Due Monday July 25 at 1:59PM

1. **Box of marbles (2/3/5 points)**

   You are given two boxes: one of them containing 900 red marbles and 100 blue marbles, the other one contains 500 red marbles and 500 blue marbles.

   (a) If we pick one of the boxes randomly, and pick a marble what is the probability that it is blue?

   (b) If we see that the marble is blue, what is the probability that it is chosen from box 1?

   (c) Suppose we pick one marble from box 1 and without looking at its color we put it aside. Then we pick another marble from box 1. What is the probability that the second marble is blue?

2. **Bayesian Inference and Pancakes (3/3/3/3/3 points)**

   Vince is making golden-brown pancakes and you are hungry!

   (a) Vince serves up a stack of 3 pancakes, but he forgot to butter the pan! Pancake A is perfect (golden-brown on both sides), pancake B is burnt on one side, and pancake C is burnt on both sides. The top of the stack is burnt. What's the probability that the other side of the top pancake is also burnt? Justify your answer.

   (b) Vince agrees that a burnt pancake on top of the stack looks un-appetizing, and suggests flipping the stack over. In the same situation as before, what's the probability that the pancake side touching the plate is burnt?

   (c) Suppose Vince makes a stack of n pancakes such that x pancakes are burnt on both sides and y pancakes are burnt on one side. If the top of the stack is burnt, what's the probability that the other side of the top pancake is also burnt? What if the top of the stack is golden-brown? Justify your answer.

   (d) You asked for chocolate chips, so Vince adds lots of chocolate chips to the batter. He makes a stack of m pancakes next to the stack of n pancakes from before. However, the k-th pancake ($1 \leq k \leq m$) in the new stack only has a k/m chance of having chocolate chips (independent from the rest of the pancakes). If you choose a pancake randomly from either stack, what's the probability that you get chocolate chips?

   (e) Vince realizes that the top few pancakes in the new stack don't really have chocolate chips in them. He shifts the top 10 pancakes from that stack (those with the smallest chance of chocolate chips) to the old stack. Given you randomly choose a pancake and it has chocolate chips, what's the probability it came from the new stack?

3. **Fundamentals (2/2/2/2 points)**

   True or false? For the following statements, provide either a proof or a simple counterexample. Let X, Y, Z be arbitrary random variables.

(a) If (X,Y) are independent and (Y,Z) are independent, then (X,Z) are independent.

(b) If (X,Y) are dependent and (Y,Z) are dependent, then (X,Z) are dependent.

(c) Assume X is discrete. If $\mathrm{Var}(X) = 0$, then X is a constant.

(d) $\mathbf{E}[X]^4 \leq \mathbf{E}[X^4]$

4. **Like a Rolling Die (2/2/2/2/2/2 points)**

   Suppose you roll a fair die n times. What is the expectations of each of the following random variables?

   (a) A is the random variable that denotes the sum of the numbers in those rolls.

   (b) B is the random variable that denotes the maximum number in the those rolls.

   (c) C is the random variable that denotes the sum of the largest two numbers in the first three rolls.

   (d) D is the random variable that denotes the number of multiples of three in those rolls.

   (e) E is the random variable that denotes the number of faces which fail to appear in those rolls.

   (f) F is the random variable that denotes the number of distinct faces that appear in those rolls.

5. **Linearity (4/4/4 points)**

   Solve each of the following problems using linearity of expectation. Explain your methods clearly.

   (a) In an arcade, you play game A 10 times and game B 20 times. Each time you play game A, you win with probability $1/3$ (independently of the other times), and if you win you get 3 tickets (redeemable for prizes), and if you lose you get 0 tickets. Game B is similar, but you win with probability $1/5$, and if you win you get 4 tickets. What is the expected total number of tickets you receive?

   (b) A monkey types at a 26-letter keyboard with one key corresponding to each of the lower-case English letters. Each keystroke is chosen independently and uniformly at random from the 26 possibilities. If the monkey types 1 million letters, what is the expected number of times the sequence "book" appears?

   (c) A building has n floors numbered $1, 2, \ldots, n$, plus a ground floor G. At the ground floor, m people get on the elevator together, and each gets off at a uniformly random one of the n floors (independently of everybody else). What is the expected number of floors the elevator stops at (not counting the ground floor)?

6. **Runs (5 points)**

   Suppose I have a biased coin which comes up heads with probability p, and I flip it n times. A "run" is a sequence of coin flips all of the same type, which is not contained in any longer sequence of coin flips all of the same type. For example, the sequence "HHHTHH" has three runs: "HHH," "T," and "HH."

