COS 341: Discrete Mathematics

Precept #6.5 — Fall 2006

For the week of: November 20

---

1. We define conditional expectation $\mathrm{E}[X|A]$, of a random variable X, given A, to be:

$$\mathrm{E}[X|A] = \sum_{x \in \mathrm{range}(X)} x \cdot \Pr[X = x|A].$$

Prove that if the sample space is a disjoint union of events $A_1, \ldots, A_n$ then

$$\mathrm{E}[X] = \sum_{i=1}^{n} \mathrm{E}[X|A_i] \cdot \Pr[A_i].$$

For example, if D is a random variable representing the outcome of the throw of an ordinary die, then $\mathrm{E}[D|D \geq 3] = (3+4+5+6)/4 = 18/4$. And the claim that you need to prove implies, for instance, that

$$\mathrm{E}[D] = \mathrm{E}[D|D < 3] \cdot \Pr[D < 3] + \mathrm{E}[D|D \geq 3] \cdot \Pr[D \geq 3].$$

2.

a. Suppose that A and B are independent events, and let C be some other event. Must it be true that A and B are conditionally independent given C? (See homework #6 for the definition of conditional independence.)

b. Suppose that $\mathrm{E}[XY] = \mathrm{E}[X] \cdot \mathrm{E}[Y]$ for some random variables X and Y. Must it be true that X and Y are independent?

c. Let X and Y be independent random variables, and let A be some event (not the same as in the previous parts). Must it be true that

$$\mathrm{E}[XY|A] = \mathrm{E}[X|A] \cdot \mathrm{E}[Y|A]?$$

If not, under what conditions does this become true?

3. A web server wants to keep track of the number of distinct web users that access it in a day. However, this is a very popular web site, and has millions of users every day. Fortunately, the web administrator has successfully completed COS 341 at Princeton. In order to save precious memory, he devised the following scheme for approximating the daily hits number:

First, suppose he knows in advance the number of distinct users will be no more then N. He then chooses π to be a random function $\pi : \{1, ..., N\} \to \{1, ..., N\}$. When there is a web access, the system determines the id of the web user, say t, computes $\pi(t)$, and keeps track of the *minimum* such $\pi(t)$ so far.

Therefore, the space usage is reduced significantly to just one number! In this question, we will see that the web admin's scheme indeed approximates the hit count well.

a. Suppose that the number of distinct people that actually surfed this web site on a given day is k. Show that the expectation of the minimum is approximately $\frac{N}{k+1}$.

b. How would you use this fact to find the number of distinct users?