Outline

- Solution of assignment 3.
- Hint of assignment 4.

Outline

- Solution of assignment 3.
- Hint of assignment 4.

Assignment 3 - problem 1

- We prove that if Z is a Poisson random variable of mean μ, where $\mu \geq 1$ is an integer, then $\Pr(Z \geq \mu) \geq 1/2$.

  1. Show that $\Pr(Z = \mu + h) \geq \Pr(Z = \mu - h - 1)$ for $0 \leq h \leq \mu - 1$.
     Ans. For any non-negative integer k, $\Pr(Z = k) = e^{-\mu}\mu^k/k!$ by definition. So,

$$
\begin{aligned}
\frac{\Pr(Z = \mu + h)}{\Pr(Z = \mu - h - 1)} &= \frac{\mu^{2h+1}}{(\mu - h)(\mu - (h-1))\cdots(\mu + h)} \\
&= \frac{\mu^2}{\mu^2 - h^2} \cdot \frac{\mu^2}{\mu^2 - (h-1)^2} \cdots \frac{\mu^2}{\mu^2 - 1^2} \cdot \frac{\mu}{\mu} \\
&\geq 1.
\end{aligned}
$$

In other words, $\Pr(Z = \mu + h) \geq \Pr(Z = \mu - h - 1)$.

Assignment 3 - problem 1 (con't)

- Using part (a), argue that $\Pr(Z \geq \mu) \geq 1/2$.
  Ans.

  $$1 = \sum_{k=1}^{\infty} \Pr(Z = k) = \sum_{k=0}^{\mu-1} \Pr(Z = k) + \Pr(Z \geq \mu).$$

  Then by part (a), we know that

  $$\sum_{k=0}^{\mu-1} \Pr(Z = k) \leq \sum_{k=\mu}^{2\mu-1} \Pr(Z = k) \leq \sum_{k=\mu}^{\infty} \Pr(Z = k) = \Pr(Z \geq \mu).$$

  Combining, we have

  $$1 = \sum_{k=0}^{\infty} \Pr(Z = k) = \sum_{k=0}^{\mu-1} \Pr(Z = k) + \Pr(Z \geq \mu) \leq 2 \Pr(Z \geq \mu).$$

  Thus, $\Pr(Z \geq \mu) \geq 1/2$.

Assignment 3 - problem 2

- In Page 15 of Lecture Notes 14, we showed that for any nonnegatvie function f,

$$E\left[f(Y_1^{(m)},\cdots,Y_n^{(m)})\right] \geq E\left[f(X_1^{(m)},\cdots,X_n^{(m)})\right]\Pr\left(\sum Y_i^{(m)} = m\right)$$

1. Now, suppose we further know that $E\left[f(X_1^{(m)},\cdots,X_n^{(m)})\right]$ is monotonically increasing in m. Show that

$$E\left[f(Y_1^{(m)},\cdots,Y_n^{(m)})\right] \geq E\left[f(X_1^{(m)},\cdots,X_n^{(m)})\right]\Pr\left(\sum Y_i^{(m)} \geq m\right).$$

- **Proof.**

$$
\begin{aligned}
& E\left[f(Y_1^{(m)},\cdots,Y_n^{(m)})\right] \\
&= \sum_{k\geq 0} E\left[f(Y_1^{(m)},\cdots,Y_n^{(m)}) \mid \sum Y_i^{(m)} = k\right] \Pr\left(\sum Y_i^{(m)} = k\right) \\
&\geq \sum_{k\geq m} E\left[f(Y_1^{(m)},\cdots,Y_n^{(m)}) \mid \sum Y_i^{(m)} = k\right] \Pr\left(\sum Y_i^{(m)} = k\right) \\
&= \sum_{k\geq m} E\left[f(X_1^{(k)},\cdots,X_n^{(k)})\right] \Pr\left(\sum Y_i^{(m)} = k\right) \\
&\geq \sum_{k\geq m} E\left[f(X_1^{(m)},\cdots,X_n^{(m)})\right] \Pr\left(\sum Y_i^{(m)} = k\right)
\end{aligned}
$$

monotonically increasing.

$$
= E\left[f(X_1^{(m)},\cdots,X_n^{(m)})\right] \Pr\left(\sum Y_i^{(m)} \geq m\right),
$$

Assignment 3 - problem 2 (con't)

- Combining part (a) with the results in Question 1, prove the monotonically increasing case of theorem in Page 20 of Lecture Notes 14.

  Proof. Let $Z = \sum Y_i^{(m)}$. Since each $Y_i^{(m)}$ is a Poisson random variable, their sum Z is also a Poisson random variable. Further, the mean value of Z is m. Thus, by results in Question 1, $\Pr(Z \geq m) \geq 1/2$. Combining this with part (a), we have

$$E\left[f(X_1^{(m)}, \cdots, X_n^{(m)})\right] \leq 2\, E\left[f(Y_1^{(m)}, \cdots, Y_n^{(m)})\right]$$

  when $E\left[f(X_1^{(m)}, \cdots, X_n^{(m)})\right]$ is monotonically increasing in m. This completes the proof of the monotonically increasing case of the desired theorem.

