Stat 110 Midterm Review, Fall 2011

Prof. Joe Blitzstein (Department of Statistics, Harvard University)

1 General Information

The midterm will be in class on Wednesday, October 12. There is no alternate time for the exam, so please be there and arrive on time! Cell phones must be off, so it is a good idea to bring a watch. No books, notes, or calculators are allowed, except that you may bring two sheets of standard-sized paper (8.5" x 11") with whatever you want written on it (two-sided): notes, theorems, formulas, information about the important distributions, etc.

There will be 4 problems, weighted equally. Many of the parts can be done quickly if you have a good understanding of the ideas covered in class (e.g., seeing if you can use Bayes' Rule or understand what independence means), and for many you can just write down an answer without needing to simplify. None will require long or messy calculations. They are *not* arranged in order of increasing difficulty. Since it is a short exam, make sure not to spend too long on any one problem.

Suggestions for studying: review all the homeworks and read the solutions, study your lecture notes (and possibly relevant sections from either book), the strategic practice problems, and this handout. Solving practice problems (which means trying hard to work out the details yourself, not just trying for a minute and then looking at the solution!) is extremely important.

2 Topics

- Combinatorics: multiplication rule, tree diagrams, binomial coefficients, permutations and combinations, inclusion-exclusion, story proofs.
- Basic Probability: sample spaces, events, axioms of probability, equally likely outcomes, inclusion-exclusion, unions, intersections, and complements.
- Conditional Probability: definition and meaning, writing $P(A_1 \cap A_2 \cap \cdots \cap A_n)$ as a product, Bayes' Rule, Law of Total Probability, thinking conditionally, independence vs. conditional independence.

- Random Variables: definition, meaning of $X = x$, stories, indicator r.v.s, probability mass functions (PMFs), probability density functions (PDFs), cumulative distribution functions (CDFs), independence, Poisson approximation.
- Expected Value and Variance: definitions, linearity, standard deviation, Law of the Unconscious Statistician (LOTUS).
- Important Discrete Distributions: Bernoulli, Binomial, Geometric, Negative Binomial, Hypergeometric, Poisson.
- Important Continuous Distributions: Uniform, Normal.
- General Concepts: stories, symmetry, discrete vs. continuous, conditional probability is the soul of statistics, checking simple and extreme cases.
- Important Examples: birthday problem, matching problem, Newton-Pepys problem, Monty Hall problem, testing for a rare disease, elk problem (capture-recapture), gambler's ruin, Simpson's paradox, St. Petersburg paradox.

3 Important Distributions

The eight most important distributions we have discussed so far are listed below, each with its PMF/PDF, mean, and variance. *This table will be provided on the last page of the midterm.* As usual, we let $0 < p < 1$ and $q = 1 - p$. Each of these distributions is important because it has a natural, useful story, so understanding these stories (and recognizing equivalent stories) is crucial. It is also important to know how these distributions are related to each other. For example, Bern(p) is the same as Bin($1, p$), and Bin(n, p) is approximately Pois(λ) if n is large, p is small and $\lambda = np$ is moderate.

Name	Param.	PMF or PDF	Mean	Variance
Bernoulli	p	$P(X = 1) = p, P(X = 0) = q$	p	pq
Binomial	n, p	$\binom{n}{k} p^k q^{n-k}$, for $k \in \{0, 1, \dots, n\}$	np	npq
Geometric	p	$q^k p$, for $k \in \{0, 1, 2, \dots\}$	q/p	q/p^2
NegBinom	r, p	$\binom{r+n-1}{r-1} p^r q^n, n \in \{0, 1, 2, \dots\}$	rq/p	rq/p^2
Hypergeom	w, b, n	$\frac{\binom{w}{k}\binom{b}{n-k}}{\binom{w+b}{n}}$, for $k \in \{0, 1, \dots, n\}$	$\mu = \frac{nw}{w+b}$	$\left(\frac{w+b-n}{w+b-1}\right) n \frac{\mu}{n}\left(1 - \frac{\mu}{n}\right)$
Poisson	λ	$\frac{e^{-\lambda}\lambda^k}{k!}$, for $k \in \{0, 1, 2, \dots\}$	λ	λ
Uniform	$a < b$	$\frac{1}{b-a}$, for $x \in (a, b)$	$\frac{a+b}{2}$	$\frac{(b-a)^2}{12}$
Normal	μ, σ^2	$\frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/(2\sigma^2)}$	μ	σ^2

4 Some Useful Formulas

4.1 De Morgan's Laws

$$(A_1 \cup A_2 \cdots \cup A_n)^c = A_1^c \cap A_2^c \cdots \cap A_n^c$$
$$(A_1 \cap A_2 \cdots \cap A_n)^c = A_1^c \cup A_2^c \cdots \cup A_n^c$$

4.2 Complements

$$P(A^c) = 1 - P(A)$$

4.3 Unions

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

$$P(A_1 \cup A_2 \cup \cdots \cup A_n) = \sum_{i=1}^{n} P(A_i), \text{ if the } A_i \text{ are disjoint}$$

$$P(A_1 \cup A_2 \cup \cdots \cup A_n) \leq \sum_{i=1}^{n} P(A_i)$$

$$P(A_1 \cup A_2 \cup \cdots \cup A_n) = \sum_{k=1}^{n} \left((-1)^{k+1} \sum_{i_1 < i_2 < \cdots < i_k} P(A_{i_1} \cap A_{i_2} \cap \cdots \cap A_{i_k}) \right) \text{ (Inclusion-Exclusion)}$$

4.4 Intersections

$$P(A \cap B) = P(A)P(B|A) = P(B)P(A|B)$$
$$P(A_1 \cap A_2 \cap \cdots \cap A_n) = P(A_1)P(A_2|A_1)P(A_3|A_1, A_2) \cdots P(A_n|A_1, \ldots, A_{n-1})$$

4.5 Law of Total Probability

If $E_1, E_2, \ldots, E_n$ are a partition of the sample space S (i.e., they are disjoint and their union is all of S) and $P(E_j) \neq 0$ for all j, then

$$P(B) = \sum_{j=1}^{n} P(B|E_j)P(E_j)$$

4.6 Bayes' Rule

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

Often the denominator $P(B)$ is then expanded by the Law of Total Probability.

