Massachusetts Institute of Technology
6.042J/18.062J: Mathematics for Computer Science
Professors Nancy Lynch and Srinivas Devadas

Handout 58
November 24, 1998

Problem Set 11

Due: Start of class on December 4, 1998.

Read sections 4.4 and 4.5 in Rosen.

Problem 1. **(10 points)** Several statements involving probability are listed below. Assume that A and B are events, S and T are random variables, and the underlying sample space is countable. Furthermore, assume that all expectations exist. For each statement, describe restrictions on A, B, S, and T sufficient to ensure that the statement is true. Make your restrictions as minimal as you can. No proofs are necessary.

(a) $\Pr(A \cup B) = \Pr(A) + \Pr(B)$

(b) $\Pr(A \cup B) = \Pr(A) + \Pr(B) - \Pr(A \cap B)$

(c) $\Pr(A \cap B) = \Pr(A) \cdot \Pr(B)$

(d) $\Pr(A \cap B) < \Pr(A)$

(e) $\mathrm{Ex}(S + T) = \mathrm{Ex}(S) + \mathrm{Ex}(T)$

(f) $\mathrm{Ex}(S \cap T) = \mathrm{Ex}(S) \cdot \mathrm{Ex}(T)$

Problem 2. **(5 points)** Suppose that I roll a 4-sided die, a 6-sided die, an 8-sided die, a 10-sided die, a 12-sided die, and a 20-sided die. What is the expected number of 6's that come up? (Assume that all of the dice are fair.)

Problem 3. **(10 points)** Suppose that I roll n dice that are 6-sided, fair, and mutually independent. What is the expected value of the largest number that comes up?

Problem 4. **(10 points)** Suppose that I choose a listing of the numbers $1, 2, \ldots, n$ uniformly at random. What is the expected number of entries that are greater than all preceding entries? For example, in the listing 4, 2, 1, 5, 3, the numbers 4 and 5 are greater than all preceding entries.

Problem 5. **(10 points)** Biologists are inexplicably obsessed with *genetic strings*, which consist of the characters A, C, T, and G. A biology student once observed that the sequence AAA occurs more often in a random genetic string than, say, CTG.

His reasoning is that once AAA has appeared, there is a $\frac{1}{4}$ chance that the next symbol will be an A. If this happens, then the sequence AAA has appeared a second time, offset one character to the right of the first occurrence. By contrast, once CTG has appeared, the next appearance of CTG must be offset at least three characters to the right of the first occurrence. Furthermore, the probability that a second CTG appears immediately to the right of the first is a miniscule $\frac{1}{4^3} = \frac{1}{64}$.

(a) Compute the expected number of occurrences of AAA and of CTG in a genetic string of length 4 chosen uniformly at random.

(b) Compute the expected number of occurrences of AAA and of CTG in a genetic string of length n chosen uniformly at random.

Problem 6. **(20 points)** Suppose that R is a random variable. Let E be an event. The *conditional expectation* of R given that event E occurs is denoted $\text{Ex}(R \mid E)$ and is defined to be

$$\text{Ex}(R \mid E) \quad = \sum_{x \in \text{range}(R)} x \cdot \Pr(R = x \mid E).$$

(a) Compute the expected value of the number rolled on a fair, 6-sided die, given that the outcome is even.

(b) Compute the expected value of the number rolled on a fair, 6-sided die, given that the outcome is odd.

(c) Prove the following identity, assuming that all expectations exist.

$$\text{Ex}(R) \quad = \quad \Pr(E) \cdot \text{Ex}(R \mid E) + \Pr(\overline{E}) \cdot \text{Ex}(R \mid \overline{E})$$

(d) Compute the expected value of the product of the numbers rolled on two dice that are 6-sided, fair, and independent, given that the numbers rolled are distinct.

Problem 7. **(10 points)** Suppose that I start throwing balls uniformly and mutually independently at random into n bins.

(a) Suppose that at some stage, exactly k of the bins contain at least one ball. Now I throw additional balls until exactly $k+1$ of the bins contain at least one ball. What is the expected number of additional balls that I throw?

(b) Suppose that the bins start out empty, and I throw balls until every bin has at least one ball. What is the expected number of balls that I throw?

Problem 8. **(15 points)** Let R be a random variable taking on non-negative integer values. Let r be the probability generating function for R; that is, define

$$r(x) = \Pr(R=0) + \Pr(R=1)\cdot x + \Pr(R=2)\cdot x^2 + \Pr(R=3)\cdot x^3 + \ldots.$$

(a) Write an expression involving r that is equal to the expected value of R.

(b) Suppose that $\Pr(R=k) = \frac{2}{3^{k+1}}$ for all $k \geq 0$. Find a closed-form expression for r, the probability generating function for R.

(c) Using your solution to part (a), determine the expected value of the random variable R as defined in part (b).

Problem 9. **(10 points)** 6.042 TA Eric Lehman has offered to play a "fun game" with Theory Pig, who rarely comes out of his drawer anymore after all the abuse that he has suffered this term.

Here are the rules. Eric flips a coin until he gets a heads. Let n be the number of tails preceding the first heads. For example, if he flips $TTTH$, then $n = 3$, and if he flips just H, then $n = 0$. Now Eric offers to pat Theory Pig on each of the next 3^n days. If Theory Pig declines the offer, then Eric repeats the coin-tossing experiment and instead pats Theory Pig on each of next 3^m days where m is the number of tails before the first heads in the second trial.

As usual, this game sounds great, but Theory Pig is understandably nervous. He wants to determine his best strategy.

(a) As a warmup, consider the following simpler variant of the game. Suppose that Eric rolls a fair, 6-sided die. He offers to pat Theory Pig for the number of days indicated on the die. If Theory Pig rejects the offer, then Eric rolls again and pats Theory Pig for the number of days shown on the second roll. What is expected number of patting-days, if Theory Pig uses an optimal strategy?

(b) Now return to the original coin-flipping game. Describe a strategy that maximizes the expected number of patting days.

WARNING: If less than 80% of the class solves this problem part correctly, then Eric will withdraw his offer to play the game, and THE PIG GETS NOTHING!