Problem Set 1

CS6230: Topics in Information Security
Due on: Aug 16, 2021
Total Points: 30

- While collaboration on other problem sets is generally encouraged, it is strongly recommended that you solve this one on your own. Familiarity with the concepts covered here will be crucial to following upcoming lectures.
- If you do collaborate, write up your submission on your own, and list the names of all of your collaborators on your submission.
- Solutions to some of the problems in this problem set are easy to find on the internet or in textbooks. Please do not look them up.
- Formal mathematical proofs are required in support of your answer to each of the problems.

1 Concentration Bounds

Concentration bounds are inequalities bounding the probability that a random variable is far from its expectation. These inequalities will be central to proofs throughout our course. The following exercise is an illustration of some of these inequalities. Below, all random variables in the problem descriptions are to be taken to be discrete.

Consider the process of tossing N balls uniformly into N bins, for some natural number N. That is, for each of the N balls, one of the N bins is selected uniformly at random, and the ball is tossed into it. We would like to provide a bound on what the maximum number of balls in any bin is likely to be at the end of this process.

Problem 1.1 (2 points)**.** *At the end of this process, what is the expected number of balls in the first bin?*

1.1 Markov's Inequality

The simplest concentration bound we will encounter is Markov's inequality. Though simple, it is used to prove many significantly stronger bounds.

Problem 1.2 (4 points)**.** *Prove that, for any positive real-valued random variable X and any $a > 0$,*

$$\Pr[X \geq a] \leq \frac{\mathrm{E}[X]}{a}$$

Often, applying Markov to an appropriate random variable already yields interesting results.

Problem 1.3 (4 points)**.** *Prove that the probability that the first bin has more than* 100 *balls is at most* $1/100$*.*

1.2 Chebyshev's Inequality

A stronger bound may be obtained by taking into account higher moments of the random variable in question. Recall that the variance of a random variable X is defined as follows:

$$\mathrm{Var}[X] = \mathrm{E}\left[(X - \mathrm{E}[X])^2\right]$$

Problem 1.4 (2 points)**.** *Prove that, for any random variable X and $a > 0$,*

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

The above quantifies the fact that random variables with small variance are more likely to be concentrated around their expectation.

Problem 1.5 (2 points)**.** *Prove that the probability that the first bin has more than $10\sqrt{N}$ balls is at most $1/100N$.*

Problem 1.6 (4 points)**.** *Prove that the probability that there exists a bin with more than $10\sqrt{N}$ balls is at most* $1/100$*.*

1.3 The Chernoff Bound

The bound that we will be using most often is stronger than the above, but only applies to a specific kind of random variable. Let the $X_1, \ldots, X_n$ be independent and identically distributed (i.i.d.) Bernoulli random variables with parameter p. That is, each X_i takes value either 0 or 1, and $\Pr[X_i = 1] = p$. Let $X = \sum_{i=1}^{n} X_i$, and $\mathrm{E}[X] = \mu$.

The Chernoff bound says that, with high probability, this sum of random variables is very close to its expectation. The following is one version of this bound that holds for any $k \geq 0$:

$$\Pr[(X - \mu) \geq k\mu] \leq e^{-\frac{k^2\mu}{2+k}}$$

Problem 1.7 (2 points)**.** *Prove that the probability that there exists a bin with more than* $10 \log N$ *balls is at most* $1/100$*.*

2 Polynomials of Low Degree

A number of our constructions of proof systems will rely on the fact that polynomials over finite fields of relatively low degree cannot have too many zeroes. In the case of univariate polynomials, this becomes the following simple fact – given any univariate polynomial f of degree d over a finite field $\mathbb{F}$, there are at most d distinct $x \in \mathbb{F}$ such that $f(x) = 0$.

2.1 Derandomising Freivald's Protocol

In class, we saw Freivald's Protocol that, given matrices $A, B, C \in \mathbb{F}^{n \times n}$, verifies that $C = A \cdot B$. It did this by picking a random vector $v \in \mathbb{F}^n$ and checking whether $Cv = (A \cdot B)v$. Consider, instead, the following slightly more complex procedure:

1. Pick a uniformly random $x \leftarrow \mathbb{F}$.
2. Construct the vector $v = (1 \; x \; x^2 \; \ldots \; x^{n-1})^T$.
3. Compute $u \leftarrow Bv$ and $w \leftarrow Au$.
4. Accept if $Cv = w$, and reject otherwise.

It is easy to see that the above verification procedure is perfectly complete (that is, it always accepts if $C = A \cdot B$). Its computational efficiency is roughly the same as before, but it uses much less randomness – just one random element from $\mathbb{F}$, as opposed to n random elements to sample v. Suppose the field $\mathbb{F}$ has size at least n^2.

Problem 2.1 (5 points)**.** *Prove that the soundness error in the above procedure is at most* $1/n$*. That is, if* $C \neq A \cdot B$*, then the probability that it accepts is at most* $1/n$*.*

2.2 Verifying Polynomial Multiplication

Consider the task of multiplying two degree-n polynomials over a finite field $\mathbb{F}$. This can be done in many fields with $O(n \log n)$ field operations using an analogue of the Fast Fourier Transform. Can this computation be *verified* faster than this?

You are given (the co-efficients of) three univariate polynomials f, g, h over a finite field $\mathbb{F}$, where the degrees of f and g are at most n, and the degree of h is at most $2n$. Your objective is to verify that $h = f \cdot g$. The field is of size more than $2n^2$.

Problem 2.2 (5 points)**.** *Design a procedure to do the above that uses* $O(n)$ *field operations. It should be perfectly complete and have soundness error at most* $1/n$*.*