Probability Theory Foundations of Cryptography

Meng Ziyu

January 15, 2022

1 Notational Conventions

Typically, the probability distribution in cryptography is only referred to the *discrete* probability distributions. And the probability space is the set of all strings over a certain string length l. Taken $l = 2$ as example, there are 4 possible strings: $00, 01, 10, 11$. As for the *random variables*, they are functions for mapping the sample space into the set of binary strings.

How to Read Probability Statements. Typically, we shall write the function $\Pr[f(X) = 1]$, where X is a random variable beforehand, and f is a function. We have to note that, all occurrences of a given symbol in a probabilistic statement refer to the same (unique) random variable. Hence we have the expression below:

$$\Pr[B(X, X)] = \sum_{x} \Pr[X = x] \cdot \chi(B(x, x))$$

Where X is a random variable, the function B is a boolean expression, the function χ is an indicator function. The expression above gives a measure of possibility that X is satisfied with boolean function B.

Given 2 independent random variables X and Y, we have the $\Pr[B(X, Y)]$ below:

$$\Pr[B(X, Y)] = \sum_{x,y} \Pr[X = x] \cdot \Pr[Y = y] \cdot \chi(B(x, y))$$

Which denotes that $B(x, y)$ holds when the pair (x, y) is chosen with probability $\Pr[X = x] \cdot \Pr[Y = y]$.

2 Tree Inequalities

Markov Inequality

Let X be a non-negative random variable and v a real number. Then

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

The Markov inequality indicates that for a **bounded random variable**, there must be a relation between the deviation of a value from the expectation and the probability assigned to this value.

Chebyshev's Inequality

Let X be a non-negative random variable and $\delta > 0$. Then

$$\Pr[|X - E(X)| \geq \delta] \leq \frac{V(X)}{\delta^2}$$

Where $V(X) = E(X^2) - E^2(X)$

Chebyshev's inequality is particularly useful for analysis of the error probability of approximation via repeated sampling.

Corollary(Pairwise - Indenpendent Sampling): Let $X_1, X_2, ..., X_n$ be pairwise-independent random variables with same expectations, denote μ, and same variance, denote σ^2. Then for every $\epsilon > 0$,

$$\Pr[|\frac{\sum_{i=1}^{n} X_i}{n} - \mu| \geq \epsilon] \leq \frac{\sigma^2}{n\epsilon^2}$$

Keywords: **pairwise-independent; same expectations**

Chernoff Bound: Let $p \leq \frac{1}{2}$,$X_1, X_2, ..., X_n$ be independent 0-1 random variables, so that $\Pr[X_i = 1]$=p for each i. Then for all $0 < \epsilon \leq p(1-p)$, we have

$$\Pr[|\frac{\sum_{i=1}^{n} X_i}{n} - p| \geq \epsilon] \leq 2e^{-\frac{\epsilon^2}{2p(1-p)}n}$$

Keywords: **independent 0-1 random variables**

Hoefding Inequality:

Let $X_1, X_2, ..., X_n$ be n independent random variables with same probability distribution, each ranging over the (real) interval $[a, b]$, and let μ denotes the expectation of each random variables. Then for $\epsilon > 0$, we have

$$\Pr[|\frac{\sum_{i=1}^{n} X_i}{n} - \mu| \geq \epsilon] \leq 2e^{-\frac{2\epsilon^2}{(b-a)^2}n}$$

Keywords: **independent; same probability distribution**