~ MathDefs ~

CS 127: Cryptography / Boaz Barak

Homework 1

Total of 141 points. (Note that while this exercise is long, 100 points are a perfect score, so you don't *have* to solve all questions if you don't have the time for it.)

0. (10 points + 5 points bonus) Log in to canvas and: **(a)** Post on the canvas discussion board a short message introducing yourself to the rest of the class- what's your background and why you are interested in cryptography. Feel free to also add something about your non academic interests and hobbies. For a bonus of 5 points include a photo of yourself. **(b)** Answer on the "Week 0" module in canvas the "background questionnaire" quiz. This is not graded and there are no wrong answers- it's just a way for me to get a better sense of people's backgrounds.

Exercises from the "mathematical background" handout.

1. (16 points) In the following exercise X, Y denote random variables over some sample space S. You can assume that the probability on S is the uniform distribution— every point s is output with probability $1/|S|$. Thus $\mathbb{E}[X] = (1/|S|)\sum_{s\in S} X(s)$. We define the variance and standard deviation of X and Y as above (e.g., $Var[X] = \mathbb{E}[(X - \mathbb{E}[X])^2]$ and the standard deviation is the square root of the variance).

   a. (2 points) Prove that $Var[X]$ is always non-negative.

   b. (2 points) Prove that $Var[X] = \mathbb{E}[X^2] - \mathbb{E}[X]^2$.

   c. (2 points) Prove that always $\mathbb{E}[X^2] \geq \mathbb{E}[X]^2$.

   d. (2 points) Give an example for a random variable X such that $\mathbb{E}[X^2] \neq \mathbb{E}[X]^2$.

   e. (2 points) Give an example for a random variable X such that its standard deviation is *not equal* to $\mathbb{E}[|X - \mathbb{E}[X]|]$.

   f. (2 points) Give an example for two random variables X, Y such that $\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y]$.

   g. (2 points) Give an example for two random variables X, Y such that $\mathbb{E}[XY] \neq \mathbb{E}[X]\mathbb{E}[Y]$.

h. (2 points) Prove that if X and Y are independent random variables (i.e., for every x, y, $\Pr[X = x \wedge Y = y] = \Pr[X = x]\Pr[Y = y]$) then $\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y]$ and $Var[X + Y] = Var[X] + Var[Y]$.

2. (15 points) Suppose that H is chosen to be a random function mapping the numbers $\{1, \ldots, n\}$ to the numbers $\{1, .., m\}$. That is, for every $i \in \{1, \ldots, n\}$, $H(i)$ is chosen to be a random number in $\{1, \ldots, m\}$ and that choice is done independently for every i. For every $i \leq j \in \{1, \ldots, n\}$, define the random variable $X_{i,j}$ to equal 1 if there was a *collision* between $H(i)$ and $H(j)$ in the sense that $H(i) = H(j)$ and to equal 0 otherwise.

   a. (3 points) For every $i \leq j$, compute $\mathbb{E}[X_{i,j}]$.
   b. (3 points) Define $Y = \sum_{i \leq j} X_{i,j}$ to be the total number of collisions. Compute $\mathbb{E}[Y]$ as a function of n and m. In particular your answer should imply that if $m < n^2/1000$ then $\mathbb{E}[Y] > 1$ and hence in expectation there should be at least one collision and so the function H will not be one to one.
   c. (3 points) Prove that if $m > 1000 \cdot n^2$ then the probability that H is one to one is at least 0.9.
   d. (3 points) Give an example of a random variable Z (unrelated to the function H) that is always equal to a non-negative integer, and such that $\mathbb{E}[Z] \geq 1000$ but $\Pr[Z > 0] < 0.001$.
   e. (3 points) Prove that if $m < n^2/1000$ then the probability that H is one to one is at most 0.1.

3. (15 points) In this exercise we we will work out an important special case of the Chernoff bound. You can take as a given the following facts:

   I) The number of $x \in \{0, 1\}^n$ such that $\sum x_i = k$ is $\binom{n}{k} = \frac{n!}{k!(n-k)!}$.
   II) Stirling's approximation formula: for every $n \geq 1$,

   $$\sqrt{2\pi n}\left(\frac{n}{e}\right)^n \leq n! \leq 2\sqrt{2\pi n}\left(\frac{n}{e}\right)^n$$

   where $e = 2.7182\ldots$ is the base of the natural logarithm.

