MATH/STAT 341: PROBABILITY: SPRING 2015
COMMENTS ON HW PROBLEMS

STEVEN J. MILLER (SJM1@WILLIAMS.EDU): MATH/STAT 341, SPRING 2015

ABSTRACT. A key part of any math course is doing the homework. This ranges from reading the material in the book so that you can do the problems to thinking about the problem statement, how you might go about solving it, and why some approaches work and others don't. Another important part, which is often forgotten, is how the problem fits into math. Is this a cookbook problem with made up numbers and functions to test whether or not you've mastered the basic material, or does it have important applications throughout math and industry? Below I'll try and provide some comments to place the problems and their solutions in context.

Date: November 6, 2018.

1. HW #2: DUE SEPTEMBER 14, 2018

1.1. **Assignment.** First assignment:

#1: Section 1.2: Modify the basketball game so that there are 2013 players, numbered 1, 2, ..., 2013. Player i always gets a basket with probability $1/2^i$. What is the probability the first player wins?

#2: Section 1.2: Is the answer for Example 1.2.1 consistent with what you would expect in the limit as c tends to minus infinity? (Note there is a typo in the book.)

#3: Section 1.2: Compute the first 42 terms of 1/998999 and comment on what you find; you may use a computer (but Mathematica or some program like that is probably better!).

#4: Section 2.2.1: Find sets A and B such that $|A| = |B|$, A is a subset of the real line and B is a subset of the plane (i.e., $\mathbb{R}^2$) but is not a subset of any line.

#5: Section 2.2.1: Write at most a paragraph on the continuum hypothesis.

#6: Section 2.2.2: Give an example of an open set, a closed set, and a set that is neither open nor closed (you may not use the examples in the book); say a few words justifying your answer.

#7: Section 2.3: Give another proof that the probability of the empty set is zero.

#8: Find the probability of rolling exactly k sixes when we roll five fair die for k = 0, 1, ..., 5. Compare the work needed here to the complement approach in the book.

#9: If f and g are differentiable functions, prove the derivative of $f(x)g(x)$ is $f'(x)g(x) = f(x)g'(x)$. Emphasize where you add zero.

1.2. **Solutions.** First assignment:

#1: Section 1.2: Modify the basketball game so that there are 2013 players, numbered 1, 2, ..., 2013. Player i always gets a basket with probability $1/2^i$. What is the probability the first player wins?

Solution: There is a very elegant way of solving this. We only care about the probability of the first person winning. Thus, we may group persons 2 through 2013 as a team. The probability the first person makes a basket is 1/2, and the probability the first person misses is 1 - 1/2 = 1/2. What about the second person? The probability they miss is $1 - 1/2^2$, and in general the probability the k^{th} person misses is $1 - 1/2^k$. Thus the probability that *everyone* misses their first shot is $\prod_{k=1}^{2013}(1 - 1/2^k)$; if we call that product r we find $r \approx 0.288788$.

We now argue as in the book. If x is the probability the first person wins, then $x = \frac{1}{2} + rx$ (if everyone misses their first shot, then the first player has the ball and it's like the game has just begun; by definition the probability they win in this configuration is just x). Solving we find $x = \frac{1/2}{1-r} \approx .703025$.

When solving a problem, it's always good to check and see if the answer is reasonable. Our answer is between 1/2 and 1. Clearly the first person can't have a greater than 100% chance of winning; further, the odds must be at least 50%, as player 1 shoots first and hits half their shots. Thus our answer passes the smell test and is reasonable.

A major theme of the class is to write simple code to see if your answer is reasonable. Here is an uncommented Mathematica code. Try to figure out the logic (hopefully the comments help!).

```
hoops[num_] := Module[{},
   win = 0; (* keeps track of wins *)
   For[n = 1, n <= num, n++, (* loops from game 1 to game num *)
    { (* starts the n loop *)

     basket = 0; (* set basket to 0, stop game when someone gets one *)
     While[basket == 0, (* do stuff below while no basket made *)
      { (* start the basket loop, keep shooting till someone gets it *)
       For[k = 1, k <= 2013, k++, (* goes through all 2013 people *)
        {
         x = Random[]; (* chooses a random number uniformly in [0,1]*)
         If[x <= 1/2^k, basket = 1]; (*player k shoots, if x < 1/2^k basket!*)
         If[basket == 1 && k == 1, win = win + 1]; (*if basket and k=1, 1st player wins*)
         If[basket == 1, k = 2222]; (* no matter what, if basket made stop game *)
         }]; (* end k *)
```

```
      }]; (* end of while *)

    }]; (* end n *)

  Print["We played ", num, " games."]; (*says how many games played)
  Print["Percent player one won: ", 1.0 win/num]; (*says percentage player 1 won*)
  ];
(* end module *)
```

Typing Timing[hoops[100000]] (which plays the game 100,000 times and records how long it takes to run), we observe the first player winning with probability 0.70203 (it took 5.64 minutes to run), which supports our prediction of $.703025$.

#2: Section 1.2: Is the answer for Example 1.2.1 consistent with what you would expect in the limit as c tends to minus infinity? (Note there is a typo in the book.)
Solution: Yes. The claimed answer is

$$\int_0^\pi e^{cx}\cos x dx \ = \ -\frac{ce^{\pi c}+c}{c^2+1}.$$

As $c \to -\infty$ the integral tends to zero because the cosine factor is bounded but the exponential function rapidly approaches zero. The right hand side looks like $-c/(c^2+1)$ for c large and negative, which also tends to zero.

#3: Section 1.2: Compute the first 42 terms of 1/998999 and comment on what you find; you may use a computer (but Mathematica or some program like that is probably better!).
Solution: Type `SetAccuracy[1/998999, 50]` in Mathematica or online at WolframAlpha (go to `http://www.wolframalpha.com/`). This yields

$$.000001001002003005008013021034055089144233377610\dots.$$

Notice the Fibonacci numbers! This is not a coincidence; we'll see why this is true when we get to generating functions. For more on this see *An Unanticipated Decimal Expansion* by Allen Schwenk in Math Horizons (September 2012), available online (you may need to move forward a few pages) at
`http://digitaleditions.walsworthprintgroup.com/publication/?i=123630&p=3`.

#4: Section 2.2.1: Find sets A and B such that $|A| = |B|$, A is a subset of the real line and B is a subset of the plane (i.e., $\mathbb{R}^2$) but is not a subset of any line.
Solution: There are many solutions. An easy one is to let $B = \{(a,a^2) : a \in A\}$ for any set A with at least 3 points (if A had just two points then we would get a line).

#5: Section 2.2.1: Write at most a paragraph on the continuum hypothesis.
Solution: The continuum hypothesis concerns whether or not there can be a set of cardinality strictly larger than that of the integers and strictly smaller than that of the reals. As the reals are essentially the powerset of the integers, the question is whether or not there is a set of size strictly between $\mathbb{N}$ and $\mathcal{P}(\mathbb{N})$. Work of Kurt Gödel and Paul Cohen proved the continuum hypothesis is independent of the other standard axioms of set theory. See `http://en.wikipedia.org/wiki/Continuum_hypothesis`.

It's interesting to think about whether or not it *should* be true. For example, if A is a finite set with n elements, then $|A| = n$ but $|\mathcal{P}(A)| = 2^n$. Note that $|A| < |\mathcal{P}(A)|$; in fact, as n increases there are *many* sets of size strictly between A and $\mathcal{P}(A)$. Should something similar hold for $|\mathbb{N}|$ and $|\mathcal{P}(\mathbb{N})|$? Note $\mathbb{N}$ is the smallest infinity, while $\mathcal{P}(\mathbb{N})$ has the same cardinality as the real numbers (to see this, consider binary expansions of numbers in $[0,1]$, and this is essentially the same as $\mathcal{P}(\mathbb{N})$, as taking integer k corresponds to having a 1 in the 2^{-k} digit of a base 2 expansion). This is a nice example where the infinite case may have a very different behavior than the finite case – can you find another such example?

#6: Section 2.2.2: Give an example of an open set, a closed set, and a set that is neither open nor closed (you may not use the examples in the book); say a few words justifying your answer.
Solution: An open set in the plane would be $\{(x,y) : 0 < x < 1\}$; this is a vertical strip. Given any point (x_0, y_0) in the set, the ball of radius $r = \frac{1}{2}\min(x_0, 1-x_0)$ is entirely contained in the strip. The set becomes closed if we include both vertical lines, and is neither open nor closed if we include only one of the vertical lines.

#7: Section 2.3: Give another proof that the probability of the empty set is zero.
Solution: As $\emptyset \cup \emptyset = \emptyset$, we have

$$\mathrm{Prob}(\emptyset) \quad = \quad \mathrm{Prob}(\emptyset \cup \emptyset) \ = \ \mathrm{Prob}(\emptyset) + \mathrm{Prob}(\emptyset) \ = \ 2\mathrm{Prob}(\emptyset).$$

Hence $\mathrm{Prob}(\emptyset) = 0$.

#8: Find the probability of rolling exactly k sixes when we roll five fair die for k = 0, 1, ..., 5. Compare the work needed here to the complement approach in the book.
Solution: Some k are straightforward: if $k = 0$ then the answer is $(5/6)^5$ as we must get a non-six each time. Similarly if $k = 5$ the answer is $(1/6)^5$ as we need to get a six each time.

This leaves us with $k = 1, 2$ and 3. There are only five 'ways' to roll five die and get exactly one six; letting $*$ denote a roll that isn't a six, the possibilities are $6 * * * *$, $*6 * **$, $* * 6 * *$, $* * *6*$, and $* * * * 6$. Each of these events has probability $(1/6)^1(5/6)^4$ (there is a one in six chance of rolling a six, which must happen for one of the five rolls, and there is a five out of six chance of rolling a non-six and that must happen four times). Thus the probability of rolling exactly one 6 is just $5 \cdot (1/6)(5/6)^4 = 3125/7776$. Similarly the probability of exactly four 6s is $5 \cdot (1/6)^4(5/6) = 25/7776$.

What about $k = 2$? There are now 10 ways to roll exactly two 6s. It's important to enumerate them in a good way so we don't miss anything. We start with all the ways where the *first* six rolled is from the first die. After we exhaust all those, we then turn to all the ways where the *first* six rolled is from the second die, and so on. Again letting $*$ denote a non-6, we find

- $66 * **$, $6 * 6 * *$, $6 * *6*$, $6 * * * 6$,
- $*66 * *$, $*6 * 6*$, $*6 * *6$,
- $* * 66*$, $* * 6 * 6$,
- $* * *66$.

Each of these occurs with probability $(1/6)^2(5/6)^3$, and thus the total probability is $10 \cdot (1/6)^2(5/6)^3 = 625/3888$.

We are left with $k = 3$. One way to do this would be to exhaustively list the possibilities again. This is *a lot* more painful, though, as we now have three 6s to move around. Fortunately, there's an easier way! There's a wonderful duality between $k = 2$ and $k = 3$ when we have five rolls. Notice that there is a one-to-one correspondence between rolling exactly two 6s in five rolls and rolling exactly three 6s in five rolls! To see this, take a set of five rolls that has exactly two 6s; change the non-6s to 6s and the 6s to non-6s! Thus, to enumerate our possibilities, we just have to take our list from $k = 2$ and change each 6 to a non-6 and each non-6 to a 6! This gives

- $* * 666$, $*6 * 66$, $*66 * 6$, $*666*$,
- $6 * *66$, $6 * 6 * 6$, $6 * 66*$,
- $66 * *6$, $66 * 6*$,
- $666 * *$.

Each of these possibilities happens with probability $(1/6)^3(5/6)^2$; as there are 10 of these the total probability of rolling exactly three 6s is $10 \cdot (1/6)^3(5/6)^2 = 125/3888$.

If we sum our five probabilities we get

$$\left(\frac{5}{6}\right)^5 + 5 \cdot \left(\frac{1}{6}\right)\left(\frac{5}{6}\right)^4 + 10 \cdot \left(\frac{1}{6}\right)^2\left(\frac{5}{6}\right)^3 + 10 \cdot \left(\frac{1}{6}\right)^3\left(\frac{5}{6}\right)^2 + 5 \cdot \left(\frac{1}{6}\right)^4\left(\frac{5}{6}\right) + \left(\frac{1}{6}\right)^5,$$

which does sum to 1!

If you know binomial coefficients, you can do this problem much faster; it's fine to have done it this way. The probability of getting exactly k of the five choices to be 6 is just $\binom{5}{k}(1/6)^k(5/6)^{5-k}$ if $k \in \{0, \dots, 5\}$.

#9: If f and g are differentiable functions, prove the derivative of $f(x)g(x)$ is $f'(x)g(x) = f(x)g'(x)$. Emphasize where you add zero.

Solution: Let f and g be differentiable functions, and set $A(x) = f(x)g(x)$. It's not unreasonable to hope that there's a nice formula for the derivative of A in terms of f, f', g and g'. A great way to guess this relationship is to take some special examples. If we try $f(x) = x^3$ and $g(x) = x^4$, then $A(x) = x^7$ so $A'(x) = 7x^6$. At the same time, $f'(x) = 3x^2$ and $g'(x) = 4x^3$. There's only two ways to combine $f(x), f'(x), g(x)$ and $g'(x)$ and get x^6: $f'(x)g(x)$ and $f(x)g'(x)$. (Okay, there are more ways if we allow divisions; there's only two ways if we restrict ourselves to addition and multiplication.) Interestingly, if we add these together we get $3x^2 \cdot x^4 + x^3 \cdot 4x^3 = 7x^6$, which is just $A'(x)$. This *suggests* that $A'(x) = f'(x)g(x)+f(x)g'(x)$. If we try more and more examples, we'll see this formula keeps working. While this is strong evidence, it's not a proof; however, it *will* suggest the key step in our proof.

From the definition of the derivative and substitution,

$$A(x) \ = \ \lim_{h\to 0} \frac{A(x+h) - A(x)}{h} \ = \ \lim_{h\to 0} \frac{f(x+h)g(x+h) - f(x)g(x)}{h}. \tag{1.1}$$

From our investigations above, we think the answer should be $f'(x)g(x) + f(x)g'(x)$. We can begin to see an $f'(x)$ and a $g'(x)$ lurking above. Imagine the last term were $f(x)g(x+h)$ instead of $f(x)g(x)$. If this were the case, the limit would equal $f'(x)g(x)$ (we pull out the $g(x+h)$, which tends to $g(x)$, and what's left is the definition of $f'(x)$). Similarly, if the first piece were instead $f(x)g(x+h)$, then we'd get $f(x)g'(x)$. What we see is that our expression is *trying* to look like the right things, but we're missing pieces. This can be remedied by adding zero, in the form $f(x)g(x+h) - f(x)g(x+h)$. Let's see what this does. In the algebra below we use the limit of a sum is the sum of the limits and the limit of a product is the product of the limits; we can use these results as all these limits exist. We find

$$\begin{aligned} A'(x) &= \lim_{h\to 0} \frac{f(x+h)g(x+h) - f(x)g(x+h) \ + \ f(x)g(x+h) - f(x)g(x)}{h} \\ &= \lim_{h\to 0} \frac{f(x+h) - f(x)}{h} g(x+h) + \lim_{h\to 0} f(x) \frac{g(x+h) - g(x)}{h} \\ &= \lim_{h\to 0} \frac{f(x+h) - f(x)}{h} \lim_{h\to 0} g(x+h) + \lim_{h\to 0} f(x) \lim_{h\to 0} \frac{g(x+h) - g(x)}{h} \ = \ f'(x)g(x) + f(x)g'(x). \end{aligned}$$

□

The above proof has a lot of nice features. First off, it's the proof of a result you should know (at least if you've taken a calculus class). Second, we were able to guess the form of the answer by exploring some special cases. Finally, the proof was a natural outgrowth of these cases. We saw terms like $f'(x)g(x)$ and $f(x)g'(x)$ appearing, and thus asked ourselves: *So, what can we do to bring out these terms from what we have?* This led to adding zero in a clever way. It's fine to add zero, as it doesn't change the value. The advantage is we ended up with a new expression where we could now do some great simplifications.

No written homework due next Friday! Instead use the time to build up your strategic reserve in the book. We will not cover most of Chapter 3 in class – read the material and if there are calculations you are having trouble with or want to see in class, let me know and I'll do. Start reading Chapter Four. Monday's class will be a quick run of the Chapter Three material. Later we will talk about some applications of probability to mathematical modeling. There will not be reading assigned for this; the purpose of this is to (1) quickly show you how useful probability can be, and (2) give you a sense of the tools and techniques we'll see later in the semester, and (3) give you plenty of time to read ahead and build up your strategic reserve (if you don't take advantage of this you will have some painful weeks down the road!).

Also use this time to make sure you can do simple, basic coding; this is why I am not giving you HW to submit. I don't care what language you use (Mathematica, R, Python, Fortran, ...), but you should be comfortable doing simple assignments. I'll post a list of basic problems you should be able to do. If you want to learn Mathematica, I have a template online and a YouTube tutorial. Just go to `http://web.williams.edu/Mathematics/sjmiller/public_html/math/handouts/latex.htm` (note you'll also get links to using LaTeX). There are also videos online from introducing coding in my problem solving class.

- 2017: `https://youtu.be/e8I5alerkOw`

- 2018: https://youtu.be/e8I5alerkOw (code here: http://web.williams.edu/Mathematics/sjmiller/public_html/331Fa18/mathematicaprograms/math331IntroCoding2018.nb)

1.3. **Assignment: HW #3: Due Friday, September 28, 2018.** #1: Imagine we have a deck with s suits and N cards in each suit. We play the game **Aces Up**, except now we put down s cards on each turn. What is the probability that the final s cards are all in different suits? Write a computer program to simulate 1,000,000 deals and compare your observed probability with your theoretical prediction; it is fine to just do the program for s=4 and N=13 (a standard hand); you may earn 15 out of 10 points if you write general code for general s, N. #2: Consider all generalized games of **Aces Up** with C cards in s suits with N cards in a suit; thus $C = sN$. What values of s and N give us the greatest chance of all the cards being in different suits? Of being in the same suit? #3: The double factorial is defined as the product of every other integer down to 1 or 2; thus $6!! = 6 \cdot 4 \cdot 2$ while $7!! = 7 \cdot 5 \cdot 3 \cdot 1$. One can write $(2n-1)!!$ as $a!/(b^c d!)$ where a, b, c and d depend on n; find this elegant formula! *Hint: b turns out to be a constant, taking the same value for all n.* #4: A regular straight is five cards (not necessarily in the same suit) of five consecutive numbers; aces may be high or low, but we are *not* allowed to wrap around. A kangaroo straight differs in that the cards now differ by 2 (for example, 4 6 8 10 Q). What is the probability someone is dealt a kangaroo straight in a hand of five cards? #5: A prisoner is given an interesting chance for parole. He's blindfolded and told to choose one of two bags; once he does, he is to reach in and pull out a marble. Each bag has 25 red and 25 blue marbles, and the marbles all feel the same. If he pulls out a red marble he is set free; if it's a blue, his parole is denied. What is his chance of winning parole? #6: The set-up is similar to the previous problem, except now the prisoner is free to distribute the marbles among the two bags however he wishes, so long as all the marbles are distributed. He's blindfolded again, chooses a bag at random again, and then a marble. What is the best probability he can do for being set free? While you can get some points for writing down the correct answer, to receive full credit you must *prove* your answer is optimal!

