EECS 376: Foundations of Computer Science
University of Michigan, Spring 2024 **Solutions to Discussion Worksheet 11**

1 Chernoff-Hoeffding Bounds

1. Suppose we flip a fair coin 376 times and want to bound the probability that at least 203 heads are flipped. Let X be the random variable representing the number of heads flipped. What is the best possible bound we can obtain using the following Chernoff-Hoeffding bound?

$$\Pr\left[\frac{1}{n}X \geq p+\varepsilon\right] \leq e^{-2\varepsilon^2 n}$$

Solution: First, observe that $X = \sum X_i$ where X_i is a 0-1 indicator taking value 1 if the i-th coin flip is heads, and 0 otherwise.

Since the coin is fair, $\mathbb{E}[X/n] = 1/376 \cdot 376/2 = 1/2$. Since we want $203/376 = 1/2+\varepsilon$, we get $\varepsilon = 203/376 - 1/2 = 15/376$. Therefore, $\Pr[X \geq 203] = \Pr[X/376 \geq 1/2 + 15/376] \leq e^{-2\cdot(15/376)^2\cdot 376}$. We then obtain the bound $\Pr[X \geq 203] \lessapprox 0.30$.

2. Suppose you have $n \geq 2$ bins. You independently toss n^2 balls at the bins, uniformly at random.

(a) Use Chernoff-Hoeffding to bound the probability that bin i has at least $2n$ balls.

Solution: Let $X =$ the number of balls in bin i, and let X_j be an indicator random variable for whether ball j lands in bin i. Thus, $X = X_1 + X_2 + ... + X_{n^2}$.
Pr(ball j lands in bin i) $= \frac{1}{n}$ (since they are all uniformly thrown at random). Then, $E[X] = (\frac{1}{n}) * n^2 =$ n balls.

We want to figure out $\Pr(X \geq 2n)$ using Chernoff-Hoeffding; thus, we must put it in the same form as the respective formula. First, we figure out $p = E[\frac{1}{n^2} * X] = (\frac{1}{n^2})E[X] = \frac{n}{n^2} = \frac{1}{n}$. Then we manipulate to look like Chernoff-Hoeffding:

$$\begin{aligned}\Pr(X \geq 2n) &= \Pr(\frac{1}{n^2}X \geq \frac{2n}{n^2})\\ &= \Pr(\frac{X}{n^2} \geq \frac{2}{n})\\ &= \Pr(\frac{X}{n^2} \geq \frac{1}{n}+\frac{1}{n})\\ &= \Pr(\frac{X}{n^2} \geq p+\frac{1}{n})\end{aligned}$$

We now know that $\epsilon = \frac{1}{n}$. Plugging into our Upper Tail Chernoff-Hoeffding formula, we get: $\Pr(X \geq 2n) \leq e^{-2(\frac{1}{n})^2 n^2} = e^{-2}$

(b) Use Chebyshev's inequality to bound the probability that bin i has at least $2n$ balls.

Solution: Let $X_i \in \{0,1\}$ be the indicator random variable for whether i'th ball lands in the first bin. Then $X = X_1 + \cdots + X_n$.
Since $\mathbf{E}(X_i) = \Pr(i\text{'th ball lands in first bin}) = 1/n$, and by linearity of expectation, $\mathbf{E}(X) = \sum_i \mathbf{E}(X_i) = n^2/n = n$. Similarly, since $\mathbf{Var}(X_i) = \mathbf{E}(X_i^2) - (\mathbf{E}(X_i))^2 = 1/n - 1/n^2 = (n-1)/n^2$, by the properties of variance for the sum of independent random variables, $\mathbf{Var}(X) = \sum_i \mathbf{Var}(X_i) = (n-1)n^2/n^2 = n-1$.

$$\begin{aligned}\Pr(X \geq 2\mathbf{E}(X)) &= \Pr(X \geq \mathbf{E}(X) + \mathbf{E}(X)) \\ &\leq \mathbf{Var}(X)/(\mathbf{E}(X))^2 \\ &= (n-1)/n^2\end{aligned}$$

3. Recall the Chernoff-Hoeffding (CH) bound: let $Y = \sum_{i=1}^n Y_i$ be the sum of independent, identically distributed random variables $Y_i \in [0,1]$. Then for any $\varepsilon > 0$, we have the upper-tail bound $\Pr[Y/n \geq \mathrm{E}[Y_i] + \varepsilon] \leq e^{-2\varepsilon^2 n}$, and similarly for the lower tail $Y/n \leq \mathrm{E}[Y_i] - \varepsilon$.

