MA 2B RECITATION 02/02/12

1. Introduction

There will be a midterm review on Friday at 8pm in Sloan 151.

Since there's going to be a midterm review, I don't want to just rehash the same things here, so let's do this by working through some examples, and discussing the best way to approach things. Most of the other skills required for the exam were tested on the homework, such as knowing how to apply things properly and remembering properties about our objects of study, or are covered well in the chapter summaries in Pitman. Your midterm review will also likely cover these things, so let's focus today on an important but often neglected skill in probability: translating problems into the language of probability in such a way that admits an easy solution.

2. Oodles of Examples

Example 1. In a country where everybody wants a girl, each family continues having babies until they have a girl. Assuming that the probability of having a boy or a girl is the same and independent of previous children, after some time, what is the proportion of boys to girls in the country?

Solution. Suppose that there are N couples in the country, so there would a total of N girls in the country. The number of boys can be calculated as follows.

Let i be an integer and let "i" denote the probability of a family having i boys. Then the number of boys is

$$B = 0 \cdot P(0) + 1 \cdot P(1) + 2 \cdot P(2) + \cdots = \sum_{n=0}^{\infty} n \cdot P(n)$$

and it turns out, with some calculation, $B = N$, so there is a $1 : 1$ ratio of boys to girls in this country. □

Example 2. You have an urn filled with red and white balls, two-thirds are one color and one-third is another. You have drawn 5 balls from the urn and saw that 4 are red and 1 is white.

Your friend has drawn 20 balls and found that 12 are red and 8 are white.

Who should feel more confident that the urn contains two-thirds red balls and one-third white balls, instead of the other way around?

Solution. Let A be the event that $2/3$ of the balls are red, and $\neg A$ be the event that $2/3$ of the balls are white. Let B be the event that the first observer sees 4 red balls of out of 5. Let C be the event that the second observer sees 12 red balls out of 20.

Recall Bayes rule

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}.$$

Date: February 2, 2011.

We have

$$P(B|A) = \binom{5}{4}(2/3)^4(1/3)^1 = 80/243$$
$$P(B|\neg A) = \binom{5}{4}(1/3)^4(2/3)^1 = 10/243.$$

Therefore from Bayes' Rule, you have a confidence level of $80/(80+10) = 8/9$ that A is true.

For the second observer,

$$P(C|A) = \binom{20}{12}(2/3)^{12}(1/3)^8 = 125970 \cdot \frac{2^{12}}{3^{20}}$$
$$P(C|\neg A) = \binom{20}{12}(1/3)^{12}(2/3)^8 = 125970 \cdot \frac{2^8}{3^{20}}.$$

Therefore, from BAyes' Rule, observer 2 has a confidence level of $2^{12}/(2^{12}+2^8) = 16/17$ that A is true. Therefore, your friend has a higher confidence level that 2/3 of the balls are red.

This phenomenon is a bit strange. Intuitively, either answer at first seems correct. You have a higher proportion of red balls, so it seems like you should be more confident. On the other hand, your friend has a larger sample size, which might make them feel more confident. The key is to understand how Bayes' Law works, and it might not sink in until after this class when you see Bayesian analysis applied to a problem you're familiar with. In fact, all that matters is the *difference* in the number of red and white balls observed. Everything else (specifically the number of total balls drawn) cancels out in the equations. □

Example 3. A true/false quiz has 20 questions. A student knows the correct answer to 14 of the Each of the 6 other questions she answers by tossing a coin: if the coin lands heads she answers "True" and if it lands tails she answers "False." One of the questions is picked at random. Given that she got it right, what is the chance that she knew the answer?

Solution. Simplest way: draw the tree. The answer is

$$\frac{0.7 \times 1}{(0.7 \times 1) + (0.3 \times 0.5)} = \frac{14}{17}.$$

□

Example 4. Same setting as above, but now suppose that the grading scheme is changed. Now 1 point will be awarded for correct answer and 1 point will be taken off for each wrong answer. Let S be the student's score on the test. Find $E(S)$ and $SD(S)$.

Solution. Let R be the number of questions she gets right among the 6 that she answers at random. Then R has a binomial distribution with parameters 6 and 0.5, and

$$S = 14 + R - (6 - R) = 8 + 2R.$$

Then

$$E(S) = 8 + 2 \cdot 3 = 14.$$

No surprise. We have

$$SD(S) = 2SD(R) = 2\sqrt{6 \cdot 0.5 \cdot 0.5} \approx 2.45.$$

□

We see that when set up the problem in the right way, everything else just falls into place, so it's worth practicing setups once you have the techniques down.

Example 5. A circle has random radius R with $E(R) = 10cm$ and $SD(R) = 2mm$. What is the expected area of the circle?

Solution. The area is

$$E(\pi R^2) = \pi E(R^2) = \pi(2^2 + 10^2) = 104\pi$$

square cm, using the fact that $E(R^2) - [E(R)]^2 = \mathrm{Var}(R) = SD(R)^2$. □

Example 6. Suppose that U_1 and U_2 are IID uniform on $\{1, 2, \ldots, 100\}$. Let $M = \max(U_1, U_2)$. Given that $U_1 + U_2 = 98$, find the distribution of M.