2. HW #3: Due Friday, September 28, 2018

2.1. **Assignment (followed by solutions):** #1: Imagine we have a deck with s suits and N cards in each suit. We play the game **Aces Up**, except now we put down s cards on each turn. What is the probability that the final s cards are all in different suits? Write a computer program to simulate 1,000,000 deals and compare your observed probability with your theoretical prediction; it is fine to just do the program for s=4 and N=13 (a standard hand); you may earn 15 out of 10 points if you write general code for general s, N. #2: Consider all generalized games of **Aces Up** with C cards in s suits with N cards in a suit; thus $C = sN$. What values of s and N give us the greatest chance of all the cards being in different suits? Of being in the same suit? #3: The double factorial is defined as the product of every other integer down to 1 or 2; thus $6!! = 6 \cdot 4 \cdot 2$ while $7!! = 7 \cdot 5 \cdot 3 \cdot 1$. One can write $(2n-1)!!$ as $a!/(b^c d!)$ where a, b, c and d depend on n; find this elegant formula! *Hint: b turns out to be a constant, taking the same value for all n.* #4: A regular straight is five cards (not necessarily in the same suit) of five consecutive numbers; aces may be high or low, but we are *not* allowed to wrap around. A kangaroo straight differs in that the cards now differ by 2 (for example, 4 6 8 10 Q). What is the probability someone is dealt a kangaroo straight in a hand of five cards? #5: A prisoner is given an interesting chance for parole. He's blindfolded and told to choose one of two bags; once he does, he is to reach in and pull out a marble. Each bag has 25 red and 25 blue marbles, and the marbles all feel the same. If he pulls out a red marble he is set free; if it's a blue, his parole is denied. What is his chance of winning parole? #6: The set-up is similar to the previous problem, except now the prisoner is free to distribute the marbles among the two bags however he wishes, so long as all the marbles are distributed. He's blindfolded again, chooses a bag at random again, and then a marble. What is the best probability he can do for being set free? While you can get some points for writing down the correct answer, to receive full credit you must *prove* your answer is optimal!

#1: Imagine we have a deck with s suits and N cards in each suit. We play the game **Aces Up**, except now we put down s cards on each turn. What is the probability that the final s cards are all in different suits? What is the probability that the final s cards are all in different suits? Write a computer program to simulate 1,000,000 deals and compare your observed probability with your theoretical prediction; it is fine to just do the program for s=4 and N=13 (a standard hand); you may earn 15 out of 10 points if you write general code for general s, N.

Solution: There are sN cards and there are $\binom{sN}{s}$ ways to choose s of them with order not mattering. How many ways are there to choose one card from each suit? It's $\binom{N}{1}\binom{N}{1}\cdots\binom{N}{1}$ a total of s times, or N^s. Thus the answer is $N^s/\binom{sN}{s} = N^s s!(sN-s)!/(sN)!$.

```
acesup[n_, S_, numiter_] := Module[{},
   (*n,S variables*)
   (*n is number of cards in a suit,
   can't use C or N as variable in Mathematica*)
   (*S is the number of suits*)
   (*for this problem only care about suits of cards,
   not numbers*)(*only care about the suit of the cards;
   creates a deck of S suits with n cards in each suit*)
   (*have suites 1, 10, 100,
   1000 and so on so can easily tell if one of each*)
   deck = {};
   For[i = 1, i <= n, i++, (*
    this goes through each of the possible numbers *)
    For[s = 1, s <= S, s++, (* this goes through each possible suit *)


     deck = AppendTo[deck, 10^(s - 1)]]];

   maxsum = Sum[10^(s - 1), {s, 1, S}];
   (* this is the value of a hand of S cards, 1 in each suit *)
   (* only way to sum to 111...111 is if one of each suit! *)
   success = 0; (* initialize number of successes to 0 *)
```

```
(* now the main loop, randomly checking s cards numiter times *)
For[i = 1, i <= numiter, i++,
 {
  (* prints an update every time do 10% *)
  If[Mod[i, numiter/10] == 0,
   Print["Have done ", 100.0 i/numiter, "%."]];
  hand = RandomSample[deck, S]; (*
  randomly chooses S cards from deck *)
  If[Sum[hand[[i]], {i, 1, S}] == maxsum, success = success + 1];
  (* if the hand has S different suits increase success by 1 *)
  }]; (* end of i loop *)

Print["Observed Percent of time last ", S, " same suit is ",
 100.0  success/numiter, "%."];
Print["Theoretical Percent of time last ", S, " same suit is ",
 100.0  n^S / Binomial[S n, S], "%."];
]; (* end of module *)
```

For example, running with four suits, 13 cards in a suit, and doing 1,000,000 simulations gave us an observed probability of 10.5343%, very close to the theoretical prediction of 10.5498%.

#2: Consider all generalized games of **Aces Up** with C cards in s suits with N cards in a suit; thus $C = sN$. What values of s and N give us the greatest chance of all the cards being in different suits? Of being in the same suit?
Solution: If $s = 1$ then all the cards are in the same suit; if $N = 1$ then all the cards are in different suits.

#3: The double factorial is defined as the product of every other integer down to 1 or 2; thus $6!! = 6{\cdot}4{\cdot}2$ while $7!! = 7{\cdot}5{\cdot}3{\cdot}1$. One can write $(2n-1)!!$ as $a!/(b^c d!)$ where a, b, c and d depend on n; find this elegant formula! *Hint: b turns out to be a constant, taking the same value for all n.*
Solution: We have

$$\begin{aligned}
(2n-1)!! &= (2n-1)(2n-3)\cdots 3\cdot 1\\
&= (2n-1)(2n-3)\cdots 3\cdot 1\cdot \frac{2n\cdot(2n-2)\cdots 4\cdot 2}{2n\cdot(2n-2)\cdots 4\cdot 2}\\
&= \frac{(2n)!}{2n\cdot(2n-2)\cdots 4\cdot 2} \;=\; \frac{(2n)!}{2^n n!};
\end{aligned}$$

thus $a = 2n$, $b = 2$, $c = n$ and $d = n$. It's natural to multiply by the even numbers; we have a product over all odd numbers, which isn't a factorial because we're missing the even numbers. The final bit is then noticing that in the denominator we have all even numbers, and by pulling out a 2 from each we have a nice factorial. Why is this problem useful? We'll see later in the semester how to approximate factorials (Stirling's formula), which then immediately yields estimates on the double factorial (which we'll also see has combinatorial significance).

#4: A regular straight is five cards (not necessarily in the same suit) of five consecutive numbers; aces may be high or low, but we are *not* allowed to wrap around. A kangaroo straight differs in that the cards now differ by 2 (for example, 4 6 8 10 Q). What is the probability someone is dealt a kangaroo straight in a hand of five cards?
Solution: The possibilities are A 3 5 7 9, 2 4 6 8 10, 3 5 7 9 J, 4 6 8 10 Q, 5 7 9 J K, 6 8 10 Q A. There are thus six possibilities. All we need to do is figure out the probability of one of these six and then multiply by six. For each fixed kangaroo straight, we have four choices for each card, for a total of 4^5 Kangaroo straights of a given pattern. Thus the total number of Kangaroo

straights is $6 \cdot 4^5 = 6144$. As the number of ways to choose five cards from 52 (with order not mattering) is $\binom{52}{5} = 2,598,960$, we see the probability of a Kangaroos straight is $128/54145 \approx 0.00236402$.

#5: A prisoner is given an interesting chance for parole. He's blindfolded and told to choose one of two bags; once he does, he is to reach in and pull out a marble. Each bag has 25 red and 25 blue marbles, and the marbles all feel the same. If he pulls out a red marble he is set free; if it's a black, his parole is denied. What is his chance of winning parole?
Solution: His chance is 50%. Probably the easiest way to see this is that there is complete symmetry here between red and blue marbles, and thus he has an equal chance of choosing either. Note there are 100 marbles in all (50 red and 50 blue).

#6: The set-up is similar to the previous problem, except now the prisoner is free to distribute the marbles among the two bags however he wishes, so long as all the marbles are distributed. He's blindfolded again, chooses a bag at random again, and then a marble. What is the best probability he can do for being set free? While you can get some points for writing down the correct answer, to receive full credit you must *prove* your answer is optimal!
Solution: Let's assume he places r red marbles and b blue marbles in the first bag; thus the second bag has $50 - r$ red marbles and $50 - b$ blue marbles. If he picks the first bag, he earns parole with probability $\frac{r}{r+b}$, while if he picks the second bag he earns parole with probability with $\frac{50-r}{100-r-b}$. As each bag has probability 1/2 of being chosen, his probability of getting parole is

$$\frac{1}{2}\frac{r}{r+b} + \frac{1}{2}\frac{50-r}{100-r-b} \ = \ \frac{b(25-r)+r(75-r)}{(100-b-r)(b+r)}.$$

It's always worthwhile checking extreme cases. If $r = b$ then the two jars are balanced, and we have a 50% chance. Can we break 50%? What if we put all the reds in one and all the blues in another? That gives us 50% as well. How about 15 red and 35 blue in one jar? That also gives 50%.

You might be thinking that, no matter what we do, we always get 50%. We've unfortunately made a *very* common mistake – we're not freely investigating all possibilities. Note that each of these cases has the same number of marbles in each jar. What if we try something else, say 20 red and 10 blue in one jar? That gives 23/42, or a tad over 54.7%. This is promising. What if we keep the 20 red and decrease to 5 blue? Doing so yields 3/5 or 60%. It now becomes natural to keep sending the number of blues in the first jar to zero. If we have 20 red in the first jar and no blue, we get 11/16 or 68.75%. Of course, it's wasteful to have 20 red in the first jar; if there are only reds then if we pick that jar we *must* get a red. This *suggests* we want to have 1 red in the first jar and all the remaining marbles in the second. If we do this we get a red with probability 74/99, or almost 75%.

Now that we have a conjectured answer, the question is how do we *prove* it? Assume we start with r red and b blue in the first jar. If we transfer some blues to the second jar we increase the chance of getting a red in the first jar but decrease the chance in the second, and we need to argue that we gain more than we lose. There are a lot of ways to try to do the algebra. One is to quantify that this movement always helps us, and then once we have no blues in the first jar we transfer the remaining reds.

Here is another approach. We break all the possibilities into cases depending on the *sum* of r and b. Thus, let's look at all pairs (r, b) such that $r + b = c$. We might as well assume $c \leq 50$ as *one* of the jars must have at most 50 marbles. This gives us the function

$$g_c(r) \ = \ \frac{c(25-r)+50r}{c(100-c)},$$

whose derivative (with respect to r) is

$$g_c'(r) \ = \ \frac{50-c}{c(100-c)}.$$

As c is positive and at most 50, we see the derivative with respect to r is positive unless $c = 50$ (in which case the derivative is zero, which corresponds to our earlier result that the probability is independent of r and b when each jar has 50 marbles). We thus see that for a given *total* number of marbles in jar 1, the best we can do is to have all the marbles in jar 1 be red. This means we only need to explore assignments where jar 1 is all red; however, now our earlier analysis is applicable. If jar 1 only has red marbles, once we have one red marble the others are useless there, and are better used in jar 2 (moving those extra red marbles over doesn't change the probability of a red in jar 1, but does increase it in jar 2). Thus, the optimal solution is jar 1 consisting of just one red marble.

We give some Mathematica code to investigate the problem.

```
f[r1_, b1_, r2_, b2_] := (1/2) (r1 / (r1 + b1)) + (1/2) (r2 / (r2 + b2));
marblejar[red_, blue_] := Module[{},
  maxprob = 0;
  maxred = 0;
  maxblue = 0;
  For[r1 = 1, r1 <= red, r1++, (*
   without loss of generality one jar has a red *)
   For[b1 = 0, b1 <= blue, b1++,
    If[r1 + b1 >= 1 && (red - r1) + (blue - b1) >= 1, (*
     this is to make sure each jar is nonempty *)
     {
      x = f[r1, b1, red - r1, blue - b1];
      If[x > maxprob,
       {
        maxprob = x;
        maxred = r1;
        maxblue = b1;
        }]; (* end of x > maxprob *)
      }]; (* end of if loop *)
    ]; (* end of b1 loop *)
   ]; (* end of r1 loop *)
  Print["Max prob is ", maxprob, " or ", 100. maxprob, "%, and have ", maxred,
  " red and ", maxblue,
   " blue in first jar."];
  ]; (* end of module *)
```

To test how fast this is, type

```
Timing[marblejar[1000, 1000]]
```

Of course, this is not necessarily the fastest code. The problem is we investigate some pairs multiple times (if the distribution of (red,blue) for the two jars are (10,40) and (40,10), that's the same as (40,10) and (10,40). We instead loop on the total number of marbles in the first jar, and we may assume without loss of generality that the first jar has at most half the total number of marbles (by the Pidgeon-hole principle, at least one jar has at most half the marbles). This leads to faster code, which for 1000 red and 1000 blue runs in a little less than half the time.

```
f[r1_, b1_, r2_, b2_] := (1/2) (r1 / (r1 + b1)) + (1/2) (r2 / (r2 + b2));
fastmarblejar[red_, blue_] := Module[{},
  maxprob = 0;
  maxred = 0;
  maxblue = 0;
  For[num = 1, num <= (red + blue)/2, num++, (*
   without loss of generality one jar has a red *)
   For[r1 = 0, r1 <= num, r1++,
    {
     b1 = num - r1;
     x = f[r1, b1, red - r1, blue - b1];
     If[x > maxprob,
      {
       maxprob = x;
       maxred = r1;
       maxblue = b1;
       }]; (* end of x > maxprob *)
     }]; (* end of r1 loop *)
   ]; (* end of num loop *)
  Print["Max prob is ", maxprob, " or ", 100. maxprob, "%, and have ", maxred,
  " red and ", maxblue,
   " blue in first jar."];
  ]; (* end of module *)
```

To test how fast this is, type

```
Timing[marblejar[1000, 1000]]
```

It is possible to solve this entirely elementarily; i.e., no calculus! The best we can do can't be better than 75%. To see this, imagine one jar has probability $p > 1/2$ of red; then there must be more blacks in the other jar than red, and it must have a probability $q < 1/2$. The best possible case is when $p = 1$ and $q = 1/2$ (it's *NOT* clear that we can do this!), which gives $(p+q)/2 = 3/4$ as our chance of winning. So we know we can't do better than 75%. How close can we come? The closer q is to 1/2, the better. We know q has to be less than 1/2; the closest it can be is if we *just* miss, which happens when have 49 red and 50 blue. Why? In this case we get 49/99, so we miss 1/2 by $1/198$; note any other fraction $r/(r+b)$ misses 1/2 by more (as the best case is when $r = b-1$, in which case we miss 1/2 by $1/(4b-2)$. Thus the best we can do is $(1+49/99)/2 = 74/99 \approx 0.747475\%$.

Assignment: HW #4: Due Friday, October 5: Note Mathematical Induction might be useful for some of these problems. #1: Let $\{A_n\}_{n=1}^{\infty}$ be a countable sequence of events such that for each n, $\mathrm{Prob}(A_n) = 1$. Prove the probability of the intersection of all the A_n's is 1. #2: Prove the number of ways to match $2n$ people into n pairs of 2 is $(2n-1)!!$ (recall the double factorial is the product of every other integer, continuing down to 2 or 1). #3: Assume $0 < \mathrm{Prob}(X), \mathrm{Prob}(Y) < 1$ and X and Y are independent. Are X^c and Y^c independent? (Note X^c is not X, or $\Omega \setminus X$). Prove your answer. #4: Using the Method of Inclusion-Exclusion, count how many hands of 5 cards have at least one ace. You need to determine what the events A_i should be. Do not find the answer by using the Law of Total Probability and complements (though you should use this to check your answer). #5: We are going to divide 15 identical cookies among four people. How many ways are there to divide the cookies if all that matters is how many cookies a person receives? Redo this problem but now only consider divisions of the cookies where person i gets at least i cookies (thus person 1 must get at least one cookie, and so on). #6: Redo the previous problem (15 identical cookies and 4 people), but with the following constraints: each person gets at most 10 cookies (it's thus possible some people get no cookies). #7: Find a discrete random variable, or prove none exists, with probability density function f_X such that $f_X(x) = 2$ for some x between 17 and 17.01. #8: Find a continuous random variable, or prove none exists, with probability density function f_X such that $f_X(x) = 2$ for all x between 17 and 17.01. #9: Let X be a continuous random variable with pdf f_X satisfying $f_X(x) = f_X(-x)$. What can you deduce about F_X, the cdf? #10: Find if you can, or say why you cannot, the first five Taylor coefficients of (a) $\log(1-u)$ at $u = 0$; (b) $\log(1-u^2)$ at $u = 0$; (c) $x\sin(1/x)$ at $x = 0$. #11: Let X be a continuous random variable. (a) Prove F_X is a non-decreasing function; this means $F_X(x) \le F_X(y)$ if $x < y$. (b) Let U be a random variable with cdf $F_U(x) = 0$ if $u < 0$, $F_U(x) = x$ if $0 < x < 1$, and $F_U(x) = 1$ if $1 < x$. Let F be any continuous function such that F is strictly increasing and the limit as x approaches negative infinity of $F(x)$ is 0 and the limit as x approaches positive infinity is 1. Prove $Y = F^{-1}(U)$ is a random variable with cdf F.

3. HW #4: Due Friday, October 5

3.1. **Assignment:** HW #4: Due Friday, October 5: Note Mathematical Induction might be useful for some of these problems. #1: Let $\{A_n\}_{n=1}^{\infty}$ be a countable sequence of events such that for each n, $\text{Prob}(A_n) = 1$. Prove the probability of the intersection of all the A_n's is 1. #2: Prove the number of ways to match $2n$ people into n pairs of 2 is $(2n-1)!!$ (recall the double factorial is the product of every other integer, continuing down to 2 or 1). #3: Assume $0 < \text{Prob}(X), \text{Prob}(Y) < 1$ and X and Y are independent. Are X^c and Y^c independent? (Note X^c is not X, or $\Omega \setminus X$). Prove your answer. #4: Using the Method of Inclusion -Exclusion, count how many hands of 5 cards have at least one ace. You need to determine what the events A_i should be. Do not find the answer by using the Law of Total Probability and complements (though you should use this to check your answer). #5: We are going to divide 15 identical cookies among four people. How many ways are there to divide the cookies if all that matters is how many cookies a person receives? Redo this problem but now only consider divisions of the cookies where person i gets at least i cookies (thus person 1 must get at least one cookie, and so on). #6: Redo the previous problem (15 identical cookies and 4 people), but with the following constraints: each person gets at most 10 cookies (it's thus possible some people get no cookies). #7: Find a discrete random variable, or prove none exists, with probability density function f_X such that $f_X(x) = 2$ for some x between 17 and 17.01. #8: Find a continuous random variable, or prove none exists, with probability density function f_X such that $f_X(x) = 2$ for all x between 17 and 17.01. #9: Let X be a continuous random variable with pdf f_X satisfying $f_X(x) = f_X(-x)$. What can you deduce about F_X, the cdf? #10: Find if you can, or say why you cannot, the first five Taylor coefficients of (a) $\log(1-u)$ at $u = 0$; (b) $\log(1-u^2)$ at $u = 0$; (c) $x \sin(1/x)$ at $x = 0$. #11: Let X be a continuous random variable. (a) Prove F_X is a non-decreasing function; this means $F_X(x) \le F_X(y)$ if $x < y$. (b) Let U be a random variable with cdf $F_U(x) = 0$ if $u < 0$, $F_U(x) = x$ if $0 < x < 1$, and $F_U(x) = 1$ if $1 < x$. Let F be any continuous function such that F is strictly increasing and the limit as x approaches negative infinity of $F(x)$ is 0 and the limit as x approaches positive infinity is 1. Prove $Y = F^{-1}(U)$ is a random variable with cdf F.

