CS 70 SPRING 2007 — DISCUSSION #11

ALEX FABRIKANT

1. Administrivia

(1) Midterm 2 will be on **this Tuesday 04/10, in-class**
(2) Midterm 2 review session will be held **4pm-6pm on Mon 04/09**
(3) Review problems for the midterm have been posted on the website. It's best to go through them before coming to the review session.
(4) Homework 9 will be posted on Tue 04/09 and will be due a week from then.

2. Warm-up

Exercise 1. What is the variance and expectation of the constant r.v. $X = C$?

Exercise 2. What is the variance of an indicator r.v. for an event with probability p?

3. Exceeding royal expectations

Recall problem 5 from the last homework:

Exercise 3 (HW8, problem 5)**.** We say that a deck has k *king adjacencies* if k of the kings are followed by another king. That is, if, in a particular deck arrangement, all the kings are next to each other, there are 3 king adjacencies. What is the expected number of king adjacencies in a standard 52-card deck, shuffled into a permutation uniformly at random?

Solution: A king adjacency can happen at any of 51 positions (1st card and 2nd card, 2nd card and 3rd card, ... , 51st card and 52nd card). Thus, we can set up our linearity-of-expectations trick by thinking of the random variable A, the number of king adjacencies, as a sum $\sum_{i=1}^{51} A_i$ where A_i is an indicator variable for the event of the i'th card and the $i+1$'th card forming a king adjacency.

There are a total of $\binom{52}{4}$ ways to place 4 kings in the deck (for the purposes of this problem, we don't distinguish between kings, or between non-king cards, so we can think of a deck as a sequence like $NNNNKNNKKN...N$ with N for non-kings and K for kings, and a total of 4 K's). For any given A_i to be 1, there must be two kings at i and $i+1$, and there are $\binom{50}{2}$ ways to place the other two kings in the remaining deck slots. Thus $E[A_i] = \Pr[A_i = 1] = \binom{50}{2}/\binom{52}{4}$. From linearity of expectation, we have $E[A] = \sum_{i=1}^{51} E[A_i] = 51 \cdot \binom{50}{2}/\binom{52}{4} \approx 0.2308$. $\square$

Exercise 4. What does Markov's inequality state about the probability that there are at least 2 king adjacencies, $\Pr[A \geq 2]$?

Exercise 5. What does Markov's inequality state about the probability that there are at least 4 king adjacencies, $\Pr[A \geq 4]$? What is the actual value of this probability?

Date: April 6, 2007.
The author gratefully acknowledges the TA's of CS70 Past for the use of their previous notes: Amir Kamil, Chris Crutchfield, David Garmire, Lorenzo Orecchia, and Ben Rubinstein. Their notes form the basis for this handout.

4. Midterm 1 statistics

As you may recall[1], the mean on midterm 1 was 43, and the standard deviation was (approximately) 7. Pick a CS70 uniformly at random, and let S be their MT1 score?

Exercise 6. What is Ω, the probability space over which random variable S is defined?

Exercise 7. What does Markov's inequality state about $\Pr[S \geq 53.5]$?

Exercise 8. What does Chebyshev's inequality state about $\Pr[\{S \geq 53.5\} \cup \{S \leq 32.5\}]$?

5. Life Insurance

As an extended example of probability, we analyze a simple life insurance system. A real system would be too cumbersome to look at, so we make many simplifications here.

Here are the basic rules for our system:

1. You pay b dollars to the insurance company when you are born. You never have to pay again.
2. If you die before age c, the company pays your beneficiaries d dollars.
3. The insurance company is non-profit, so just wants to break even.

Given these rules, what should the insurance company set as the values of b and d, in terms of c? Let X be the age at which a person dies. The fraction of its customers the insurance pays is then the fraction of those that die before age c, or $\Pr[X < c]$. Then b and d are related by $b = d \cdot \Pr[X < c]$.

Lets do a detailed example, where $c = 60$ and $d = \$1,000,000$. We need to compute $\Pr[X < 60]$.

5.1. Distribution of Death.

Before we can calculate $\Pr[X < 60]$, we need to know what the distribution of X looks like. First, lets assume that nobody lives past 100. Now we cant just take the distribution to be uniform in the range $\{1, \ldots, 100\}$, since a person is more likely to die as they get older. So let's assume a linear distribution of the form $\Pr[X = k] = \frac{k}{N}$ for $k \in \{1, \ldots, 100\}$.

Exercise 9. Calculate the constant N in order to ensure the probabilities sum to 1.

5.2. Life Expectancy.

The first thing we should calculate is the expected age at which a person dies.

Exercise 10. Calculate $\mathbb{E}[X]$. Use the identity $\sum_{i=1}^{n} i^2 = \frac{n(n+1)(2n+1)}{6}$.

Knowing just the expectation is not enough to calculate $\Pr[X < 60]$. Consider the two distributions (A) where $\Pr[X = 67] = 1$ and (B) where $\Pr[X = 55] = \Pr[X = 79] = 0.5$. In (A), $\Pr[X < 60] = 0$, whereas in (B) $\Pr[X < 60] = 0.5$. Notice that in both cases $\mathbb{E}[X] = 67$.

The variance is what makes the difference in the above distributions. It is variance that makes insurance useful. If there were no variance, everyone would know when they would die and thus no one would need or provide life insurance.

5.3. Variance and Chebyshevs Inequality.

We proceed by calculating the variance of the age at which a person dies.

Exercise 11. Calculate $\mathrm{Var}[X]$, using the identity $\sum_{i=1}^{n} i^3 = \left(\sum_{i=1}^{n} i\right)^2$.

Now recall Chebyshevs inequality

$$\Pr[|X - \mathbb{E}[X]| > r] \quad \leq \quad \frac{\mathrm{Var}[X]}{r^2} \ .$$

Exercise 12. Now use Chebyshev's inequality to upper-bound $\Pr[X < 60]$. What is the problem with this bound?

Exercise 13. When does Chebyshev's give a bound less than 1?

Even in general, Chebyshevs still gives us a weak bound. Its usefulness is due to the fact that it is easy to compute and only requires knowledge of the expectation and variance of a random variable.

[1] If you actually remember these numbers, you're worrying about grades way too much. You shouldn't. Really.

5.4. **Exact Solution.** In this case, since the distribution is so simple, we can compute $\Pr[X < 60]$ directly.

Exercise 14. Compute this probability directly, and thus determine the appropriate b that the insurance company should charge clients at birth.

6. Challenge problems

Exercise 15. Suppose you have n random variables X_i which are mutually independent and have the same *distribution* (i.e. $\Pr[X_i = a] = \Pr[X_j = a]$ for all i, j, a). Let $X = \sum_i X_i$. Use Markov's inequality on the random variable e^X to obtain a bound on $\Pr[X - E[X]] > k\sqrt{\mathrm{Var}[X]}$ which is exponential in k. This will be solved in class next Thursday (or perhaps the following Tuesday), so do it before then. You can check your work by googling for "Chernoff bound" (oh, those Russians...).

Exercise 16. Under what circumstances can Chebyshev's inequality give you a "tight bound" (i.e. when does it exactly predict what $\Pr[|X - \mu| > \alpha]$ is)?