CSCI2244-Randomness and Computation
First Exam with Solutions

March 1, 2018

Each part of each problem is worth 5 points. There are actually two parts to Problem 2, since you are asked to compute two probabilities. Problems 1 and 4 require numerical answers, except as indicated, but as always, you have to show your work: If the right answer is 7, then just '7' is the wrong answer! The calculations are entirely rudimentary, involving arithmetic with one-digit integers for the most part, so you do not need to use a calculator.

1 A Little Box of Chocolates

A little box of chocolates contains five chocolate truffles: two dark chocolates–the kind I like–and three milk chocolates, the kind I don't care for. I close my eyes grab two of the truffles at random.

(a) Model the outcomes of the experiment as the set S of two-element subsets of $\{1, 2, 3, 4, 5\}$, where 1 and 2 represent the dark chocolate truffles. What is $|S|$? (Give an exact numerical answer, as well as an expression involving binomial coefficients.)

Solution. $|S| = \binom{5}{2} = 10.$

(b) Let E be the event 'at least one of the chocolates I pull out is a dark chocolate'. Write the event E explicitly as a subset of S. What is $|E|$?

Solution.

$$E = \{\{1, 2\}, \{1, 3\}, \{1, 4\}, \{1, 5\}, \{2, 3\}, \{2, 4\}, \{2, 5\}\}$$

so $|E| = 7$.

(c) Assume that each outcome in S has equal probability. What is $P(E)$?

Solution. $P(E) = \frac{|E|}{|S|} = 0.7$.

Most common errors: This was supposed to be a 'gift' problem, but surprisingly there were a lot of errors even on the part of students who did well on what I considered harder problems. The most common error was to confuse notation about sets versus ordered pairs. Strictly speaking, writing the solution to (b) as

$$E = \{(1,2),(1,3),(1,4),(1,5),(2,3),(2,4),(2,5)\}$$

is not correct If the error was limited to this and you were consistent about treating 'a subset with two elements' as identical with 'a sorted ordered pair with two elements', I did not take points off for this.

The idea here was that you could solve this problem knowing nothing beyond the definitions. If you wanted to be really fancy about this you could compute the answer to part (b) using the technique for Problem 2, and write

$$|E| = \binom{2}{1} \cdot \binom{3}{1} + \binom{2}{2} \cdot \binom{3}{0} = 2 \cdot 3 + 1 \cdot 1 = 7,$$

but this seems like an instance of the art of making simple things difficult!

2 A Bigger Box of Chocolates

Same problem, scaled up: A bigger box of chocolates contains twenty truffles, 8 of which are dark chocolate, and 12 of which are milk chocolate. I reach in and pull out ten truffles. What is the probability that *exactly* five of them are dark chocolate? What is the probability that *at least* five of them are dark chocolate? Your answer should contain binomial coefficients and summations, but you do not need to evaluate these.

Solution. Exactly 5:

$$\frac{\binom{8}{5} \cdot \binom{12}{5}}{\binom{20}{10}}.$$

At least 5:

$$\frac{1}{\binom{20}{10}} \cdot \sum_{j=5}^{8} \binom{8}{j} \cdot \binom{12}{10-j}.$$

You could also use the complementary probability here:

$$1 - \frac{1}{\binom{20}{10}} \cdot \sum_{j=0}^{4} \binom{8}{j} \cdot \binom{12}{10-j}.$$

Most common error. By *far* the most common error on the exam was to treat this sampling-wthout-replacement problem (hypergeometric distribution) as one of sampling-with-replacement (binomial distribution), leading to the incorrect answer

$$\binom{10}{5} \cdot 0.4^5 \cdot 0.6^5.$$

This evaluates to 0.20, while the correct answer is 0.24, so this is not even a particularly close approximation. Worse still was an odd mix of the two that I observed on a number of papers, writing this expression with $\binom{20}{10}$ rather than $\binom{10}{5}$.

Some students wrote the summation in the second part as $\sum_{j=5}^{10}$ rather than $\sum_{j=5}^{8}$. This means that you are including terms that have binomial coefficients $\binom{8}{9}$ and $\binom{8}{10}$. While we 'officially' defined $\binom{n}{k}$ only in the case where $0 \le k \le n$, in fact $\binom{8}{9}$ has a perfectly reasonable interpretation as the number of of 9-element subsets of an 8-element set. This number is 0, of course, so the formula remains correct when you extend the summation in this way. I accordingly accepted this answer as correct. Other students, presumably uncomfortable with the $\sum$ notation, wrote out all four complicated terms of the sum. I accepted this answer as well, because it is perfectly correct, but make sure you get comfortable with $\sum$!