3.2. **Solutions:**

#1: Let $\{A_n\}_{n=1}^{\infty}$ be a countable sequence of events such that for each n, $\text{Prob}(A_n) = 1$. Prove the probability of the intersection of all the A_n's is 1.
Solution: If $\mathbb{P}(A_n) = 1$ then $\mathbb{P}(A_n^c) = 0$. If we look at the intersection of the events A_n, we see this is all elements in *each* A_n. In other words, it is the complement of the union of the events A_n^c. Let's prove this carefully.

Claim: $\cap_{n=1}^{\infty} A_n = (\cup_{n=1}^{\infty} A_n^c)^c$. Proof: A standard way to establish set-theoretic identities such as this is to show every element in the left hand side is in the right, and every element in the right hand side is in the left. This implies the two sets have the same elements, and are therefore equal.

Imagine now $x \in \cap_{n=1}^{\infty} A_n$. Then $x \in A_n$ for each n, which means $x \notin A_n^c$ for each n, which implies $x \notin \cup_{n=1}^{\infty} A_n^c$, and then taking complements yields $x \in (\cup_{n=1}^{\infty} A_n^c)^c$.

For the other direction, imagine $x \in (\cup_{n=1}^{\infty} A_n^c)^c$. Then $x \notin \cup_{n=1}^{\infty} A_n^c$, which means that for each n we have $x \notin A_n^c$. This immediately implies $x \in A_n$ for all n, and hence $x \in \cap_{n=1}^{\infty} A_n$ as desired. □

The reason we did this is we have results relating the probability of a union to the sum of the probabilities. It's thus natural to try and recast the problem in terms of unions.

Returning to the proof, we have $\mathbb{P}(A_n^c) = 0$. Imagine we could prove that the probability of the union of the A_n^c's is 0; i.e., $\mathbb{P}(\cup_{n=1}^{\infty} A_n^c) = 0$. Then the complement of this union is 1, but from our analysis above the complement is $\cap_{n=1}^{\infty} A_n$. In other words, if we prove $\mathbb{P}(\cup_{n=1}^{\infty} A_n^c) = 0$ then we deduce $\mathbb{P}(\cap_{n=1}^{\infty} A_n) = 1$, which is our goal.

We're thus left with proving $\mathbb{P}(\cup_{n=1}^{\infty} A_n^c) = 0$. If the events A_n^c were disjoint we would be done, as the probability of a countable union of disjoint events is the sum of the probabilities, and each probability is 0. Unfortunately the events $\{A_n^c\}$ need not be disjoint, so some care is needed. When we look at $A_1^c \cup A_2^c$, what matters is what is in A_2^c that is not in A_1^c, as anything in A_1^c is already included. We can write the union of these two events as $A_1^c \cup (A_2^c \cap A_1)$. This is a disjoint union, and since $A_2^c \cap A_1 \subset A_2^c$, it still has probability zero since A_2^c has probability zero. What we're doing is we're throwing away anything in A_2^c that's in A_2.

To help highlight what's going on, let $B_1 = A_1^c$, $B_2 = A_2^c \cap B_1^c$ (the items in A_2^c not in $B_1 = A_1^c$). We then let $B_3 = A_3^c \cap (A_1^c \cup A_2^c)$, the new items that are in A_3^c but not in A_1^c or A_2^c. As $B_3 \subset A_3^c$ and $\mathbb{P}(A_3^c) = 0$, we find $\mathbb{P}(B_3) = 0$.

We continue in this manner, setting $B_n = A_n^c \cap (A_1^c \cup \cdots \cup A_{n-1}^c)$, and $\mathbb{P}(B_n) = 0$. The B_n's are now a disjoint union of events of probability zero, and thus $\mathbb{P}(\cup_{n=1}^\infty B_n) = \sum_{n=1}^\infty \mathbb{P}(B_n) = 0$. This completes the proof, as $\cup_{n=1}^\infty B_n = \cup_{n=1}^\infty A_n^c$. $\Box$

Actually, we don't need all the arguments above. We can avoid introducing the B_n's by using $\mathbb{P}(\cup_{n=1}^\infty A_n^c) \le \sum_{n=1}^\infty \mathbb{P}(A_n^c)$. As each summand on the right hand side is zero, the double counting is harmless and we again find this union has probability zero.

Remark: What we really want to do is remark that $\lim_{N\to\infty} \mathbb{P}(\cap_{n=1}^N A_n) = \mathbb{P}(\lim_{n\to\infty} \cap_{n=1}^N A_n)$. In other words, we want to interchange the limit and the probability. The arguments above are to help us make such a justification. *Assume* you knew this fact from other sources. We present another proof that the probability of the intersection is 1. Notice the right hand side is readily analyzed, as $\mathbb{P}(\lim_{n\to\infty} \cap_{n=1}^N A_n) = \mathbb{P}(\cap_{n=1}^\infty A_n)$. For the left hand side, if we can show for each finite N that $\mathbb{P}(\cap_{n=1}^N A_n) = 1$ then the limit is 1.

The simplest way to prove this is by induction. If X and Y happen with probability one, then $\mathbb{P}(X \cap Y) = \mathbb{P}(X) + \mathbb{P}(Y) - \mathbb{P}(X \cup Y)$. Note every probability on the right hand side equals 1 (no event can have probability greater than 1, and $X \subset X \cup Y$ so $\mathbb{P}(A \cup Y) = 1$). This implies $\mathbb{P}(X \cap Y) = 1$. We've thus shown if two events each have probability one then their union has probability one. We now proceed by induction, setting $X = \cup_{n=1}^{N-1} A_n$ and $Y = A_N$ to get $\mathbb{P}(\cap_{n=1}^N A_n) = 1$ for all N. So for any finite N we have $\mathbb{P}(\cap_{n=1}^N A_n) = 1$. We now take the limit as $N \to \infty$, and we get $\lim_{N\to\infty} \mathbb{P}(\cap_{n=1}^N A_n) = 1$.

We could have argued slightly differently above. The key is proving $\mathbb{P}(X \cap Y) = 1$; another approach is to use partitions, and observe $\mathbb{P}(X) = \mathbb{P}(X \cap Y) + \mathbb{P}(X \cap Y^c)$. As $\mathbb{P}(Y) = 1$, $\mathbb{P}(Y^c) = 0$ and thus $\mathbb{P}(X \cap Y^c) = 0$ (as $X \cap Y^c \subset Y^c$). Thus $\mathbb{P}(X) = \mathbb{P}(X \cap Y)$, and as $\mathbb{P}(X) = 1$ we finally deduce $\mathbb{P}(X \cap Y) = 1$. Note how important in this problem the $n = 2$ case is in the inductive proof. Frequently in induction proofs we just need to use the result with n to prove $n + 1$; however, a sizable number of times the general proof basically just reduces to understanding the $n = 2$ case.

#2: Prove the number of ways to match $2n$ people into n pairs of 2 is $(2n - 1)!!$ (recall the double factorial is the product of every other integer, continuing down to 2 or 1).
Solution: As anyone can be matched with anyone, there are $(2n - 1)!!$ ways to do this, where the double factorial means we take the product of every other term ($6!! = 6 \cdot 4 \cdot 2$ and $5!! = 5 \cdot 3 \cdot 1$). One way to see this is to note this is just

$$\binom{2n}{2}\binom{2n-2}{2}\cdots\binom{4}{2}\binom{2}{2}\cdot\frac{1}{n!};$$

we divide by $n!$ as we have attached labels to each pair of people, and there aren't supposed to be labels. We now do some algebra. Noting $\binom{2i}{2} = \frac{2i(2i-1)}{2}$, we get our product is

$$\begin{aligned} & \frac{2n(2n-1)}{2}\,\frac{(2n-2)(2n-3)}{2}\cdots\frac{2\cdot 1}{2}\,\frac{1}{n!} \\ = \quad & \frac{n(2n-1)\cdot(n-1)(2n-3)\cdots 1(1)}{n(n-1)\cdots 2\cdot 1} \\ = \quad & (2n-1)(2n-3)\cdots 1. \end{aligned} \tag{3.1}$$

We could also proceed by induction. The first person must be matched with someone; there are $2n - 1$ ways to do this. We now pair off the remaining $2n - 2$ people, which by induction happens $(2n - 3)!!$ ways, so there are $(2n - 1) \cdot (2n - 3)!! = (2n - 1)!!$ ways. If you must be matched with someone from the opposite side, there are only $n!$ ways.

#3: Assume $0 < \mathrm{Prob}(X), \mathrm{Prob}(Y) < 1$ and X and Y are independent. Are X^c and Y^c independent? (Note X^c is not X, or $\Omega \setminus X$). Prove your answer.
Solution: Since X and Y are independent, we have $\mathbb{P}(X \cap Y) = \mathbb{P}(X)\mathbb{P}(Y)$. We want to show $\mathbb{P}(X^c \cap Y^c) = \mathbb{P}(X^c)\mathbb{P}(Y^c)$. We have $\mathbb{P}(X^c) = 1 - \mathbb{P}(X)$ and $\mathbb{P}(Y^c) = 1 - \mathbb{P}(Y)$, therefore

$$\begin{aligned} \mathbb{P}(X^c)\mathbb{P}(Y^c) &= (1 - \mathbb{P}(X))(1 - \mathbb{P}(Y)) \\ &= 1 - \mathbb{P}(X) - \mathbb{P}(Y) + \mathbb{P}(X)\mathbb{P}(Y) \\ &= 1 - \mathbb{P}(X) - \mathbb{P}(Y) + \mathbb{P}(X \cap Y). \end{aligned}$$

We need to work on the right hand side to make it look like $\mathbb{P}(X^c \cap Y^c)$. Using

$$\mathbb{P}(X \cup Y) = \mathbb{P}(X) + \mathbb{P}(Y) - \mathbb{P}(X \cap Y)$$

(essentially the double counting formula), we recognize the right hand side as $1 - \mathbb{P}(X \cup Y)$. We're making progress, and we now have

$$\mathbb{P}(X^c)\mathbb{P}(Y^c) \ = \ 1 - \mathbb{P}(X \cup Y) \ = \ \mathbb{P}\left((X \cup Y)^c\right)$$

from the Law of Total Probability. Now we just need to observe that $(X \cup Y)^c = X^c \cap Y^c$. This is a basic fact in set theory, and completes the proof. Thus, X^c and Y^c are independent.

Appendix to the problem: The proof that $(X \cup Y)^c = X^c \cap Y^c$ involves a nice technique; you didn't need to prove this, but it's good to see. We show that any t in the set on the left hand side is in the set on the right hand side, and any s in the right hand side is in the left hand side. Thus the two sets have the same elements, and must be equal.

Imagine $t \in (X \cup Y)^c$. This means t is *not* in $X \cup Y$, so $t \in X^c$ and $t \in Y^c$, hence $t \in X^c \cap Y^c$.

What if $s \in X^c \cap Y^c$? Then $s \in X^c$ and $s \in Y^c$ (it must be in each if it is in the intersection), so s is not in $X \cup Y$, which means s is in $(X \cup Y)^c$.

Note: Alternative proof: There is another way to do this problem, and it illustrates a nice technique. The philosopher David Hume asked what happens if each day you replace a different board on a ship. Clearly it's still the same point when you go from day n to day $n+1$, but at some point there are no longer any of the original boards left! What does that have to do with this problem? We can move in stages. We start with X and Y are independent. Step 1 is to show that if A and B are independent events then so too are A and B^c. Why does this help? We then go from (X, Y) independent to (X, Y^c) are independent, which of course is the same as (Y^c, X) are independent. We now apply this observation again (so now $A = Y^c$ and $B = X$) and find (Y^c, X^c) are independent! We thus "move" to the point we want in successive stages. You might have seen this method before in a calculus class (it can occur in the multidimensional chain rule, trying to figure out where to evaluate points).

#4: Using the Method of Inclusion-Exclusion, count how many hands of 5 cards have at least one ace. You need to determine what the events A_i should be. Do not find the answer by using the Law of Total Probability and complements (though you should use this to check your answer).

Solution: Let A_i be the event that the ace in suit i is in our hand (we'll let spades be the first suit, hearts the second suit, diamonds the third and clubs the fourth, though of course these labels don't matter). The event $A_1 \cup A_2 \cup A_3 \cup A_4$ is the event that our hand has at least one ace. We give two "proofs" of this result. *Read carefully below. Are they both right? If not, which one is wrong? It is good to occasionally see wrong answers, as these can highlight subtle issues.*

Note that $\#A_i = \binom{1}{1}\binom{48}{4}$ for each i (we have to have a specific ace, cannot have any other aces, and then must choose 4 cards from the 48 non-aces).

Similarly, for all pairs $i \neq j$, we have $\#(A_i \cap A_j) = \binom{1}{1}\binom{1}{1}\binom{48}{3}$. Continuing along these lines we find for each triple $i < j < k$ we have $\#(A_i \cap A_j \cap A_k) = \binom{1}{1}^3\binom{48}{2}$ and finally $\#(A_1 \cap A_2 \cap A_3 \cap A_4) = \binom{48}{1}$.

The Inclusion-Exclusion Formula is "nice" to use here, as all that matters is how many pairs (or triples or quadruples) of indices we have, as all the options have the same count. If N_{aces} is the number of hands, we find

$$\begin{aligned} N_{\text{aces}} \ &= \ \sum_{i=1}^{4} \#A_i - \sum_{1 \le i < j \le 4} \#(A_i \cap A_j) + \sum_{1 \le i < j < k \le 4} \#(A_i \cap A_j \cap A_k) \\ &\quad -\#(A_1 \cap A_2 \cap A_3 \cap A_4). \end{aligned}$$

The number of pairs with $1 \le i < j \le 4$ is the number of ways to choose two elements from four with order not mattering, or $\binom{4}{2} = 6$. Similarly the number of triples with $1 \le i < j < k \le 4$ is $\binom{4}{3} = 4$. We find

$$\begin{aligned} N_{\text{aces}} \ &= \ 4\#A_1 - 6\#(A_1 \cap A_2) + 4\#(A_1 \cap A_2 \cap A_3) - \#(A_1 \cap A_2 \cap A_3 \cap A_4) \\ &= \ 4\binom{48}{4} - 6\binom{48}{3} + 4\binom{48}{2} - \binom{48}{1} \ = \ 679,008. \end{aligned}$$

Equivalently, as there are $\binom{52}{5} = 2,598,960$ possible hands, we see the probability of getting a hand with at least one ace is $\frac{679008}{2598960} \approx 26.1\%$.

Is this answer reasonable? The probability of getting one specific ace in five cards is $\binom{1}{1}\binom{48}{4}/\binom{52}{5} \approx 7.5\%$; if we multiply this by 4 we get about 30%. This is close to the correct answer. Further, it's off in the right way – we expect to be over-estimating, as multiplying by 4 leads to double (and triple and quadruple) counting.

Let's also check by using complements. The probability of getting no aces in a hand of five cards is just $\binom{48}{5}/\binom{52}{5}$. Thus 1 minus this, which is $1 - \binom{48}{5}/\binom{52}{5} = \frac{18472}{54145}$, is the probability of getting at least one ace. If we approximate the fraction, we get about 34.1%.

Something must be wrong – we can't have two different answers! While you should get in the habit of running computer simulations to get a feel for the answer, it's a very important skill to be able to do this when you have two different answers. Let's do a simulation and see if we can determine which answer is right. The code is

```
acetest[num_] := Module[{},
   count = 0;
   For[n = 1, n <= num, n++,
    {
     x =
      RandomSample[{1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15,
         16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30,
        31, 31, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45,
        46, 47, 48, 49, 50, 51, 52}, 5];
     isin = 0;
     For[j = 1, j <= 4, j++,
      If[MemberQ[x, j] == True, isin = 1];
      ];
     If[isin > 0, count = count + 1];
     }];
   Print[100. count / num];
   ];
```

Simulating 100,000 times yielded about 34.077%, very close to the second method. Thus, it's quite likely we made a mistake in the first method, but where?

The problem is that the event A_i is just getting a specific ace; it doesn't mean we can't get ***more*** aces. Thus $\#A_i = \binom{1}{1}\binom{51}{4}$ instead of $\binom{1}{1}\binom{48}{4}$. We have to continue down the line and correct all the probabilities. We get

$$\begin{aligned} N_{\text{aces}} &= 4\#A_1 - 6\#(A_1 \cap A_2) + 4\#(A_1 \cap A_2 \cap A_3) - \#(A_1 \cap A_2 \cap A_3 \cap A_4) \\ &= 4\binom{51}{4} - 6\binom{50}{3} + 4\binom{49}{2} - \binom{48}{1} \\ &= \frac{18472}{54145} \approx 34.1\%, \end{aligned}$$

exactly as before.

We also must revisit our calculation as to whether or not our answer is reasonable. The probability of getting one specific ace in five cards is $\binom{1}{1}\binom{51}{4}/\binom{52}{5} \approx 9.6\%$; if we multiply this by 4 we get about 38%. This is close to the correct answer. Further, it's off in the right way – we expect to be over-estimating, as multiplying by 4 leads to double (and triple and quadruple) counting.

#5: We are going to divide 15 identical cookies among four people. How many ways are there to divide the cookies if all that matters is how many cookies a person receives? Redo this problem but now only consider divisions of the cookies where person i gets at least i cookies (thus person 1 must get at least one cookie, and so on).

Solution: From the cookie problem, if there are C identical cookies and P people, the number of ways to divide is $\binom{C+P-1}{P-1}$; thus the answer to the first part is $\binom{15+4-1}{4-1} = \binom{18}{3} = 816$. We have just solved the equation $x_1 + x_2 + x_3 + x_4 = 816$. We now let $x_j = y_j + j$ (to deal with the constraints), and find $y_1 + y_2 + y_3 + y_4 + 10 = 15$, or equivalently $y_1 + y_2 + y_3 + y_4 = 5$. This is thus the same as looking at a cookie problem with 5 cookies and 4 people, so the answer is $\binom{5+4-1}{4-1} = \binom{8}{3} = 56$.

#6: Redo the previous problem (15 identical cookies and 4 people), but with the following constraints: each person gets at most 10 cookies (it's thus possible some people get no cookies).

Solution: We can use the Law of Total Probability (or of complementary events). We find out the number of ways without the restriction (which is just $\binom{15+4-1}{4-1} = \binom{18}{3} = 816$), then subtract off the number of ways when the restriction is violated. The key observation is that it's impossible for two or more people to each get at least 11 cookies, as there are only 15 cookies. Thus we just need to break into cases based on who gets the 11 cookies. We might as well assume the first person gets the eleven or more cookies, and then multiply by 4 for the remaining cases.

