 1.25. *For any random variable X and real number a, we have*

$$V(a \cdot X) = a^2 \cdot V(X).$$

2 Probability theory in cryptography

> We do not assume anything about the distribution of the instances of the problem to be solved. Instead we incorporate randomization into the algorithm itself.
>
> Michael O. Rabin, *Probabilistic Algorithms* [6]

[2] Or sometimes σ^2; $\sigma = \sqrt{V(X)}$ is called the *standard deviation*.

2.1 Probabilistic algorithms

In this section we assume familiarity with ordinary (*deterministic*) Turing machines, as described for example in [1, Chapter 1], and we present a basic treatment of *probabilistic* ones (see Chapter 7 of the same text for a more extensive one). We always assume that Turing machines run in *polynomial time*. That is, when we consider a Turing machine, it is always implied that there is some polynomial t such that for every input x, the machine always halts after at most $t(|x|)$ steps. In fact it is possible to assume that the machine halts after *exactly* $t(|x|)$ steps.

There are many equivalent ways to define probabilistic Turing machines as a "modification" of deterministic ones. A very concise one would be the following.

Definition 2.1 (Probabilistic Turing machines)**.** A *probabilistic Turing machine* is a Turing machine with two transition functions δ_0 and δ_1. At each step of the computation, one or the other transition function is followed, each with probability 1/2.

However, the following one is more customary in the theory of cryptography, and that is the one we will use.

Definition 2.2 (Probabilistic Turing machines)**.** A *probabilistic Turing machine* is a Turing machine with an extra tape called the *random tape*. Before the computation begins on input x, a string of length $t(|x|)$ is chosen uniformly (among all the strings of length $t(|x|)$) and written on the random tape. Execution then proceeds normally.

Remark 2.3. In informal descriptions of probabilistic algorithms it is often said that the algorithm "randomly generates" a string or some other object. Formally, this means that the machine reads some of the random bits written on its random tape.

The execution of a probabilistic Turing machine M on inputs of length n naturally induces a probability space with sample space $\{0,1\}^{t(n)}$ and uniform probability distribution. For any $x \in \{0,1\}^n$, we can then define a random variable by mapping any $r \in \{0,1\}^{t(n)}$ to the string $M(x;r)$ representing the output of M on input x when the string r is written on its random tape before execution. This random variable naturally represents the output of M on input x, and we note it $M(x)$; thus for example the probability that M outputs 1 on input x is

$$\begin{aligned} P(M(x) = 1) &= P(\{r \in \{0,1\}^{t(n)} \mid M(x;r) = 1\}) \\ &= \frac{1}{2^{t(n)}} \cdot \left|\{r \in \{0,1\}^{t(n)} \mid M(x;r) = 1\}\right|. \end{aligned}$$

Example 2.4. Let M be a probabilistic Turing machine which, on input x of length n, reads the first bit of its random tape and outputs $x+1$ if it equals 1 and x otherwise (we can assume that the input is not empty, or treat an empty input as 0). Intuitively, it is clear that, for any input x, we have $P(M(x) = x) = 1/2$ and $P(M(x) = x+1) = 1/2$. Formally, letting $S = \{r \in \{0,1\}^{t(n)} \mid M(x;r) = x\}$ we have

$$P(M(x) = x) = \frac{1}{2^{t(n)}} \cdot |S|.$$

S is the set of all strings of length $t(n)$ whose first bit equals 0, and it is clear that $|S| = 2^{t(n)}/2$, which shows that $P(M(x) = x)$ indeed equals 1/2.

Probabilistic Turing machines naturally give rise to several complexity classes, the most important of which is the complexity class **BPP** (which stands for *bounded probability polynomial time*). **BPP** is essentially the probabilistic analogue of **P**: it captures problems which can be solved efficiently by probabilistic algorithms.

Definition 2.5 (The complexity class **BPP**)**.** A language $L \subseteq \{0,1\}^*$ is in the complexity class **BPP** if there is a probabilistic polynomial-time Turing machine M such that

- for all strings $x \in L$, $P(M(x) = 1) \geq 2/3$; and
- for all strings $x \notin L$, $P(M(x) = 0) \geq 2/3$.

Remark 2.6 (Open questions). It is clear that **P** is contained in **BPP**, but the converse is still an open question. The relationship between **BPP** and **NP** is even more mysterious: both directions are still open questions.

