Prof. David Draper
Department of
Applied Mathematics and Statistics
University of California, Santa Cruz

AMS 7: Discussion Section 4

1. A big problem facing banks and merchants alike is determining potential customers' credit-worthiness. You've probably had the experience of trying to use a credit card to pay for something — store owners typically have a little electronic machine they pass your card through, and information stored in the magnetic strip on the back (mainly your credit card account number) is transmitted over a telephone line to a central *credit-verification system* somewhere. This system uses a computer program to decide if your purchase should be approved, based on factors like how often it thinks your card has been used lately and how recently it thinks you've paid your bill. If you pass this invisible screening, a little green light goes on back in the store, and you walk out with your purchase.

Of course, any computer-based system of this kind makes mistakes sometime, because of faulty information or bad programming: sometimes credit cards that are good are declared bad, sometimes vice versa. Events of this type look random to the people at the bank trying to figure out why they happen — at least until the causes of the mistakes are determined — so it makes sense to talk about the *probability* that a bad credit card is declared good, and the probability that a good card is judged bad. Standard terminology is to call the kind of mistake in which you declare a good credit card bad a *false positive*, and the other kind of mistake — in which you label a bad credit card good — a *false negative*. (Evidently the people who made up this terminology were thinking of "positive" in this context as equivalent to "calling a credit card bad" and "negative" as amounting to "calling the card good," which is a little perverse, but there it is.)

People evaluate the quality of credit-screening systems of this type by running tests in which (a) they attempt a number of fake "purchases" with some credit cards that are known to be good and some others that are known to be bad, and (b) they look to see how often the system gets it right. Suppose that in one test of this type on the system we're going to look at, 97% of the test "purchases" with cards known to be good were labeled good by the system, and 98% of the "purchases" in which the test cards were bad were declared bad by the system. Suppose further that the system is to be used in a market in which about 1% of all attempted purchases are with bad credit cards. Somebody now walks into a store using this credit verification system and tries to make a purchase by credit card, and the system comes back with a negative opinion about this person's credit-worthiness. Show that the conditional probability the card is indeed bad, given the system's "diagnosis," is only about 25%! How can you make sense of this result, assuming that the people who designed the screening system aren't stupid? Explain briefly.

2. As you may know, there are a number of different subway lines in London, some of which run in parallel under the same streets. The builders of the Underground arranged this by stacking the subway tunnels for the lines under each other, sometimes two or three deep. At many Underground stations, to get to the deepest tunnels you take a long escalator ride. For instance, at the Pimlico station the escalator down to the deepest tunnel is like a

R-59 [redacted] [redacted]

moving stairway with 96 steps. During rush hour every single escalator step typically has two people on it side by side, so that the escalator had to be designed to carry 192 people without overloading.

(a) The population of Underground riders at rush hour is almost exclusively made up of adult British men and women, whose weight averages about 158 pounds with an SD of about 33 pounds (these figures are from a 2002 study published in the *International Journal of Obesity* (BTW, the corresponding mean for American adults, from a 2006 study, was 30 pounds heavier)). If the engineers who planned the Pimlico station designed the escalator to carry 31,400 pounds worth of people without breaking, what proportion of the time when it's fully loaded with 192 people would it break down? Explain briefly. Express your answer in the form "about 1 in every k fully loaded trips."

Be explicit about your probability model (in other words, relate this setup explicitly to the population-and-sample framework we've been talking about in class, writing out the population, sample, and imaginary datasets), and comment briefly on all assumptions you make and whether you think they're reasonable.

(b) Given that the morning and evening peak traffic periods together last about an hour and a half, and the turnover on the escalator is such that it's like having a new trip with 192 new people about every minute during this period, do you regard this failure rate as acceptably low? Explain briefly. (Hint: At this rate about how often would it break down?)

(c) If the engineers had wanted the escalator to fail only about once in every 10,000 fully-loaded trips (which would still mean it would overload more than 3 times a year), how much weight should they have designed it to carry? Explain briefly. (The official place on the standard normal curve with 0.01% of the area beyond that place is 3.72.)

[redacted] R-60