PS 200b: Midterm Exam Answers
Winter 2003

1. According to de Moivre, in eighteenth century England people played a game similar to Roulette. It was called Royal Oak. There were 32 "points" or numbered pockets on a table. A ball was thrown in such a way that it landed in each pocket with an equal chance, 1 in 32.

   If a pound was bet on a point, the bet returned 27 pounds (plus the wagered pound) if the point came up and 0 otherwise.

   (a) How much would you expect to lose if you played this game 10 times?

   There was some confusion here about exactly what was wagered, gained, lost, and so forth. I accepted any interpretation of the payoffs and marked you on the logic from that point on. I would work it this way. With probability 1/32 you are 27 pounds ahead and with probability 31/32 you are 1 pound behind. Thus your expected winnings (X) from one play are:

   $$\mathrm{E}(X) = (1/32)27 + (31/32)(-1) = -0.125$$

   Over ten plays you end up on average with the expectation of $X_1 + X_2 + \ldots + X_{10}$, which is $10\mathrm{E}(X) = -1.25$.

   (b) de Moivre claims that gamblers complained about the odds they were given, they said they should get 31 pounds back (plus their pound wagered). Why is a 31 pound payoff (plus the pound wagered) fair?

   Here, the idea is that the expected value of the game is equal to the price. The price is 1. I copied this problem out of a book, but I see why it is confusing. If you win you have 32 pounds instead of the one you started with. If you lose you have 0 pounds instead of the 1 that you started with. The expected payoff is:

   $$\mathrm{E}(P) = 32(1/32) + 0(31/32) = 1$$

   which is equal to the cost of playing the game.

   (c) de Moivre writes:

   > The Master of the Ball maintained they [the gamblers] had no reason to complain; he would undertake that any

> particular point ... should come up in Two-and-Twenty [22] throws: of this he would offer to lay a Wager [at even odds], and actually laid it when required. This seeming contradiction between the Odds of One-and-Thirty [31] to One, and Two-and-Twenty Throws for any point to come up, so perplexed the Adventures [gamblers], that they begun to think the Advantage was on their side: for which reason they played on and continued to lose.

What is the "apparent contradiction" and why it is not a really a contradiction?

The apparent contradiction is if that if the player wins within 22 plays for sure, the game is a definite money maker for the player. That is, the player would know that for every 22 pounds bet (22 plays of the game), he was sure to win at least 27 pounds! The longer you play the more you win apparently.

Of course, it is not true that the player will always win within 22 plays. It is only true that she has (as we will see below) a better than 1/2 chance of doing so. Thus, there is no contradiction here.

(d) What is the probability that the Master of the Ball would win his bet (the one involving a particular point occurring within 22 throws)? Was he smart to give even odds for this bet?

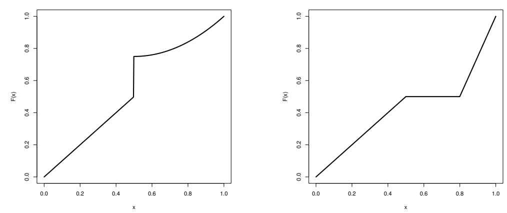
This probability is calculated as one minus the probability for not winning within 22 rolls. The reason for this is that the probability of winning one or more times is much harder to calculate. The probability of not winning in 22 rolls is $(31/32)^{22} \approx 0.497$ which is the probability of not winning in any given roll multiplied together 22 times. This works because each play is assumed independent. So the probability,

$$\mathbb{P}(LL\ldots L) = \mathbb{P}(L)\mathbb{P}(L)\ldots\mathbb{P}(L).$$

Thus, the probability of a point coming up within 22 plays is $1 - (31/32)^{22} \approx 1 - 0.497 = 0.503$.

Several folks calculated the answer to be $22(1/32)$ which would be true if the events (wins on each roll) were disjoint, which they are not. One way to have seen this error would be to suppose that the question had been what's the probability that a point comes up in 64 rolls. Using the same approach, this probability would be $64/32 = 2$!