Thus, let $x_1 = 11 + y_1$ and $x_j = y_j$ for $j \geq 3$. The number of ways when the first person has at least 11 cookies is $11 + y_1 + y_2 + y_3 + y_4 = 15$, or $y_1 + y_2 + y_3 + y_4 = 4$. This is the same as a cookie problem with 4 cookies and 4 people; the solution to that is $\binom{4+4-1}{4-1} = \binom{7}{3} = 35$. Remembering to multiply by 4 (as there are four different people who could have 11 or more) we get 140.

Thus the number of ways to divide 15 identical cookies among 4 distinct people such that each person gets at most 10 cookies is $\binom{18}{3} - 4\binom{7}{3} = 816 - 140 = 676$.

As always, it's good to write some simple code to check. The following is *not* the most efficient, but it runs very fast as the numbers are small, and it's easily coded.

```
count = 0;
For[x1 = 0, x1 <= 10, x1++,
  For[x2 = 0, x2 <= 10, x2++,
   For[x3 = 0, x3 <= 10, x3++,
    If[15 - x1 - x2 - x3 >= 0 && 15 - x1 - x2 - x3 <= 10,
     count = count + 1]
    ]]];
Print[count];
```

#7: Find a discrete random variable, or prove none exists, with probability density function f_X such that $f_X(x) = 2$ for some x between 17 and 17.01.

Solution: There is no such discrete random variable. The probability of any event is at most 1, and this would assign a probability of 2 to an event.

#8: Find a continuous random variable, or prove none exists, with probability density function f_X such that $f_X(x) = 2$ for all x between 17 and 17.01.

Solution: While we cannot assign a probability greater than 1 to an event, for a continuous random variable it is possible for the probability density function to exceed 1, as probabilities are found by integrating the pdf over intervals. Thus, so long as the interval is short, we can have $f_X(x) = 2$. The simplest example is a uniform distribution on the interval $[17, 17.5]$. If we take

$$f_X(x) \;=\; \begin{cases} 2 & \text{if } 17 \leq x \leq 17.5 \\ 0 & \text{otherwise} \end{cases}$$

then $\int_{-\infty}^{\infty} f_X(x)dx = 1$, $f_X(x) \geq 0$ and $f_X(x) = 2$ from 17 to 17.01.

#9: Let X be a continuous random variable with pdf f_X satisfying $f_X(x) = f_X(-x)$. What can you deduce about F_X, the cdf?

Solution: We must have $F_X(0) = 1/2$. The reason is the evenness of the pdf implies that half the probability is before 0, and half after. To see this mathematically, note

$$1 \;=\; \int_{-\infty}^{\infty} f_X(x)dx \;=\; \int_{-\infty}^{0} f_X(x)dx + \int_{0}^{\infty} f_X(x)dx \;=\; \int_{-\infty}^{0} f_X(-x)dx + \int_{0}^{\infty} f_X(x)dx.$$

Let's change variables in the first integral. Letting $t = -x$ we see $dx = -dt$ and the integration runs from $t = \infty$ to $t = 0$; we can thus use the minus sign in $-dt$ to have the integral range from 0 to infinity, and we find

$$1 \;=\; \int_{0}^{\infty} f_X(t)dt + \int_{0}^{\infty} f_X(x)dx \;=\; 2\int_{0}^{\infty} f_X(x)dx;$$

thus half the probability is after zero (and similarly half the probability is before).

#10: Find if you can, or say why you cannot, the first five Taylor coefficients of (a) $\log(1-u)$ at $u = 0$; (b) $\log(1-u^2)$ at $u = 0$; (c) $x\sin(1/x)$ at $x = 0$.

Solution: (a) Taking derivatives we find that if $f(u) = \log(1-u)$ then $f'(u) = -(1-u)^{-1}$, $f''(u) = -(-1)(1-u)^{-2}(-1) = -(1-u)^{-2}$, $f'''(u) = -2(1-u)^{-3}$ and $f''''(u) = -3\cdot 2(1-u)^{-4}$. At $u = 0$ we find $f(0) = 0$, $f'(0) = -1$, $f''(0) = -1$,

$f'''(0) = -2$ and $f''''(0) = -6$. Thus the fourth order Taylor series is

$$T_4(u) \quad = \quad f(0) + f'(0)u + \frac{f''(0)}{2!}u^2 + \frac{f'''(0)}{3!}u^3 + \frac{f''''(0)}{4!}u^4 \ = \ -u - \frac{u^2}{2} - \frac{u^3}{3} - \frac{u^4}{4}.$$

(b) There are lots of ways to do this problem, but my favorite is the 'lazy' way. If instead of $\log(1-u)$ we had $\log(1+u)$, all that we need do is replace u with $-u$ above, and we get

$$\log(1+u) \ = \ u - \frac{u^2}{2} + \frac{u^3}{3} - \frac{u^4}{4} + \cdots .$$

Why is this helpful? Note

$$\log(1-u^2) \ = \ \log\left((1-u)(1+u)\right) \ = \ \log(1-u) + \log(1+u).$$

Thus there's no need to go through the Taylor series arguments again; we can simply combine our two expansions and we find

$$\log(1-u^2) \ = \ -u^2 - \frac{u^4}{2} - \cdots .$$

Of course, there's another way we could have found this; we could take the Taylor series expansion for $\log(1-u)$ and substitute u^2 for u. The point is that if we spend some time thinking about our problem, we can often eliminate the need to do a lot of tedious algebra; however, if you don't see these simplifications you can still solve the problem, just with more work. For example, if we let $f(u) = \log(1-u^2)$ then $f'(u) = -2u/(1-u^2)$, and then the quotient rule and some algebra gives $f''(u) = -2(1+u^2)/(1-u^2)$, $f'''(u) = -4u(3+u^2)/(1-u^2)$ and so on.

(c) This function is *not* differentiable at the origin, though it is continuous at zero (as $x \to 0$, $x\sin(1/x) \to 0$ as $|\sin(1/x)| \le 1$ and $|x| \to 0$). The only way to make this function continuous at zero is to define it to be zero there; this is reasonable as $x\sin(1/x)$ does go to zero as long as $x \to 0$. To find the derivative of $f(x) = x\sin(1/x)$ at the origin we use the limit formula:

$$\lim_{h\to 0} \frac{f(0+h)-f(0)}{h} \ = \ \lim_{h\to 0} \frac{h\sin(1/h)}{h} \ = \ \lim_{h\to 0} \sin(1/h);$$

however, this last limit does not exist and thus our function is not differentiable. If it isn't differentiable, it can't have a Taylor series expansion.

#11: Let X be a continuous random variable. (a) Prove F_X is a non-decreasing function; this means $F_X(x) \le F_X(y)$ if $x < y$. (b) Let U be a random variable with cdf $F_U(x) = 0$ if $u < 0$, $F_U(x) = x$ if $0 < x < 1$, and $F_U(x) = 1$ if $1 < x$. Let F be any continuous function such that F is strictly increasing and the limit as x approaches negative infinity of $F(x)$ is 0 and the limit as x approaches positive infinity is 1. Prove $Y = F^{-1}(U)$ is a random variable with cdf F.

Solution: (a) Assume $x < y$. Then $F_X(y) = \text{Prob}(X \le y)$ while $F_X(x) = \text{Prob}(X \le x)$. Therefore

$$F_X(y) - F_X(x) \ = \ \text{Prob}(x < X \le y) \ \ge \ 0$$

(it is non-negative as it is a probability, and probabilities are non-negative). If $F_X(y) - F_X(x) \ge 0$ then $F_X(y) \ge F_X(x)$, which proves F_X is a non-decreasing function.

(b) This is perhaps one of the most important problems in the entire course! As F is continuous and strictly increasing, it has a continuous inverse F^{-1}. Note $\mathbb{P}(Y \le y) = \mathbb{P}(F^{-1}(U) \le y)$; however, $F^{-1}(U) \le y$ means $U \le F(y)$. Then $\mathbb{P}(Y \le y)$ equals $\mathbb{P}(U \le F(y))$; as $F(y) \in [0,1]$, from the givens of the problem $\mathbb{P}(U \le F(y)) = F(y)$, which completes the proof.

Why is this problem so important? One way of interpreting the result is to say that if we can simulate any random variable that is uniformly distributed (or equidistributed) on $[0,1]$, then we can simulate any random variable whose cumulative distribution function is strictly increasing. Of course, how does one generate a random number uniformly? This is a very hard question. See for instance `http://www.random.org/`.

Let's do a bit more with this problem. Consider the Cauchy distribution, where the density is $f_Y(y) = \frac{1}{\pi}\frac{1}{1+y^2}$. The cumulative distribution function is the integral of f_Y from $-\infty$ to y:

$$F_Y(y) \ = \ \int_{-\infty}^{y} \frac{1}{\pi}\frac{dt}{1+t^2} \ = \ \frac{\arctan(y) - \arctan(-\infty)}{\pi} \ = \ \frac{\arctan(y) + \pi/2}{\pi}.$$

We need F_Y^{-1}; setting $F_Y(y) = u$ we can solve for y in terms of u:

$$u \ = \ \frac{\arctan(y) + \pi/2}{\pi} \quad \Rightarrow \quad y \ = \ \tan(\pi u - \pi/2) \ = \ F_Y^{-1}(u).$$

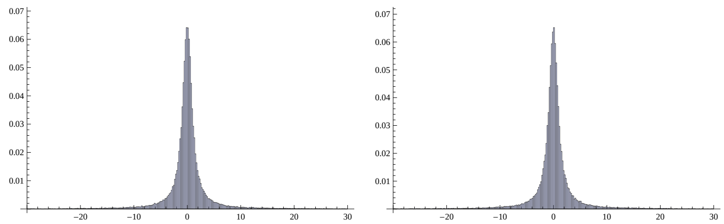

FIGURE 1. Comparing methods to simulate Cauchy random variables. Left is using the inverse CDF method, right is using Mathematica's built in function.

We have our strictly increasing function F_Y^{-1}, and can now simulate from a Cauchy. This is amazing, as the Cauchy has infinite variance!

Below is some Mathematica code to simulate from the Cauchy. We go through a few ways to display the data, as there are issues in comparing a histogram of discrete data to continuous data drawn from a Cauchy.

```
Finv[u_] := Tan[Pi u - Pi/2];
temp = {};
prunedtemp = {};
truncatetemp = {};
num = 100000;
For[n = 1, n <= num, n++,
  {
    y = Finv[Random[]];
    temp = AppendTo[temp, y];
    If[Abs[y] <= 30,
     {
      prunedtemp = AppendTo[prunedtemp, y];
      t = Floor[y - .5]/1 + .5;
      truncatetemp = AppendTo[truncatetemp, t];
      }];

    };];
Print[Length[prunedtemp] 100. / Length[temp]];
Print[Histogram[temp, Automatic, "Probability"]];
Print[Histogram[prunedtemp, Automatic, "Probability"]];
Print[Histogram[truncatetemp, Automatic, "Probability"]];
Print[Plot[{.2, (1/Pi) 1/(1 + x^2)}, {x, -30, 30}]];
```

Of course, Mathematica has the ability to directly simulate from Cauchy distributions.

```
ctemp = {};
ptemp = {};
For[n = 1, n <= 100000, n++,
  {
   y = Random[CauchyDistribution[0, 1]];
   ctemp = AppendTo[ctemp, y];
   If[Abs[y] < 30, ptemp = AppendTo[ptemp, y]];
   }];
Print[Histogram[ptemp, Automatic, "Probability"]]
```

We compare the two methods in Figure 1.

3.3. **Assignment #5: Due October 19, 2018:** #1: We toss n fair coins. Every coin that lands on heads is tossed again. What is the probability density function for the number of heads after the second set of tosses (i.e., after we have retossed all the coins that landed on heads)? If you want, imagine you have left the room and return after all the tossing is done; what is the pdf for the number of heads you see? #2: Is there a C such that $f(x) = C\exp(-x - \exp(-x))$ is a probability density function? Here $-\infty < x < \infty$. #3: Let X be a discrete random variable. Prove or disprove: $\mathbb{E}[1/X] = 1/\mathbb{E}[X]$. #4: Let $X_1, \dots, X_n$ be independent, identically distributed random variables that have zero probability of taking on a non-positive value. Prove $\mathbb{E}[(X_1 + \cdots + X_m)/(X_1 + \cdots + X_n)] = m/n$ for $1 \le m \le n$. Does this result seem surprising? Write a computer program to investigate when the random variables are drawn from a uniform distribution on $[0,1]$. #5: Let X and Y be two continuous random variables with densities f_X and f_Y. (a) For what c is $cf_X(x) + (1-c)f_Y(x)$ a density? (b) Can there be a continuous random variable with pdf equal to $f_X(x)f_Y(x)$? #6: The standard normal has density $\frac{1}{\sqrt{2\pi}} \exp(-x^2/2)$ (this means this integrates to 1). Find the first four moments. #7: Find the errorS in the following code:

```
hoops[p_, q_, need_, num_] := Module[{},
   birdwin = 0;
   For[n = 1, n <= num,
    {
     If[Mod[n, num/10] == 0, Print["We have done ", 100. n/num, "%."]];
     birdbasket = 0;
     magicbasket = 0;
     While[birdbasket  < need || magicbasket < need,
      {
       If[Random[] <= p, birdbasket = birdbasket + 1];
       If[Random[] <= q, magicbasket = magicbasket + 1];
       }]; (* end of while loop *)
     If[birdbasket == need, birdwin == birdwin + 1];
     }]; (* end  of for loop *)
   Print["Bird  wins ", 100. birdwin/num, "%."];
   Print["Magic wins ", 100. - 100. birdwin/num, "%."];
   ];

hoops[.32, .33, 5, 100]
```

4. HW #5: Due October 19, 2018

4.1. **Assignment:** #1: We toss n fair coins. Every coin that lands on heads is tossed again. What is the probability density function for the number of heads after the second set of tosses (i.e., after we have retossed all the coins that landed on heads)? If you want, imagine you have left the room and return after all the tossing is done; what is the pdf for the number of heads you see? #2: Is there a C such that $f(x) = C\exp(-x - \exp(-x))$ is a probability density function? Here $-\infty < x < \infty$. #3: Let X be a discrete random variable. Prove or disprove: $\mathbb{E}[1/X] = 1/\mathbb{E}[X]$. #4: Let $X_1, \dots, X_n$ be independent, identically distributed random variables that have zero probability of taking on a non-positive value. Prove $\mathbb{E}[(X_1 + \cdots + X_m)/(X_1 + \cdots + X_n)] = m/n$ for $1 \le m \le n$. Does this result seem surprising? Write a computer program to investigate when the random variables are drawn from a uniform distribution on $[0,1]$. #5: Let X and Y be two continuous random variables with densities f_X and f_Y. (a) For what c is $cf_X(x) + (1-c)f_Y(x)$ a density? (b) Can there be a continuous random variable with pdf equal to $f_X(x)f_Y(x)$? #6: The standard normal has density $\frac{1}{\sqrt{2\pi}}\exp(-x^2/2)$ (this means this integrates to 1). Find the first four moments. #7: Find the errorS in the following code:

```
hoops[p_, q_, need_, num_] := Module[{},
   birdwin = 0;
   For[n = 1, n <= num,
    {
     If[Mod[n, num/10] == 0, Print["We have done ", 100. n/num, "%."]];
     birdbasket = 0;
     magicbasket = 0;
     While[birdbasket  < need || magicbasket < need,
      {
       If[Random[] <= p, birdbasket = birdbasket + 1];
       If[Random[] <= q, magicbasket = magicbasket + 1];
       }]; (* end of while loop *)
     If[birdbasket == need, birdwin == birdwin + 1];
     }]; (* end  of for loop *)
   Print["Bird  wins ", 100. birdwin/num, "%."];
   Print["Magic wins ", 100. - 100. birdwin/num, "%."];
   ];

hoops[.32, .33, 5, 100]
```

4.2. **Solutions:**

#1: We toss n fair coins. Every coin that lands on heads is tossed again. What is the probability density function for the number of heads after the second set of tosses (i.e., after we have retossed all the coins that landed on heads)?
Solution: We solve this problem two ways. The first is the 'natural' approach. It has the advantage of being a reasonable method to try, but leads to a very messy formula. It's not that much more work to solve when the coin isn't fair, so let's assume there's a probability p of heads and $1-p$ of tails. There's another advantage to this. If the coin is fair, $(1/2)^m(1/2)^{n-m} = (1/2)^n$, and behavior is blended; if the coin is biased, we have $p^m(1-p)^m$, and this might focus our thoughts on the process a bit more.

Our first solution uses conditional probability. Let's say we want to compute all the ways of having m heads on the second toss, with clearly $0 \le m \le n$. We can express this probability as

$$\mathbb{P}(m \text{ heads at end}) \ = \ \sum_{k=m}^{n} \mathbb{P}(m \text{ heads on second toss}|k \text{ heads on first}) \cdot \mathbb{P}(k \text{ heads on first toss}).$$

Why? We must have tossed some number of heads on the first toss, which we denote by k. Clearly $k \ge m$ as otherwise we can't have m heads on the second. The answer is thus

$$\sum_{k=m}^{n} \binom{k}{m} p^m (1-p)^{k-m} \cdot \binom{n}{k} p^k (1-p)^{n-k}.$$

It is worth asking what would happen if we forgot about the restriction that $m < n$; for example, what if $n = 4$ and $m = 6$? We would have the binomial coefficient $\binom{4}{6}$ – how is this defined? We might at first expect it to be $\frac{4!}{6!(-2)!}$; this works but you need to know that $(-2)!$ is defined to be infinity! We'll discuss this later when we talk about the Gamma function, which generalizes the factorial function. There is another way to 'see' what the definition should be. We expect the answer to be zero, as the combinatorial interpretation is: *how many ways are there to choose 6 objects from 4 when order doesn't matter?* Clearly there are *no* such ways, and thus the answer should be zero. Another way of defining $\binom{n}{k}$ is

$$\frac{n(n-1)\cdots(n-(k-1))}{k(k-1)\cdots 1}.$$

In our case, we would have

$$\binom{4}{6} \ = \ \frac{4\cdot 3\cdot 2\cdot 1\cdot 0\cdot(-1)}{6\cdot 5\cdot 4\cdot 3\cdot 2\cdot 1} \ = \ 0$$

as we have a 0 in the numerator.

Remember, in mathematics we can make almost any definition we want – the question is when our definition is useful. The above is a great way to define the choose function when the bottom exceeds the top, and agrees with our combinatorial intuition.

Let's see if we can simplify the sum a bit. We have

$$\begin{aligned}
\mathbb{P}(m \text{ heads at end}) &= \sum_{k=m}^{n}\binom{k}{m}p^m(1-p)^{k-m}\cdot\binom{n}{k}p^k(1-p)^{n-k}\\
&= \sum_{k=m}^{n}\frac{k!}{m!(k-m)!}\frac{n!}{k!(n-k)!}p^m(1-p)^{n-m}p^k\\
&= p^m(1-p)^{n-m}\sum_{k=m}^{n}\frac{n!}{m!(k-m)!(n-k)!}p^k\\
&= p^m(1-p)^{n-m}\sum_{k=m}^{n}\frac{n!}{m!(n-m)!}\frac{(n-m)!}{(k-m)!(n-k)!}p^k,
\end{aligned}$$

where in the last step we multiplied by 1 in the form $1=(n-m)!/(n-m)!$. Why would we do this? When looking at the ratio of the factorials we notice an $n!/m!$; this is *almost* $\binom{n}{m}$. It would be, except it's missing an $(n-m)!$ in the denominator. Thus, we must multiply by $(n-m)!/(n-m)!$ so we can recognize the binomial coefficient. Notice that at the end of the day we want exactly m heads out of n coins, and thus we should be thinking of an $\binom{n}{m}$ somewhere. Further, the factor $p^m(1-p)^{n-m}$ outside is right in line with such an interpretation.

