Math 20, Spring 2005, Test 2 Solutions

Instructions: Problems 1–7 count 12 points each, while the last problem counts 16 points. You may use a calculator to help with arithmetic, including logs, exponentiation, and factorials (if there is a factorial button).

1. A typesetter makes on average 1 typographical error per 1000 words. A book he is typesetting has on average 300 words per page and is 500 pages long. What is the probability that a random page has 2 or more errors?

Solution 1. We model this with a Poisson distribution with $\lambda = .3$, the expected number of errors per page. Thus, the probability that a page has 0 or 1 errors is $e^{-.3}(1 + .3) \approx .963$. Thus the probability of 2 or more errors is $\approx .037$.

Solution 2. Viewing this as a binomial distribution with $p = .001$ for the probability of a word being typed in error, we are being asked for $1 - b(300, 0, p) - b(300, 1, p)$. This is $1 - (.999)^{300} - (300)(.999)^{299}(.001)$. This also works out to be $\approx .037$.

2. In problem 1, how many errors and how many error-free pages would one expect in the book?

Solution 1. The book has $500 \times 300 = 150000$ words. Thus, the expected number of errors is $1/1000$ times this, or 150. Using the Poisson model, the probability that a page is error free is $e^{-.3}$, so the expected number of error-free pages is $500e^{-.3} \approx 370$.

Solution 2. The first part is done as in Solution 1. For the second part we use the binomial distribution, so the probability for an error-free page is $(.999)^{300}$ and the expected number of such pages is 500 times this, which is ≈ 370.

3. An urn contains 2 gold balls and 3 silver balls. You draw balls at random without replacing them until you've drawn both of the gold balls. Each time you draw a ball you win a dollar if it is gold and lose a dollar if it is silver. What is the expectation for this game?

Let p_j be the probability that we draw the second gold ball on the jth draw. We compute these probabilities for $j = 2, 3, 4, 5$. We have $p_2 = (2/5)(1/4) = 1/10$. To have 3 draws, the picks must be GSG or SGG. The chance for the first is $(2/5)(3/4)(1/3) = 1/10$, and the chance for the second is $(3/5)(2/4)(1/3) = 1/10$. Thus, $p_3 = 1/5$. For 4 draws, the picks must be GSSG, SGSG, or SSGG. The chances for each of these are $(2/5)(3/4)(2/3)(1/2) = 1/10$, $(3/5)(2/4)(2/3)(1/2) = 1/10$, $(3/5)(2/4)(2/3)(1/2) = 1/10$, respectively, so $p_4 = 3/10$. Finally, since the probabilities add to 1, we have $p_5 = 2/5$. Now we find the expected value of the game. This is

$$2 \cdot p_2 + 1 \cdot p_3 + 0 \cdot p_4 + (-1) \cdot p_5 = 0.$$

Here's another way, perhaps more clever, to figure the probabilities. After you finish drawing both gold balls, there are either 0, 1, 2, or 3 silver balls left. Let's then compute the probabilities of picking these numbers of silver balls and never picking a gold ball. That is, we pick all 5 balls and ask how long the run of silver is at the end. This would be the same as the probability of the same run of silver at the start. The chance for 0 is

2/5, and this is p_5. The chance for 1 is $(3/5)(2/4) = 3/10$, and this is p_4. The chance for 2 is $(3/5)(2/4)(2/3) = 1/5$, and this is p_3. And the chance for 3 is $(3/5)(2/4)(1/3) = 1/10$, and this is p_2.

4. Alice and Bob play "heads and tails" (I'm not making this up, it's in the book) where a fair coin is fairly flipped n times. Each time it comes up heads, Alice wins a penny from Bob, and each time it comes up tails, she loses a penny to Bob. Let A be Alice's winnings (which may be negative if she loses money). Find $E(A)$ and $V(A)$.

Clearly $E(A) = 0$ since this holds for one coin flip and expectation is additive. Variance is also additive for independent events, so it suffices to find the variance for one coin flip, and then multiply this by n. This one-coin-flip variance is $(1-0)^2(1/2) + (-1-0)^2(1/2) = 1$. So $V(A) = n$. (Units are "pennies-squared".)

5. What is Chebyshev's inequality? How is it proved?

If X is a numerically valued random variable with mean μ and standard deviation σ, and ϵ is a positive number, then $P(|X-\mu| \geq \epsilon) \leq \sigma^2/\epsilon^2$. (In many of your statements, you just gave this last inequality without introducing the cast of characters involved. I didn't grade off for this, but it is much better to be explicit when letters are introduced.) Proof: Let x run over the values of X. We have

$$\sigma^2 = \sum_x (x-\mu)^2 m(x) \geq \sum_{|x-\mu|\geq\epsilon} (x-\mu)^2 m(x) \geq \epsilon^2 \sum_{|x-\mu|\geq\epsilon} m(x) = \epsilon^2 P(|X-\mu| \geq \epsilon),$$

from which the result follows. Here are details on the above display. The first equality in this chain is essentially the definition of variance, and the last equality is the definition of the probability involved. To get from the first sum to the second sum, the only change is that possibly some positive terms are removed, thus the first sum is greater than or equal to the second one. The next inequality follows since for each x in the second sum, the expression $(x-\mu)^2$ is at least ϵ^2, and the corresponding factor $m(x)$ is positive. In this transition, we also have factored out the constant ϵ^2 from the sum.