Solution. Given $U_1 + U_2 = 98$, the possible values of M are 49, 50, 51, …, 97. Every probability in the distribution will have as its denominator

$$P(U_1 + U_2 = 98) = \sum_{k=1}^{97} P(U_1 = k, U_2 = 98 - k) = 97/100^2.$$

For $k = 49$,

$$P(M = k \mid U_1 + U_2 = 98) = \frac{P(U_1 = 49, U_2 = 49}{P(U_1 + U_2 = 98} = \frac{1/100^2}{97/100^2} = 1/97.$$

For $k = 50, 51, \ldots, 97$,

$$\begin{aligned} P(M = k \mid U_1 + U_2 = 98) &= \frac{P(U_1 = k, U_2 = 98 - k) + P(U_1 = 98 - k, U_2 = k)}{P(U_1 + U_2 = 98)} \\ &= \frac{2/100^2}{97/100^2} \\ &= 2/97. \end{aligned}$$

As a sanity check, let's check that this is indeed a distribution. We add up the probabilities as see that

$$1/97 + 48(2/97) = 1.$$

Note that if we had an even number of possible values, then all of them would have uniform probability. □

Example 7. Suppose that a coin (not necessarily fair) which lands heads with probability p is tossed repeatedly. Let T be the number of tosses until the first head. Let N be another random variable with the Poisson (μ) distribution, independent of T. What is $P(N < T)$?

Solution. We partition the event according to the value of N. We have

$$\begin{aligned} P(N < T) &= \sum_{k=0}^{\infty} P(N = k)P(T > k) \quad \text{(by independence)} \\ &= \sum_{k=0}^{\infty} e^{-\mu}\frac{\mu^k}{k!} \cdot q^k \\ &= e^{-\mu}\sum_{k=0}^{\infty} \frac{(\mu q)^k}{k!} \\ &= e^{-\mu}e^{\mu q} = e^{-\mu(1-q)} = e^{-\mu p}. \end{aligned}$$

□

Example 8. Suppose we are at a tennis tournament where singles matches are played to a "best of three" basis: sets are played until one player wins two sets. (A tennis match consists of a number of sets that can be won or lost.) Assume that Player A has probability p of winning each set independently of all other sets.

Let W_A be the number of sets won by Player A and let S be the total number of sets played. What is the distribution of W_A/S, the proportion of sets won by Player A?

Solution. The joint distribution table for (W_A, S) is given below. The columns represent W_A and the rows represent S:

	0	1	2
2	q^2	0	p^2
3	0	$2pqq$	$2pqp$

Then the distribution of W_A/S is as follows:

$$\begin{aligned} P(W_A/S = 0) &= q^2 \\ P(W_A/S = \frac{1}{3}) &= 2pq^2 \\ P(W_A/S = \frac{2}{3}) &= 2p^2q \\ P(W_A/S = 1) &= p^2. \end{aligned}$$

□

Example 9. A biased coin lands heads with probability 1/3. The coin is tossed 100 times and the sequences of heads (H) and tails (T) is recorded. Let N be the number of times the pattern HT appears in the sequence. Find $E(N)$.

Solution. Once you figure out how to set up this problem, it's incredibly simple. There is a long brute-force way of doing it, but here is an elegant solution. We define a random variable

$$N = I_2 + I_3 + \cdots + I_{100}$$

where I_j is an indicator random variable for the event "toss $j-1$ is H and toss j is T." This is the sum of 99 indicators, each of which has expectation (1/3)(2/3), and so

$$E(N) = 99 \cdot \frac{1}{3} \cdot \frac{2}{3} = 22.$$

□

Example 10. Repeated measurements of something are IID with expectation 500 gm and SD 10 gm. What is the chance that the average of 10 measurements is greater than 501 gm?

Solution. We apply the Central Limit Theorem, either to the sum of the measurements—which has expectation 50,000 grams and standard deviation $\sqrt{100} \times 10 = 100$ grams—or to the average of the measurements, which has expectation 500 grams and SD $100/\sqrt{100} = 1$ gram.

The chance is approximately

$$1 - \Phi(1) \approx 0.16.$$

Remember that we don't use continuity correction (the "$\frac{1}{2}$" adjustment) for continuous valued RVs. □

Finally, let's end with something interesting.

Example 11. *(For Fun and Profit)* Suppose that team A (say, the L.A. Lakers) meets team B (say, the Miami Heat) in the NBA finals. To make a little money to pay for college, you're working at the memorabilia stand of team A. Suppose that you will have a profit of two million dollars if team A comes out as the winner, but only break even (zero profit, but no losses) if team B wins. Therefore, you want to hedge your risk by making bets on the outcome of each match, using the valuable life skills you've learned in Math 2b.

Assume that the odds are fair, that is, for every dollar you bet on a team, you will end up with 2 dollars if you win and nothing if you lose. Devise a betting strategy so that no matter the outcome, the combined profit of your gambling and business is always 1 million dollars.

Proof. Think about this for a little bit. There is indeed such a strategy, and it doesn't even involve having infinite amounts of money like the martingale method! There's even a way to do it if the finals are changed to a "best of n" series for arbitrary n.

Have fun with this one. I'll share a solution at the next recitation. Good luck on the midterm! □