2.2 Probabilistic arguments in cryptography

Cryptographic arguments usually involve more complex experiments than just running a probabilistic algorithm on a given input. Namely, the input is usually not fixed but is chosen from some set according to some probability distribution, and this needs to be reflected in the sample space when such experiments are studied. A typical example of such an experiment is found in the definition of one-way permutations.

Definition 2.7 (One-way permutations)**.** A function $f : \{0,1\}^* \to \{0,1\}^*$ is a *one-way permutation* if it satisfies the following four conditions.

- **Length-preserving.** For all $x \in \{0,1\}^*$, $|f(x)| = |x|$.
- **Injective.** For all $x, y \in \{0,1\}^*$, if $x \neq y$, then $f(x) \neq f(y)$.
- **Easy to compute.** There is a deterministic polynomial-time algorithm M such that for all $x \in \{0,1\}^*$, $M(x) = f(x)$.
- **Hard to invert.** For any probabilistic polynomial-time algorithm A, any polynomial p, and all sufficiently large n, we have[3]

$$P\left[x \leftarrow \{0,1\}^n : A(f(x)) = x\right] < \frac{1}{p(n)}.$$

Remark 2.8. We note that a one-way permutation is indeed a permutation of $\{0,1\}^*$, *i.e.*, a bijection from $\{0,1\}^*$ to itself.

The probabilistic expression in the hardness-to-invert condition represents the following experiment. Given a function f (which can be assumed to satisfy the first three conditions), a probabilistic polynomial-time Turing machine A, and an integer n, a string x is first chosen uniformly among all the strings of length n. A is then invoked on input $f(x)$, and we are interested in the probability that it outputs x. If this is the case, we say that A *succeeds*. This probability may be different for each choice of x: for example letting $s = f(0^n)$, we can incorporate this in the algorithm A so that on input s, it always outputs 0^n. Then, if $x = 0^n$, A succeeds with probability 1, but this certainly cannot be the case for all x, otherwise one-way permutations do not exist (but it is widely believed that they do).

So we need to include the choice of x in our sample space: for the above experiment, the sample space is $\Omega = \{0,1\}^n \times \{0,1\}^{t(n)}$, *i.e.*, the set of pairs (x, r) where $x \in \{0,1\}^n$ and $r \in \{0,1\}^{t(n)}$. We then define the random variable $X(x, r) = A(f(x); r)$, and we are interested in the probability that $X(x, r) = x$, which equals

$$P(X(x,r) = x) = \frac{1}{2^{n+t(n)}} \cdot \left|\{(x,r) \in \Omega \mid A(f(x);r) = x\}\right|.$$

[3]Or equivalently, $P\left[y \leftarrow \{0,1\}^n : A(y) = f^{-1}(y)\right]$.

Remark 2.9. In informal descriptions of experiments like the above, it is common to say that "the probability is taken over the choice of such-and-such". What this means is that all the possible values for the objects in question are included in the sample space over which the probability is defined. For example, for the experiment above one would say that the probability is taken over the choice of x and of the random tape of A.

One example of how thinking in terms of probability spaces can help understanding is the proof of the following result, which says informally that a one-way permutation can have infinitely many *fixed points*. That is, there can be infinitely many x such that $f(x) = x$.

Proposition 2.10. *Let f be a one-way permutation. Then the function g defined by*

$$g(x) = \begin{cases} 0^n & \text{if } x = 0^n; \\ f(0^n) & \text{if } x = f^{-1}(0^n); \text{ and} \\ f(x) & \text{otherwise} \end{cases}$$

is a one-way permutation.

Remark 2.11. We note that no ambiguity exists in the definition of g: if $x = 0^n = f^{-1}(0^n)$, then both the first and the second conditions yield $g(x) = 0^n$.

Proof. The proof that g satisfies the first three requirements for a one-way permutation is left to the reader, and we turn to what interests us here: to show that g is hard to invert if f is. This is, as usual, proved by contradiction, so we assume that g is easy to invert, and show that then f is easy to invert as well. To say that g is easy to invert means that there exists a probabilistic polynomial-time algorithm A and a polynomial p such that for infinitely many n, we have

$$P\left[x \leftarrow \{0,1\}^n : A(g(x)) = x\right] \geq \frac{1}{p(n)}.$$

In the following, we restrict our attention to such n. We must show that there exists a probabilistic polynomial-time algorithm A' and a polynomial q such that for infinitely many n, we have

$$P\left[x \leftarrow \{0,1\}^n : A'(f(x)) = x\right] \geq \frac{1}{q(n)},$$

which contradicts our assumption that f is hard to invert. We show that this is true for $A' = A$ and $q(n) = 2 \cdot p(n)$.