   Do the following:

   a. (5 points) Prove that for every n, $\Pr_{x \leftarrow_R \{0,1\}^n}[\sum x_i \geq 0.6n] < 2^{-n/1000}$

   The above shows that if you were given a coin of bias at least 0.6, you should only need some constant number of samples to be able to reject the "null hypothesis" that the coin is completely unbiased with extremely high confidence. In the following somewhat more challenging questions (which can be considered as bonus exercise) we try to show a converse to this:

   b. Let P be the uniform distribution over $\{0, 1\}^n$ and Q be the $1/2 + \epsilon$-biased distribution corresponding to tossing n coins in which each one has a probability of $1/2 + \epsilon$ of equalling 1 and probability $1/2 - \epsilon$

of equalling 0. Namely the probability of $x \in \{0,1\}^n$ according to Q is equal to $\prod_{i=1}^{n}(1/2 - \epsilon + 2\epsilon x_i)$.

i. (5 points) Prove that for every threshold θ between 0 and n, if $n < 1/(100\epsilon)^2$ then the probabilities that $\sum x_i \leq \theta$ under P and Q respectively differ by at most 0.1. Therefore, one cannot use the test whether the number of heads is above or below some threshold to reliably distinguish between these two possibilities unless the number of samples n of the coins is at least some constant times $1/\epsilon^2$.

ii. (5 points) Prove that for *every* function F mapping $\{0,1\}^n$ to $\{0,1\}$, if $n < 1/(100\epsilon)^2$ then the probabilities that $F(x) = 1$ under P and Q respectively differ by at most 0.1. Therefore, if the number of samples is smaller than a constant times $1/\epsilon^2$ then there is simply *no test* that can reliably distinguish between these two possiblities.

Exercises from Lecture 1

4. (20 points) Prove that every encryption scheme (E, D) is perfectly secret if and only if for every plaintexts $m, m' \in \{0,1\}^\ell$, the two random variables $\{E_k(m)\}$ and $\{E_{k'}(m')\}$ (for randomly chosen keys k and k') have precisely the same distribution.

5. (20 points- a bit harder bonus question) In the lecture we saw that any perfectly secret private key encryption scheme needs to use a key as large as the message. But we actually made an implicit subtle assumption: that the encryption process is *deterministic*. In a *probabilistic encryption scheme*, the encryption function E may be probabilistic: that is, given a message m and a key k, the value $E_k(x)$ is not fixed but is distributed according to some distribution $C_{x,k}$. The decryption function is still given only the key k and not the internal randomness used by E, and we require that for every message m, $\Pr[D_k(E_k(m)) = m] > 0.99$ where this probability is taken both over the choice of the key k and the internal randomness used by E. Prove that even a probabilistic encryption scheme cannot be perfectly secret with a key that's significantly shorter than the message. That is, show that for every probabilistic encryption scheme (E, D) using n-length keys and $n + 10$-length messages, there exist two messages $m, m' \in \{0,1\}^{n+10}$ such that the distributions $\{E_k(m)\}$ and $\{E_{k'}(m')\}$ are not identical.

Exercises from Lecture 2

6. (20 points) Prove the Computational Indistinguishability phrasing of computational security Theorem.

7. (20 points) Give a direct proof (not going through computational indistinguishability) in your own words for the length extension theorem in the special case $t = 2$ and when the messages are $m^0 = 00$ and $m^1 = 01$. That is, show how to transform an adversary *Eve* that can distinguish between the distribution $C^0 = (E'_{k_0}(k_1, 0), E'_{k_1}(k_2, 0))$ and $C^1 = (E'_{k_0}(k_1, 0), E'_{k_1}(k_2, 1))$ (for random k_0, k_1, k_2) with advantage ϵ into an adversay Eve' that runs in time polynomial in the running time of *Eve* and can distinguish between $E'_k(m')$ and $E'_k(m'')$ for two messages $m', m'' \in \{0,1\}^{n+1}$ with advantage at least, say, $\epsilon/10$.