Assigned: Wednesday, September 8, 1999
Due: Wednesday, September 15, 1999
Reading: Ross, Chapter 4.1, 4.3–4.5, 4.7, and Chapter 3
Noncredit Exercises: (Do not turn these in) Ross, pp. 173-184: 2, 7, 13, 20, 28, 35(a), 39, 40-43, 57, 59; pp. 184-188: 11, 13, 15-18

Problems:

1. An ice-cream store manufactures unflavored ice-cream and then mixes one or more of five flavor essences (vanilla, chocolate, fudge, mint, and strawberry) into the unflavored ice-cream to make its specialty ice-cream products. Thus, chocolate-mint ice-cream is produced by mixing chocolate and mint essences into the unflavored ice-cream, spumoni by mixing vanilla, chocolate, and strawberry essences into the unflavored ice-cream, etc.

(a) How many different specialty ice-creams can the store offer for sale?
(**Optional noncredit exercise:** Identify the store!)

(b) An experiment consists of picking one of the specialty ice-creams on sale at random (that is, each is equally likely to be picked.) What is the probability that your choice

- **(i)** has chocolate essence (and possibly others?) **(ii)** has *only* chocolate essence in it?
- **(iii)** has chocolate and vanilla essences (and possibly other essences as well?)
- **(iv)** has chocolate and mint essences (and possibly other essences as well?)
- **(v)** has chocolate and either vanilla or mint essence or both (and possibly others as well?)
- **(vi)** has chocolate and either vanilla or mint essence but not both (and possibly others?)
- **(vii)** has chocolate but neither vanilla nor mint essence (but might have the others?)
- **(viii)** has exactly two essences? **(ix)** has at least three essences?
- **(x)** has exactly two essences neither of which is strawberry?
- **(xi)** has exactly two essences neither of which is strawberry or fudge?

(**Further optional noncredit exercise:** Identify the professor's favorite flavor!)

2. The four parts of this problem are to be solved separately. **Do not** use information from one part in another part.

Find $P(A \quad (B^c \quad C^c)^c)$ in each of the following four cases:

(a) A, B, and C are mutually exclusive events and P(A) = 1/3.
(b) P(A) = 2P(BC) = 4P(ABC) = 1/2. **(c)** P(A) = 1/2, P(BC) = 1/3, and P(AC) = 0.
(d) $P(A^c \cap (B^c \quad C^c)) = 0.6$.

3. Use a spreadsheet/Mathematica/MATLAB for this problem.
Let A denote an event of probability p.

(a) For p = 0.1, 0.25. 0.4, 0.5, 0.6, 0.75, and 0.9, find the numerical values of the probabilities that A occurs 0, 1, 2, … , 10 times on 10 trials of the experiment.
(b) You have, in effect, computed the probability mass function for a binomial random variable **X** with parameters (10, p) for seven choices of p. For each value of p, draw a bar graph of the pmf. (For p = 0.5, the answer is shown on page 151 of Ross!)
(c) What is the relationship between the pmfs for the cases p = 0.1 and p = 0.9? for the cases p = 0.25 and p = 0.75? for the cases p = 0.4 and 0.6?
(d) Prove mathematically that if **X** is a binomial random variable with parameters (n, p), then **Y** = n – **X** is a binomial random variable with parameters (n, 1–p).
(e) From each of the seven graphs of part (b), find the value of k for which P{**X** = k} is maximum. Compare your results to the prediction of Proposition 7.1, p. 150, of Ross.

4. Let **X** denote a binomial random variable with parameters (N, p). What is the probability that **X** is an even integer? Remember that 0 is an even integer.
[Hint: What is $(x+y)^N + (x-y)^N$?]

