April 13, 2010

Math 102 / Core 143 BX — Exam II

An unsimplified answer like $12\sqrt{3.51} + 6/7$ is usually worth more than 23.3, because it is easier to understand where it came from.

1. *(25 points)* A regular icosahedron (a 20-sided solid) has its sides numbered 1 through 20.
   (a) What is the conditional probability, given that a roll of this 20-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 <u>or</u> the second is less than 3 (or both statements are true)?
   (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 6 are less than 3?

2. *(10 points)* Which term describes each of the following probability distributions, related to thousands of rolls of the 20-sided die in Question 1? The choices are: "(approximately) normal", "(approximately) uniform" and "other"; they may be used more than once or not at all.
   (a) The numbers rolled
   (b) The averages of every 100 rolls of the die
   (c) The products of every 100 rolls of the die

3. *(15 points)* The aces, deuces, treys, and 4's through 9's are selected from a straight (= bridge = poker) deck, making a 36-card deck in which the average card value is 5. 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 or neither, would the following be more likely?
   (a) The average value of the cards chosen is at least 4.
   (b) The average value of the cards chosen is at least 8.
   (c) The sum of the cards chosen is at least 8 more than 5 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. *(25 points)* A person stands in a shopping mall. She stops 400 people and asks each how many times per month they go out for hamburgers, and which of two national hamburger chains they prefer. For the 400, the average number of burger runs per month is 20, with an SD of 15, and 250 say they prefer MacTavish's hamburgers to Burger Empire's.

   (a) What is a 95% confidence interval for the average number of visits to a hamburger restaurant for a person in that city?

   (b) How many people would she have had to question to reduce the width of her confidence interval by a factor of 3 (assuming a similar average and SD in her sample)?

   (c) What is an 80% confidence interval for the fraction of the population who prefer MacTavish's burgers to Burger Empire's?

5. *(15 points)* Suppose we have a large population whose incomes we know; they do not follow the normal curve (as with most income distributions, they are strongly skewed to the right), but they average \$35,000, with an SD of \$24,000. From this population we randomly choose 900 people and average their incomes. We repeat this 400 times.

   (a) Should we expect the distribution of these 400 averages to be close to a normal curve? Why or why not?

   (b) What should we expect the average of the 400 averages to be?

   (c) How much variability should we expect in the 400 averages? (Answer in the form "________ percent of them are likely to be in the range ________ to ________ .")

   (d) If instead of 400 times we did this only 100 times, how would the answers to (a), (b) and (c) change, if at all?

   (e) If instead of using samples of 900, we used samples of 1600 (but again 400 samples), how would the answers to (a), (b) and (c), if at all?

6. *(10 points)* Related to the articles "Why Doctors Hate Science" and "The Myth of Early Detection", both by Sharon Begley: After these columns were written, the FDA changed its recommendations on how frequently women should get mammograms. In view of these columns, do you think they recommended more or less frequently, and why?

Some Possibly Useful Formulas:

$$\frac{n!}{k!(n-k)!}\, p^k (1-p)^{n-k}$$

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

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

$$SD = (\text{larger} - \text{smaller})\sqrt{(\text{fraction with larger})(\text{fraction with smaller})}$$

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

Answers: **Name:** ______________________________

1. (a)

   (b)

   (c)

   (d)

   (e)

2.

3. (a)

   (b)

   (c)

   (d)

   (e)

4. (a)

   (b)

   (c)

5. (a)

   (b)

   (c)

   (d)

   (e)

Answer Question 6 on the back of this page.

6.

Math 102 / Core 143 BX — Solutions to Exam II

1. (a) Because 8 of the 15 numbers 6 through 20 are even, the probability is 8/15.
   (b) The probability that the first is even is 1/2, and the probability that the second is less than 3 is $2/20 = 1/10$, and the events are independent, so the probability that both happen is $(1/2)(1/10) = 1/20$.
   (c) $1/2 + 1/10 - 1/20 = 11/20$.
   (d) This is the complement of the event "getting no number less than 3 on 5 rolls", so: $1 - (9/10)^5 \approx 0.41$.
   (e) $C(8,6)(1/10)^6(9/10)^2 + C(8,7)(1/10)^7(9/10)^1 + C(8,8)(1/10)^8 = [(8 \cdot 7/2 \cdot 1)9^2 + (8)9 + 1]/10^8 = 0.00002341$.

2. (a) (approximately) uniform. (b) (approximately) normal. (c) other.

3. (a) There is more variability in the average with smaller samples, so the average is more likely to be at least 4 with 40 cards: II.
   (b) Because 8 is above 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 I.
   (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 (b), 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. (a) Roughly $20 \pm 2(15/\sqrt{400}) = 20 \pm 1.5$ visits per month.
   (b) To reduce the size of the interval to 20 ± 0.5 visits per month, she would have had to interview 9 times as many people, or 3600.
   (c) Using the normal table for an area of 80%:

$$\frac{250}{400} \pm 1.3 \left(\frac{(1-0)\sqrt{\frac{250}{400} \cdot \frac{150}{400}}}{\sqrt{400}} \right) = 62.5\% \pm 3.1\% \ .$$

5. (a) Yes: By the Central Limit Theorem, no matter what distribution you start with, the average of many draws from the distribution will have a distribution that is close to a normal curve.
   (b) The EV for the average, i.e., the average income of the population, \$35,000.
   (c) The SE for the average is the SD of the population divided by the square root of the number in the sample: \$24,000/30 = \$800. So we should expect about 68% of the averages to be between \$34,200 and \$35,800. (We could also say that we expect 95% are between \$33,400 and \$36,600.)

(d) None of the answers would change: The 400 did not enter into these calculations; we were just looking at 400 numbers from the distribution of averages.

(e) The 900 did enter into the computation in (c): The new SE for the average is \$24,000/40 = \$600, so about 68% of the averages are between \$34,400 and \$35,600.

6. The FDA advised fewer screenings, at longer intervals, because early detection often meant unnecessary and possibly dangerous treatment, like removal of tumors that were not threatening.