2. Consider the following two cumulative distribution functions for random variables defined on the $[0, 1]$ interval.



(a) (b)

What do can you infer about each random variable from the jump in the first (a) function and the flat area in the second (b)? Is the first random variable, shown in graph (a), continuous, discrete, or something else? What does the PDF function look like that corresponds to the CDF in figure (b)?

The jump in graphs A indicates that the the probability of the random variable taking on exactly the value associated with the jump is equal to the height of the jump. That is, the probability of X being exactly equal to or less than x (where x is the point where the jump occurs) is greater than that than X being strictly less than x by the height of the jump.

Jumps of this sort are you would expect to see if you have a discrete random variable. Whereas a smoothly increasing CDF is indicative of a continuous CDF. Strictly speaking the graph should not have plotted the vertical bit (the jump). The graph should have just had two separate parts ending and starting with the jump. So, if you said that the graph was not a function that was ok, too. Basically, this random variable had both discrete and continuous features.

The second graphs has a **flat spot** which corresponds to an interval of values of X with have zero probability of occurring. The PDF function should be made up of rectangles separated by a blank region corresponding to the where the CDF is flat. Because the CDF is more steeply sloped in on the right of the flat spot, the PDF rectangle on the right should be taller than the one on the left.

3. Find the expected number of times two dice are rolled before the sum of seven comes up. [Extra credit: What is the expected number of

rolls required to get a sum of seven five times. (*Do this one only after finishing off the rest*).]

First thing is to find the probability of rolling two dice and getting a sum of 7. This can occur 6 ways $\{(1,6),(2,5),(3,5),(4,3),(5,2),(6,1)\}$ out of $6 \times 6 = 36$ equally likely outcomes, so we have the probability of rolling 7 equal to $6/36 = 1/6$.

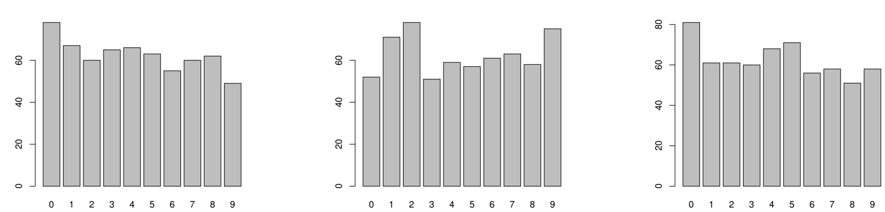
Working out the expected number of rolls needed to get a seven would be hard to do directly, but is easy to find if you recognize this distribution of waiting times as following a geometric distribution where $\pi = 1/6$ (remember the bombing example from lecture). Here the PDF is

$$f(x) = (1-\pi)^{x-1}\pi$$

finding $\mathrm{E}(X) = \sum_{x=1}^{\infty} x f(x)$ would still be unpleasant, but others have worked out that answer to be $\mathrm{E}(X) = 1/\pi$ (as noted in class), so we have the expected number of rolls equal to $1/(1/6) = 6$.

The extra credit asks about how long you would expect to wait for 10 sevens. This is just the sum of the waiting time for the first 7, the second 7, the third 7 and so forth. So the expectation of the sum is the sum of the expectations, or $5 \times 6 = 60$.

4. I collected data on the outcome of the Ohio "pick three" lottery for the years 2000 and 2001. In the "pick three" lottery, a number between 000 and 999 is picked at random if you match all three balls in order you win. Below are barplots of the number of times each digit appeared in each of the three positions:



First digit Second digit Third digit

Looking at these graphs, it seems like maybe low numbers are more likely than high numbers in the first and third positions. Across the

1250 digits drawn in the first and third positions, the mean digit was $\bar{x} = 4.25$. If lottery numbers are really uniformly distributed over $\{0, 1, 2, 3, 4, 5, 6, 7, 8, 9\}$ and if the digits are drawn independently, what is the probability of observing a mean that is 4.25 or less across 1250 digits? What can you conclude from this? (*Proceed in small steps. Derive the expectation of $\bar{x}$. Find the variance of $\bar{x}$. Note that $\bar{x}$ is the sum of 1250 independent random variables (divided by a constant), say what that implies about the distribution of $\bar{x}$, using all of the these facts find the probability of finding $\bar{X}$ to be less than* 4.25.)