Our underlying sample space is again $\Omega = \{0,1\}^n \times \{0,1\}^{t(n)}$. We consider the two events

$$S_f = \{(x,r) \in \Omega \mid A(f(x);r) = x\} \quad \text{and} \quad S_g = \{(x,r) \in \Omega \mid A(g(x);r) = x\},$$

and the two probabilities above are then respectively

$$P(S_g) = \frac{|S_g|}{2^{n+t(n)}} \geq \frac{1}{p(n)} \quad \text{and} \quad P(S_f) = \frac{|S_f|}{2^{n+t(n)}}.$$

It is clear that S_f and S_g are "almost equal". Namely, if x does not equal 0^n or $f^{-1}(0^n)$, then $f(x) = g(x)$, and then $A(g(x);r) = A(f(x);r)$ for all r, and (x,r) is in S_f if and only if it is in S_g. We will use this observation to derive a lower bound on the cardinality of S_f.

By the observation above, S_f contains (at least) all the pairs (x,r) such that $(x,r) \in S_g$ and $x \notin \{0^n, f^{-1}(0^n)\}$. Let S be the set of such pairs; then $|S_f| \geq |S|$. Further, let S' be the set of

pairs (x, r) such that x equals 0^n or $f^{-1}(0^n)$. Then $S = S_g \setminus S'$, and so $|S| \geq |S_g| - |S'|$. We know that

$$|S_g| \geq \frac{2^{n+t(n)}}{p(n)},$$

and it is clear that

$$|S'| \leq 2^{t(n)+1}.$$

Thus in the end we have

$$|S_f| \geq |S| \geq |S_g| - |S'| \geq \frac{2^{n+t(n)}}{p(n)} - 2^{t(n)+1},$$

and so

$$P(S_f) = \frac{|S_f|}{2^{n+t(n)}} \geq \frac{1}{p(n)} - 2^{-(n-1)}.$$

But $2^{-(n-1)}$ is a *negligible* quantity, which means that if n is sufficiently large, it is smaller than the inverse of any polynomial. In particular, it is smaller than $1/(2 \cdot p(n))$, thus finally

$$P(S_f) \geq \frac{1}{p(n)} - \frac{1}{2 \cdot p(n)} \geq \frac{1}{2 \cdot p(n)}$$

for infinitely many n. □

3 Some useful probabilistic inequalities

> Les guerres, il faut les gagner. Survivre. Avoir les bons outils. Le logarithme juste. Le reste, poésie. Fausses promesses.
>
> Virginie Despentes, *Vernon Subutex*, tome 1

3.1 The difference lemma

The proof of Proposition 2.10 is a special case of the following result; the reader is encouraged to rewrite the proof using this lemma. Pushing this idea further leads to the *sequences of games* methodology introduced by Shoup [7] (see also the similar *code-based games* of Bellare and Rogaway [2]).

Proposition 3.1 (The difference lemma)**.** *Let A, B, F be events on a probability space (Ω, Σ, P), and suppose that $A \cap F^c = B \cap F^c$. Then $|P(A) - P(B)| \leq P(F)$.*

Proof. We have

$$\begin{aligned} |P(A) - P(B)| &= |P(A \cap F) + P(A \cap F^c) - P(B \cap F) - P(B \cap F^c)| \\ &= |P(A \cap F) - P(B \cap F)|, \end{aligned}$$

and since $P(A \cap F)$ and $P(B \cap F)$ are both between 0 and $P(F)$, they cannot differ by more than $P(F)$. □

3.2 The averaging argument

Another simple fact which is surprisingly useful is that if the average of n real numbers $x_1, \ldots, x_n$ is μ, then there is a k such that $x_k \geq \mu$. In probabilistic terms, this idea yields the following.

Proposition 3.2 (The averaging argument)**.** *If X is a random variable on a probability space (Ω, Σ, P) and $E(X) = \mu$, then $P(X \geq \mu) > 0$. That is, there is an $\omega \in \Omega$ such that $P(\omega) > 0$ and $X(\omega) \geq \mu$.*

Proof. Assuming for contradiction that $P(X \geq \mu) = 0$, we have $P(X = x) = 0$ for all $x \geq \mu$, and also $P(X < \mu) = 1$. Now,

$$\begin{aligned} E(X) &= \sum_{x \in \mathbf{R}} x \cdot P(X = x) \\ &= \sum_{x < \mu} x \cdot P(X = x) + \sum_{x \geq \mu} x \cdot P(X = x) \\ &= \sum_{x < \mu} x \cdot P(X = x) \\ &< \sum_{x < \mu} \mu \cdot P(X = x) \\ &< \mu \cdot \sum_{x < \mu} P(X = x) \\ &< \mu \cdot P(X < \mu) \\ &< \mu, \end{aligned}$$

contradicting the assumption that $E(X) = \mu$. □

3.3 The union bound

The *union bound* says that given finitely many events, the probability that at least one of them occurs can be no greater than the sum of their probabilities.