4.7 Expected Value and Variance

Expected value is linear: for any random variables X and Y and constant c,

$$E(X+Y) = E(X) + E(Y)$$

$$E(cX) = cE(X)$$

It is *not* true in general that $\mathrm{Var}(X+Y) = \mathrm{Var}(X) + \mathrm{Var}(Y)$. For example, let X be Bernoulli(1/2) and $Y = 1 - X$ (note that Y is also Bernoulli(1/2)). Then $\mathrm{Var}(X) + \mathrm{Var}(Y) = 1/4 + 1/4 = 1/2$, but $\mathrm{Var}(X+Y) = \mathrm{Var}(1) = 0$ since $X+Y$ is always equal to the constant 1. (We will see later exactly when the variance of the sum is the sum of the variances.)

Constants come out from variance as the constant squared:

$$\mathrm{Var}(cX) = c^2\mathrm{Var}(X)$$

The variance of X is defined as $E(X - EX)^2$, but often it is easier to compute using the following:

$$\mathrm{Var}(X) = E(X^2) - (EX)^2$$

4.8 Law of the Unconscious Statistician (LOTUS)

Let X be a discrete random variable and h be a real-valued function. Then $Y = h(X)$ is a random variable. To compute $E(Y)$ using the definition of expected value, we would need to first find the PMF of Y, and then use $E(Y) = \sum_y yP(Y=y)$. The Law of the Unconscious Statistician says we can use the PMF of X directly:

$$E(h(X)) = \sum_x h(x)P(X=x),$$

where the sum is over all possible values of X. Similarly, for X a continuous r.v. with PDF f_X, we can find the expected value of $Y = h(X)$ using the PDF of X, without having to find the PDF of Y:

$$E(h(X)) = \int_{-\infty}^{\infty} h(x)f_X(x)dx$$

5 Stat 110 Midterm from 2007

1. Alice and Bob have just met, and wonder whether they have a mutual friend. Each has 50 friends, out of 1000 other people who live in their town. They think that it's unlikely that they have a friend in common, saying "each of us is only friends with 5% of the people here, so it would be very unlikely that our two 5%'s overlap."

Assume that Alice's 50 friends are a random sample of the 1000 people (equally likely to be any 50 of the 1000), and similarly for Bob. Also assume that knowing who Alice's friends are gives no information about who Bob's friends are.

(a) Compute the expected number of mutual friends Alice and Bob have (simplify).

(b) Let X be the number of mutual friends they have. Find the PMF of X.

(c) Is the distribution of X one of the important distributions we have looked at and if so, which one? Note: even without solving (b), you can get credit by giving clear reasons for or against each of the important distributions.

2. Two coins are in a hat. The coins look alike, but one coin is fair (with probability 1/2 of Heads), while the other coin is biased, with probability 1/4 of Heads. One of the coins is randomly pulled from the hat, without knowing which of the two it is. Call the chosen coin "Coin C".

(a) Coin C is tossed twice, showing Heads both times. Given this information, what is the probability that Coin C is the fair coin? (Simplify.)

(b) Are the events "first toss of Coin C is Heads" and "second toss of Coin C is Heads" independent? Explain briefly.

(c) Find the probability that in 10 flips of Coin C, there will be exactly 3 Heads. (The coin is equally likely to be either of the 2 coins; do not assume it already landed Heads twice as in (a). Do not simplify.)

3. Five people have just won a \$100 prize, and are deciding how to divide the \$100 up between them. Assume that whole dollars are used, not cents. Also, for example, giving \$50 to the first person and \$10 to the second is different from vice versa.

(a) How many ways are there to divide up the \$100, such that each gets at least \$10?

Hint: there are $\binom{n+k-1}{k}$ ways to put k indistinguishable balls into n distinguishable boxes; you can use this fact without deriving it.

(b) Assume that the \$100 is randomly divided up, with all of the possible allocations counted in (a) equally likely. Find the expected amount of money that the first person receives (justify your reasoning).

(c) Let A_j be the event that the jth person receives more than the first person (for $2 \leq j \leq 5$), when the \$100 is randomly allocated as in (b). Are A_2 and A_3 independent? (No explanation needed for this.) Express $I_{A_2 \cap A_3}$ and $I_{A_2 \cup A_3}$ in terms of I_{A_2} and I_{A_3} (where I_A is the indicator random variable of any event A).

4. (a) Let $X \sim \text{Pois}(\lambda)$, with $\lambda > 0$. Find $E(X!)$, the average factorial of X. (Simplify, and specify what condition on λ is needed to make the expectation finite.)

(b) The owner of a certain website is studying the distribution of the number of visitors to the site. Every day, a million people independently decide whether to visit the site, with probability $p = 2 \times 10^{-6}$ of visiting. Give a good, simple approximation of the probability of getting *at least two* visitors on a particular day (simplify; your answer should not involve series).

(c) In the scenario of (b), approximately how many days will it take on average until there is a day with at least two visitors (including the day itself)?

5. Alice flips a fair coin n times and Bob flips another fair coin $n+1$ times, resulting in independent $X \sim \text{Bin}(n, \frac{1}{2})$ and $Y \sim \text{Bin}(n+1, \frac{1}{2})$.