We now continue simplifying the algebra. We change summation variables and let $\ell = k-m$ (so $k=m+\ell$). Since k runs from m to n we have ℓ runs from 0 to $n-m$, and p^k becomes $p^{\ell+m}$. We find

$$\begin{aligned}
\mathbb{P}(m \text{ heads at end}) &= \binom{n}{m}p^m(1-p)^{n-m}\sum_{\ell=0}^{n-m}\frac{(n-m)!}{\ell!(n-m-\ell)!}p^{m+\ell}\\
&= \binom{n}{m}p^m(1-p)^{n-m}p^m\sum_{\ell=0}^{n-m}\binom{n-m}{\ell}p^\ell 1^{n-m-\ell}\\
&= \binom{n}{m}p^m(1-p)^{n-m}p^m(1+p)^{n-m},
\end{aligned}$$

where we wrote 1^{n-m} to highlight the application of the binomial theorem. Note $1-p$ and $1+p$ are both to the $n-m$ power; combining them gives $(1-p^2)^{n-m}$, and we obtain our final simplification:

$$\mathbb{P}(m \text{ heads at end}) \ = \ \binom{n}{m}(p^2)^m(1-p^2)^{n-m}.$$

Notice that this is the density of a binomial random variable with probability p^2 of success and thus $1-p^2$ of failure. As this is such a beautiful answer with such a nice interpretation, it is highly suggestive that there is a *much better approach to this problem* then the algebraic nightmare we did above!

We now give an alternate solution. While the problem says we only re-flip the coins that landed heads initially, we can re-flip all if we want, but only count as a 'heads' coins that are heads on both tosses. A much better way to look at this problem is to think what must happen for a coin to end up heads after two tosses. The only way this can occur is if the first and second tosses are heads, which (since the coin lands on heads with probability p) happens with probability $p \cdot p = p^2$. Our situation turns out to be equivalent to the following: *Toss a biased coin (with probability p^2 of landing on heads) a total of n times; what is the probability mass function?* The answer is just

$$\mathbb{P}(m \text{ heads}) \ = \ \binom{n}{m} \left(p^2\right)^m \left(1-p^2\right)^{n-m} \ = \ \binom{n}{m} p^{2m}(1-p^2)^{n-m}.$$

The above analysis illustrates one of the most common ways to prove combinatorial identities. Namely, we calculate a given quantity two different ways. As both count the same object, they must be equal. Typically one is easily computed, and thus the other, harder combinatorial expression must equal the easier one. For example, in our case above the second approach was fairly easy to compute. If we take $p = 1/2$ and set the first and second solutions equal to each other, we find

$$\sum_{k=m}^{n} \binom{k}{m}\binom{n}{k}\left(\frac{1}{2}\right)^{n+k} \ = \ \binom{n}{m}\frac{3^{n-m}}{2^{2n}}.$$

We can verify this identity for any choices of $m \le n$; however, is there a way of proving this directly (and not relying on us being clever and noticing this counting problem was equivalent to another)?

#2: Is there a C such that $f(x) = C\exp(-x - \exp(-x))$ is a probability density function? Here $-\infty < x < \infty$.
Solution: Our proposed density is again non-negative, so the question is just whether or not it will integrate to 1 for some choice of C. We have

$$\int_{-\infty}^{\infty} C\exp(-x - \exp(-x))dx \ = \ C\int_{-\infty}^{\infty} \exp(-x)\exp(-\exp(-x))dx.$$

We do a u substitution. Let

$$u \ = \ \exp(-\exp(-x))$$

so

$$du \ = \ \exp(-x)\exp(-\exp(-x))dx,$$

and $x : -\infty \to \infty$ becomes $u : 0 \to 1$. Thus our integral is

$$C\int_0^1 du \ = \ 1.$$

There are other change of variables we could make, but this is the simplest. The integral is thus equal to 1 if $C = 1$.

#3: Let X be a discrete random variable. Prove or disprove: $\mathbb{E}[1/X] = 1/\mathbb{E}[X]$.
Solution: Usually $\mathbb{E}[1/X]$ is not $1/\mathbb{E}[X]$. Almost anything is a counter-example. A trivial one is to take $X = \pm 1$ with probability 1/2 for each. Another example is to take $X = 2$ or 4 with probability 1/2 for each, as

$$\mathbb{E}[1/X] \ = \ \frac{1}{2}\cdot\frac{1}{2} + \frac{1}{4}\cdot\frac{1}{2} \ = \ \frac{3}{8},$$

while

$$\frac{1}{\mathbb{E}[X]} \ = \ \frac{1}{2\cdot\frac{1}{2} + 4\cdot\frac{1}{2}} \ = \ \frac{1}{3}.$$

It is possible for them to be equal – this is always the case if $X = x$ with probability 1 for some non-zero x. Assume we have $X = x_i$ with probability p_i for $i \in \{1, 2\}$ and we want these two to be equal. As $p_2 = 1 - p_1$, letting $p = p_1$ that then requires

$$\frac{p}{x_1} + \frac{1-p}{x_2} \ = \ \frac{1}{x_1 p + x_2(1-p)}$$

or

$$\frac{x_1(1-p) + px_2}{x_1x_2} \ = \ \frac{1}{x_1p + x_2(1-p)},$$

which simplifies to

$$(x_1(1-p) + px_2)\left(x_1p + x_2(1-p)\right) - x_1x_2 \ = \ 0.$$

Are there any non-trivial solutions to this? We have three unknowns and only one equation, so this should be solvable. Of course, we do have restrictions: $0 < p < 1$ and $x_1 \neq x_2$. (We take $p \neq 0, 1$ as otherwise this reduces to the trivial solution.)

By symmetry, so long as $x_1 \neq 0$ we can take $x_1 = 1$ (this is just multiplying both sides by x_1). This reduces our equation to

$$-x_2 + (p + (1-p)x_2)(1 - p + px_2) \ = \ 0;$$

unfortunately this equation has a double root at 1 and thus there are no non-zero solutions where the probability is concentrated on two distinct masses.

#4: Let $X_1, \dots, X_n$ be independent, identically distributed random variables that have zero probability of taking on a non-positive value. Prove $\mathbb{E}[(X_1 + \cdots + X_m)/(X_1 + \cdots + X_n)] = m/n$ for $1 \leq m \leq n$. Does this result seem surprising? Write a computer program to investigate when the random variables are drawn from a uniform distribution on $[0, 1]$.

Solution: This is one of my favorite problems. At first the answer seems too good to be true, as it is independent of the distribution of the X_i's! All that matters is that they are identically distributed and that the sum is non-zero (so the division makes sense). Let X have the same distribution as the X_i's. The key technique here is to multiply by 1. We start with

$$\mathbb{E}\left[\frac{1}{1}\right] \ = \ 1;$$

this trivial observation is the key to the proof. We now write 1/1 in a clever way, and use **linearity of expectation**:

$$\begin{aligned} 1 &= \mathbb{E}\left[\frac{X_1 + \cdots + X_n}{X_1 + \cdots + X_n}\right] \\ &= \mathbb{E}\left[\sum_{k=1}^{n} \frac{X_k}{X_1 + \cdots + X_n}\right] \\ &= \sum_{k=1}^{n} \mathbb{E}\left[\frac{X_k}{X_1 + \cdots + X_n}\right] \\ &= n\mathbb{E}\left[\frac{X}{X_1 + \cdots + X_n}\right], \end{aligned}$$

and so

$$\mathbb{E}\left[\frac{X}{X_1 + \cdots + X_n}\right] \ = \ \mathbb{E}\left[\frac{X_k}{X_1 + \cdots + X_n}\right] \ = \ \frac{1}{n}.$$

The key step above is that as the X_k's are identically distributed, the expected value of any one of them over the sum is the same as that of any other over the sum. We now calculate the quantity of interest:

$$\mathbb{E}\left[\frac{X_1 + \cdots + X_m}{X_1 + \cdots + X_n}\right] \ = \ \sum_{k=1}^{m} \mathbb{E}\left[\frac{X_k}{X_1 + \cdots + X_n}\right] \ = \ \frac{m}{n}.$$

Note an alternative way to view our solution is to do the case $m = n$ first; this is a natural choice, as then the fraction is just 1.

Here is some code.

```
ratiotest[m_, n_, numiter_, listwork_] := Module[{},
  (* m and n parameters from problem *)
  (* numiter is number of times do it *)
  (* if listwork = 1 we save each run and do a histogram at end *)
  (* initialize list and sum of ratios to 0 *)
  list = {};
  sumratio = 0;
  (* loop numiter times *)
  For[i = 1, i <= numiter, i++,
   {
    (* print out every ten percent *)
    If[Mod[i, numiter/10] == 0,
     Print["Have done ", 100. i/numiter, "%."]];
    (* calculates the numerator and denominator, sums of unif rvs *)
```

```
    numerator = Sum[Random[], {k, 1, m}];
    denominator =
     numerator + If[m == n, 0, Sum[Random[], {j, 1, n - m}]];
    (* calculates the ratio, updates sum of ratios *)
    ratio  = numerator 1.0 / denominator;
    sumratio  = sumratio + ratio;
    (* if listwork is 1 saves to list *)
    If[listwork == 1, list = AppendTo[list, ratio]];
    }]; (* end of i loop *)
  (* calculates average ratio, prints results *)
  averatio = sumratio / numiter;
  Print["Average ratio for m = ", m, " and n = ", n, " is ",
   1.00 averatio];
  Print["Compare to m/n = ", m/n, " = ", 1.0 m/n];
  If[listwork == 1,
   {
    Print[Histogram[list, Automatic, "Probability"]];
    }];
  ]
```

#5: Let X and Y be two continuous random variables with densities f_X and f_Y. (a) For what c is $cf_X(x) + (1-c)f_Y(x)$ a density? (b) Can there be a random variable with pdf equal to $f_X(x)f_Y(x)$?

Solution: (a) It is definitely a density when $c \in [0,1]$, as then the function is non-negative and integrates to 1:

$$\int_{-\infty}^{\infty} [cf_X(x) + (1-c)f_Y(x)]\, dx \;=\; c\int_{-\infty}^{\infty} f_X(x)dx + (1-c)\int_{-\infty}^{\infty} f_Y(x)dx \;=\; c + (1-c) \;=\; 1.$$

It's possible for it to work for all c (it does if $f_X = f_Y$). If, however, $c \notin [0,1]$ then it is always possible to find a pair of densities such that $cf_X(x) + (1-c)f_Y(x)$ is *not* a density. To see this, just take

$$f_X(x) \;=\; \begin{cases} 1 & \text{if } 0 \le x \le 1 \\ 0 & \text{otherwise} \end{cases}$$

$$f_Y(x) \;=\; \begin{cases} 1 & \text{if } 2 \le x \le 3 \\ 0 & \text{otherwise.} \end{cases}$$

Note that if $c \notin [0,1]$ then this density is negative for some x. For example, if $c < 0$ then the $cf_X(x)$ term is negative for $0 \le x \le 1$, while if $c > 1$ the second factor is negative for $2 \le x \le 3$. Thus, while it is possible to be a density for certain choices of f_X and f_Y, the only choices of c such that it is *always* a density are $0 \le c \le 1$.

(b) It's not always the case that $f_X(x)f_Y(x)$ is a density. A nice example is f is the uniform density on $[0,1]$ and g the uniform density on $[2,3]$. Then

$$f_X(x) \;=\; \begin{cases} 1 & \text{if } 0 \le x \le 1 \\ 0 & \text{otherwise} \end{cases}$$

and

$$f_Y(x) \;=\; \begin{cases} 1 & \text{if } 2 \le x \le 3 \\ 0 & \text{otherwise.} \end{cases}$$

Then $f(x)g(x) = 0$ for all x. It's often a good idea to play around searching for counterexamples, or seeing what makes examples succeed. Just because f and g are non-negative and integrate to 1, nothing implies the same must be true for their product.

Of course, the problem only asks whether or not there *can* be a random variable with pdf equal to the product $f_X(x)f_Y(x)$, not whether or not the product *must* be a density. There are examples where this is a density. The simplest is $f_X(x) = f_Y(x)$ for $0 \le x \le 1$ and 0 otherwise; note in this case $f_X(x)f_Y(x) = 1$.

A more interesting example is

$$f_X(x) \ = \ \begin{cases} 2 & \text{if } 0 \le x \le 1/2 \\ 0 & \text{otherwise} \end{cases} \qquad f_Y(x) \ = \ \begin{cases} 2 & \text{if } 1/4 \le x \le 3/4 \\ 0 & \text{otherwise.} \end{cases}$$

Note

$$f_X(x) f_Y(x) \ = \ \begin{cases} 4 & \text{if } 1/4 \le x \le 1/2 \\ 0 & \text{otherwise,} \end{cases}$$

which is a density function (it is non-negative and integrates to 1). The continuous case is very different from the discrete case. In the discrete case, the only way we can have a solution is if all the mass of each is concentrated at one point. The reason is the probabilities are multiplied and *must* decrease as the probabilities are at most 1; in the continuous case, the density can be greater than 1 at a point or in short intervals. This is one of the reasons for the earlier problem on whether or not densities can exceed 1 at a point.

#6: The standard normal has density $\frac{1}{\sqrt{2\pi}} \exp(-x^2/2)$ (this means this integrates to 1). Find the first four moments.
Solution: We need to compute

$$\int_{-\infty}^{\infty} x^k \cdot \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx$$

for $k \in \{1, 2, 3, 4\}$. There are several ways to proceed. First, notice that since the integrand is odd when k is odd and the region is symmetric about the symmetry point, the integral vanishes for $k = 1$ or 3, while for the even values it's just double the integral from 0 to ∞.

One way to finish the problem is to use u-substitution. Given

$$\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} x^{2\ell} e^{-x^2/2} dx$$

(for $k = 2\ell$) we see if we let $u = x^2$ then $du = 2xdx$ and the integral equals

$$\frac{2}{\sqrt{2\pi}} \int_{-\infty}^{\infty} x^{\ell-1} e^{-u/2} 2du.$$

There's no problem if $\ell = 1$. IF $\ell = 2$ we integrate by parts. Doing the algebra we find the second moment is 1 and the fourth moment is 3. The algebra isn't too bad because we already did the u-substitution by replacing x^2 with u; this lead to terms like $e^{-u/2}$ instead of $e^{-x^2/2}$. This is very important, as there is no closed form for an anti-derivative of $e^{-x^2/2}$. We discuss this issue a bit in some of the integration exercises in the book.

Now that we have *one* solution, let's look for another. We define

$$I(k) \ := \ \int_{-\infty}^{\infty} x^k \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx \ = \ \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} x^k e^{-x^2/2}.$$

While we could do the u-substitution, let's see what happens if we don't. We try to integrate by parts. This is a good thing to try. The reason is that our integrand is *rapidly* decaying as $x \to \pm\infty$, so we won't have to worry about boundary terms. We have to decide what to make u and what to make dv. We want the polynomial to go down in degree, so it's natural to think of setting $u = x^k$, but this doesn't work. The issues is then $dv = e^{-x^2/2} dx$, and we can't integrate that. We need xdx not x (really, $-xdx$). So, we pull off one factor of x from x^k, and write

$$I(k) \ = \ \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} x^{k-1} \cdot e^{-x^2/2} x dx.$$

We now set

$$u \ = \ x^{k-1}, \quad dv \ = \ e^{-x^2/2} x dx$$

and find

$$du \ = \ (k-1) x^{k-2}, \quad v \ = \ -e^{-x^2/2}.$$

Using

$$\int_{-\infty}^{\infty} u dv \ = \ uv \Big|_{-\infty}^{\infty} - \int_{-\infty}^{\infty} v du,$$

we obtain

$$I(k) \;=\; \frac{1}{\sqrt{2\pi}}\left[-x^{k-1}e^{-x^2/2}\Big|_{-\infty}^{\infty} + \int_{-\infty}^{\infty}(k-1)x^{k-2}e^{-x^2/2}dx\right] \;=\; (k-1)I(k-2)$$

(note the boundary terms vanish as the exponential decay dwarfs the polynomial growth). We've discovered a recurrence relation:

$$I(k) \;=\; (k-1)I(k-2).$$

We can keep iterating this. Each time we decrease the index by 2. We know that $I(1) = 0$ (as the integrand is odd) and $I(0) = 1$ (as it is a probability density function, it integrates to 1). Thus $I(k) = 0$ if k is odd, while if $k = 2\ell$ then

$$I(2\ell) \;=\; (2\ell-1)I(2\ell-2) \;=\; (2\ell-1)(2\ell-3)I(2\ell-4) \;=\; (2\ell-1)(2\ell-3)(2\ell-5)I(2\ell-6) \;=\; \cdots.$$

We continue until we hit $I(0) = 1$. Recalling the definition of the double factorial ($(2m)!! = (2m-1)(2m-3)(2m-5)\cdots 3\cdot 1$), we see $I(2\ell) \;=\; (2\ell-1)!!$; in particular, $I(2) \;=\; 1!! \;=\; 1, \qquad I(4) \;=\; 3!! \;=\; 3\cdot 1 \;=\; 3.$

It's interesting to see a *combinatorial* quantity arising in the moments of our density; it turns out this has *profound* implications. In other words, this was not a busy-work problem!