Proposition 3.3 (The union bound)**.** *Let $A_1, \ldots, A_n$ be events on a probability space (Ω, Σ, P). Then*

$$P\left(\bigcup_{i=1}^{n} A_i\right) \leq \sum_{i=1}^{n} P(A_i).$$

Proof. The proof is by induction on n. If $n = 1$, the two probabilities are just $P(A_1)$, so the inequality holds (and is in fact an equality). If $n = 2$, then

$$\begin{aligned} P(A_1 \cup A_2) &= P(A_1) + P(A_2) - P(A_1 \cap A_2) \\ &\leq P(A_1) + P(A_2). \end{aligned}$$

Suppose now that the inequality is true for some n. Then

$$\begin{aligned} P\left(\bigcup_{i=1}^{n+1} A_i\right) &= P\left(\left(\bigcup_{i=1}^{n} A_i\right) \cup A_{n+1}\right) \\ &\leq P\left(\bigcup_{i=1}^{n} A_i\right) + P(A_{n+1}) \\ &\leq \left(\sum_{i=1}^{n} P(A_i)\right) + P(A_{n+1}) \\ &\leq \sum_{i=1}^{n+1} P(A_i). \end{aligned}$$

□

3.4 Markov's inequality

Given a random variable with some fixed lower bound, *Markov's inequality* gives a relationship between a value v and the probability that the variable takes a value larger than v.

Proposition 3.4 (Markov's inequality)**.** *Let X be a random variable taking non-negative real values, and let v be any positive real number. Then*

$$P(X \geq v) \leq \frac{E(X)}{v}.$$

Proof. By definition of $E(X)$ we have

$$\begin{aligned} E(X) &= \sum_{x \geq 0} x \cdot P(X = x) \\ &= \sum_{0 \leq x < v} x \cdot P(X = x) + \sum_{x \geq v} x \cdot P(X = x) \\ &\geq \sum_{0 \leq x < v} 0 \cdot P(X = x) + \sum_{x \geq v} v \cdot P(X = x) \\ &\geq v \cdot \sum_{x \geq v} P(X = x) \\ &\geq v \cdot P(X \geq v). \end{aligned}$$

□

3.5 Chebychev's inequality and approximation by repeated sampling

Chebychev's inequality helps us estimate the probability that X deviates from $E(X)$ by a prescribed amount $\varepsilon > 0$, if we know $V(X)$ or a good upper bound thereof.

Proposition 3.5 (Chebychev's inequality)**.** *Let X be a random variable and ε be a positive real number. Then*

$$P\left(|X - E(X)| \geq \varepsilon\right) \leq \frac{V(X)}{\varepsilon^2}.$$

Proof. We define a random variable $Y = (X - E(X))^2$, and we note that $E(Y) = V(X)$ and that Y is non-negative. Applying Markov's inequality to Y, we obtain

$$P\left((X - E(X))^2 \geq \varepsilon^2\right) \leq \frac{V(X)}{\varepsilon^2},$$

and since $(X - E(X))^2 \geq \varepsilon^2$ if and only if $|X - E(X)| \geq \varepsilon$, this completes the proof. □

The bound of Chebychev's inequality can be rendered even lower by taking the average of several samples, if the samples are taken in a pairwise independent manner.

