Assigned: Wednesday, February 17, 1999
Due: Wednesday, February 24, 1999
Reading: Ross, Chapter 3
Noncredit Exercises: Ross pp. 104-117: 53, 58, 59, 62, 63, 70-74, 78, 81
Problems:

1. Ross, #12, p. 105

2. Ross, #19, p. 106

3. Ross #5, p. 123. In addition to the probability asked for, find the probability that the second ball is red, and determine if it is smaller or larger or the same as the probability that the first ball is red.

4. (Remember: 99.44% of all statistics are made up by the writer)
The experiment consists of picking a flight at random from all the United Airlines and America West flights landing at Chicago, Los Angeles, Phoenix, San Diego, or San Francisco. Let U and W respectively denote the event that the chosen flight is an United Airlines or an America West flight, let C, L, X, D, and F respectively denote the event that the chosen flight is landing at Chicago, Los Angeles, Phoenix, San Diego, or San Francisco, and let T denote the event that the chosen flight is on time. The conditional probabilities of on-time arrival are as follows:

$P(T|UC) = 0.85$, $P(T|UL) = 0.92$, $P(T|UX) = 0.95$, $P(T|UD) = 0.91$, $P(T|UF) = 0.83$,
$P(T|WC) = 0.78$, $P(T|WL) = 0.88$, $P(T|WX) = 0.92$, $P(T|WD) = 0.85$ $P(T|WF) = 0.73$.

(a) Based on this data, which airline would you say has better on-time performance? Does the answer depend on which airport you are talking about?

(b) Use the fact that {C, L, X, D, F} form a partition of the sample space to show that the average on-time arrival probability $P(T|U)$ for United flights is given by

$$P(T|U) = P(T|UC)P(C|U)+P(T|UL)P(L|U)+P(T|UX)P(X|U)+P(T|UD)\,P(D|U)+P(T|UF)P(F|U)$$

where $P(C|U)$ is the conditional probability that the flight is landing at Chicago given that it is a United flight etc. State a similar expression for $P(T|W)$. (cf. Ross pp. 98-99)

(c) 60% of United Airlines flights land at its hub (snowy Chicago), 15% at each of LA and San Francisco, and 5% at each of Phoenix and San Diego. 75% of America West flights land at **its** hub (sunny Phoenix), 10% at LA, and 5% at each of the other three airports. Use these numbers in the formula of part (b) and show that $P(T|U) < P(T|W)$, i.e., United has a worse average on-time performance even though it beats America West at all the five airports! Write a short explanation of the discrepancy between the per-airport on-time performance and the overall on-time performance.

5. A baseball pitcher's repertoire is limited to fastballs (event F), curveballs (event C), or sliders (event S). It is known that $P(C) = 2P(F)$. Also, the event H that the batter hits the ball has probabilities $P(H|F) = 2/5$, $P(H|C) = 1/4$, and $P(H|S) = 1/6$.

(a) If $P(H) = 1/4$, what is $P(C)$?

(b) A fan sitting in the bleachers sees the batter getting a hit, i.e. the event H. He knows the values of $P(H|F)$, $P(H|C)$, and $P(H|S)$, but is sitting too far away to tell whether the pitch was a fastball, a curveball or a slider. What is his maximum-likelihood decision as to what kind of pitch it was?

(c) After cheering the hit, the fan finds that P(F), P(C), and P(S) are listed in the program guide. What is his Bayesian (that is, maximum *a posteriori* probability or minimum-error-probability) decision as to the kind of pitch it was?