#7: Find the errorS in the following code:

```
hoops[p_, q_, need_, num_] := Module[{},
   birdwin = 0;
   For[n = 1, n <= num,
    {
     If[Mod[n, num/10] == 0, Print["We have done ", 100. n/num, "%."]];
     birdbasket = 0;
     magicbasket = 0;
     While[birdbasket  < need || magicbasket < need,
      {
       If[Random[] <= p, birdbasket = birdbasket + 1];
       If[Random[] <= q, magicbasket = magicbasket + 1];
       }]; (* end of while loop *)
     If[birdbasket == need, birdwin == birdwin + 1];
     }]; (* end  of for loop *)
   Print["Bird  wins ", 100. birdwin/num, "%."];
   Print["Magic wins ", 100. - 100. birdwin/num, "%."];
   ];

hoops[.32, .33, 5, 100]
```

Solution: The for loop needs an n++, it should be && (for and) not || (for or) in the while statement, and in the If statement involving birdbasket use a single = to assign birdwin + 1 to birdwin. The correct code is:

```
hoops[p_, q_, need_, num_] := Module[{},
   birdwin = 0;
   For[n = 1, n <= num, n++,
    {
     birdbasket = 0;
     magicbasket = 0;
     While[birdbasket  < need && magicbasket < need,
      {
       If[Random[] <= p, birdbasket = birdbasket + 1];
       If[Random[] <= q, magicbasket = magicbasket + 1];
       }]; (* end of while loop *)
     If[birdbasket == need, birdwin = birdwin + 1];
     }]; (* end  of for loop *)
   Print["Bird  wins ", 100. birdwin/num, "%."];
   Print["Magic wins ", 100. - 100. birdwin/num, "%."];
   ];
```

4.3. **Assignment #6:** Due October 26, 2018: #1: Calculate the second and third moments of X when $X \sim \mathrm{Bin}(n, p)$ (a binomial random variable with parameters n and p). #2: We toss N coins (each of which is heads with probability p), where the number N is drawn from a Poisson random variable with parameter lambda. Let X denote the number of heads. What is the probability density function of X? Justify your answer. #3: Find the probability density function of Y when $Y = \exp(X)$ for $X \sim N(0, 1)$. #4: Each box of cereal is equally likely to have exactly one of a set of c prizes. Thus, every time you open a box you have a $1/c$ chance of getting prize 1, a $1/c$ chance of getting prize 2, ... How many boxes to you expect to have to open before you have at least one of each of the c prizes? If you have having trouble, do $c = 2$ for half credit. #5: Let $X_1, \dots, X_n$ be independent Bernoulli random variables where $X_k \sim \mathrm{Bern}(p_k)$ (you can think of this as n independent coin tosses, where coin k is heads with probability p). If $Y = X_1 + \cdots + X_n$, what is the mean and what is the variance of Y? Assume $p_1 + \cdots + p_n = \mu$; what choice or choices of the p_k's lead to the variance of Y being the largest possible? #6: State anything you learned or enjoyed in Arms' talk. One or two sentences suffice. #7: The kurtosis of a random variable X is defined by $\mathrm{kur}(X) := \mathbb{E}[(X-\mu)^4]/\sigma^4$, where μ is the mean and σ is the standard deviation. The kurtosis measures how much probability we have in the tails. If $X \sim \mathrm{Poiss}(\lambda)$, find the kurtosis of X. #8: Consider a coin with probability p of heads. Find the probability density function for X_1, where X_1 is how long we must weight before we get our *first* head. #9: Consider a coin with probability p of heads. Find the probability density function for X_2, where X_2 is how long we must weight before we get our *second* head. #10: Alice, Bob and Charlie are rolling a fair die in that order. They keep rolling until one of them rolls a 6. What is the probability each of them wins? #11: Alice, Bob and Charlie are rolling a fair die in that order. What is the probability Alice is the first person to roll a 6, Bob is the second and Charlie is the third? #12: Alice, Bob and Charlie are *still* rolling the fair die. What is the probability that the first 6 is rolled by Alice, the second 6 by Bob and the third 6 by Charlie? #13: What are the mean and variance of a chi-square distribution with 2 degrees of freedom? If $X \sim \chi^2(2)$, what is the probability that X takes on a value at least twice its mean? What is the probability X takes on a value at most half of its mean?

5. HW #6: Due October 26, 2018:

5.1. **Assignment #6:** Due October 26, 2018: #1: Calculate the second and third moments of X when $X \sim \mathrm{Bin}(n,p)$ (a binomial random variable with parameters n and p). #2: We toss N coins (each of which is heads with probability p), where the number N is drawn from a Poisson random variable with parameter lambda. Let X denote the number of heads. What is the probability density function of X? Justify your answer. #3: Find the probability density function of Y when $Y = \exp(X)$ for $X \sim N(0,1)$. #4: Each box of cereal is equally likely to have exactly one of a set of c prizes. Thus, every time you open a box you have a $1/c$ chance of getting prize 1, a $1/c$ chance of getting prize 2, ... How many boxes to you expect to have to open before you have at least one of each of the c prizes? If you have having trouble, do $c = 2$ for half credit. #5: Let $X_1, \dots, X_n$ be independent Bernoulli random variables where $X_k \sim \mathrm{Bern}(p_k)$ (you can think of this as n independent coin tosses, where coin k is heads with probability p). If $Y = X_1 + \cdots + X_n$, what is the mean and what is the variance of Y? Assume $p_1 + \cdots + p_n = \mu$; what choice or choices of the p_k's lead to the variance of Y being the largest possible? #6: State anything you learned or enjoyed in Arms' talk. One or two sentences suffice. #7: The kurtosis of a random variable X is defined by $\mathrm{kur}(X) := \mathbb{E}[(X-\mu)^4]/\sigma^4$, where μ is the mean and σ is the standard deviation. The kurtosis measures how much probability we have in the tails. If $X \sim \mathrm{Poiss}(\lambda)$, find the kurtosis of X. #8: Consider a coin with probability p of heads. Find the probability density function for X_1, where X_1 is how long we must weight before we get our *first* head. #9: Consider a coin with probability p of heads. Find the probability density function for X_2, where X_2 is how long we must weight before we get our *second* head. #10: Alice, Bob and Charlie are rolling a fair die in that order. They keep rolling until one of them rolls a 6. What is the probability each of them wins? #11: Alice, Bob and Charlie are rolling a fair die in that order. What is the probability Alice is the first person to roll a 6, Bob is the second and Charlie is the third? #12: Alice, Bob and Charlie are *still* rolling the fair die. What is the probability that the first 6 is rolled by Alice, the second 6 by Bob and the third 6 by Charlie? #13: What are the mean and variance of a chi-square distribution with 2 degrees of freedom? If $X \sim \chi^2(2)$, what is the probability that X takes on a value at least twice its mean? What is the probability X takes on a value at most half of its mean?

5.2. **Solutions:** #1: Calculate the second and third moments of X when $X \sim \mathrm{Bin}(n,p)$ (a binomial random variable with parameters n and p).

Solution: The problem only asks us to find $\mathbb{E}[X^2]$ and $\mathbb{E}[X^3]$, but we'll compute the centered moments $\mathbb{E}[(X-\mu)^2]$ and $\mathbb{E}[(X-\mu)^3]$ below, as this allows us to highlight more techniques and discuss more issues.

One natural way to compute these quantities is from the definition. To evaluate the second moment, we either need to compute $\mathbb{E}[(X-\mu)^2]$ or $\mathbb{E}[X^2] - \mathbb{E}[X]^2$. In the latter, this leads us to finding

$$\sum_{k=0}^{n} k^2 \cdot \binom{n}{k} p^k (1-p)^{n-k}.$$

While we can do this through differentiating identities, it is faster to use linearity of expectation. Let $X_1, \dots, X_n$ be i.i.d.r.v. (independent identically distributed random variables) with the Bernoulli distribution with parameter p. Note these are independent, and we have the probability X_i is 1 is p and the probability X_i is 0 is $1-p$. Let $X = X_1 + \cdots + X_n$. As they are independent, the variance of the sum is the sum of the variances:

$$\mathrm{Var}(X_1 + \cdots + X_n) \ = \ \mathrm{Var}(X_1) + \cdots + \mathrm{Var}(X_n) \ = \ np(1-p),$$

as the variance of each X_i is just $p(1-p)$. To see this, note

$$\mathbb{E}[(X_i - \mu_i)^2] \ = \ \mathbb{E}[(X_i - p)^2] \ = \ (1-p)^2 p + (0-p)^2(1-p) \ = \ p(1-p)(1-p+p) \ = \ p(1-p).$$

We redo the calculations in a way that will help with the analysis of the third moment. We have

$$\begin{aligned}
\mathbb{E}[X^2] &= \mathbb{E}[(X_1 + \cdots + X_n)^2] \\
&= \mathbb{E}[X_1^2 + \cdots + X_n^2 + 2X_1X_2 + 2X_2X_3 + \cdots + 2X_{n-1}X_n] \\
&= \sum_{i=1}^{n} \mathbb{E}[X_i^2] + \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \mathbb{E}[X_iX_j].
\end{aligned}$$

As the X's are independent, $\mathbb{E}[X_iX_j] = \mathbb{E}[X_i]\mathbb{E}[X_j] = p^2$ (so long as $i \neq j$); note there are $\binom{n}{2}$ pairs (i,j) with $1 \leq i < j \leq n$. What about $\mathbb{E}[X_i^2]$? That is readily seen to be just $1^2 \cdot p + 0^2 \cdot (1-p) = p$. Substituting gives

$$\mathbb{E}[X^2] \quad = \quad \sum_{i=1}^{n} p + 2\sum_{i=1}^{n-1}\sum_{j=i+1}^{n} p^2 \ = \ np + \binom{n}{2}p^2.$$

Thus the variance is

$$\mathbb{E}[X^2] - \mathbb{E}[X]^2 \quad = \quad np + 2\frac{n(n-1)p^2}{2} - (np)^2 \ = \ np - np^2 \ = \ np(1-p).$$

We thus recover our result from above.

How should we handle the third moment? As $\mathbb{E}[X] = np$ and $\mathbb{E}[X^2] = p$, we have

$$\begin{aligned}\mathbb{E}[(X-\mu)^3] &= \mathbb{E}[X^3 - 3X^2\mu + 3X\mu^2 - \mu^3] \\ &= \mathbb{E}[X^3] - 3np\mathbb{E}[X^2] + 3(np)^2\mathbb{E}[X] - (np)^3 \\ &= \mathbb{E}[X^3] - 3n^2p^2(1-p) + 3n^3p^3 - n^3p^3.\end{aligned}$$

We can complete the analysis in a similar manner as above, namely expanding out

$$X^3 \ = \ (X_1 + \cdots + X_n)^3$$

and then using linearity of expectation. At this point, differentiating identities isn't looking so bad!

To solve this with differentiating identities, we must evaluate a sum such as

$$\sum_{k=0}^{n} k^3 \cdot \binom{n}{k} p^k (1-p)^{n-k}.$$

We start with the identity

$$(x+y)^n \ = \ \sum_{k=0}^{n}\binom{n}{k}x^k y^{n-k}.$$

We apply the operator $x\frac{d}{dx}$ three times to each side, and find (after some tedious but straightforward algebra and calculus) that the left hand side equals

$$nx(x+y)^{n-3}\left(n^2x^2 + 3nxy - y(x-y)\right).$$

Setting $y = 1-x$ and $x = p$ yields

$$np\left(1 + 3(n-1)p + (n^2 - 3n + 2)p^2\right) \ = \ \sum_{k=0}^{n} k^3 \cdot \binom{n}{k}p^k(1-p)^{n-k}.$$

The above is quite messy, and there is a very good chance we have made an algebra mistake. Thus, let's see if we can find another approach which will lead to cleaner algebra. Instead of applying $x\frac{d}{dx}$ three times, let's apply $x^3\frac{d^3}{dx^3}$. Applying this to $(x+y)^n$ is very easy, giving $x^3 \cdot n(n-1)(n-2)(x+y)^{n-3}$; applying it to the combinatorial expansion gives not k^3 and $k(k-1)(k-2)$. Collecting, we find

$$\begin{aligned}n(n-1)(n-2)x^3(x+y)^{n-3} &= x^3\sum_{k=0}^{n} k(k-1)(k-2)\binom{n}{k}x^{k-3}y^{n-k} \\ &= \sum_{k=0}^{n}\left(k^3 - 3k^2 + 2k\right)\binom{n}{k}x^k y^{n-k} \\ &= \sum_{k=0}^{n} k^3\binom{n}{k}x^k y^{n-k} - 3\sum_{k=0}^{n} k^2\binom{n}{k}x^k y^{n-k} \\ &\quad + 2\sum_{k=0}^{n} k\binom{n}{k}x^k y^{n-k}.\end{aligned}$$

Setting $x = p$ and$y = 1-p$ yields

$$n(n-1)(n-2)p^3 \quad = \quad \mathbb{E}[X^3] - 3\mathbb{E}[X^2] + 2\mathbb{E}[X].$$

We have made a lot of progress, as we already know $\mathbb{E}[X]$ and $\mathbb{E}[X^2]$ and can thus solve for $\mathbb{E}[X^3]$. **The point is that it is easier *not* to try and find $\mathbb{E}[X]$ directly, but rather to find a related quantity.** Note, of course, that this method requires us to know $\mathbb{E}[X]$ and $\mathbb{E}[X^2]$ before we can deduce the value of $\mathbb{E}[X^3]$; this is not an unreasonable request, as typically we want to know all the moments up to a certain point.

The general principle here is that algebra can be hard, painful and tedious, but if you look at a problem the right way, you can minimize how much algebra you need to do. It's worthwhile to spend a few minutes thinking about how we can try and approach a problem, as often this leads to a way with significantly less messy computations.

#2: We toss N coins (each of which is heads with probability p), where the number N is drawn from a Poisson random variable with parameter lambda. Let X denote the number of heads. What is the probability density function of X? Justify your answer.

Solution: We toss N coins (each of which is heads with probability p), where $N \sim \text{Poisson}(\lambda)$, and let X denote the number of heads. What is the probability mass function of X? We compute it by calculating the probability of getting m heads when we toss n coins, and weight that by the probability of having n coins to toss. Thus the answer is

$$\begin{aligned}
\text{Prob}(X = m) &= \sum_{n=m}^{\infty} \text{Prob}(X = m | N = n) \cdot \text{Prob}(N = n) \\
&= \sum_{n=m}^{\infty} \binom{n}{m} p^m (1-p)^{n-m} \cdot \frac{\lambda^n e^{-\lambda}}{n!} \\
&= p^m e^{-\lambda} \sum_{n=m}^{\infty} \frac{n!}{m!(n-m)!} (1-p)^{n-m} \frac{\lambda^n}{n!} \\
&= \frac{p^m e^{-\lambda}}{m!} \sum_{n=m}^{\infty} \frac{(1-p)^{n-m} \lambda^n}{(n-m)!}.
\end{aligned}$$

We need to be 'clever' here to simplify the algebra and get a nice, clean expression, but note the very large 'hints' we get by looking at the expression so far. First off, we have a factor of $p^m e^{-\lambda}/m!$ outside. This looks a bit like the mass function of a Poisson, but not quite. Second, the sum above has two pieces that depend on $n - m$ and one piece that depends on n. This suggests we should add zero, and write

$$\lambda^n \ = \ \lambda^{n-m+m} \ = \ \lambda^{n-m} \cdot \lambda^m.$$

We can then pull the λ^m outside of the sum and we find

$$\text{Prob}(X = m) \quad = \quad \frac{p^m \lambda^m e^{-\lambda}}{m!} \sum_{n=m}^{\infty} \frac{(1-p)^{n-m} \lambda^{n-m}}{(n-m)!}.$$

We now let $k = n - m$ so the sum runs from 0 to ∞. We also combine the factors, and obtain

$$\begin{aligned}
\text{Prob}(X = m) &= \frac{(p\lambda)^m e^{-\lambda}}{m!} \sum_{k=0}^{\infty} \frac{((1-p)\lambda)^k}{k!} \\
&= \frac{(p\lambda)^m e^{-\lambda}}{m!} e^{(1-p)\lambda}
\end{aligned}$$

from the definition of e^x as

$$e^x \ = \ \sum_{k=0}^{\infty} \frac{x^k}{k!}.$$

Simplifying the above expression, we finally obtain

$$\text{Prob}(X = m) \ = \ \frac{(p\lambda)^m e^{-p\lambda}}{m!},$$

which is the probability mass function for a Poisson random variable with parameter $p\lambda$.

It takes awhile to become proficient and fluent with such algebraic manipulations. A good guiding principle is that we want to manipulate the expressions towards some known end, which guides us in how to multiply by 1 or add 0. Here the key step was writing λ^n and $\lambda^{n-m}\lambda^m$.

The following is not needed for the problem, but provides another opportunity to review some of the concepts we've seen, and their application. Let's compute the average value of a random variable Y with the Poisson distribution with parameter λ. We have

$$\begin{aligned}\mathbb{E}[Y] &= \sum_{n=0}^{\infty} n \cdot \frac{\lambda^n e^{-\lambda}}{n!} \\ &= \sum_{n=1}^{\infty} n \cdot \frac{\lambda^n e^{-\lambda}}{n!} \\ &= e^{-\lambda} \sum_{n=1}^{\infty} \frac{\lambda^n}{(n-1)!}.\end{aligned}$$

To finish the evaluation, it's natural to write λ^n and $\lambda^{n-1}\lambda$. The reason for this is that we have a sum where the denominator involves $n-1$, and thus it is helpful to make the numerator depend on $n-1$ as well. If we let $k = n-1$, then as n runs from 1 to ∞ we have k runs from 0 to ∞, and we find

$$\mathbb{E}[Y] \;=\; e^{-\lambda} \sum_{k=0}^{\infty} \frac{\lambda^k \cdot \lambda}{k!} \;=\; \lambda e^{-\lambda} \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} \;=\; \lambda e^{-\lambda} e^{\lambda} \;=\; \lambda,$$

where again we made use of the series expansion of e^x.

Using this fact, we can find the expected number of heads in the assigned problem *without* actually proving that X is given by the Poisson distribution with parameter λp. To see this, we claim that if

$$\text{Prob}(X = m) \;=\; \sum_{n=m}^{\infty} \text{Prob}(X = m | N = n) \cdot \text{Prob}(N = n),$$

then

$$\mathbb{E}[X] \;=\; \sum_{n=0}^{\infty} \mathbb{E}[X | N = n] \cdot \text{Prob}(N = n),$$

which leads to

$$\begin{aligned}\mathbb{E}[X] &= \sum_{n=0}^{\infty} np \cdot \frac{\lambda^n e^{-\lambda}}{n!} \\ &= p \sum_{n=0}^{\infty} n \cdot \frac{\lambda^n e^{-\lambda}}{n!};\end{aligned}$$

the last sum is just the expected value of the Poisson distribution with parameter λ, which we know is λ. Thus $\mathbb{E}[X] = p\lambda$.

#3: Find the probability density function of Y when $Y = \exp(X)$ for $X \sim N(0, 1)$.

Solution: We want to compute the density of $Y = e^X$, where $X \sim N(0, 1)$. The latter means that X has the standard normal distribution, namely that the density function of X, f_X, satisfies

$$f_X(x) \;=\; \frac{1}{\sqrt{2\pi}}\, e^{-x^2/2}.$$

One very easy way to compute the answer to problems like this is by using cumulative distribution functions, and noting the probability density is the derivative. Let F_X and F_Y represent the cumulative distribution functions of X and Y, and let f_X and f_Y denote their densities. We have

$$\begin{aligned}F_Y(y) &= \text{Prob}(Y \le y) \\ &= \text{Prob}(e^X \le y) \\ &= \text{Prob}(X \le \log y) \\ &= F_X(\log y).\end{aligned}$$

We now differentiate, using the chain rule.

$$f_Y(y) \quad = \quad F_X'(\log y) \cdot (\log y)' \ = \ f_X(\log y) \cdot \frac{1}{y}.$$

Substituting for f_X, we obtain

$$f_Y(y) \ = \ \frac{1}{\sqrt{2\pi}} \frac{1}{y} e^{-\frac{\log^2(y)}{2}}.$$

#4: Each box of cereal is equally likely to have exactly one of a set of c prizes. Thus, every time you open a box you have a $1/c$ chance of getting prize 1, a $1/c$ chance of getting prize 2, ... How many boxes to you expect to have to open before you have at least one of each of the c prizes? If you have having trouble, do $c = 2$ for half credit.