(a) Let $V = \min(X, Y)$ be the smaller of X and Y, and let $W = \max(X, Y)$ be the larger of X and Y. (If $X = Y$, then $V = W = X = Y$.) Find $E(V) + E(W)$ in terms of n (simplify).

(b) Is it true that $P(X < Y) = P(n - X < n + 1 - Y)$? Explain why or why not.

(c) Compute $P(X < Y)$ (simplify). Hint: use (b) and that X and Y are integers.

6 Stat 110 Midterm from 2008

1. The gambler de Méré asked Pascal whether it is more likely to get at least one six in 4 rolls of a die, or to get at least one double-six in 24 rolls of a pair of dice. Continuing this pattern, suppose that a group of n fair dice is rolled $4 \cdot 6^{n-1}$ times.

(a) Find the expected number of times that "all sixes" is achieved (i.e., how often among the $4 \cdot 6^{n-1}$ rolls it happens that all n dice land 6 simultaneously). (Simplify.)

(b) Give a simple but accurate approximation of the probability of having at least one occurrence of "all sixes", for n large (in terms of e but not n).

(c) de Méré finds it tedious to re-roll so many dice. So after one normal roll of the n dice, in going from one roll to the next, with probability $6/7$ he leaves the dice in the same configuration and with probability $1/7$ he re-rolls. For example, if $n = 3$ and the 7th roll is $(3, 1, 4)$, then $6/7$ of the time the 8th roll remains $(3, 1, 4)$ and $1/7$ of the time the 8th roll is a new random outcome. Does the expected number of times that "all sixes" is achieved stay the same, increase, or decrease (compared with (a))? Give a short but clear explanation.

2. To battle against spam, Bob installs two anti-spam programs. An email arrives, which is either legitimate (event L) or spam (event L^c), and which program j marks as legitimate (event M_j) or marks as spam (event M_j^c) for $j \in \{1, 2\}$. Assume that 10% of Bob's email is legitimate and that the two programs are each "90% accurate" in the sense that $P(M_j|L) = P(M_j^c|L^c) = 9/10$. Also assume that given whether an email is spam, the two programs' outputs are conditionally independent.

(a) Find the probability that the email is legitimate, given that the 1st program marks it as legitimate (simplify).

(b) Find the probability that the email is legitimate, given that both programs mark it as legitimate (simplify).

(c) Bob runs the 1st program and M_1 occurs. He updates his probabilities and then runs the 2nd program. Let $\tilde{P}(A) = P(A|M_1)$ be the updated probability function after running the 1st program. Explain briefly in words whether or not $\tilde{P}(L|M_2) = P(L|M_1 \cap M_2)$: is conditioning on $M_1 \cap M_2$ in one step equivalent to first conditioning on M_1, then updating probabilities, and then conditioning on M_2?

3. (a) Let $X_1, X_2, \dots$ be independent $\mathcal{N}(0, 4)$ r.v.s., and let J be the smallest value of j such that $X_j > 4$ (i.e., the index of the first X_j exceeding 4). In terms of the standard Normal CDF Φ, find $E(J)$ (simplify).

(b) Let f and g be PDFs with $f(x) > 0$ and $g(x) > 0$ for all x. Let X be a random variable with PDF f. Find the expected value of the ratio $\frac{g(X)}{f(X)}$ (simplify).

(c) Define $F(x) = e^{-e^{-x}}$. This is a CDF (called the *Gumbel distribution*) and is a continuous, strictly increasing function. Let X have CDF F, and define $W = F(X)$. What are the mean and variance of W (simplify)?

4. (a) Find $E(2^X)$ for $X \sim \text{Pois}(\lambda)$ (simplify).

(b) Let X and Y be independent $\text{Pois}(\lambda)$ r.v.s, and $T = X + Y$. Later in the course, we will show that $T \sim \text{Pois}(2\lambda)$; here you may use this fact. Find the conditional distribution of X given $T = n$, i.e., find the conditional PMF $P(X = k|T = n)$ (simplify). Which "important distribution" is this conditional distribution, if any?

(c) Again let X and Y be $\text{Pois}(\lambda)$ r.v.s, and $T = X + Y$, but now assume now that X and Y are *not* independent, and in fact $X = Y$. Prove or disprove the claim that $T \sim \text{Pois}(2\lambda)$ in this scenario.

7 Stat 110 Midterm from 2009

1. (a) Let $X \sim \text{Pois}(\lambda)$. Find $E(e^X)$ (simplify).

(b) The numbers $1, 2, 3, \ldots, n$ are listed in some random order (with all $n!$ permutations equally likely). An *inversion* occurs each time a pair of numbers is out of order, i.e., the larger number is earlier in the list than the smaller number. For example, $3, 1, 4, 2$ has 3 inversions (3 before 1, 3 before 2, 4 before 2). Find the expected number of inversions in the list (simplify).

2. Consider four nonstandard dice (the *Efron dice*), whose sides are labeled as follows (the 6 sides on each die are equally likely).

A: $4, 4, 4, 4, 0, 0$
B: $3, 3, 3, 3, 3, 3$
C: $6, 6, 2, 2, 2, 2$
D: $5, 5, 5, 1, 1, 1$

These four dice are each rolled once. Let A be the result for die A, B be the result for die B, etc.

(a) Find $P(A > B), P(B > C), P(C > D)$, and $P(D > A)$.

(b) Is the event $A > B$ independent of the event $B > C$? Is the event $B > C$ independent of the event $C > D$? Explain.

3. A discrete distribution has the *memoryless property* if for X a random variable with that distribution, $P(X \geq j+k|X \geq j) = P(X \geq k)$ for all nonnegative integers j, k.