3 The Chocolate Factory

The Chocolate Factory makes millions of chocolate truffles every day. Approximately one of every thousand chocolates that leave the factory contains a golden ticket. You order a shipment of 2500 chocolates.

(a) What is the probability that your shipment contains NO golden tickets? Give a (nearly) exact answer, based on treating this as a problem of sampling with replacement, along with an easy-to-compute approximation.

Solution. 0.999^{2500} for the exact value, $e^{-2.5}$ for the approximate value, since $0.999 \approx e^{-0.001}$.

(b) What is the probability that your shipment contains at least two golden tickets? Just write the easy-to-compute approximation.

Solution.

$$1 - e^{-\lambda}(1+\lambda) = 1 - 3.5 \cdot e^{-2.5}.$$

($P(X > 1)$ for a Poisson random variable with $\lambda = 2.5$.)

(c) You repeat this experiment many times, and keep track of how many golden tickets are in each shipment. What is the average number of golden tickets that you get in a shipment? Justify your answer—you can cite any result that was taught in class.

Solution. 2.5. You can get this either from the expected value λ of a Poisson distribution, or the expected value np of a binomial distribution with n trials and success probability p.

Common errors. If you treat the problem as sampling with replacement, then the exact distribution is binomial, and the approximate distribution is Poisson. A few students apparently thought that anything with e in it is hard to compute, but I maintain that $e^{-2.5}$ is an awful lot easier to compute than 0.999^{2500}. Be that as it may, it should still be clear which one is the exact answer and which one is the approximation.

There were a lot of correct, but infelicitously expressed, answers for which I did not take off points. For example, I accepted

$$\binom{2500}{0} 0.001^0 \cdot 0.999^{2500}$$

for the exact answer in (a). However, I did take off a point for writing

$$\sum_{j=2}^{\infty} e^{-2.5} \frac{2.5^j}{j!}$$

for the answer in (b), since using the complementary probability here is an essential key to computing the answer easily.

4 I couldn't figure out how to make this problem about chocolate, so it's about dice.

Roll two fair 6-sided dice. Let the random variable X_1 denote the number of spots showing on the first die, and let X_2 be the number of spots showing on the second die. Let $Z = |X_1 - X_2|$; that is, Z is the absolute value of the *difference* between the two dice. (For instance, $Z = 5$ if you roll a 1 and a 6.)

(a) Formally, Z is a real-valued function on the sample space $S = \{1,2,3,4,5,6\} \times \{1,2,3,4,5,6\}$. What is $Z((2,4))$?

Solution. $|2-4| = 2$.

(b) What is $P(Z = 0)$? What is is $P_Z(1)$? What is $F_Z(2)$? (Recall that P_Z denotes the PMF of Z, and F_Z the CDF of Z.)

Solution. There are 6 outcomes $\{(1,1),(2,2),\cdots,(6,6)\}$ for which $Z = 0$, so $P(Z = 0) = \frac{6}{36} = \frac{1}{6}$. There are 10 outcomes

$$\{(j, j+1) : 1 \le j \le 5\} \cup \{(j+1, j) : 1 \le j \le 5\}$$

for which $Z = 1$, so $P_Z(1) = \frac{10}{36} = \frac{5}{18}$. A similar enumeration gives 8 outcomes for which $Z = 2$, so $P_Z(2) = \frac{8}{36} = \frac{2}{9}$, and thus

$$F_Z(2) = \frac{6+10+8}{36} = \frac{2}{3}.$$

(c) Determine all the values of P_Z and sketch the graph of this function. (The preceding problem will get you halfway there.)

Solution. The remaining counts of outcomes are 6,4,2, respectively for $Z = 3, 4, 5$, found just as in the tabulation above for $Z = 1$. This gives the plot in the figure below.

(d) Find $E(Z)$. You can leave the answer as an unevaluated expression. (The numerical answer is about 1.944, which should help tell you if you did the problem correctly.)