6. A fair coin is fairly flipped 10,000 times. What is the approximate probability that it lands heads exactly 4971 times?

Solution 1. We use the approximation theorem, which is part of the central limit theorem. Here we have $\mu = 5000$ and $\sigma = \sqrt{2500} = 50$. Thus, we are $29/50 = .58$ standard deviations below the mean. The answer then is about $(1/50)(1/\sqrt{2\pi})e^{-(-.58)^2/2} \approx .00674$.

Solution 2. This is given by the binomial probability and is exactly $\binom{10,000}{4971}2^{-10,000}$. The binomial coefficient may be approximated with Stirling's formula, giving

$$\frac{\sqrt{2\pi}10,000^{10,000.5}}{\sqrt{2\pi}4971^{4971.5}\sqrt{2\pi}5029^{5029.5}},$$

since the powers of e cancel. The log of this expression is about 6926.473. And the log of $2^{-10,000}$ is about -6931.472. Adding, the log of our probability is about -4.999, so the probability is about $e^{-4.999} \approx .00674$.

7. In problem 6, what is the approximate probability that the coin lands heads fewer than 4971 times?

This may be done with the "Bernoulli tails" method we discussed near the beginning of the course, but an easier and more accurate method is to use the central limit theorem. This probability is approximated by

$$\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{-.6} e^{-t^2/2}\, dt = .5 - \mathrm{NA}(.6).$$

Looking up NA(.6) from the table, we see the answer is about .2743. The number $-.6$ in this solution comes from using that the standard deviation is 50, so that 4970 is .6 standard deviations below the mean. It may be more accurate to "split the difference" between 4970 and 4971 and so say we are $29.5/50 = .59$ standard deviations below the mean. Interpolating in the table (that is taking a number 9/10 of the way from .1915 to .2257), we have that our probability is about $.5 - .2223 = .2777$.

8. Consider a lottery where there is a 0.9 chance of not winning anything, a 0.099 chance of winning \$5, and a 0.001 chance of winning \$250. The lottery ticket costs \$1. What is the expected value of this game? What is the probability of breaking even or better if you buy 100 tickets?

The expected value is

$$(-1)(.9) + 4(.099) + 249(.001) = -.255.$$

(Or one can figure $(0)(.9) + 5(.099) + 250(.001) = .745$, and then subtract 1 for the cost of playing the game.) The mean square is

$$(-1)^2(.9) + 4^2(.099) + 249^2(.001) = 64.485.$$

Thus the variance for one game is

$$64.485 - (-.255)^2 \approx 64.42.$$

Hence the variance for 100 games is about 6442, and the standard deviation about 80.26. The mean for 100 games is -25.5. Thus, breaking even, or 0, is $25.5/80.26 \approx .318$ standard deviations above the mean. The chance for reaching at least this is then about

$$\frac{1}{\sqrt{2\pi}} \int_{.318}^{\infty} e^{-t^2/2}\, dt = .5 - \mathrm{NA}(.318).$$

Interpolating in the table, this is about $.5 - .1246 = .3754$. Using $-.5$ instead of 0 in the above calculation, gives $25/80.26 \approx .311$ standard deviations above the mean, resulting in a final probability of about .3780.

This is the way I expected you to do the problem, but there is a cautionary tale here. The central limit theorem tells what happens as $n \to \infty$. Is $n = 100$ close enough to infinity that the central limit theorem gives useful numbers? We can try to figure the odds on this problem another way. Clearly you will at least break even if you win the \$250 jackpot at least once, or if you win \$5 at least 20 times. The first has probability about $1 - e^{-.1} \approx .095$. The second has probability

$$\sum_{j \geq 20} \binom{100}{j} (.099)^j (.901)^{100-j} \approx .002.$$

(The chance that both should occur is negligible.) Thus, the actual probability of at least breaking even is about .097, which is quite a bit different from what the central limit theorem "tells" us. (I checked out the instructor's solution manual for problem 9.2.12, which is similar, and is one we discussed in class. It too used the central limit theorem and got answers significantly off from the truth. I've informed the authors.) The moral of the story is that the central limit theorem can give a useful approximation for finite problems with large variance when the number of trials is very large. If the variance is not so large, then the number of trials need not be quite so big. Of course, a quantitative version of this rule-of-thumb is called for!