(a) If X has a memoryless distribution with CDF F and PMF $p_i = P(X = i)$, find an expression for $P(X \geq j + k)$ in terms of $F(j), F(k), p_j, p_k$.

(b) Name one important discrete distribution we have studied so far which has the memoryless property. Justify your answer with a clear interpretation in words or with a computation.

4. The book *Red State, Blue State, Rich State, Poor State* (by Andrew Gelman) discusses the following election phenomenon: within any U.S. state, a wealthy voter is more likely to vote for a Republican than a poor voter; yet the wealthier states tend to favor Democratic candidates! In short: rich individuals (in any state) tend to vote for Republicans, while states with a higher percentage of rich people tend to favor Democrats.

(a) Assume for simplicity that there are only 2 states (called Red and Blue), each of which has 100 people, and that each person is either rich or poor, and either a Democrat or a Republican. Make up numbers consistent with the above, showing how this phenomenon is possible, by giving a 2 by 2 table for each state (listing how many people in each state are rich Democrats, etc.).

(b) In the setup of (a) (not necessarily with the numbers you made up there), let D be the event that a randomly chosen person is a Democrat (with all 200 people equally likely), and B be the event that the person lives in the Blue State. Suppose that 10 people move from the Blue State to the Red State. Write P_{old} and P_{new} for probabilities before and after they move. Assume that people do not change parties, so we have $P_{\text{new}}(D) = P_{\text{old}}(D)$. Is it possible that *both* $P_{\text{new}}(D|B) > P_{\text{old}}(D|B)$ and $P_{\text{new}}(D|B^c) > P_{\text{old}}(D|B^c)$ are true? If so, explain how it is possible and why it does not contradict the law of total probability $P(D) = P(D|B)P(B) + P(D|B^c)P(B^c)$; if not, show that it is impossible.

8 Stat 110 Midterm from 2010

1. A family has two children. The genders of the first-born and second-born are independent (with boy and girl equally likely), and which seasons the children were born in are independent, with all 4 seasons equally likely.

(a) Find the probability that both children are girls, given that a randomly chosen one of the two is a girl who was born in winter (simplify).

(b) Find the probability that both children are girls, given that at least one of the two is a girl who was born in winter (simplify).

2. In each day that the "Mass Cash" lottery is run in Massachusetts, 5 of the integers from 1 to 35 are chosen (randomly and without replacement).

(a) When playing this lottery, find the probability of guessing exactly 3 numbers right, given that you guess at least 1 of the numbers right (leave your answer in terms of binomial coefficients).

(b) Find an exact expression for the expected number of days needed so that all of the $\binom{35}{5}$ possible lottery outcomes will have occurred (leave your answer as a sum, which can involve binomial coefficients).

(c) Approximate the probability that after 50 days of the lottery, every number from 1 to 35 has been picked at least once (don't simplify, but your answer shouldn't involve a sum).

3. Let $U \sim \text{Unif}(0,1)$, and $X = \ln\left(\frac{U}{1-U}\right)$.
(a) Write down (but do not compute) an integral giving $E(X^2)$.

(b) Find the CDF of X (simplify).

(c) Find $E(X)$ *without using calculus* (simplify).
Hint: $1-U$ has the same distribution as U.

4. Let $X_1, X_2, X_3, \ldots, X_{10}$ be the total number of inches of rain in Boston in October of 2011, 2012, 2013, ..., 2020, with these r.v.s independent $\mathcal{N}(\mu, \sigma^2)$. (Of course, rainfall can't be negative, but μ and σ are such that it is extremely likely that all the X_j's are positive.) We say that a *record value* is set in a certain year if the rainfall is greater than all the previous years (going back to 2011; so by definition, a record is always set in the first year, 2011).

(a) On average, how many of these 10 years will set record values? (Your answer can be a sum but the terms should be simple.)

(b) Is the indicator of whether the year 2013 sets a record independent of the indicator of whether the year 2014 sets a record? (Justify briefly.)

(c) Later in the course, we will show that if $Y_1 \sim \mathcal{N}(\mu_1, \sigma_1^2)$ and $Y_2 \sim \mathcal{N}(\mu_2, \sigma_2^2)$ are independent, then $Y_1 - Y_2 \sim \mathcal{N}(\mu_1 - \mu_2, \sigma_1^2 + \sigma_2^2)$. Using this fact, find the probability that the October 2014 rainfall will be more than double the October 2013 rainfall in Boston, in terms of Φ.

Stat 110 Midterm Review Solutions, Fall 2011

Prof. Joe Blitzstein (Department of Statistics, Harvard University)

Here are solutions to the midterm review problems. Please try your best to solve a problem *before* reading the solution. Good luck!

1 Stat 110 Midterm from 2007

1. Alice and Bob have just met, and wonder whether they have a mutual friend. Each has 50 friends, out of 1000 other people who live in their town. They think that it's unlikely that they have a friend in common, saying "each of us is only friends with 5% of the people here, so it would be very unlikely that our two 5%'s overlap."

Assume that Alice's 50 friends are a random sample of the 1000 people (equally likely to be any 50 of the 1000), and similarly for Bob. Also assume that knowing who Alice's friends are gives no information about who Bob's friends are.

(a) Compute the expected number of mutual friends Alice and Bob have (simplify).

Let I_j be an indicator r.v. for the jth person being a mutual friend. Then

$$E(\sum_{j=1}^{1000} I_j) = 1000E(I_1) = 1000P(I_1 = 1) = 1000 \cdot (\frac{5}{100})^2 = 2.5.$$

(b) Let X be the number of mutual friends they have. Find the PMF of X.