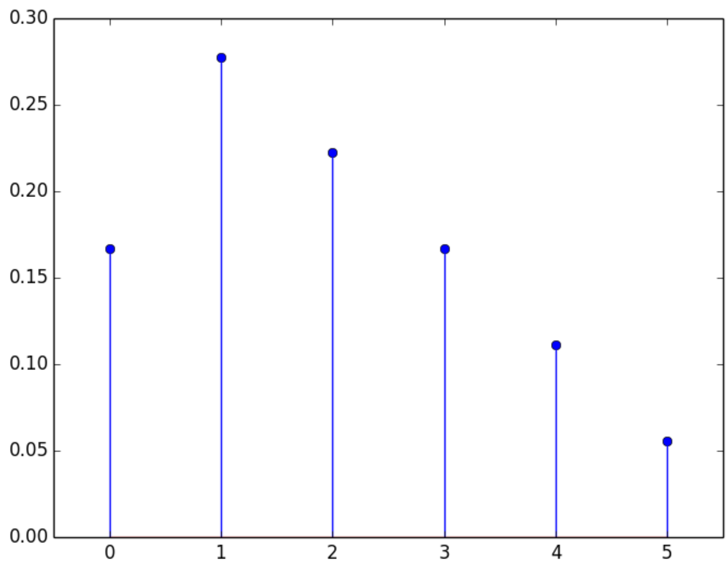

Figure 1: PMF of Z

Solution. Linearity of expectation is not terribly helpful here; there's not really a simpler strategy than going directly to the definition of expected value:

$$E(Z) = \frac{1}{36} \cdot (0 \cdot 6 + 1 \cdot 10 + 2 \cdot 8 + 3 \cdot 6 + 4 \cdot 4 + 5 \cdot 2) = \frac{35}{18}.$$

(e) Let W be the random variable denoting the larger of the values on the two dice. That is

$$W = \max(X_1, X_2).$$

Determine $E(W)$, by applying the handy fact

$$\max(x, y) = \frac{x + y + |x - y|}{2}.$$

You can use the answer to the preceding question, even if you did not solve it.

Solution. Now we can use linearity, along with the simple fact, shown in class, that $E(X_1) = E(X_2) = 3.5$:

$$E(W) = \frac{1}{2} \cdot (E(X_1) + E(X_2) + E(Z)) = \frac{1}{2} \cdot (\frac{7}{2} + \frac{7}{2} + \frac{35}{18}) = \frac{161}{36} = 4.472.$$

Most common incorrect answer. Apart from a lot of arithmetic errors, the biggest issue was students who correctly computed the expected value of W in the last part of the problem by brute force application of the definition, without using linearity and the already-computed value of $E(Z)$. I considered the connection between these two as an essential part of the problem. At least one student did apply the linearity, but took it too far, writing, incorrectly,

$$E(|X_1 - X_2|) = |E(X_1) - E(X_2)| = |3.5 - 3.5| = 0 \text{ WRONG!}.$$

5 How'd You Do?

Aw, it was too *easy*.

If you got 50 or above (26 out of the 35 students who took the exam), you're ready to move on to the more challenging material ahead, but make sure that you understand what you did wrong.

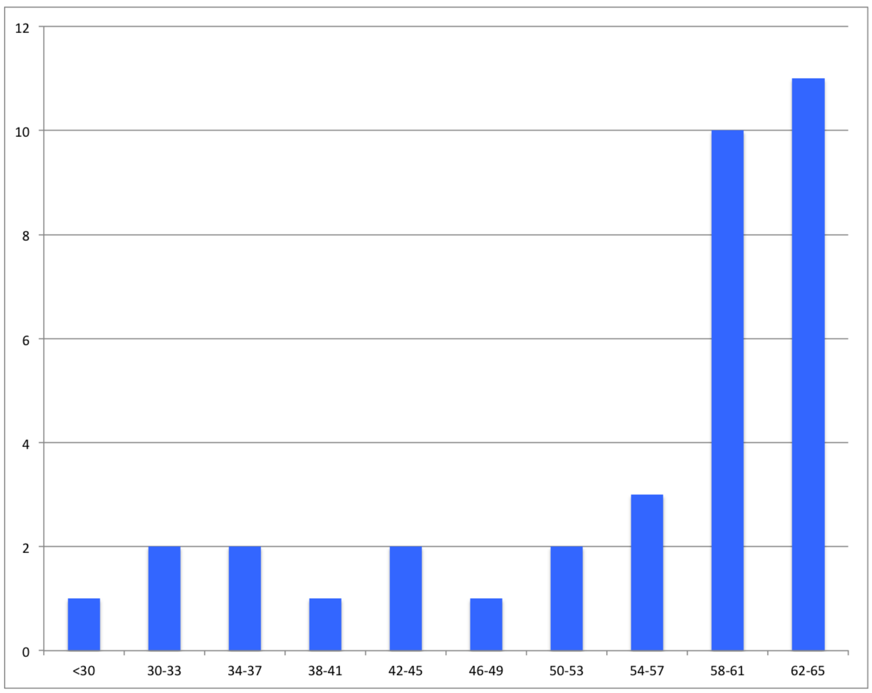

Figure 2: Distribution of grades: The median was 59/65 (91%), the mean 53.3/65 (82%)

If your score was in the 40's, you're probably missing more than one or two fundamental concepts. You'll need to review this material carefully and work similar problems to be sure you've mastered it.

A score below 40 is cause for serious concern. Come talk to me!