CPSC 351 — Tutorial Exercise #10

Discrete Probability for Computer Science — Part Two

These questions are intended to give you practice in using discrete probability theory to study problems in computer science.

Problems To Be Solved

1. Consider, again, a ***random walk*** along a line, as in the previous tutorial exercise. Suppose the ***sample space*** Ω_n, the ***random variables*** LS_n, RS_n, Pos_n, Dis_n, and SqDis_n are all as defined in that exercise. Furthermore, let us consider the ***uniform probability distribution*** $P_n : \Omega_n \to \mathbb{R}$, as in (most of) the previous tutorial exercise, as well.

   Once again, let n be a positive integer and consider where (and how far away from the origin) a (rambling) walker is, after taking n steps, as described in the previous exercise.

   (a) Recall that the expected value of the ***square*** of the distance of the walker from the origin, after taking n steps, is n.

   Let a be a real number such that $a > 1$. Use this, along with an applicable technique from Lecture #11, to state and prove the best bound that you can for the probability that $\text{Dis}_n \geq a\sqrt{n}$.

   Hint: How large must the value of SqDis_n be, when $\text{Dis}_n \geq a\sqrt{n}$?

   Your choices of techniques to use, here, might be somewhat limited, because you do not know the ***variance*** of any of the (relevant) random variables that are being considered. With that noted let $\text{FthDis}_n : \Omega_n \to \mathbb{R}$ be the *fourth* power of the distance of the walker from the origin after n steps — so that

$$\text{FthDis}_n(\vec{\alpha}) = \text{SqDis}_n(\vec{\alpha})^2 = \text{Dis}_n(\vec{\alpha})^4$$

   for every outcome $\vec{\alpha} \in \Omega_n$, and let $\textit{fthdist}_n$ be the expected value of the random variable FthDis_n with respect to the uniform probability distribution, P_n — so $\textit{fthdist}_n = 1$.

It is possible to modify the solution for Question #4 on the previous tutorial exercise to prove that

$$\textit{fthdist}_{n+1} = \textit{fthdist}_n + 6n + 1$$

for every positive integer n — and this can be used to prove that

$$\textit{fthdist}_n = 3n^2 - 2n$$

for every positive integer n. You may use this fact, without proving it, when solving the rest of this problem.

(b) Compute the ***variance*** of the random variable SqDis_n with respect to the uniform probability distribution.

(c) Use this, along with techniques from Lecture #11 that can now be applied, to state and prove the best bound that you can for the probability that $\mathsf{Dis}_n \geq \alpha\sqrt{n}$.

(d) Identify the random variables, that have been introduced for this problem, for which the ***Chernoff Bound*** can be used to establish "tail bounds" for these random variables. Can these be used (probably, indirectly, as part of a longer proof) to establish additional bounds for the probability that $\mathsf{Dis}_n \geq \sqrt{n}$? If this is true then are these bounds useful?

2. The preparatory material for Lecture #11 the following result.

Theorem 8 (Cantelli's Inequality)**.** *Let Ω be a* **finite** *sample space with probability distribution P : $\Omega \to \mathbb{R}$, let $X : \Omega \to \mathbb{R}$ be a random variable, and let $a \in \mathbb{R}$ such that $a > 0$. Then*

$$\mathsf{P}(X - \mathsf{E}[X] \geq a) \leq \frac{\mathsf{var}(X)}{a^2 + \mathsf{var}(X)}.$$

Prove this result.

Hint: Consider the random variable $Y = E - \mathsf{E}[X]$ and the other results, implying "tail bounds", that were included in the preparatory material for Lecture #11.