Condition on who Alice's friends are, and then count the number of ways that Bob can be friends with exactly k of them. This gives

$$P(X = k) = \frac{\binom{50}{k}\binom{950}{50-k}}{\binom{1000}{50}}$$

for $0 \leq k \leq 50$ (and 0 otherwise).

(c) Is the distribution of X one of the 5 important distributions we have looked at and if so, which one? Note: even without solving (b), you can get credit by giving clear reasons for or against each of the 5 distributions.

Yes, it is the Hypergeometric distribution, as shown by the PMF from (b) or by thinking of "tagging" Alice's friends (like the elk) and then seeing how many tagged people there are among Bob's friends.

2. Two coins are in a hat. The coins look alike, but one coin is fair (with probability 1/2 of Heads), while the other coin is biased, with probability 1/4 of Heads. One of the coins is randomly pulled from the hat, without knowing which of the two it is. Call the chosen coin "Coin C".

(a) Coin C is tossed twice, showing Heads both times. Given this information, what is the probability that Coin C is the fair coin? (Simplify.)

By Bayes' Rule,

$$P(\text{fair}|HH) = \frac{P(HH|\text{fair})P(\text{fair})}{P(HH)} = \frac{(1/4)(1/2)}{(1/4)(1/2) + (1/16)(1/2)} = \frac{4}{5}.$$

(b) Are the events "first toss of Coin C is Heads" and "second toss of Coin C is Heads" independent? Explain briefly.

They're not independent: the first toss being Heads is evidence in favor of the coin being the fair coin, giving information about probabilities for the second toss.

(c) Find the probability that in 10 flips of Coin C, there will be exactly 3 Heads. (The coin is equally likely to be either of the 2 coins; do not assume it already landed Heads twice as in (a). Do not simplify.)

Let X be the number of Heads in 10 tosses. By the Law of Total Probability (conditioning on which of the two coins C is),

$$\begin{aligned} P(X=3) &= P(X=3|\text{fair})P(\text{fair}) + P(X=3|\text{biased})P(\text{biased}) \\ &= \binom{10}{3}(1/2)^{10}(1/2) + \binom{10}{3}(1/4)^3(3/4)^7(1/2) \\ &= \frac{1}{2}\binom{10}{3}\left(\frac{1}{2^{10}} + \frac{3^7}{4^{10}}\right). \end{aligned}$$

3. Five people have just won a \$100 prize, and are deciding how to divide the \$100 up between them. Assume that whole dollars are used, not cents. Also, for example, giving \$50 to the first person and \$10 to the second is different from vice versa.

(a) How many ways are there to divide up the \$100, such that each gets at least \$10?

Hint: there are $\binom{n+k-1}{k}$ ways to put k indistinguishable balls into n distinguishable boxes; you can use this fact without deriving it.

Give each person \$10, and then distribute the remaining \$50 arbitrarily. By the hint (thinking of people as boxes and dollars as balls!), the number of ways is

$$\binom{5+50-1}{50} = \binom{54}{50} = \binom{54}{4}.$$

(b) Assume that the \$100 is randomly divided up, with all of the possible allocations counted in (a) equally likely. Find the expected amount of money that the first person receives (justify your reasoning).

Let X_j be the amount that j gets. By symmetry, $E(X_j)$ is the same for all j. But $X_1 + \cdots + X_5 = 100$, so by linearity $100 = 5EX_1$. Thus, EX_1 is \$20.

(c) Let A_j be the event that the jth person receives more than the first person (for $2 \le j \le 5$), when the \$100 is randomly allocated as in (b). Are A_2 and A_3 independent? (No explanation needed for this.) Express $I_{A_2 \cap A_3}$ and $I_{A_2 \cup A_3}$ in terms of I_{A_2} and I_{A_3} (where I_A is the indicator random variable of any event A).

The events A_2 and A_3 are not independent.

By definition, $I_{A_2 \cap A_3}$ is 1 exactly when I_{A_2}, I_{A_3} are both 1. So

$$I_{A_2 \cap A_3} = I_{A_2} I_{A_3}.$$

As in inclusion-exclusion,

$$I_{A_2 \cup A_3} = I_{A_2} + I_{A_3} - I_{A_2 \cap A_3}.$$

By the above, this is $I_{A_2} + I_{A_3} - I_{A_2} I_{A_3}$.

4. (a) Let $X \sim \text{Pois}(\lambda)$, with $\lambda > 0$. Find $E(X!)$, the average factorial of X. (Simplify, and specify what condition on λ is needed to make the expectation finite.)

By LOTUS,

$$E(X!) = \sum_{k=0}^{\infty} k! e^{-\lambda} \lambda^k / k! = e^{-\lambda} \sum_{k=0}^{\infty} \lambda^k = \frac{e^{-\lambda}}{1-\lambda}$$

for $0 < \lambda < 1$.

(b) The owner of a certain website is studying the distribution of the number of visitors to the site. Every day, a million people independently decide whether to visit the site, with probability $p = 2 \times 10^{-6}$ of visiting. Give a good, simple approximation of the probability of getting *at least two* visitors on a particular day (simplify; your answer should not involve series).

Let X be the number of visitors. Here $n = 10^6$ is large, p is small, and $np = 2$ is moderate. So the Pois(2) distribution gives a good approximation, and

$$P(X \geq 2) = 1 - P(X < 2) \approx 1 - e^{-2} - e^{-2} \cdot 2 = 1 - \frac{3}{e^2}.$$

(c) In the scenario of (b), approximately how many days will it take on average until there is a day with at least two visitors (including the day itself)?