Assignment 3 - problem 3

- Bloom filters can be used to estimate set differences. Suppose you have a set X and I have a set Y, both with n elements. For example, the sets might represent our 100 favorite songs. We both create Bloom filters of our sets, using the same number of bits m and the *same* k hash functions. Determine the expected number of bits where our Bloom filters differ as a function of m, n, k, and $|X \cap Y|$.

Assignment 3 - problem 3 (con't)

- **Ans.** Let Z be a random variable denoting the number of bits where the Bloom filters differ. Let Z_i be an indicator such that

  $Z_i = 1$ if the ith bit of the Bloom filters differ

  $Z_i = 0$ otherwise.

  Thus, $Z = Z_1 + Z_2 + \cdots + Z_m$

Assignment 3 - problem 3 (con't)

- When $|X \cap Y| = r$, $Z_i = 1$ only happens when each of the r common elements are not mapped to the ith bit, together with exactly one of the following cases (that causes the ith bit different):
    1. Some elements of $X - (X \cap Y)$ is mapped to the ith bit, but all elements of $Y - (X \cap Y)$ are not;
    2. Some elements of $Y - (X \cap Y)$ is mapped to the ith bit, but all elements of $X - (X \cap Y)$ are not.

Assignment 3 - problem 3 (con't)

- Let Q_i denote the event that the r common elements are not mapped to the ith bit. By assuming that the hash functions we choose will map elements independently and uniformly at random to one of the m bits, we have

$$\begin{aligned}\Pr(Z_i = 1) &= \Pr(Q_i \cap (\text{Case (a) or Case (b)})) \\ &= \Pr(Q_i)\,(\Pr(\text{Case (a)}) + \Pr(\text{Case (b)})) \\ &= \left(1 - \frac{1}{m}\right)^{rk} (\Pr(\text{Case (a)}) + \Pr(\text{Case (b)}))\end{aligned}$$

Assignment 3 - problem 3 (con't)

- $$\begin{aligned} &= \left(1-\frac{1}{m}\right)^{rk} \times 2 \times \underbrace{\left(1-\left(1-\frac{1}{m}\right)^{(n-r)k}\right)}_{\text{some elements mapped to } i\text{th bit}} \\ &\quad \times \underbrace{\left(\left(1-\frac{1}{m}\right)^{(n-r)k}\right)}_{\text{no elements mapped to } i\text{th bit}} \\ &= 2\left(1-\frac{1}{m}\right)^{nk}\left(1-\left(1-\frac{1}{m}\right)^{(n-r)k}\right). \end{aligned}$$

Assignment 3 - problem 3 (con't)

- As Z_i is an indicator, $E[Z_i] = \Pr(Z_i = 1)$. Thus,

$$E[Z] = \sum_{i=0}^{m-1} E[Z_i] = m * E[Z_i]$$

$$= 2m\left(1 - \frac{1}{m}\right)^{nk}\left(1 - \left(1 - \frac{1}{m}\right)^{(n-r)k}\right).$$

Assignment 3 - problem 4

- For the leader election problem briefly introduced in Lecture Notes 15, we have n users, each with an identifier. Suppose that we have a *good* hash function (that looks uniform and independent), which outputs a b-bit hash value for each identifier. One way to solve the leader election problem is as follows: Each user obtains the hash value from its identifier, and the leader is the user with the smallest hash value. Give lower and upper bounds on the number of bits b necessary to ensure that a unique leader is successfully chosen with probability p. Make your bounds as tight as possible.

Assignment 3 - problem 4 (con't)

- **Ans.** (Lower Bound:) First, a unique leader is determined if and only if for some identifier i, it is mapped exactly by one user, while all other users are mapped to the identifiers larger than i. This implies that:

$$\begin{aligned} p &= \sum_{i=0}^{2^b-2} \Pr(\text{“identifier } i \text{ is mapped exactly by one user”} \\ &\qquad \cap \text{“all other users are larger than } i\text{”}) \\ &= \sum_{i=0}^{2^b-2} \frac{1}{2^b} \cdot \left(1 - \frac{i+1}{2^b}\right)^{n-1} \qquad (1) \\ &\leq \sum_{i=0}^{2^b-2} \frac{1}{2^b} \cdot \left(1 - \frac{1}{2^b}\right)^{n-1} \\ &\leq \left(1 - \frac{1}{2^b}\right)^{n-1}. \end{aligned}$$

Assignment 3 - problem 4 (con't)

