MATHEMATICAL TOOLS IN CS - PROBLEM SET 1

Due Sunday, November 20th, 20:00.

Random Variables. In the following, all random variables are assumed to have finite mean and variance. All graphs are finite. $\mathbb{P}$ denotes probability and $\mathbb{E}$ denotes expectation.

As a public service, here are Markov and Chebychev's respective inequalities:

Proposition. *Let X be a non-negative random variable with $\mathbb{E}[X] > 0$, and let $c > 0$. Then $\mathbb{P}[X \geq c\mathbb{E}[X]] \leq \frac{1}{c}$.*

Proposition. *Let X be a random variable with $Var[X] > 0$, and let $c > 0$. Then $\mathbb{P}\left[|X - \mathbb{E}[X]| \geq c\sqrt{Var[X]}\right] \leq \frac{1}{c^2}$.*

Problem 1.

(1) Let X, Y be independent random variables. Show that $\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y]$.

(2) Show that the converse need not hold: There exist random variables X, Y s.t. $\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y]$ but X and Y aren't independent.

(3) Let $X_1, \ldots, X_n$ be pairwise independent random variables (that is, for $1 \leq i < j \leq n$, X_i and X_j are independent). Prove that $Var\left[\sum_{i=1}^n X_i\right] = \sum_{i=1}^n Var[X_i]$.

(4) Give an example of random variables X, Y, Z that are pairwise independent but not independent.

Reminder: Random variables $X_1, X_2, \ldots, X_n$ are called *independent* if for all $x_1, x_2, \ldots, x_n$:

$$\mathbb{P}[X_1 = x_1, X_2 = x_2, \ldots, X_3 = x_3] = \mathbb{P}[X_1 = x_1] \cdot \mathbb{P}[X_2 = x_2] \cdot \ldots \cdot \mathbb{P}[X_n = x_n]$$

.

Problem 2.

(1) Show that the Markov and Chebychev bounds are optimal. To do this, define random variables X and Y s.t. there exist $a, b > 0$ s.t:

$$\mathbb{P}[X \geq a\mathbb{E}[X]] = \frac{1}{a}$$

$$\mathbb{P}\left[|Y - \mathbb{E}[Y]| \geq b\sqrt{Var[Y]}\right] = \frac{1}{b^2}$$

(2) Show that the Markov and Chebychev bounds are not necessarily tight: Define random variables X and Y s.t. there exist $a \geq 1, b > 0$ s.t:

$$\mathbb{P}[X \geq a\mathbb{E}[X]] < \frac{1}{a}$$

$$\mathbb{P}\left[|Y - \mathbb{E}[Y]| \geq b\sqrt{Var[Y]}\right] < \frac{1}{b^2}$$

Note that here you're asked to find $a \geq 1$. Would your have changed your answer if the requirement was just $a > 0$?

Problem 3.

(1) Let $p \in (0,1)$ and let $X_n \sim Bin\left(n,p\right)$. Let $0 \leq \alpha < p < \beta \leq 1$. Prove:
$$\lim_{n\to\infty} \mathbb{P}\left[X_n \leq \alpha n\right] = \lim_{n\to\infty} \mathbb{P}\left[X_n \geq \beta n\right] = 0$$
(2) Set $n = 100, p = \frac{1}{2}, \alpha = \frac{1}{4}$. Bound $\mathbb{P}\left[X_{100} \leq \frac{1}{4} \cdot 100\right]$ in a way analogous to what you did in part 1.
(3) Calculate $\mathbb{P}\left[X_{100} \leq \frac{1}{4} \cdot 100\right]$ to two decimal digits of precision (for this you might want to use a computer). How close is the result to what you got in part 2? If the results aren't close, why do you think this is the case?
(4) Fix $p \in [0,1]$ and $\varepsilon > 0$. For every n, let A_n be the random variable denoting the number of edges in the $G\left(n,p\right)$ model of random graphs. What is $\mathbb{E}\left[A_n\right]$? Show that:
$$\lim_{n\to\infty} \mathbb{P}\left[\left(1-\varepsilon\right)\mathbb{E}\left[A_n\right] \leq A_n \leq \left(1+\varepsilon\right)\mathbb{E}\left[A_n\right]\right] = 1$$

Problem 4. A *triangle* in a graph $G = (V,E)$ is a set $\{a,b,c\} \subseteq V$ of cardinality 3 s.t. $ab, ac, bc \in E$. For a graph G, let $T\left(G\right)$ be the number of triangles in G. A graph is called *triangle-free* if $T\left(G\right) = 0$. If G_n is sampled from $G\left(n, \frac{1}{n}\right)$, then $T\left(G_n\right)$ is a sequence of random variables. Let A_n be the number of edges in G_n.

(1) Show that $\mathbb{P}\left[T\left(G_n\right) = 0\right] > \frac{5}{6}$.
(2) Show that $\mathbb{P}\left[A_n \leq \frac{n}{2} - \sqrt{n}\right] \leq \frac{5}{6}$.
(3) Conclude that there exist triangle-free graphs with n vertices, and at least $\frac{n}{2} - \sqrt{n}$ edges.
(4) Can you give an example of a triangle-free graph with at least $\frac{n^2}{4} - 1$ edges?