Let T be the number of days needed, so $T - 1$ is Geometric with parameter the probability found in (b) (using the convention that the Geometric starts at 0). Then $E(T-1) = (1-p_2)/p_2$, where p_2 is the probability from (b). Thus,

$$E(T) = 1/p_2 \approx (1 - 3/e^2)^{-1}.$$

5. Alice flips a fair coin n times and Bob flips another fair coin $n+1$ times, resulting in independent $X \sim \text{Bin}(n, \frac{1}{2})$ and $Y \sim \text{Bin}(n+1, \frac{1}{2})$.

(a) Let $V = \min(X, Y)$ be the smaller of X and Y, and let $W = \max(X, Y)$ be the larger of X and Y. (If $X = Y$, then $V = W = X = Y$.) Find $E(V) + E(W)$ in terms of n (simplify).

Note that $V + W = X + Y$ (adding the smaller and larger is the same as adding both numbers). So by linearity,

$$E(V) + E(W) = E(V + W) = E(X + Y) = E(X) + E(Y) = (2n + 1)/2 = n + \frac{1}{2}.$$

(b) Is it true that $P(X < Y) = P(n - X < n + 1 - Y)$? Explain why or why not.

Yes: $n - X \sim \text{Bin}(n, 1/2)$ (if a fair coin is flipped n times, then the number of Heads and the number of Tails have the same distribution). Similarly, $n+1-Y$ has the same distribution as Y, so the equation is true.

(c) Compute $P(X < Y)$ (simplify). Hint: use (b) and that X and Y are integers.

Simplifying,

$$P(X < Y) = P(n - X < n + 1 - Y) = P(Y < X + 1) = P(Y \leq X)$$

since X and Y are integers (e.g., $Y < 5$ is equivalent to $Y \leq 4$). But $Y \leq X$ is the complement of $X < Y$, so $P(X < Y) = 1 - P(X < Y)$. Thus, $P(X < Y) = 1/2$.

2 Stat 110 Midterm from 2008

1. The gambler de Méré asked Pascal whether it is more likely to get at least one six in 4 rolls of a die, or to get at least one double-six in 24 rolls of a pair of dice. Continuing this pattern, suppose that a group of n fair dice is rolled $4 \cdot 6^{n-1}$ times.

(a) Find the expected number of times that "all sixes" is achieved (i.e., how often among the $4 \cdot 6^{n-1}$ rolls it happens that all n dice land 6 simultaneously). (Simplify.)

Let I_j be an indicator r.v. for the event "all sixes" on the jth roll. Then $E(I_j) = 1/6^n$, so the expected value is $4 \cdot 6^{n-1}/6^n = 2/3$.

(b) Give a simple but accurate approximation of the probability of having at least one occurrence of "all sixes", for n large (in terms of e but not n).

By a Poisson approximation with $\lambda = 2/3$ (the expected value from (a)), the probability is approximately $1 - e^{-2/3}$.

(c) de Méré finds it tedious to re-roll so many dice. So after one normal roll of the n dice, in going from one roll to the next, with probability 6/7 he leaves the dice in the same configuration and with probability 1/7 he re-rolls. For example, if $n = 3$ and the 7th roll is $(3, 1, 4)$, then 6/7 of the time the 8th roll remains $(3, 1, 4)$ and 1/7 of the time the 8th roll is a new random outcome. Does the expected number of times that "all sixes" is achieved stay the same, increase, or decrease (compared with (a))? Give a short but clear explanation.

The answer stays the same, by the same reasoning as in (a), since linearity of expectation holds even for dependent r.v.s.

2. To battle against spam, Bob installs two anti-spam programs. An email arrives, which is either legitimate (event L) or spam (event L^c), and which program j marks as legitimate (event M_j) or marks as spam (event M_j^c) for $j \in \{1, 2\}$. Assume that 10% of Bob's email is legitimate and that the two programs are each "90% accurate" in the sense that $P(M_j|L) = P(M_j^c|L^c) = 9/10$. Also assume that given whether an email is spam, the two programs' outputs are conditionally independent.

(a) Find the probability that the email is legitimate, given that the 1st program marks it as legitimate (simplify).

By Bayes' Rule,

$$P(L|M_1) = \frac{P(M_1|L)P(L)}{P(M_1)} = \frac{\frac{9}{10} \cdot \frac{1}{10}}{\frac{9}{10} \cdot \frac{1}{10} + \frac{1}{10} \cdot \frac{9}{10}} = \frac{1}{2}.$$

(b) Find the probability that the email is legitimate, given that both programs mark it as legitimate (simplify).

By Bayes' Rule,

$$P(L|M_1, M_2) = \frac{P(M_1, M_2|L)P(L)}{P(M_1, M_2)} = \frac{(\frac{9}{10})^2 \cdot \frac{1}{10}}{(\frac{9}{10})^2 \cdot \frac{1}{10} + (\frac{1}{10})^2 \cdot \frac{9}{10}} = \frac{9}{10}.$$

(c) Bob runs the 1st program and M_1 occurs. He updates his probabilities and then runs the 2nd program. Let $\tilde{P}(A) = P(A|M_1)$ be the updated probability function after running the 1st program. Explain briefly in words whether or not $\tilde{P}(L|M_2) = P(L|M_1 \cap M_2)$: is conditioning on $M_1 \cap M_2$ in one step equivalent to first conditioning on M_1, then updating probabilities, and then conditioning on M_2?

Yes, they are the same. If this were not the case, conditional probability would be incoherent, since both are valid methods for updating probability. The probability of an event given various pieces of evidence does not depend on the order in which the pieces of evidence are incorporated into the updated proabilities.

3. (a) Let $X_1, X_2, \dots$ be independent $\mathcal{N}(0,4)$ r.v.s., and let J be the smallest value of j such that $X_j > 4$ (i.e., the index of the first X_j exceeding 4). In terms of the standard Normal CDF Φ, find $E(J)$ (simplify).