- By re-arranging terms, we have:

$$b \geq \log_2 \left(\frac{1}{1 - p^{\frac{1}{n-1}}} \right).$$

Assignment 3 - problem 4 (con't)

- (Upper Bound:) Next, p is greater than the probability that user 1 is the unique leader. This happens when user 1 is mapped to uniquely to some number, and this number is smallest among all numbers mapped by other users. Thus,

$$\begin{aligned} p &\geq \Pr(\text{“user 1 is mapped uniquely”} \cap \text{“this number is smallest”}) \\ &= \Pr(\text{“this number is smallest”} \mid \text{“user 1 is mapped uniquely”}) \\ &\quad \times \Pr(\text{“user 1 is mapped uniquely”}) \\ &= \Pr(\text{“this number is smallest”} \mid \text{“user 1 is mapped uniquely”}) \\ &\geq \frac{1}{n}\left(1 - \frac{1}{2^b}\right)^{n-1}. \end{aligned}$$

By re-arranging terms, we have:

$$b \leq \log_2\left(\frac{1}{1-(np)^{\frac{1}{n-1}}}\right).$$

Assignment 3 - problem 4 (con't)

- In conclusion,

$$\log_2\left(\frac{1}{1-p^{\frac{1}{n-1}}}\right) \leq b \leq \log_2\left(\frac{1}{1-(np)^{\frac{1}{n-1}}}\right).$$

Assignment 4 - problem 1

- Consider an instance of SAT with m clauses, where every clause has exactly k literals.
  1. Give a Las Vegas algorithm that finds an assignment satisfying at least $m(1 - 2^{-k})$ clauses, and analyze its expected running time.
  2. Give a derandomization of the randomized algorithm using the method of conditional expectations.

Assignment 4 - problem 2

- 1. Prove that, for every integer n, there exists a coloring of the edges of the complete graph K_n by two colors so that the total number of monochromatic copies of K_4 is at most $\binom{n}{4}2^{-5}$.
  2. Give a randomized algorithm for finding a coloring with at most $\binom{n}{4}2^{-5}$ monochromatic copies of K_4 that runs in expected time polynomial in n.
  3. Show how to construct such a coloring deterministically in polynomial time using the method of conditional expectations.

Assignment 4 - problem 2 (con't)

- hint:
  1. The expectation of the total number of monochromatic copies of K_4.
  2. Las Vegas algorithm
  3. You can compare the expectation numbers of monochromatic copies of K_4 when a edge is assigned to different color.

Assignment 4 - problem 3

- Given an n-vertex undirected graph $G = (V, E)$, consider the following method of generating an independent set. Given a permutation σ of the vertices, define a subset $S(\sigma)$ of the vertices as follows: for each vertex i, $i \in S(\sigma)$ if and only if no neighbor j of i precedes i in the permutation σ.
  1. Show that each $S(\sigma)$ is an independent set in G.
  2. Suggest a natural randomized algorithm to produce σ for which you can show that the expected cardinality of $S(\sigma)$ is

     $$\sum_{i=1}^{n} \frac{1}{d_i + 1}$$

     where d_i denotes the degree of vertex i.
  3. Prove that G has an independent set of size at least $\sum_{i=1}^{n} 1/(d_i + 1)$.

Assignment 4 - problem 3 (con't)

- Choose a random permutation σ from S_n, the set of all $n!$ permutations of V.
- A vertex i is good if $\sigma(i) < \sigma(j)$ for every j adjacent to i.
- Let $S(\sigma)$ be the collection of good vertices and output $S(\sigma)$.

Assignment 4 - problem 3 (con't)

- hint:
  1. You can think about the definition of "good vertices", and proof it by contradiction.
  2. The expected cardinality of $S(\sigma)$ is equal to the expected number of good vertices in G.

Assignment 4 - problem 4

- We have shown using the probabilistic method that, if a graph G has n nodes and m edges, then there exists a partition of the n nodes into sets A and B such that at least $m/2$ edges cross the partition. Improve this result slightly: show that there exists a partition such that at least $mn/(2n-1)$ edges cross the partition.

Assignment 4 - problem 4 (con't)

- hint:
  1. You can fix a point and find the probability of the crossing edge.
  2. You can discuss from two cases.
     (a) n is odd.
     (b) n is even.

Assignment 4 - problem 5

- Use the general form of the Lovasz local lemma to prove that the symmetric version of Theorem 6.11 can be improved by replacing the condition $4dp \leq 1$ by the weaker condition $ep(d+1) \leq 1$.

Assignment 4 - problem 5 (con't)

- hint:
  1. Set $x_i = 1/(d+1)$ to the general case Lovasz local lemma.
  2. You have already had $\Pr(E_i) \leq p$ from the symmetric version of Lovasz local lemma and you want to prove $\Pr(E_i) \leq x_i \prod_{(i,j)\in E}(1 = x_j)$.