CS203 Endsem

April 29, 2022

Instructions:

- Please do not use any immoral means. Please write clearly and cut out any rough work.
- You can use
  - $\sum_{i=1}^{n} \frac{1}{i} \approx \ln n; \quad \sum_{i=1}^{\infty} \frac{1}{i^2} = \frac{\pi^2}{6}$
  - $\sum_{i=0}^{\infty} \binom{i+2}{2} q^i = (1-q)^{-3}$

1. (15+15 points) Assume none of the probabilities considered can be zero. Define two events A, B to be *conditionally independent* iff $P[A \cap B|C] = P[A|C]P[B|C]$.
   1. Show that the definition is equivalent to $P[A|C] = P[A|B \cap C]$.
   2. Construct an example to show that independence of two events DOES NOT imply conditional independence. Hint: you can consider the sample space of two coin tosses.

Solution:

1. $P[A \cap B|C] = \frac{P[A \cap B \cap C]}{P[C]} = \frac{P[C]P[B|C]P[A|B \cap C]}{P[C]}$.
   Simplifying $P[A \cap B|C] = P[B|C]P[A|B \cap C]$.
   Putting $P[A|C]P[B|C]$ on LHS will give the result.
2. Let the two events be first coin is head and second coin is head. They are not conditionally independent under condition that the two coin tosses are same.

2. (30 points) A scientific hypothesis predicts that an experiment (say E_1) will come out to be positive with probability .8 and a different experiment (say E_2) will be positive with probability .7. Hypothesis also predicts that at least one of the above experiments will definitely turn out to be true.

   The unconditional probability that both experiments are true is .4. Suppose both experiments turn out to be positive, if your prior belief (probability of being true) in the scientific hypothesis was .5, what is the new (posterior) probability of the scientific hypothesis being true?

Solution: $P[H|E_1 \cap E_2] = \frac{P[E_1 \cap E_2|H]P[H]}{P[E_1 \cap E_2]}$. We know $P[E_1 \cap E_2|H] = .8 + .7 - 1 = .5$ So $P[H|E_1 \cap E_2] = .5 \times .5/.4 = .625$

3. (20+20+10 points) Let M be a symmetric transition matrix for an Ergodic homogeneous Markov chain. Define $T := pI + (1-p)M$ for a $p \in (0,1)$.

1. Show that T is a symmetric transition matrix for a Markov chain.
2. Show that T is a regular Markov chain.
3. Show that the stationary distribution of T is the eigenvalue 1 eigenvector of M.

Solution:

1. All entries are positive and column sum is 1.
2. All entries of T^{n+1} (n is the number of states) will be positive.
3. Eigenvalue 1 eigenvector of M will be an eigenvalue 1 eigenvector of T.

4. (15+15+20 points) Your friend is trying to invest in NFTs. In particular, he is investing in a scheme where you get a random NFT from a set of n NFTs for opening a Rs. 50 pack. The probability of getting a particular NFT in a pack is $\frac{1}{n}$. If you get a new NFT that wasn't in your collection already, then it is added to your collection, otherwise that pack is a waste. The company has an offer where they give Rs. 100 for each NFT if you have the complete set, and hence "double" the money. You have decided to show your friend that this is a losing deal and save him from the scam:

   1. Calculate a close approximation for the expected amount of money he has to spend to collect all the n NFTs.
   2. Show that the variance of a geometric random variable with probability p is $\frac{1-p}{p^2}$. (The expectation of a geometric random variable is $1/p$.)
   3. (Difficult) Your Friend is still stubborn and thinks he is lucky enough to get a profit out of this. Show that the probability to make profit out of the scheme goes to 0 as n tends to ∞.

Solution:

1. money = 50* number of draws. X=no. of draws= $\sum_{i=0}^{n-1} X_i$, X_i no. of draws to go from i elements in collection to $i+1$. each X_i is a geometric random variable with probability $\frac{n-i}{n}$. $E[X] = nH_n \approx n \ln n$.
2. Compute $E[X(X-1)]$. It is $2(1-p)/p^2$. Find variance using $Var[X] = E[X^2] - E[X]^2$.
3. As X is sum of independent grv's (not Bernoulli), $Var[X] = \sum_i Var[X_i]$ where $Var[X_i] = \frac{i/n}{((n-i)/n)^2}$. So $Var[X] = n^2\pi^2/6 - n \ln n \approx n^2\pi^2/12$

   $Pr[50X < 100n] = Pr[X < 2n] = Pr[n \ln n - X > n(\ln n - 2)] = Pr[E[X] - X > n(\ln n - 2)]$.

   Applying Chebyshev, we get a bound of $1/\ln n^2$.

5. (40 points) (Difficult) Let n be an even integer, define

$$S_{n,k} := \sum_{i=\frac{n}{2}-k}^{\frac{n}{2}+k} \binom{n}{i}.$$

For example, $S_{n,\frac{n}{2}} = 2^n$. Show that there exist a constant c (not dependent on n) such that

$$\frac{S_{n,c\sqrt{n}}}{2^n} \geq \frac{1}{2}.$$

Hint: Can you relate this ratio to probability of an event?

Solution: Let X be a Bernoulli random variable with success probability $1/2$ and X_i's be i.i.d's of X. Notice that $\frac{S_{n,k}}{2^n}$ is the probability that $S = \sum_i X_i$ is between $n/2 - k$ and $n/2 + k$.
By Chernoff bound,

$$P(|S - n/2| \geq (n/2)(2k/n)) \quad \leq \quad 2e^{-(n/2)(2k/n)^2(1/3)}.$$

To make the exponent constant, we need $k = c\sqrt{n}$.
Same can be done using Chebyshev.