We have $J - 1 \sim \text{Geom}(p)$ with $p = P(X_1 > 4) = P(X_1/2 > 2) = 1 - \Phi(2)$, so $E(J) = 1/(1 - \Phi(2))$.

(b) Let f and g be PDFs with $f(x) > 0$ and $g(x) > 0$ for all x. Let X be a random variable with PDF f. Find the expected value of the ratio $\frac{g(X)}{f(X)}$ (simplify).

By LOTUS,

$$E\frac{g(X)}{f(X)} = \int_{-\infty}^{\infty} \frac{g(x)}{f(x)} f(x)dx = \int_{-\infty}^{\infty} g(x)dx = 1.$$

(c) Define $F(x) = e^{-e^{-x}}$. This is a CDF (called the *Gumbel distribution*) and is a continuous, strictly increasing function. Let X have CDF F, and define $W = F(X)$. What are the mean and variance of W (simplify)?

Note that W is obtained by plugging X into its own CDF. The CDF of W is

$$P(W \le w) = P(F(X) \le w) = P(X \le F^{-1}(w)) = F(F^{-1}(w)) = w$$

for $0 < w < 1$, so $W \sim \text{Unif}(0,1)$. Thus, $E(W) = 1/2$ and $\text{Var}(W) = 1/12$.

4. (a) Find $E(2^X)$ for $X \sim \text{Pois}(\lambda)$ (simplify).

By LOTUS, $E(2^X) = \sum_{k=0}^{\infty} 2^k e^{-\lambda}\lambda^k/k! = e^{-\lambda}\sum_{k=0}^{\infty}(2\lambda)^k/k! = e^{-\lambda}e^{2\lambda} = e^{\lambda}$.

(b) Let X and Y be independent $\text{Pois}(\lambda)$ r.v.s, and $T = X + Y$. Later in the course, we will show that $T \sim \text{Pois}(2\lambda)$; here you may use this fact. Find the conditional distribution of X given $T = n$, i.e., find the conditional PMF $P(X = k|T = n)$ (simplify). Which "important distribution" is this conditional distribution, if any?

$$\frac{P(X = k, X + Y = n)}{P(T = n)} = \frac{P(X = k)P(Y = n - k)}{P(T = n)} = \frac{e^{-\lambda}\lambda^k}{k!}\frac{e^{-\lambda}\lambda^{n-k}}{n-k!}\frac{e^{2\lambda}n!}{(2\lambda)^n} = \binom{n}{k}\frac{1}{2^n},$$

which is the PMF of the $\text{Bin}(n, 1/2)$ distribution.

(c) Again let X and Y be $\text{Pois}(\lambda)$ r.v.s, and $T = X + Y$, but now assume now that X and Y are *not* independent, and in fact $X = Y$. Prove or disprove the claim that $T \sim \text{Pois}(2\lambda)$ in this scenario.

The r.v. $T = 2X$ is *not* Poisson: it can only take even values $0, 2, 4, 6, \ldots$, whereas any Poisson r.v. has positive probability of being any of $0, 1, 2, 3, \ldots$.

Alternatively, we can compute the PMF of $2X$, or note that $\text{Var}(2X) = 4\lambda \neq 2\lambda = E(2X)$, whereas for any Poisson r.v. the variance equals the mean.

3 Stat 110 Midterm from 2009

1. (a) Let $X \sim \text{Pois}(\lambda)$. Find $E(e^X)$ (simplify).

By LOTUS and the Taylor series for e^x,

$$E(e^X) = \sum_{k=0}^{\infty} e^k e^{-\lambda} \lambda^k / k! = e^{-\lambda} \sum_{k=0}^{\infty} (\lambda e)^k / k! = e^{-\lambda} e^{\lambda e} = e^{\lambda(e-1)}.$$

(b) The numbers $1, 2, 3, \dots, n$ are listed in some random order (with all $n!$ permutations equally likely). An *inversion* occurs each time a pair of numbers is out of order, i.e., the larger number is earlier in the list than the smaller number. For example, $3, 1, 4, 2$ has 3 inversions (3 before 1, 3 before 2, 4 before 2). Find the expected number of inversions in the list (simplify).

Let I_{ij} be the indicator of i and j being out of order, for each pair (i, j) with $i < j$. There are $\binom{n}{2}$ such indicators, each of which has expected value $1/2$ by symmetry (i before j and j before i are equally likely). So by linearity, the expected number of inversions is $\binom{n}{2}/2 = \frac{n(n-1)}{4}$.

2. Consider four nonstandard dice (the *Efron dice*), whose sides are labeled as follows (the 6 sides on each die are equally likely).

A: $4, 4, 4, 4, 0, 0$
B: $3, 3, 3, 3, 3, 3$
C: $6, 6, 2, 2, 2, 2$
D: $5, 5, 5, 1, 1, 1$

These four dice are each rolled once. Let A be the result for die A, B be the result for die B, etc.

(a) Find $P(A > B), P(B > C), P(C > D)$, and $P(D > A)$.

$$P(A > B) = P(A = 4) = 2/3$$
$$P(B > C) = P(C = 2) = 2/3$$
$$P(C > D) = P(C = 6) + P(C = 2, D = 1) = 2/3$$
$$P(D > A) = P(D = 5) + P(D = 1, A = 0) = 2/3$$

So the probability of each die beating the next is 2/3, going all the way around in a cycle (these are "nontransitive dice").

(b) Is the event $A > B$ independent of the event $B > C$? Is the event $B > C$ independent of the event $C > D$? Explain.

$A > B$ is independent of $B > C$ since $A > B$ is the same thing as $A = 4$, knowledge of which gives no information about $B > C$ (which is the same thing as $C = 2$). On the other hand, $B > C$ is *not* independent of $C > D$ since $P(C > D|C = 2) = 1/2 \neq 1 = P(C > D|C \neq 2)$.