Corollary 3.6 (Approximation by repeated sampling)**.** *Let $X_1, \ldots, X_n$ be pairwise independent random variables with the same expectation μ and variance σ^2. Then for any positive integer n and positive real number ε, we have*[4]

$$P\left(|\frac{\sum_{i=1}^{n} X_i}{n} - \mu| \geq \varepsilon\right) \leq \frac{\sigma^2}{n \cdot \varepsilon^2}.$$

Proof. For all i, we define a random variable $\tilde{X}_i = X_i - E(X_i)$, and we note that the variables $\tilde{X}_i$ are pairwise independent and have zero expectation, and that $E(\tilde{X}_i^2) = \sigma^2$ for all i. We apply Chebychev's inequality to the random variable $\sum_{i=1}^{n} \frac{X_i}{n}$ (which clearly has expectation μ), and we obtain

$$\begin{aligned} P\left(|\frac{\sum_{i=1}^{n} X_i}{n} - \mu| \geq \varepsilon\right) &\leq \frac{V\left(\sum_{i=1}^{n} \frac{X_i}{n}\right)}{\varepsilon^2} \\ &\leq \frac{V\left(\sum_{i=1}^{n} X_i\right)}{n^2 \cdot \varepsilon^2}. \end{aligned}$$

Now,

$$\begin{aligned} V\left(\sum_{i=1}^{n} X_i\right) &= E\left[\left(\sum_{i=1}^{n} X_i - E\left(\sum_{i=1}^{n} X_i\right)\right)^2\right] \\ &= E\left[\left(\sum_{i=1}^{n} \tilde{X}_i\right)^2\right] \\ &= \sum_{i=1}^{n}\left(E\left(\tilde{X}_i^2\right) + \sum_{\substack{j=1 \\ j \neq i}}^{n} E\left(\tilde{X}_i \cdot \tilde{X}_j\right)\right). \end{aligned}$$

If $i \neq j$, $\tilde{X}_i$ and $\tilde{X}_j$ are independent, so we have $E(\tilde{X}_i \cdot \tilde{X}_j) = E(\tilde{X}_i) \cdot E(\tilde{X}_j)$, and recalling that $E(\tilde{X}_i) = 0$ for all i, we obtain finally

$$\begin{aligned} V\left(\sum_{i=1}^{n} X_i\right) &= \sum_{i=1}^{n} E\left(\tilde{X}_i^2\right) \\ &= \sum_{i=1}^{n} V\left(X_i\right) \\ &= n \cdot \sigma^2. \end{aligned}$$

□

The following example illustrates that approximation by repeated sampling is a very powerful tool to reduce the error probability of probabilistic algorithms. We recall the definition of the complexity class **BPP**.

[4] We can note that Chebychev's inequality is the special case $n = 1$.

Definition 3.7 (The complexity class **BPP**)**.** A language $L \subseteq \{0,1\}^*$ is in the complexity class **BPP** if there is a probabilistic polynomial-time machine M such that

- for all strings $x \in L$, $P(M(x) = 1) \geq 2/3$; and
- for all strings $x \notin L$, $P(M(x) = 0) \geq 2/3$.

Reading the previous definition, one wonders what is so special about the number 2/3 so that we use it (instead of, say, 3/4) in the definition of **BPP**. The answer is that there is nothing special at all about 2/3, and we obtain the same class if we replace 2/3 by any constant between 1/2 and 1 (exclusive). In fact, we can do even more: we can replace 2/3 by $1/2 + \varepsilon$ or by $1 - \varepsilon$, where ε is a positive quantity which decreases polynomially with the length of x. As an example of how this can be proved, we prove the following.

Proposition 3.8. *Let L be a language, p be a polynomial, and M be a probabilistic polynomial-time algorithm such that for all $x \in L$, $P(M(x) = 1) = 1/2 + 1/p(|x|)$. Then there is a probabilistic polynomial-time algorithm M' such that for all $x \in L$, $P(M'(x) = 1) \geq 2/3$.*

Proof. We can suppose that M always outputs 0 or 1. M' simply runs $M(x)$ polynomially many times in a pairwise independent manner, and outputs the value (0 or 1) which occurs a strict majority of the time (we can assume that we run $M(x)$ an odd number of times). By "in a pairwise independent manner", we mean that before each run, a new string r is generated independently of any previous string and written on the random tape of M'. This implies that the random variables representing the output of M on each run are pairwise independent. The difficult part is to show that running $M(x)$ polynomially many times is indeed sufficient. Letting n be the number of runs and X_i be a random variable representing the output of M on the ith run, we want to find an n such that

$$P\left(\frac{\sum_{i=1}^{n} X_i}{n} > \frac{1}{2}\right) \geq \frac{2}{3},$$

or equivalently

$$P\left(\frac{\sum_{i=1}^{n} X_i}{n} \leq \frac{1}{2}\right) \leq \frac{1}{3}.$$