Statisticians are often hired to check out randomization devices such as the lottery. In one famous example, it was shown that the military service draft lottery (I believe during the early Vietnam War lottery) was not truly random, people with birthdays during the end of the year had a greater chance of being selected.

At any rate, following the steps the problem is fairly straight forward. Under the assumption that the digits are equally likely and letting X be a random digit, we have:

$$\mathrm{E}(X) = \sum_{x=0}^{9} \frac{1}{10} x = 4.5$$

$$\mathrm{V}(X) = \sum_{x=0}^{9} \frac{1}{10} x^2 - \mathrm{E}(X)^2 = 8.25$$

Now the expectation of $\bar{X}$ is the same the expectation of X, so $\mathrm{E}(\bar{X}) = 4.5$ and the variance of $\bar{X}$ is $\frac{\mathrm{V}(X)}{n} = 8.25/1250 = 0.0066$ (assuming the digits are sampled independently).

Because the sample is large, we know that $\bar{X}$ will be approximately normal so the we can find

$$\mathbb{P}(|\bar{X} - 4.5| \geq 0.25)$$

by making Z scores and looking up the probabilities on a normal table. Here we have $Z = \frac{4.25-4.50}{\sqrt{0.0066}} \approx -3$ and $\mathbb{P}(|Z| > 3) \approx 0.003$. Thus there is only about a 3 in a thousand chance of seeing a mean this far from the population mean if the true process placed equal probability on each digit.

However, there are a many many "unusual" patterns we might have observe in the data. If we look for say 250 different possible "unusual" patterns (ones that would occur with probability equal to 0.003), we have a pretty good chance (better than 50 percent) of finding a least one!

5. Suppose you know that traffic court judges are of two types: lenient and tough. Tough judges convict 90 percent of defendants whereas lenient judges only convict 50 percent of defendants. Rushing to get to class, you get a summons for reckless driving. You are pretty worried because you know that 99 percent of judges in LA are tough. Waiting for your case to be tried you sit in the courtroom watching other cases come before the judge that will be trying your case. To your surprise and glee, the judge finds the first five cases of the day in favor of the defendants. What is the probability, that your case is coming before a lenient judge? What is the probability that you will be convicted? (*Assume that the cases dispositions are determined independently and that each case, including yours, as the same merit. That is the probability of conviction is only a function what kind of judge you come before.*)

   This is basically just a Bayes' rule problem that requires a bit of preprocessing. Let L be the event "lenient judge", T be the event "tough judge", A be the event "acquitted," and C be the event "convicted."

   From the given info we have, $\mathbb{P}(L) = 0.01$, $\mathbb{P}(T) = 0.99$, $\mathbb{P}(C|L) = 0.5$, $\mathbb{P}(C|T) = 0.9$, $\mathbb{P}(A|L) = 0.5$, and $\mathbb{P}(A|T) = 0.1$. Finally, we assume that each decision is independent of the last. Now, we are told that we see the judge hand out five acquittals. We can find

$$\mathbb{P}(AAAAA|L) = 0.5^5 = 0.03125$$

   and

$$\mathbb{P}(AAAAA|T) = 0.1^5 = 0.00001$$

   Now, by Bayes' rule, we have

$$\mathbb{P}(T|AAAAA) = \frac{0.00001 \times 0.99}{0.00001 \times 0.99 + 0.03125 \times 0.01} \approx 0.03$$

   thus,

$$\mathbb{P}(L|AAAAA) \approx 0.97$$

   Finally, the probability that you will be convicted is:

$$0.9 \times 0.03 + 0.97 \times 0.5 = 0.512.$$

   So, even though the last 5 folks got off, your chances of getting off are less than 1/2.