   Now let $X = \sum_{i=1}^n X_i$ for independent, identically distributed random variables $X_i \in [a,b]$ for some (known) $a < b$.[1] Generalize the CH bound to prove as tight of an upper bound as you can for the upper tail $\Pr[X/n \geq \mathrm{E}[X_i] + \varepsilon]$, for $\varepsilon > 0$.

   (Your argument should also apply symmetrically for the lower tail, though you do not need to repeat it.)

   Hint: Introduce suitable alternative random variables, and apply the ordinary CH bound.

Solution: Define the random variables $Y_i = (X_i - a)/(b-a) \in [0,1]$ and let

$$Y = \sum_{i=1}^n Y_i = \sum_{i=1}^n (X_i - a)/(b-a) = (X - na)/(b-a).$$

Then $\mathrm{E}[Y_i] = (\mathrm{E}[X_i] - a)/(b-a)$ and $\mathrm{E}[Y] = (\mathrm{E}[X] - na)/(b-a)$ by linearity of expectation. So, by substitution, basic algebraic manipulations, and the upper-tail Chernoff bound on Y, we have that

$$\begin{aligned}\Pr\left[\frac{X}{n} \geq \mathrm{E}[X_i] + \varepsilon\right] &= \Pr\left[\frac{(b-a)Y}{n} + a \geq (b-a)\mathrm{E}[Y_i] + a + \varepsilon\right] \\ &= \Pr\left[\frac{Y}{n} \geq \mathrm{E}[Y_i] + \frac{\varepsilon}{b-a}\right] \leq e^{-2(\varepsilon/(b-a))^2 n}.\end{aligned}$$

Observe that the ultimate bound is the same as in ordinary CH, except that ε has been "rescaled" to $\varepsilon/(b-a)$. This makes intuitive sense, since the variables X_i are in an interval of width $b-a$ rather than 1. (The corresponding lower-tail bound holds symmetrically, by the same substitution.)

[1] For example, if the X_i are indicator variables, then $a = 0$ and $b = 1$.

2 Union Bound

4. In a computer system equipped with 50 processors, each engaged in concurrent multithreading tasks, there exists a probability of 0.001 for an individual processor to experience failure. Determine the likelihood that at least one processor encounters failure.

Solution: Define the indicator random variable X_i, where i represents the number of the processor where:

$$X_i = \begin{cases} 1 & \text{if the } i\text{-th processor fails} \\ 0 & \text{otherwise} \end{cases}.$$

$$\begin{aligned} \Pr[X_1 \cup X_2 \cup \cdots \cup X_n] &\leq \sum_{i=1}^{n} \Pr[X_i] \\ &\leq \sum_{i=1}^{50} \Pr[X_i] \\ &= 50(0.001) \\ &= 0.05 \end{aligned}$$

5. (True/False) In order to apply the union bound formula, each event must be independent.

Solution: False, the union bound formula can be applied to dependent events. Independence is a property that simplifies probability calculations, but it is not a requirement for the Union Bound to hold. The Union Bound is a conservative bound that holds true regardless of the relationships between the events. If the events are independent, the Union Bound can be tighter than when they are dependent, but it is still a valid upper bound in either case.

6. A hash function maps data objects to the indices of an array where they are stored; we would like the function to distribute the objects roughly evenly among the indices. Suppose we need to hash a large number n of objects to k indices. Consider an "ideal" function f that maps each object to a uniformly random and independent index in $\{1, 2, \ldots, k\}$.

   (a) Let X_i be a random variable for the number of objects that f hashes to index i. Find, with proof, a closed-form expression for $\mathrm{E}[X_i]$.

   Solution: Let $X_{i,j}$ be the indicator random variable that is 1 precisely when the jth object is hashed to bucket i. Then, since f is an ideal hash function, $\mathrm{E}[X_{i,j}] = \frac{1}{k}$ for every i, j. Now, $X_i = \sum_{j=1}^{n} X_{i,j}$. So, by linearity of expectation, $\mathrm{E}[X_i] = \sum_{j=1}^{n} \mathrm{E}[X_{i,j}] = \sum_{j=1}^{n} \frac{1}{k} = \frac{n}{k}$.

   (b) Find, with proof, an upper bound on the probability that f maps at least $2 \cdot \mathrm{E}[X_i]$ different objects to index i:

i. Using Markov's inequality

Solution: By Markov's inequality (which is valid because X_i is a nonnegative random variable), $\Pr[X_i \geq 2\mathrm{E}[X_i]] \leq \frac{\mathrm{E}[X_i]}{2\mathrm{E}[X_i]} = \frac{1}{2}$.

ii. Using Chernoff bounds

Solution: Applying the Chernoff bound (which is valid because X_i is the sum of indicator variables $X_{i,j}$ for $j = 1, \ldots, n$, which are independent because each object is hash independently of all others), we have

$$\begin{aligned}\Pr[X_i \geq 2\mathrm{E}[X_i]] &= \Pr\left[X_i \geq \frac{2n}{k}\right] \\ &= \Pr\left[\frac{X_i}{n} \geq \frac{1}{k} + \frac{1}{k}\right] \\ &\leq e^{-2(1/k)^2 n} \\ &= e^{-2n/k^2}.\end{aligned}$$

iii. Under what condition on n and k does each method yield a tighter bound?