5. There are N multiple-choice questions on a certain examination. A student knows the answer to K of these and marks the answer sheet accordingly. For the remaining N – K questions, the student guesses randomly among the five choices. The examiner can easily determine C, the number of correct answers on the answer sheet, but is more interested in estimating the value of K, since K is a better measure of the student's knowledge than C. (Educators like to nitpick about such subtle differences!). Note that the number of **w**rong answers can be modeled as a binomial random variable **W** with parameters (N – K, 0.8).

(a) The examiner notes that n questions have been answered incorrectly by the student, i.e. the event {**W** = n} is observed. Write an expression for P{**W** = n} in terms of N, K, and n.

(b) Obviously, $0 \le K \le N - n$. Now, use the method used in the proof of Proposition 7.1, p. 150 of Ross to show that of all possible assumptions K = 0, K = 1, K = 2, … , K = N–n that the examiner might make, the assumption that K is the largest integer not exceeding N – 1.25n + 1, i.e. estimating K as $\hat{K} = \lfloor N - 1.25n + 1 \rfloor$ maximizes P{**W** = n}. $\hat{K}$ is the called the maximum-likelihood estimate of K. Find the numerical value of $\hat{K}$ for the case N = 100 and n = 8.

(c) Since C = N – n, examiners generally subtract one-fourth of the wrong answers from C and estimate the value of K as $\tilde{K} = C - \lfloor 0.25n \rfloor \approx N - 1.25n$. This is called applying the guessing penalty, and it *can* hurt scores slightly in the sense that the examiner's estimate $\tilde{K}$ *might* be smaller than the maximum-likelihood estimate $\hat{K}$ found in part (b).
Compare $\hat{K}$ and $\tilde{K}$ for the case N = 100 and n = 8.
Compare $\hat{K}$ and $\tilde{K}$ for the case N = 100 and n = 10.

(d) If N = 100 and K = 90, which of the events {**W** = 0}, {**W** = 1}, … , {**W** = 10} has the largest probability? (Hint: See Proposition 7.1, p. 150 of Ross). Suppose that this largest probability event actually occurred. Does the examiner's estimate $\tilde{K}$ correctly estimate K? Does the maximum-likelihood estimate $\hat{K}$ correctly estimate K?

(e) Continuing to assume that N = 100 and K = 90, what happens if by sheer dumb luck the student manages to guess right on 6 problems so that the event {**W** = 4} occurs? Compare this to the case when the event {**W** = 10} occurs (as in part (c)).

(f) Continuing to assume that N = 100 and K = 90, find the probabilities that the examiner's estimate $\tilde{K}$ respectively overestimates, underestimates, and correctly estimates K.

(g) **Optional Noncredit Exercise:** Suppose that for each of the 10 questions to which the answer is not known to the student, the student can nonetheless correctly eliminate three answers as being obviously wrong. The student then chooses at random between the other two answers. Which of the events {**W** = 0}, {**W** = 1}, … , {**W** = 10} is the most probable? If this is the event which actually occurs, what is the examiner's estimate $\hat{K}$? (Note that this, in essence, gives *some* "partial credit" by rewarding the student for the partial knowledge that three of the five answers to each problem are wrong.)

(g) **Optional Noncredit Exercise:** Write a 500-word essay on why it is more important to be lucky than smart.

6. Let **X** denote a Poisson random variable with unknown parameter λ. Suppose that the event {**X** = k} occurs.

(a) What is the maximum–likelihood estimate of λ? That is, what value of λ maximizes the probability of the observed event {**X** = k}?

(b) Consider a binomial random variable **Y** with parameters (N, p) where the parameter p is unknown. If the event {**Y** = k} is observed (e.g. heads occurs k times on N tosses of a biased coin with P(Heads) = p), then we showed in class that $\hat{p} = k/N$ is the maximum–likelihood estimate of p. Since for large N and small p, the binomial random variable **Y** can be approximated by a Poisson random variable **X** with parameter λ = Np, it would seem reasonable that the maximum-likelihood estimate of λ would be $\hat{\lambda} = N\hat{p} = k$. Does your answer to part (a) give this result?