6. Let H_0, H_1, and H_2 respectively denote the hypotheses that a UIUC student is excellent, good, or average (by definition, there are no poor students at UIUC). A certain professor does not give any F's in his 3-hour course (and no ± letter grades either!). Thus, the discrete random variable **X** denoting the number of grade points earned by the student in the course takes on values 3, 6, 9, and 12 only. From previous experience, the professor knows that the pmf of **X** when H_0 is the true hypothesis is

$p_0(12) = 0.75,\quad p_0(9) = 0.15,\quad p_0(6) = 0.08,\quad p_0(3) = 0.02,$

that is, an excellent student has 75% chance of doing well enough on the exam to get an A, 15% chance of a B, 8% chance of a C, and 2% chance that the performance on the exam will warrant the award of a D. Similarly, when H_1 is the true hypothesis, the pmf of **X** is

$p_1(12) = 0.15,\quad p_1(9) = 0.6,\quad p_1(6) = 0.15,\quad p_1(3) = 0.1$

and when H_2 is the true hypothesis, the pmf of **X** is

$p_2(12) = 0.05,\quad p_2(9) = 0.1,\quad p_2(6) = 0.65,\quad p_2(3) = 0.2.$

The professor observes the grade earned by a student and must decide which of the three hypotheses H_0, H_1, and H_2 is true.

(a) For each of the four values of **X**, find the professor's maximum-likelihood decision as to whether the student is excellent, or good or average.

(b) What is the probability that an excellent student is mistakenly labeled as good by the professor? What is the probability that an excellent student is mistakenly labeled as average by the professor? What is the probability that an average student is classified as being above average (that is, as good or excellent)?

(c) If 20% of the students are excellent, 55% are good, and 25% are average (that is, $P(H_0) = 0.2$, $P(H_1) = 0.55$, and $P(H_2) = 0.25$, what is the probability that the professor's maximum-likelihood decision rule mis-classifies students?

(d) What is the Bayes' decision rule corresponding to these probabilities and what is the probability that the Bayes' decision rule mis-classifies students?

(e) At the Lake Wobegon campus of the University of Illinois, 95% of students are excellent and 5% are good (and thus they are all above average!) What is Bayes' decision rule in this case? That is, what does the Bayesian professor decide about a student based on the four possible results of the student's exam?

(f) For what range of values of $P(H_0)$ does the Bayes' decision rule always decide that H_0 is true? Note that the decision rule should be applicable regardless of the values of $P(H_1)$ and $P(H_2)$ (subject, of course to the constraints that $P(H_0) \geq 0$, $P(H_1) \geq 0$, $P(H_2) \geq 0$, and that $P(H_0) + P(H_1) + P(H_2) = 1$)

7. The probability that you can hear the sound of a pin dropping onto a table during a long-distance telephone call from Champaign-Urbana to Hollywood is p_0 if the call is being carried over the AT&T network and p_1 if the call is being carried over the Sprint network. Assume that both p_0 and p_1 are quite small and that $p_1 > p_0$. You call Miss Candice Bergen from a payphone owned by Sleazo Telecom Corporation which happens to lease its long-distance lines either from AT&T or Sprint (but you don't know which!), and she agrees to drop pins one by one onto a table until you hear the sound of one dropping. Suppose that you hear her counting "One, two, three, … " as she drops the pins, but you hear nothing else until she says "thirtyfour" and you finally hear the sound made by the pin dropping onto the table.

(a) Let H_0 and H_1 denote the hypotheses that the call is being carried by AT&T and Sprint respectively. What is the likelihood ratio LR for the observation that the 34th pin dropped was the first one heard?

(b) The maximum-likelihood decision compares LR to the threshold 1 and announces in favor of H_0 and H_1 according as $LR < 1$ or $LR > 1$. Show that this decision rule can be expressed in terms of a threshold test on k, the number of the first pin that you heard being dropped.

(c) If $p_1 = 0.04$ and $p_0 = 0.02$, what is the maximum-likelihood decision if the 34th pin is the first one heard?

(d) AT&T is the lessor of 95% of all long-distance telephone lines while Sprint is the lessor of the remaining 5%, and thus it is reasonable to assume that $P(H_0) = 0.95$. What is the Bayesian decision if the 34th pin is the first one heard?

(e) Noncredit exercise: Whom do you think is *my* long-distance carrier?