November 13, 2007

Math 102 / Core 143 DX — Exam II

Put your answers to all the problems on the numbered sheet, except for the essays (#6 and #7), which go on the back. The blue book is for your scratch work and will NOT be graded.

1. *(25 points)* A regular dodecahedron (a 12-sided solid) has its sides numbered 1 through 12.
   (a) What is the conditional probability, given that a roll of this 12-sided die is at least 6, that it is even?
   (b) What is the probability that, on 2 rolls of this die, the first is even and the second is less than 3 (i.e., 1 or 2)?
   (c) What is the probability that, on 2 rolls of this die, the first is even or the second is less than 3 (or both)?
   (d) What is the probability that, on 5 rolls, at least one is less than 3?
   (e) What is the probability that, on 8 rolls, at least 7 are less than 3?

2. *(10 points)* This question refers to the same dodecahedral die as in Question 1. The standard deviation of the numbers from 1 to 12 is about 3.45. Using this value, find the approximate probability that the average of 100 rolls of this die will lie between 6 and 7.

3. *(15 points)* The deuces, treys, and 4's through 10's are selected from a staight (=bridge=poker) deck, making a 36-card deck in which the average card value is 6. We could (I) pick 20 cards from the deck (replacing the card and shuffling between picks) or (II) pick 40 cards (under the same conditions). In which of these two scenarios, I or II, would the following be more likely?
   (a) The average value of the cards chosen is at least 8.
   (b) The average value of the cards chosen is at least 5.
   (c) The sum of the cards chosen is at least 8 more than 6 times the number of cards chosen.
   (d) The sum of the cards chosen is at least 8 times the number of cards chosen.
   (e) More than 20% of the cards chosen are 6's (exactly). (Warning: This should change your "box model" from the one you were using in (a) through (d).)

4. *(6 points)* The test for the disgusting disease lantzeria has a sensitivity of 85% (i.e., if you have it, the test will say you have it 85% of the time) and a specificity of 90% (i.e., if you don't have it, the test will say you don't 90% of the time). Suppose that 5% of the population have this disease, and that you test positive for it: what is the probability that you actually have it? (Steps in the solution: 1. Percent of the population who are sick and test positive. 2. Percent who are well and still test positive. 3. Percent who test positive. 4. Percent who are sick and test positive divided by percent who test positive.)

5. *(24 points)* For each of the following data sources, should we expect the data to follow a normal distribution, a uniform distribution (all outcomes equally frequent), or neither. For the "neither" responses, give a rough sketch.
   (a) salaries in a large population
   (b) averages salaries of large samples from a large population
   (c) rolls of an icosahedral (20-sided) die, with the numbers 1 through 20 on its faces
   (d) heights in a large population

6. *(10 points)* Relative to the article "Cancer coincidence? An overwhelmed state office tries to connect disease clusters to pollution – but some say the agency is wasting its time", by Beth Daley: If each census tract is thought of as one survey, what is the proportion of tracts that you would expect to be significantly higher than normal? (Later in the course, "significantly" becomes a technical term, but here it means only your general understanding of the word.)

7. *(10 points)* Relative to the article "What is the chance of your being guilty?" by John Kay: What does Kay mean by the "prosecutor's fallacy" and how can this mislead jurors?

Some possibly useful formulas:

$$\frac{n!}{k!(n-k)!}\,p^k(1-p)^{n-k} \qquad \sqrt{\frac{\text{total population} - \text{sample size}}{\text{total population} - 1}}$$

$$\sigma = (\text{larger} - \text{smaller})\sqrt{(\text{fraction with larger})(\text{fraction with smaller})} = \sqrt{\text{average of } (x-\overline{x})^2}$$

$$EV \text{ of sum (or count)} = n \cdot (AV \text{ of box}) \qquad SE \text{ of sum (or count)} = \sqrt{n} \cdot (SD \text{ of box})$$

$$EV \text{ of avg (or \%)} = AV \text{ of box} \qquad SE \text{ of avg (or \%)} = \frac{SD \text{ of box}}{\sqrt{n}}$$

Normal table (Area between $-z$ and z)									
z	Area(%)	z	Area(%)	z	Area(%)	z	Area(%)	z	Area(%)
0.0	0.0	0.9	63.19	1.8	92.81	2.7	99.31	3.6	99.968
0.05	3.99	0.95	65.79	1.85	93.57	2.75	99.4	3.65	99.974
0.1	7.97	1	68.27	1.9	94.26	2.8	99.49	3.7	99.978
0.15	11.92	1.05	70.63	1.95	94.88	2.85	99.56	3.75	99.982
0.2	15.85	1.1	72.87	2	95.45	2.9	99.63	3.8	99.986
0.25	19.74	1.15	74.99	2.05	95.96	2.95	99.68	3.85	99.988
0.3	23.58	1.2	76.99	2.1	96.43	3	99.73	3.9	99.99
0.35	27.37	1.25	78.87	2.15	96.84	3.05	99.771	3.95	99.992
0.4	31.08	1.3	80.64	2.2	97.22	3.1	99.806	4	99.9937
0.45	34.73	1.35	82.3	2.25	97.56	3.15	99.837	4.05	99.9949
0.5	38.29	1.4	83.85	2.3	97.86	3.2	99.863	4.1	99.9959
0.55	41.77	1.45	85.29	2.35	98.12	3.25	99.885	4.15	99.9967
0.6	45.15	1.5	86.64	2.4	98.36	3.3	99.903	4.2	99.9973
0.65	48.43	1.55	87.89	2.45	98.57	3.35	99.919	4.25	99.9979
0.7	51.61	1.6	89.04	2.5	98.76	3.4	99.933	4.3	99.9983
0.75	54.67	1.65	90.11	2.55	98.92	3.45	99.944	4.35	99.9986
0.8	57.63	1.7	91.09	2.6	99.07	3.5	99.953	4.4	99.9989
0.85	60.47	1.75	91.99	2.65	99.2	3.55	99.961	4.45	99.9991