Solution: This is a beautiful problem illustrating the power of expectation. Not surprisingly, it starts off as another geometric series problem (i.e., waiting for the first success). Let Y_j be the random variable which denotes how much time we need to wait to get the next new prize given that we have j distinct prizes (of the c prizes). For each pick, the probability we get one of the j prizes we already have is $\frac{j}{c}$, and thus the probability p we get a new prize is $p = 1 - \frac{j}{c} = \frac{c-j}{c}$. Thus, letting $p = \frac{c-j}{c}$ we find the probability that we get the next new prize on pick n is just $(1-p)^{n-1}p$, so the expected value is

$$\sum_{n=1}^{\infty} n \cdot (1-p)^{n-1} p \ = \ \sum_{n=1}^{\infty} \left(\frac{j}{c}\right)^{n-1} \frac{c-j}{c};$$

as $p = \frac{c-j}{c}$ and the expected value is $1/p$, we have $\mathbb{E}[Y_j] = \frac{c}{c-j}$. Note the answer is reasonable. When $j = 0$ the expected wait is just one pick (which makes sense, as we have no prizes so anything is new). When $j = c - 1$ we are missing only one prize, and the answer is an expected wait of c (also reasonable!).

If Y is the random variable which denotes how long we must wait to get all the prizes, then $Y = Y_0 + \cdots + Y_{c-1}$. As expectation is linear,

$$\mathbb{E}[Y] \ = \ \mathbb{E}[Y_0] + \cdots + \mathbb{E}[Y_{c-1}] \ = \ \frac{c}{c-0} + \cdots + \frac{c}{c-(c-1)}.$$

If we read the sum in reverse order and factor out a c, we notice it is

$$\mathbb{E}[Y] \ = \ c\left(1 + \frac{1}{2} + \frac{1}{3} + \cdots + \frac{1}{c}\right) \ \approx \ c \log c,$$

as the sum is the c^{th} harmonic number H_c, which is about $\log c$ (a better approximation is $\log c + \gamma$, where γ is the Euler-Mascheroni constant and is about .5772156649). See

`http://en.wikipedia.org/wiki/Harmonic_number`

for more information.

As it's often hard to see how to attack the general case immediately, it's a good idea to try a simple case first and detect the pattern. Let's try $c = 2$. Our first box has to give us a prize we don't have; without loss of generality let's say we got the first prize. We keep picking until we get the second prize. Each box we open from this point onward has a 50% chance of getting us that second prize and ending our picking. Thus the probability we need one more box (or two total) is 1/2, that we need two more boxes (or three total) is $(1/2)^2 = 1/4$, that we need three more boxes (or four total) is $(1/2)^3 = 1/8$ and so on. If Y_1 denotes how long we have to wait from getting the first prize to getting the second, we see $\text{Prob}(Y_1 = n) = (1/2)^n$. Thus Y_1 is a geometric random variable with parameter 1/2, and the total wait to get both prizes is $1 + Y_1$. As the expected value of a geometric random variable with parameter p is $1/p$, $\mathbb{E}[1 + Y_1] = 1 + 2 = 3$.

#5: Let $X_1, \dots, X_n$ be independent Bernoulli random variables where $X_k \sim \text{Bern}(p_k)$ (you can think of this as n independent coin tosses, where coin k is heads with probability p). If $Y = X_1 + \cdots + X_n$, what is the mean and what is the variance of Y? Assume $p_1 + \cdots + p_n = \mu$; what choice or choices of the p_k's lead to the variance of Y being the largest possible?

Solution: By linearity of expectation, $\mathbb{E}[Y] = \mathbb{E}[X_1] + \cdots + \mathbb{E}[X_n]$ and $\text{Var}(Y) = \text{Var}(X_1) + \cdots + \text{Var}(X_n)$. As $\mathbb{E}[X_k] = p_k$ and $\text{Var}(X_k) = p_k(1-p_k)$ we find $\mathbb{E}[Y] = p_1 + \cdots + p_n$ and $\text{Var}(Y) = p_1(1-p_1) + \cdots + p_n(1-p_n)$. We now turn to finding the choice that leads to the largest possible variance.

We first claim that there must be at least one choice which gives a maximum variance. To see this, we appeal to a result from real analysis: a continuous function on a compact set (i.e., a set that is closed and bounded) attains its maximum and minimum values. If you're not familiar with this, look at the third proof below (the one using Lagrange multipliers) and think about how that was presented in your Calc III class.

It turns out to be sufficient to study the special case when $n = 2$; before explaining why, we'll analyze this case in detail. We give the 'standard' proof using techniques from calculus. While the idea is simple, the algebra quickly gets involved and tedious, though everything does work out if we're patient enough. As this much algebra is unenlightening, we give an alternate, simpler proof below as well.

First proof: long algebra. We first give the standard proof that one might give after taking a calculus class. Namely, we convert everything to a function of one variable, and just plow ahead with the differentiation, finding the critical points and comparing the values at the critical points to the end-points. While this is exactly what we've been taught to do in calculus, we'll quickly see the algebra becomes involved and unenlightening, and thus we will give *many* alternate proofs afterwards!

Our situation is that we have $p_1 + p_2 = \mu$ and we want to maximize $p_1(1-p_1) + p_2(1-p_2)$. As $p_2 = \mu - p_1$, we must maximize

$$\begin{aligned} g(p_1) &= p_1(1-p_1) + (\mu - p_1)(1-\mu+p_1) \\ &= p_1 - p_1^2 + \mu(1-\mu) - p_1(1-\mu) + p_1\mu - p_1^2 \\ &= 2p_1\mu - 2p_1^2 + \mu(1-\mu). \end{aligned}$$

To find the maximum, calculus tells us to find the critical points (the values of p_1 where $g'(p_1) = 0$) and compare that value to the endpoints (which for this problem would be $p_1 = \max(0, \mu - 1)$ and $p_1 = \min(\mu, 1)$). We have $g'(p_1) = 2\mu - 4p_1$, so the critical point is $p_1 = \mu/2$ which gives $g(\mu/2) = \mu - \frac{\mu^2}{2}$. Straightforward algebra now shows that this is larger than the boundary values. As $g(p_1) = g(1-p_1)$, it suffices to check the lower bounds. If $p_1 = 0$ that means $0 \le \mu \le 1$, and in this case $p_2 = \mu$ so $g(0) = \mu(1-\mu) = \mu - \mu^2$, which is clearly smaller than $g(\mu/2) = \mu - \frac{\mu^2}{2}$. Similarly if $p_1 = \mu - 1$ (which implies $1 \le \mu \le 2$) then $p_2 = 1$ and thus $g(\mu - 1) = (\mu - 1)(2-\mu) + 0 = -\mu^2 + 3\mu - 2$. If this were larger than $g(\mu/2)$, we would have the following chain:

$$\begin{aligned} -\mu^2 + 3\mu - 2 &> \mu - \frac{\mu^2}{2} \\ 0 &> \frac{\mu^2}{2} - 2\mu + 2 \\ 0 &> \mu^2 - 4\mu + 4 \ > \ (\mu-2)^2, \end{aligned}$$

which is impossible. Thus, after tedious but straightforward algebra, we see the maximum value occurs not at a boundary point but at the critical point $p_1 = \mu/2$, which implies $p_2 = \mu/2$ as well.

We now consider the case of general n. Imagine we are at the maximum variance with values $\mathfrak{p}_1, \cdots, \mathfrak{p}_n$. If any two of the $\mathfrak{p}_k$'s were unequal (say the i and j values), by the argument above (in the case of just two values) we could increase the variance by replacing $\mathfrak{p}_i$ and $\mathfrak{p}_j$ with $\frac{\mathfrak{p}_i + \mathfrak{p}_j}{2}$. Thus the maximum value of the variance occurs when all are equal.

Second proof: cleaner algebra. As the algebra is a bit tedious, we give another approach. Imagine (back in the $n = 2$ case) that $p_1 \neq p_2$. Let's write $p_1 = \frac{\mu}{2} + x$ and $p_2 = \frac{\mu}{2} - x$. We need to show the variance is maximized when $x = 0$. If $x = 0$ the variance is just $\mu - \frac{\mu^2}{2}$, while for general x it is

$$\left(\frac{\mu}{2} + x\right)\left(1 - \frac{\mu}{2} - x\right) + \left(\frac{\mu}{2} - x\right)\left(1 - \frac{\mu}{2} + x\right) \ = \ \mu - \frac{\mu^2}{2} - 2x^2,$$

where the last step follows from multiplying everything out. Thus the variance is maximized in this case when $x = 0$. Note how much faster this approach is. We included the first approach as this is what we're taught in calculus, namely find the critical points and check the boundary points; however, especially in instances where we have some intuition as to what the answer should be, there are frequently better ways of arranging the algebra.

Third proof: Lagrange multipliers. We give one more proof, though here the pre-requisites are more. We use Lagrange multipliers: we want to maximize $f(p_1, p_2) = p_1(1-p_1) + p_2(1-p_2)$ subject to $g(p_1, p_2) = p_1 + p_2 - \mu = 0$. We need

$\nabla f = \nabla g$, so

$$\begin{aligned} f(p_1,p_2) &= p_1 - p_1^2 + p_2 - p_2^2 \\ g(p_1,p_2) &= p_1 + p_2 - \mu \\ \nabla f(p_1,p_2) &= (1-2p_1, 1-2p_2) \\ \nabla g(p_1,p_2) &= (1,1). \end{aligned}$$

As $\nabla f = \lambda g$ and $\nabla g(p_1,p_2) = (1,1)$, we find $1-2p_1 = 1-2p_2$ or $p_1 = p_2$ as claimed. Note how readily this generalizes to n variables, as in this case we would have

$$\begin{aligned} \nabla f(p_1,\dots,p_n) &= (1-2p_1,\dots,1-2p_n) \\ \nabla g(p_1,\dots,p_n) &= (1,\dots,1), \end{aligned}$$

which implies all the p_i's are equal.

Fourth proof: geometry. We give yet another proof in the case $n=2$ and $p_1+p_2=\mu$. We are trying to maximize

$$p_1(1-p_1) + p_2(1-p_2) \ = \ p_1 - p_1^2 + p_2 - p_2^2 \ = \ \mu - (p_1^2+p_2^2).$$

As we are subtracting $p_1^2+p_2^2$, we want that to be as small as possible. We may interpret this as the distance of the point (p_1,p_2) from the origin, given that $p_1+p_2=\mu$. Geometrically it should be clear that the closest point to the origin is the midpoint of the line from $(0,\mu)$ to $(\mu,0)$; if not and if we need to resort to calculus, this is at least an easier problem. Namely, let $p_2 = \mu - p_1$ so we are trying to minimize

$$\mu - (p_1^2 + (\mu-p_1)^2) \ = \ \mu - \mu^2 - (2p_1^2 - 2\mu p_1) \ = \ \mu - \mu^2 - 2p_1(p_1-\mu).$$

We thus need to minimize the value of the quadratic $p_1(p_1-\mu)$; as the roots of this are 0 and μ, the minimum is at the vertex which is at the midpoint of the roots, namely $p_1 = \mu/2$. In general, we are trying to minimize the function $\mu - (p_1^2+\cdots+p_n^2)$ subject to $0 \le p_1,\dots,p_n \le 1$ and $p_1+\cdots+p_n = \mu$. This is equivalent to finding the point on the hyperplane closest to the origin in n-dimensional space, which is given by the point where they are all equal.

Finally, is this result surprising? If ever a $p_k = 0$ or 1, then there would be no variation in the contribution from X_k. Thus the variance will be smallest when all the p_k's are in $\{0,1\}$.

#6: State anything you learned or enjoyed in Arms' talk. One or two sentences suffice.

Solution: Anything should be fine!

#7: The kurtosis of a random variable X is defined by $\mathrm{kur}(X) := \mathbb{E}[(X-\mu)^4]/\sigma^4$, where μ is the mean and σ is the standard deviation. The kurtosis measures how much probability we have in the tails. If $X \sim \mathrm{Poiss}(\lambda)$, find the kurtosis of X.
Solution: Let $X \sim \mathrm{Poiss}(\lambda)$, so the mass function is $f(n) = \lambda^n e^{-\lambda}/n!$ for $n \ge 0$ and 0 otherwise. For a Poisson random variable with parameter λ, the mean is λ and the standard deviation is $\sqrt{\lambda}$ (or equivalently the variance is λ), and thus

$$\mathrm{kur}(X) \quad = \quad \frac{\sum_{n=0}^{\infty}(n-\lambda)^4\lambda^n e^{-\lambda}/n!}{\lambda^2}.$$

There are several ways to try and analyze this. One way is to expand out $(n-\lambda)^4$. Whenever we have an n, we can cancel that with the n in $n!$, and we are left with terms such as $n^k\lambda^j/(n-1)!$. We could then write n as $(n-1)+1$, expand and do some more canceling. While this will work, the algebra becomes tedious. The point of this exercise is to see that, while there are numerous ways to solve a problem, it is important to weigh their advantages and disadvantages. For instance, we can either make the linear combinations easy at the cost of more involved differentiation, or we can have easier combinations at the expense of more tedious differentiation. For this problem, it seems as if the easiest algebra is when we make the differentiation hard but the combinations easy. It takes awhile to develop a feel for which approach will be most tractable for a given problem. This is one reason why we provide so many different solutions.

First solution. One of the best ways to compute the moments of Poisson (and other discrete) random variables is through differentiating identities. Consider the identity

$$e^x \;=\; \sum_{n=0}^{\infty} \frac{x^n}{n!}.$$

We could keep applying the operator $x\frac{d}{dx}$ to this and obtain the moments, and then by expanding $(n-\lambda)^4$ piece everything together. A faster way is to apply the operator $-\lambda + x\frac{d}{dx}$ four times and then set $x = \lambda$. If we do that we obtain

$$\left(-\lambda + x\frac{d}{dx}\right)\left(-\lambda + x\frac{d}{dx}\right)\left(-\lambda + x\frac{d}{dx}\right)\left(-\lambda + x\frac{d}{dx}\right) e^x\Big|_{x=\lambda} \;=\; \sum_{n=0}^{\infty}(n-\lambda)^4 \cdot \frac{\lambda^n}{n!}.$$

After some long but standard differentiation, we find the derivative above equals

$$e^x\left(\lambda^4 - 4\lambda^3 x + 6\lambda^2 x(1+x) - 4\lambda x(1+3x+x^2) + x(1+7x+6x^2+x^3)\right);$$

setting $x = \lambda$ gives

$$\lambda e^{\lambda} + 3\lambda^2 e^{\lambda} \;=\; \sum_{n=0}^{\infty}(n-\lambda)^4 \cdot \frac{\lambda^n}{n!},$$

which means the kurtosis is

$$\mathrm{kur}(X) \;=\; \frac{e^{-\lambda}}{\lambda^2}\left(\lambda e^{\lambda} + 3\lambda^2 e^{\lambda}\right) \;=\; 3 + \frac{1}{\lambda}.$$

Second solution. In terms of keeping the algebra simple, it might be easier to expand $(n-\lambda)^4$ and apply the operator $x\frac{d}{dx}$ four times.

Third solution. Another possibility is to apply d/dx four times and then build back. For example, we start with

$$e^x \;=\; \sum_{n=0}^{\infty} \frac{x^n}{n!}.$$

Differentiating with respect to x once gives

$$e^x \;=\; \sum_{n=0}^{\infty} n \cdot \frac{x^{n-1}}{n!}.$$

Taking $x = \lambda$ and multiplying both sides by $\lambda e^{-\lambda}$ gives

$$\lambda e^{-\lambda} \cdot e^{\lambda} \;=\; \sum_{n=0}^{\infty} n \cdot \frac{\lambda^n e^{-\lambda}}{n!} \;=\; \mathbb{E}[X],$$

which implies the mean is λ. If we differentiate e^x twice with respect to x, we find

$$e^x \;=\; \sum_{n=0}^{\infty} n(n-1) \cdot \frac{x^{n-2}}{n!} \;=\; \sum_{n=0}^{\infty} n^2 \cdot \frac{x^{n-2}}{n!} - \sum_{n=0}^{\infty} n \cdot \frac{x^{n-2}}{n!}.$$

Taking $x = \lambda$ again and multiplying both sides by $\lambda e^{-\lambda}$ gives

$$\lambda^2 e^{-\lambda} e^{\lambda} \;=\; \sum_{n=0}^{\infty} n^2 \cdot \frac{\lambda^n e^{-\lambda}}{n!} - \sum_{n=0}^{\infty} n \cdot \frac{\lambda^n e^{-\lambda}}{n!};$$

as the last sum is λ, we find

$$\mathbb{E}[X^2] \;=\; \sum_{n=0}^{\infty} n^2 \cdot \frac{\lambda^n e^{-\lambda}}{n!} \;=\; \lambda^2 + \lambda.$$

Continuing in this way we can get $\mathbb{E}[X^3]$ and $\mathbb{E}[X^4]$, and then substitute into

$$\mathbb{E}[(X-\mu)^4] \;=\; \mathbb{E}[X^4] - 4\mu\mathbb{E}[X^3] + 6\mu^2\mathbb{E}[X^2] - 4\mu^3\mathbb{E}[X] + \mu^4.$$

Fourth solution. For our fourth solution, we use some ideas from linear algebra. We start, as always, with the identity $e^x = \sum_{n=0}^{\infty} x^n/n!$, and we differentiate this 4 times:

$$\begin{aligned}
e^x &= \sum_{n=0}^{\infty} \frac{x^n}{n!} \\
e^x &= \sum_{n=0}^{\infty} n \cdot \frac{x^{n-1}}{n!} \\
e^x &= \sum_{n=0}^{\infty} n(n-1) \cdot \frac{x^{n-2}}{n!} \\
e^x &= \sum_{n=0}^{\infty} n(n-1)(n-2) \cdot \frac{x^{n-3}}{n!} \\
e^x &= \sum_{n=0}^{\infty} n(n-1)(n-2)(n-3) \cdot \frac{x^{n-4}}{n!}.
\end{aligned}$$

We take $x = \lambda$ and multiply the k^{th} equation above by λ^k, and find

$$\begin{aligned}
e^\lambda &= \sum_{n=0}^{\infty} \frac{\lambda^n}{n!} \\
\lambda e^\lambda &= \sum_{n=0}^{\infty} n \cdot \frac{\lambda^n}{n!} \\
\lambda^2 e^\lambda &= \sum_{n=0}^{\infty} (n^2 - n) \cdot \frac{\lambda^n}{n!} \\
\lambda^3 e^\lambda &= \sum_{n=0}^{\infty} (n^3 - 3n^2 + 2n) \cdot \frac{\lambda^n}{n!} \\
\lambda^4 e^\lambda &= \sum_{n=0}^{\infty} (n^4 - 6n^3 + 11n^2 - 6n) \cdot \frac{\lambda^n}{n!}.
\end{aligned}$$

We want to evaluate

$$\frac{e^{-\lambda}}{\lambda^2} \sum_{n=0}^{\infty} (n-\lambda)^4 \cdot \frac{\lambda^n}{n!} \;=\; \frac{e^{-\lambda}}{\lambda^2} \sum_{n=0}^{\infty} (n^4 - 4n^3\lambda + 6n^2\lambda^2 - 4n\lambda^3 + \lambda^4) \cdot \frac{\lambda^n}{n!}.$$

We write $n^4 - 4n^3\lambda + 6n^2\lambda^2 - 4n\lambda^3 + \lambda^4$ as a linear combination of the terms above. This is just solving a system of equations (for example, we may regard $n^4 - 4n^3\lambda + 6n^2\lambda^2 - 4n\lambda^3 + \lambda^4$ as the vector $(1, -4, 6, -4, 1, 0)$, with the last component 0 as there is no constant term). Solving the associated system of equations gives

$$n^4 - 4n^3\lambda + 6n^2\lambda^2 - 4n\lambda^3 + \lambda^4$$

equals

$$\begin{aligned}
&1 \cdot (n^4 - 6n^3 + 11n^2 - 6n) \;+\; (6 - 4\lambda) \cdot (n^3 - 3n^2 + 2n) \;+\; (7 - 12\lambda + 6\lambda^2) \cdot (n^2 - n) \\
&+ (1 - 4\lambda + 6\lambda^2 - 4\lambda^3) \cdot n \;+\; a^4 \cdot 1
\end{aligned}$$

and thus the kurtosis is

$$\begin{aligned}
&\frac{e^{-\lambda}}{\lambda^2} \Big[1 \cdot \lambda^4 e^\lambda + (6 - 4\lambda)\lambda^3 e^\lambda + (7 - 12\lambda + 6\lambda^2)\lambda^2 e^\lambda + \\
&\qquad (1 - 4\lambda + 6\lambda^2 - 4\lambda^3)\lambda e^\lambda + 1 e^\lambda \Big] \;=\; \frac{1}{\lambda^2}\left[3\lambda^2 + \lambda\right] \;=\; 3 + \frac{1}{\lambda}.
\end{aligned}$$

#8: Consider a coin with probability p of heads. Find the probability density function for X_1, where X_1 is how long we must wait before we get our *first* head.