   Compute the expected number of runs in a sequence of n flips.

7. **Coupon Collection (6 points)**

   Suppose you take a deck of n cards and repeatedly perform the following step: take the current top card and put it back in the deck at a uniformly random position. (I.e., the probability that the card is placed in any of the n possible positions in the deck — including back on top — is $1/n$.) Consider the card that starts off on the bottom of the deck. What is the expected number of steps until this card rises to the top of the deck? (Hint: Let T be the number of steps until the card rises to the top. We have $T = T_n + T_{n-1} + \cdots + T_2$, where the random variable T_i is the number of steps until the bottom card rises from position i to position $i-1$. Thus, for example, T_n is the number of steps until the bottom card rises off the bottom of the deck, and T_2 is the number of steps until the bottom card rises from second position to top position. What is the distribution of T_i?) (More hints: You may use the fact that $\sum_{i=1}^{n} \frac{1}{i} \approx \ln n$.)

8. **Markov's Inequality and Chebyshev's Inequality (2/2/2/2/2/2 points)**

   A random variable X has variance $\mathrm{Var}(X) = 9$ and expectation $\mathbb{E}(X) = 2$. Furthermore, the value of X is never greater than 10. Given this information, provide either a proof or a counterexample for the following statements.

   (a) $\mathbb{E}(X^2) = 13$.
   (b) $\Pr[X = 2] > 0$.
   (c) $\Pr[X \geq 2] = \Pr[X \leq 2]$.
   (d) $\Pr[X \leq 1] \leq 8/9$.
   (e) $\Pr[X \geq 6] \leq 9/16$.
   (f) $\Pr[X \geq 6] \leq 9/32$.

9. **Casino wins (2/2/3/3 points)**

   A gambler plays 120 hands of draw poker, 60 hands of black jack, and 20 hands of stud poker per day. He wins a hand of draw poker with probability $1/6$, a hand of black jack with probability $1/2$, and a hand of stud poker with probability $1/5$. Assume the outcomes of the card games are mutually independent.

   (a) What is the expected number of hands the gambler wins in a day?
   (b) What is the variance in the number of hands won per day?
   (c) What would the Markov bound be on the probability that the gambler will win 108 hands on a given day?
   (d) What would the Chebyshev bound be on the probability that the gambler will win 108 hands on a given day?

10. **Those 3407 Votes (2/3/5 points)**

    In the aftermath of the 2000 US Presidential Election, many people have claimed that unusually large number of votes cast for Pat Buchanan in Palm Beach County are statistically highly significant, and thus of dubious validity. In this problem, we will examine this claim from a statistical viewpoint.

    The total percentage votes cast for each presidential candidate in the entire state of Florida were as follows:

Gore	Bush	Buchanan	Nader	Browne	Others
48.8%	48.9%	0.3%	1.6%	0.3%	0.1%

    In Palm Beach County, the actual votes cast (before the recounts began) were as follows:

Gore	Bush	Buchanan	Nader	Browne	Others	Total
268945	152846	3407	5564	743	781	432286

To model this situation probabilistically, we need to make some assumptions. Let's model the vote cast by each voter in Palm Beach County as a random variable X_i, where X_i takes on each of the six possible values (five candidates or "Others") with probabilities corresponding to the Florida percentages. (Thus, e.g., $\Pr[X_i = \text{Gore}] = 0.488$.) There are a total of $n = 432286$ voters, and their votes are assumed to be mutually independent. Let the r.v. B denote the total votes cast for Buchanan in Palm Beach County (i.e., the number of voters i for which $X_i = \text{Buchanan}$).

(a) Compute the expectation $\mathbf{E}[B]$ and the variance $\text{Var}(B)$.

(b) Use Chebyshev's inequality to compute an *upper bound* b on the probability that Buchanan receives at least 3407 votes, i.e., find a number b such that

$$\Pr[B \geq 3407] \leq b.$$

Based on this result, do you think Buchanan's vote is significant?

(c) Suppose that your bound b in part (b) is exactly accurate, i.e., assume that $\Pr[X \geq 3407]$ is exactly equal to b. [*In fact the true value of this probability is much smaller*] Suppose also that all 67 counties in Florida have the same number of voters as Palm Beach County, and that all behave independently according to the same statistical model as Palm Beach County. What is the probability that in *at least one* of the counties, Buchanan receives at least 3407 votes? How would this affect your judgment as to whether the Palm Beach tally is significant?