Name: ______________________________

Your answers to Exam II

1. (a)
   (b)
   (c)
   (d)
   (e)
2.
3. (a)
   (b)
   (c)
   (d)
   (e)
4.
5. (a)
   (b)
   (c)
   (d)
6. and 7. on back.

6.

7.

Math 102 / Core 143 — Solutions to Exam II

1. (a) Because four of the seven numbers 6 through 12 are even, the probability is 4/7.
   (b) The probability that the first is even is 1/2, and the probability that the second is less than 3 is $2/12 = 1/6$, and the events are independent, so the probability that both happen is $(1/2)(1/6) = 1/12$.
   (c) $1/2 + 1/6 - 1/12 = 7/12$.
   (d) This is the complement of the event "getting no number less than 3 on 5 rolls", so: $1 - (5/6)^5 \approx 0.6$.
   (e) $C(8,7)(1/6)^7(5/6)^1 + C(8,8)(1/6)^8 = [8(5)+1]/6^8 \approx 0.000024$.

2. The averages of 100 rolls have a nearly normal distribution, with an EV of 6.5 (the average of the possible values) and an SE of about $3.45/\sqrt{100} = .345$, so, using the normal table,

$$P(6 < \text{avg} < 7) = P\left(\frac{6-6.5}{.345} < z < \frac{7-6.5}{.345}\right) = P(-1.45 < z < 1.45) \approx .85\ .$$

3. (a) There is more variability in the average with smaller samples so the average is more likely to be above 8 with 20 cards: I.
   (b) Because 5 is below the expected value of the average, and the sample average tends to be closer to the expected value of the average when the samples are larger, this is more likely with II.
   (c) Because there is more variability in the sum with a larger sample, this is more likely with II.
   (d) Despite the fact that it is phrased in terms of sum, this is really the same as part (a), so the answer is I.
   (e) The "deck" is now essentially 1's and 0's, meaning 6's (of which there are 4) and non-6's (of which there are 32). With this box model, we should expect only $1/9 \approx 11\%$ of the cards to be 6's, so 20% is higher than expected; and variability in percentage is larger with small samples, so this is more likely with I.

4. Of the 95% of the population who do not have the disease, the test will erroneously say that 10%, or 9.5% of the population, actually have it. Of the 5% who have the disease, the test will say that 85% of them, or 4.25% of the population, have it. Thus, of the 13.75% of the population who test positive, only 4.25% actually have the disease: $4.25\%/13.75\% \approx 31\%$. So you only have about a 31% chance of actually having the disease even though you tested positive.

5. (a) We know this is usually neither normal nor uniform; it has a long tail (skew) to the right.
   (b) Averages of large samples, even from a non-normal population, are normally distributed, by the Central Limit Theorem.
   (c) Uniform distribution (at least roughly).
   (d) Normally distributed (the cumulative effect of many small influences, usually genetic). Other answers are acceptable, for example taking into account that men's and women's average heights differ.

6. Just because of chance variation, some census tracts will have higher than average cases of cancer, and some will have lower. In fact, there will be some that have very high incidences,

and of course it is possible that there is some carcinogen in the environment of such a tract. But some of the high-incidence tracts may be high-incidence just by accident — even some that seem significantly (i.e., meaningfully) high, say 80 or 100 percent above the average rate. Deciding whether it is chance or some undiscovered carcinogen is what overwhelms that office. What level seems "significant" to you is of course your own decision. But for example, if you thought 80% above average is significant, then you should be guessing how much variability there is in the cancer rate, to decide how often that high a cancer rate appears.

7. Here is the relevant section of the article:

   > The prosecutor's fallacy is the assertion that, because the story before the court is highly improbable, the defendant's innocence is equally improbable. But all accounts of events in high-profile legal cases are highly improbable. That is why they are high-profile legal cases. The courts do not hear reports of happy families and normal behaviour. Their services are required only for bizarre and unlikely incidents — such as the saga of Nicole and O.J. Simpson and the tragedy of the Clark family.
   >
   > So juries are not asked to decide whether the events before them are out of the run of everyday experience. They are asked to decide on the most probable explanation of improbable events.

   We might rephrase this as follows: The prosecutor's fallacy is to encourage the jury to assume that the probability of the offense being committed is the same as the probability that the defendant is somehow associated with the offense (by the investigation) but did not commit it. Kay points out that the offense <u>is</u> improbable (fortunately for civilization), but the jury should be trying to judge the <u>conditional</u> probability that, given that the offense has occurred (as it has), the defendant was not the one who committed it.