Solution: Clearly $\mathrm{Prob}(X_1 = n) = 0$ unless $n \in \{1, 2, 3, \dots\}$. For $n \in \{1, 2, 3, \dots\}$ we have to start with $n-1$ tails (each happening independently with probability $1-p$) and then end with a head (which happens with probability p. Thus

$$\mathrm{Prob}(X_1 = n) \ = \ \begin{cases} (1-p)^{n-1}p & \text{if } n \in \{1, 2, 3, \dots\} \\ 0 & \text{otherwise.} \end{cases}$$

#9: Consider a coin with probability p of heads. Find the probability density function for X_2, where X_2 is how long we must weight before we get our *second* head.

Solution: The solution is similar to the previous problem. There are two small changes. First, the non-zero probabilities are for $n \in \{2, 3, 4, \dots\}$. Second, in the first $n-1$ tosses we now have $n-2$ tails and 1 head; there are $\binom{n-1}{1} = n-1$ ways to choose which of the first $n-1$ tosses is the head. Each of these $n-1$ possibilities happens with probability $(1-p)^{n-2}p^2$, and we find

$$\mathrm{Prob}(X_2 = n) \ = \ \begin{cases} (n-1)(1-p)^{n-2}p^2 & \text{if } n \in \{2, 3, 4, \dots\} \\ 0 & \text{otherwise.} \end{cases}$$

#10: Alice, Bob and Charlie are rolling a fair die in that order. They keep rolling until one of them rolls a 6. What is the probability each of them wins?

Solution: Let x be the probability Alice rolls the first six. We have

$$x \ = \ \frac{1}{6} + \frac{5}{6} \cdot \frac{5}{6} \cdot \frac{5}{6} \cdot x;$$

to have Alice win, she either rolls a six on her first turn (which happens with probability 1/6) or all three don't roll a six on their first turn (which happens with probability $5/6 \cdot 5/6 \cdot 5/6$), at which point Alice now wins with probability x (we have a memoryless process, and it's as if we just started the game again. Thus $x = \frac{1}{6} + \frac{125}{216}x$, or $x = 36/91$.

If we let y be the probability Bob wins the game, clearly Bob cannot win if Alice rolls a six on her first try. Thus we have

$$y \ = \ \frac{5}{6} \cdot x;$$

this is because once Alice rolls a non-six (which happens 5/6 of the time), the probability Bob wins is just x. We find $y = \frac{5}{6}\frac{36}{91} = 30/91$.

Similarly, if z is the probability Charlie wins, then

$$z \ = \ \frac{5}{6} \cdot \frac{5}{6} \cdot x,$$

as both Alice and Bob must roll non-sixes (which happens with probability $5/6 \cdot 5/6$), at which point the probability of Charlie winning is just x. We find $z = \frac{5}{6}\frac{5}{6}\frac{36}{91} = \frac{25}{91}$.

Note there was no need to find z; we could have found it by noting $z = 1 - x - y$. It's good to calculate it from scratch as this provides a check. Does $x + y + z = 1$? We have $36/91 + 30/91 + 25/91 = 91/91$, which is 1.

#11: Alice, Bob and Charlie are rolling a fair die in that order. What is the probability Alice is the first person to roll a 6, Bob is the second and Charlie is the third?

Solution: After A throws a 6 we do not care if she (A is obviously named Alice) throws another 6 before B (clearly Bob) or C (surely Charlie) does; all we care about is that B then throws a 6 before Charlie. Let x be the probability that A rolls the first 6. Then

$$x \ = \ \frac{1}{6} + \left(\frac{5}{6}\right)^3 x;$$

this is because she either rolls a 6 on her first try, or she and B and C all miss, and then it is as if we've started the game fresh. (Note how important the memoryless feature is in solving these problems!) We thus find $x = \frac{1}{6} + \frac{125}{216}x$, or after some algebra $x = \frac{36}{91}$. We now keep rolling, and we only care about the rolls of B and C. It suffices to determine the probability B gets the next 6, as clearly C will then be the last to roll. Let w be the probability B rolls a 6 before C, given that B rolls first. A similar analysis gives

$$w \ = \ \frac{1}{6} + \left(\frac{5}{6}\right)^2 w,$$

or $w = \frac{1}{6} + \frac{25}{36}y$, which gives $w = \frac{6}{11}$. Thus the probability that A is first, then B and then C is just

$$\frac{36}{91} \cdot \frac{6}{11} \cdot 1 \ = \ \frac{216}{1001}.$$

As always, we should ask if this answer is reasonable. There are $3! = 6$ ways to order 3 people. As the denominator is 1001, if all six orderings were equally likely we would get a fraction of (approximately) 167/1001. Thus our answer is a bit higher than the case where all outcomes are equally likely. This is reasonable, as we do expect A to get the first six....

#12: Alice, Bob and Charlie are *still* rolling the fair die. What is the probability that the first 6 is rolled by Alice, the second 6 by Bob and the third 6 by Charlie?

Solution: Using the notation and results from the earlier problems, we now want A to roll the first 6, and then the next 6 *must* be rolled by B, and then the next *must* be rolled by C; thus, we now care about A's subsequent rolls. Fortunately we've already solved this problem! In the analysis above, we may interpret $x = 36/91$ as the probability that the first 6 is rolled by the person currently rolling. Thus the answer here is just $x^3 = (36/91)^3$; the reason is that once A rolls a six, it is now B's turn to roll.

#13: What are the mean and variance of a chi-square distribution with 2 degrees of freedom? If $X \sim \chi^2(2)$, what is the probability that X takes on a value at least twice its mean? What is the probability X takes on a value at most half of its mean?
Solution: The probability density function of a χ^2 random variable with ν degrees of freedom is

$$(2^{\nu/2}\Gamma(\nu/2))^{-1}x^{\nu/2-1}e^{-x/2}$$

for $x \geq 0$ and 0 otherwise. Notice that if $\nu = 2$ then the density is

$$f_2(x) \ = \ \begin{cases} \frac{1}{2}e^{-x/2} & \text{if } x \geq 0 \\ 0 & \text{otherwise;} \end{cases}$$

this is an exponential random variable with $\lambda = 2$. We've shown in class that the mean of an exponential random variable with parameter λ is λ. The answer to the first question is thus

$$\text{Prob}(X \geq 2 \cdot 2) \quad = \quad \int_4^\infty \frac{1}{2}e^{-x/2}dx \ = \ -e^{-x/2}\Big|_4^\infty \ = \ e^{-2}.$$

For the second question, we have

$$\text{Prob}(X \leq \frac{1}{2} \cdot 2) \quad = \quad \int_0^1 \frac{1}{2}e^{-x/2} \ = \ -e^{-x/2}\Big|_0^1 \ = \ 1 - e^{-1/2}.$$

5.3. **HW #7: Due Friday, Nov 9, 2018:** #1: 10% of the numbers on a list are 15, 20% are 25, and the rest are 50. What is the average? #2: All 100 numbers in a list are non-negative and the average is 2. Prove that at most 25 exceed 8. #3: A and B are independent events with indicator random variables I_A and I_B; thus $I_A(x) = 1$ with probability $\Pr(A)$ and is 0 with probability $1 - \Pr(A)$. (a) What is the distribution of $(I_A + I_B)^2$? (b) What is $\mathbb{E}[(I_A + I_B)^2]$? #4: Consider a random variable X with expectation 10 and standard deviation 5. (a) Find the smallest upper bound you can for $\mathbb{P}X \geq 20$. (b) Could X be a binomial random variable? #5: Suppose average family income is \$10,000. (a) Find an upper bound for the percentage of families with income over \$50,000. (b) Redo (a) but with the added knowledge that the standard deviation is \$8,000. #6: (a) Let X be a random variable with $0 \leq X \leq 1$ and $\mathbb{E}[X] = \mu$. Show that $0 \leq \mu \leq 1$ and $0 \leq \text{Var}(X) \leq \mu(1-\mu) \leq 1/4$. (b) Generalize and consider the case $a \leq X \leq b$. (c) Assume $0 \leq X \leq 9$. Find a random variable where the variance is as large as possible.

6. HW #7: Due Friday, Nov 9, 2018

6.1. **HW #7: Due Friday, Nov 9, 2018:** #1: 10% of the numbers on a list are 15, 20% are 25, and the rest are 50. What is the average? #2: All 100 numbers in a list are non-negative and the average is 2. Prove that at most 25 exceed 8. #3: A and B are independent events with indicator random variables I_A and I_B; thus $I_A(x) = 1$ with probability $\Pr(A)$ and is 0 with probability $1 - \Pr(A)$. (a) What is the distribution of $(I_A + I_B)^2$? (b) What is $\mathbb{E}[(I_A + I_B)^2]$? #4: Consider a random variable X with expectation 10 and standard deviation 5. (a) Find the smallest upper bound you can for $\mathbb{P}X \geq 20$. (b) Could X be a binomial random variable? #5: Suppose average family income is \$10,000. (a) Find an upper bound for the percentage of families with income over \$50,000. (b) Redo (a) but with the added knowledge that the standard deviation is \$8,000. #6: (a) Let X be a random variable with $0 \leq X \leq 1$ and $\mathbb{E}[X] = \mu$. Show that $0 \leq \mu \leq 1$ and $0 \leq \mathrm{Var}(X) \leq \mu(1-\mu) \leq 1/4$. (b) Generalize and consider the case $a \leq X \leq b$. (c) Assume $0 \leq X \leq 9$. Find a random variable where the variance is as large as possible.

6.2. **Solutions:** #1: 10% of the numbers on a list are 15, 20% are 25, rest are 50. What is the average?
Solution: Let there be n numbers. The mean is $\frac{10n}{100n} \cdot 15 + \frac{20n}{100n} \cdot 25 + \frac{70n}{100n} \cdot 50 = \frac{150+500+3500}{100} = 41.5$. Note the answer is greater than 15 (the smallest number on our list) and smaller than 50 (the largest on our list). Also 70% of the numbers are 50, so we expect the mean to be close to 50.

#2: All 100 numbers in list are non-negative and average is 2. Prove that at most 25 exceed 8.
Solution: Imagine there were 26 that were greater than 8. What would these contribute to the mean? Well, if the 26 numbers were 8, we would have a contribution to the mean of $\frac{26}{100} \cdot 8 = 2.08$; as the other numbers are non-negative, the mean would have to be at least 2.08, contradicting the fact that the mean is 2. In a sense, this question is poorly phrased, as we can do better. We can show that there are at most 24 that exceed 8. The average is smallest when 25 exceed 8 if 75 are 0 and 25 exceed 8, which gives a mean exceeding 2.

#3: A and B independent events with indicator random variables I_A and I_B; thus $I_A(x) = 1$ with probability $\Pr(A)$ and is 0 with probability $1 - \Pr(A)$. (a) What is the distribution of $(I_A + I_B)^2$? (b) What is $\mathbb{E}[(I_A + I_B)^2]$?
Solution: Squaring, it is $I_A^2 + 2I_AI_B + I_B^2 = I_A + 2I_AI_B + I_B$ as the square of an indicator random variable is just the indicator. It can only take on the values 0, 1, and 4. It is zero when A and B don't happen, or it is 0 with probability $(1 - \mathbb{P}A)(1 - \mathbb{P}B)$. It is 1 if exactly one of A and B happens, so it is 1 with probability $\mathbb{P}A(1 - \mathbb{P}B) + (1 - \mathbb{P}A)\mathbb{P}B$. It is 2 if both happens, or it is 2 with probability $\mathbb{P}A\mathbb{P}B$.

(b) We use $I_A^2 + 2I_AI_B + I_B^2 = I_A + 2I_AI_B + I_B$ and the linearity of expectation to see that this is $\mathbb{E}[I_A] + 2\mathbb{E}[I_AI_B] + \mathbb{E}[I_B]$. The middle term is just $\mathbb{E}[I_A]\mathbb{E}[I_B] = \mathbb{P}A\mathbb{P}B$ as the random variables are independent, and so this answer is just $\mathbb{P}A + 2\mathbb{P}A\mathbb{P}B + \mathbb{P}B$. As an aside, if we had $(I_A + I_B)^n$, what do you think this will approximately equal for n large?

#4: Consider a random variable X with expectation 10 and standard deviation 5. (a) Find the smallest upper bound you can for $\mathbb{P}X \geq 20$. (b) Could X be a binomial random variable?
Solution: (a) Note that 20 is 2 standard deviations above the mean, and thus by Chebyshev's inequality the probability of being at least 20 is at most 1/4. (b) If yes, it would have some n and a probability p. We would have to solve $\mathbb{E}[X] = np = 10$, $\mathrm{Var}(X) = np(1-p) = 25$. There are many ways to do the algebra. Substitute for np, which must be 10, in the variance equation to find $10(1-p) = 25$, so $1 - p = 2.5$ or $p = -1.5$, which is impossible. What if instead we were told the mean was 10 and the variance was 5? In that case we would have $10(1-p) = 5$, which gives $p = 1/2$ and then from $np = 10$ we get $n = 20$, so in this case it is possible to have a binomial random variable.

#5: Suppose average family income is \$10,000. (a) Find upper bound for percentage of families with income over \$50,000. (b) Redo (a) but with the added knowledge that the standard deviation is \$8,000.
Solution: (a) Note that income is non-negative (we hope!), so let's try Markov's inequality. So $\mathbb{P}I \geq \$50,000 \leq \mathbb{E}[I]/\$50,000$ $= 10000/50000 = 1/5$. (b) If we know the standard deviation is \$8,000, then we see that we are 4 standard deviations from the mean, so by Chebyshev the probability of being at least 5 standard deviations away from the mean is at most $1/5^2$. Not surprisingly, we can do much better when we know more.

#6: (a) Let X be a random variable with $0 \leq X \leq 1$ and $\mathbb{E}[X] = \mu$. Show that $0 \leq \mu \leq 1$ and $0 \leq \mathrm{Var}(X) \leq \mu(1-\mu) \leq 1/4$. (b) Generalize and consider the case $a \leq X \leq b$. (c) Assume $0 \leq X \leq 9$. Find a random variable where the variance is as large as possible.

Solution: (a) As $0 \le X \le 1$, $0 \le \mathbb{E}[X] = \mu \le 1$. For the second claim, note $0 \le X^2 \le X \le 1$ as $0 \le X \le 1$. As $\mathrm{Var}(X) = \mathbb{E}[X^2] - \mathbb{E}[X]^2 = \mathbb{E}[X^2] - \mu^2$, and $\mathbb{E}[X^2] \le \mathbb{E}[X] \le \mu$, we have $\mathrm{Var}(X) \le \mu - \mu^2 = \mu(1-\mu)$. Since $\mu \in [0,1]$, a calculus exercise shows the maximum of the function $g(\mu) = \mu(1-\mu)$ occurs when $\mu = 1/2$, leading to the value 1/4. Another way to see this is to note $\mu(1-\mu) = -\mu^2 - \mu = -(\mu - 1/2)^2 - 1/4$; as $(\mu - 1/2)^2 \ge 0$ the minimum value is 1/4. *Note: the variance bound should be a function of μ. If we let $Y = 1 - X$ then the mean of Y is $1 - \mu$ but the variance of Y is the same as the variance of X; thus we expect our variance bound to be a function of $1 - \mu$. Thus the final result should be a function of $\mu(1-\mu)$, as we can't tell X apart from Y if we only care about the variance.* Still another way is to note $\mathrm{Var}(X) = \mathbb{E}[X^2] - \mathbb{E}[X]^2 = \mathbb{E}[X^2] - \mu^2$, and then note

$$\mathbb{E}[X^2] \;=\; \int_0^1 x^2 p(x)dx \;\le\; \int_0 xp(x)dx \;=\; \mathbb{E}[X];$$

this is essentially the same calculation, just written differently.

(b) The argument proceeds similarly. As $a \le X \le b$, $\int_a^b ap(x)dx \le \int_a^b xp(x)dx \le \int_a^b bp(x)dx$, so $a \le \mathbb{E}[X] \le b$. For the variance, we could use $\mathrm{Var}(X) = \mathbb{E}[X^2] - \mathbb{E}[X]^2$, but it's better to reduce to part (a). Let $Y = (X-a)/(b-a)$. Note $0 \le Y \le 1$ and $\mu_Y = (\mu_X - a)/(b-a)$. By part (a), the variance of Y is at most $\mu_Y(1-\mu_Y)$, which gives

$$\mathrm{Var}(X) \;\le\; \frac{\mu_X - a}{b-a}\left(1 - \frac{\mu_X - a}{b-a}\right) \;=\; \frac{(\mu_X - a)(b - \mu_X)}{(b-a)^2}.$$

Note that $\mathrm{Var}(Y) = \mathrm{Var}((X-a)/(b-a)) = \mathrm{Var}(X)/(b-a)^2$. Thus $\mathrm{Var}(X) \le (\mu_X - a)(b - \mu_X)$. Using calculus, we see this is largest when $\mu_X = \frac{b-a}{2}$, which gives after some algebra $\mathrm{Var}(X) \le \frac{1}{4}(b-a)^2$. (To see this, let $f(u) = (u-a)(b-u) = -u^2 + (b-a)u - ab$, so $f'(u) = -2u + (b-a)$, so the critical point is where $u = (b-a)/2$.)

(c) If half the numbers are 9 and half are 0, then the mean is 4.5 and the standard deviation is 4.5 (so the variance is 4.5^2), as everything is 4.5 units from the mean. From part (b), the maximum the variance of X can be is $\frac{1}{4}(9-0)^2 = 20.25 = 4.5^2$. Thus the variance is as large as possible. This forces the mean to be 4.5, and then the variance is maximized when half are 0 and half are 9. It's not surprising that parts (a) and (b) are useful here.