Now, it is easily seen that

$$E(X_i) = P(X_i = 1) = \frac{1}{2} + \frac{1}{p(|x|)},$$

and that

$$\begin{aligned} V(X_i) &= E(X_i) \cdot (1 - E(X_i)) \\ &= \left(\frac{1}{2} + \frac{1}{p(|x|)}\right) \cdot \left(\frac{1}{2} - \frac{1}{p(|x|)}\right) \\ &= \frac{1}{4} - \frac{1}{(p(|x|))^2}. \end{aligned}$$

In the following, let $\mu = E(X_i)$. Now, if $\frac{\sum_{i=1}^{n} X_i}{n} \leq \frac{1}{2}$, then $|\frac{\sum_{i=1}^{n} X_i}{n} - \mu| \geq \frac{1}{p(|x|)}$, and so

$$P\left(\frac{\sum_{i=1}^{n} X_i}{n} \leq \frac{1}{2}\right) \leq P\left(|\frac{\sum_{i=1}^{n} X_i}{n} - \mu| \geq \frac{1}{p(|x|)}\right),$$

and we can now apply Corollary 3.6. We obtain

$$\begin{aligned} P\left(\frac{\sum_{i=1}^{n} X_i}{n} \leq \frac{1}{2}\right) &\leq \frac{\frac{1}{4} - \frac{1}{(p(|x|))^2}}{n \cdot \frac{1}{(p(|x|))^2}} \\ &\leq \frac{\frac{(p(|x|))^2}{4} - 1}{n} \\ &\leq \frac{(p(|x|))^2 - 4}{4 \cdot n}. \end{aligned}$$

We want $\frac{(p(|x|))^2-4}{4\cdot n} \leq \frac{1}{3}$, so $4 \cdot n \geq 3 \cdot (p(|x|))^2 - 12$, and so finally we find that we must take

$$n \geq \frac{3 \cdot (p(|x|))^2 - 12}{4},$$

which is indeed polynomial in $|x|$. □

We can prove similarly that if there is a probabilistic polynomial-time algorithm M such that $P(M(x) = 1) = 2/3$ for all $x \in L$, then for any polynomial p there is a probabilistic polynomial-time algorithm M' such that $P(M'(x) = 1) \geq 1 - 1/p(|x|)$ for all $x \in L$. Thus in the end for any polynomials p and q, an algorithm with "success probability" $1 - 1/p(|x|)$ exists if and only if an algorithm with success probability $1/2 + 1/q(|x|)$ does.

In fact, using a stronger inequality known as the *Chernoff bound* (which we will not prove here; see [1, Theorem A.14] and references given there), we can even replace $p(|x|)$ by $2^{p(|x|)}$ if we take totally independent samples (rather than just pairwise independent). We also note that this method is not applicable if we start with an algorithm whose success probability is exactly $1/2$, since that would require us to apply Chebychev's inequality with $\varepsilon = 0$, which is impossible. (In an intuitive sense, an algorithm with success probability $1/2$ is a random guess, and it would be odd indeed if polynomially many random guesses were sufficient to obtain arbitrarily high success probability.)

References

[1] S. Arora and B. Barak, *Computational Complexity: A Modern Approach*. Cambridge University Press, New York, 2009. Preliminary draft available online at `http://theory.cs.princeton.edu/complexity/`.

[2] M.Bellare and P. Rogaway, *The Security of Triple Encryption and a Framework for Code-Based Game-Playing Proofs*. In *Advances in Cryptology - EUROCRYPT 2006*, S. Vaudenay (Ed.), *Lecture Notes in Computer Science*, vol. 4004, pp. 409-426, Springer, Berlin Heidelberg, 2006.

[3] J. B. Fraleigh, *A First Course in Abstract Algebra*, seventh edition. Pearson, 2003.

[4] O. Goldreich, *Foundations of Cryptography* (in two volumes). Cambridge University Press, Cambridge, 2001-2004. Preliminary drafts available online at `http://www.wisdom.weizmann.ac.il/~oded/foc-drafts.html`.

[5] A. Gut, *Probability: A Graduate Course*, second edition. *Springer Texts in Statistics*, Springer, New York, 2013. May be available online (depending on institutional access) at `http://link.springer.com/book/10.1007/978-1-4614-4708-5`.

[6] M. O. Rabin, *Probabilistic Algorithms.* In *Algorithms and Complexity: New Directions and Recent Results*, J. F. Traub (Ed.), pp. 21-39, Academic Press, New York, 1976.

[7] V. Shoup, *Sequences of Games: A Tool for Taming Complexity in Security Proofs. Cryptology ePrint Archive*, Report 2004/332, 2004. `http://eprint.iacr.org/2004/332`