Solution: The Chernoff bound is better exactly when $e^{-2n/k^2} < 1/2$, or equivalently, when $n > \frac{1}{2} \cdot k^2 \ln 2$. This is because, for an upper bound, a lower value yields a tighter bound; for example, if the real value is 0.5 and we have two upper bounds that are 0.6 and 0.7, the bound of 0.6 is closer to the real value and so is tighter.

(c) Assuming that n, k are such that the Chernoff bound is the tighter of the two bounds above, give an upper bound on the probability that f hashes at least $2 \cdot \mathrm{E}[X_i]$ objects to index i for *some* index i.

Solution: Let A_i be the event that f hashes at least $2\mathrm{E}[X_i] = 2n/k$ objects to index i. From part (b) and the assumption that the Chernoff bound is tighter, we know that $\Pr[A_i] \leq e^{-2n/k^2}$. Now, let A be the union of all of the A_i. In words, A is the event that f hashes at least $2n/k$ objects to *some* index. Then, by the union bound,

$$\begin{aligned}\Pr[A] &= \Pr\left[\bigcup_{i=1}^{k} A_i\right] \\ &\leq \sum_{i=1}^{k} \Pr[A_i] \\ &\leq \sum_{i=1}^{k} e^{-2n/k^2} \\ &= ke^{-2n/k^2}.\end{aligned}$$

3 Fingerprinting

7. Recall the fingerprinting protocol discussed in lecture. For each value of x and y (i.e. the number that Alice and Bob have, respectively), state whether the algorithm is correct for all choices of p, or give *all* examples of p that cause the protocol to result in the incorrect decision

   (a) $x = 342, y = 342$

   (b) $x = 64, y = 70$

   (c) $x = 600, y = 719$

> **Solution:**
>
> (a) They are the same number, so $x \bmod p \equiv y \bmod p \ \forall p$
>
> (b) Incorrect for $p = 2$ and $p = 3$
>
> (c) Incorrect for $p = 7$ and $p = 17$

4 Cryptography Intro

8. Suppose an attacker knows that Alice's password is either `ABCD` or `ABEG` and has the encryption of Alice's password under the Caesar cipher. Can the attacker determine Alice's password?

> **Solution:** Yes. Caesar ciphers preserve distances between characters in the alphabet from the plaintext to the ciphertext. Specifically, since the second and third characters are either BC or BE, an attacker can determine which of the two passwords is correct by checking the distance between the second and third characters in the ciphertext. If the distance is 1, then the password is ABCD. If the distance is 3, then the password is ABEG.

9. Suppose an attacker knows that Alice's password is either `ABCD` or `ABEG` and has the encryption of Alice's password under a one-time pad. Can the attacker determine Alice's password?

> **Solution:** No. If a message is encrypted using a one-time pad, no insight can be gained about the plaintext from the ciphertext.

10. The value $g = 3$ is a generator of $\mathbb{Z}_{11}^*$.

    A. True

    B. False

    C. Unknown

Solution: False. Notice that since 11 is prime, the elements of $\mathbb{Z}_{11}^*$ are $\{1, 2, \dots 10\}$. Any generator of $\mathbb{Z}_{11}^*$, when raised (modulo p) to the powers 1 through 10, must produce all these elements. The mod-p powers of $g = 3$ are $g^1 = 3, g^2 = 9, g^3 = 5, g^4 = 4, g^5 = 1, g^6 = 3$. Since we have cycled back to $g^6 = g^1$ without generating all the elements in $\mathbb{Z}_{11}^*$, we conclude that $g = 3$ is not a generator for $\mathbb{Z}_{11}^*$.

11. Use fast modular exponentiation to calculate $3^{57} \mod 14$

Solution: First calculate the powers of 2 up to 32. $3^1 \mod 14 = 3$ $3^2 \mod 14 = 9$ $3^4 \mod 14 = 81 \mod 14 = 11$ $3^8 \mod 14 = 121 \mod 14 = 9$ $3^{16} \mod 14 = 11$ $3^{32} \mod 14 = 9$ Since $57 = 111001_2 = 32+16+8+1$ We have, $3^{57} \mod 14 = (3^1 * 3^8 * 3^{16} * 3^{32}) \mod 14 = (3 * 9 * 11 * 9) \mod 14 = 13$