3. A discrete distribution whose possible values are nonnegative integers has the *memoryless property* if for X an r.v. with that distribution, $P(X \geq j + k | X \geq j) = P(X \geq k)$ for all nonnegative integers j, k.

(a) If X has a memoryless distribution with CDF F and PMF $p_i = P(X = i)$, find an expression for $P(X \geq j + k)$ in terms of $F(j), F(k), p_j, p_k$.

By the memoryless property,

$$P(X \geq k) = P(X \geq j + k | X \geq j) = \frac{P(X \geq j + k, X \geq j)}{P(X \geq j)} = \frac{P(X \geq j + k)}{P(X \geq j)},$$

so

$$P(X \geq j + k) = P(X \geq j)P(X \geq k) = (1 - F(j) + p_j)(1 - F(k) + p_k).$$

(b) Name one important discrete distribution we have studied so far which has the memoryless property. Justify your answer with a clear interpretation in words or with a computation.

The Geometric distribution is memoryless (in fact, it turns out to be essentially the *only* discrete memoryless distribution!). This follows from the story of the Geometric: consider Bernoulli trials, waiting for the first success (and defining waiting time to be the number of failures before the first success). Say we have already had j failures without a success. Then the additional waiting time from that point forward has the same distribution as the original waiting time (the Bernoulli trials neither are conspiring against the experimenter nor act as if he or she is "due" for a success: the trials are independent). A calculation agrees: for $X \sim \text{Geom}(p)$,

$$P(X \geq j + k | X \geq j) = \frac{P(X \geq j + k)}{P(X \geq j)} = \frac{q^{j+k}}{q^j} = q^k = P(X \geq k).$$

4. The book *Red State, Blue State, Rich State, Poor State* (by Andrew Gelman) discusses the following election phenomenon: within any U.S. state, a wealthy voter is more likely to vote for a Republican than a poor voter; yet the wealthier states tend to favor Democratic candidates! In short: rich individuals (in any state) tend to vote for Republicans, while states with a higher percentage of rich people tend to favor Democrats.

(a) Assume for simplicity that there are only 2 states (called Red and Blue), each of which has 100 people, and that each person is either rich or poor, and either a Democrat or a Republican. Make up numbers consistent with the above, showing how this phenomenon is possible, by giving a 2 by 2 table for each state (listing how many people in each state are rich Democrats, etc.).

Red	Dem	Rep	**Total**
Rich	5	25	30
Poor	20	50	70
Total	25	75	100

Blue	Dem	Rep	**Total**
Rich	45	15	60
Poor	35	5	40
Total	80	20	100

The above tables are as desired: within each state, a rich person is more likely to be a Republican than a poor person; but the richer state has a higher percentage of Democrats than the poorer state. Of course, there are many possible tables that work.

Just giving tables all that was needed for this part, but note that the above example is a form of *Simpson's paradox*: aggregating the two tables seems to give different conclusions than conditioning on which state a person is in. Letting D, W, B be the events that a randomly chosen person is a Democrat, wealthy, and from the Blue State (respectively), for the above numbers we have $P(D|W, B) < P(D|W^c, B)$ and $P(D|W, B^c) < P(D|W^c, B^c)$ (controlling for whether the person is in the Red State or the Blue State, a poor person is more likely to be a Democrat than a rich person), but $P(D|W) > P(D|W^c)$ (stemming from the fact that the Blue State is richer and more Democratic).

(b) In the setup of (a) (not necessarily with the numbers you made up there), let D be the event that a randomly chosen person is a Democrat (with all 200 people equally likely), and B be the event that the person lives in the Blue State. Suppose that 10 people move from the Blue State to the Red State. Write P_{old} and P_{new} for probabilities before and after they move. Assume that people do not change parties, so we have $P_{\text{new}}(D) = P_{\text{old}}(D)$. Is it possible that *both* $P_{\text{new}}(D|B) > P_{\text{old}}(D|B)$ and $P_{\text{new}}(D|B^c) > P_{\text{old}}(D|B^c)$ are true? If so, explain how it is possible and why it does not contradict the law of total probability $P(D) = P(D|B)P(B) + P(D|B^c)P(B^c)$; if not, show that it is impossible.

Yes, it is possible. Suppose with the numbers from (a) that 10 people move from the Blue State to the Red State, of whom 5 are Democrats and 5 are Republicans. Then $P_{\text{new}}(D|B) = 75/90 > 80/100 = P_{\text{old}}(D|B)$ and $P_{\text{new}}(D|B^c) = 30/110 > 25/100 = P_{\text{old}}(D|B^c)$. Intuitively, this makes sense since the Blue State has a higher percentage of Democrats initially than the Red State, and the people who move have a percentage of Democrats which is between these two values.

This result does not contradict the law of total probability since the weights $P(B), P(B^c)$ also change: $P_{\text{new}}(B) = 90/200$, while $P_{\text{old}}(B) = 1/2$. The phenomenon could not occur if an equal number of people also move from the Red State to the Blue State (so that $P(B)$ is kept constant).

4 Stat 110 Midterm from 2010

1. A family has two children. The genders of the first-born and second-born are independent (with boy and girl equally likely), and which seasons the children were born in are independent, with all 4 seasons equally likely.

(a) Find the probability that both children are girls, given that a randomly chosen one of the two is a girl who was born in winter (simplify).

Once we specify the random child and learn she is a girl, we just need the other child to be a girl; this has probability 1/2.

To write this more precisely, let G_j be the event that the jth born is a girl and W_j be the event that the jth born is winter-born, for $j = 1, 2$. Define G_3, W_3 similarly for the randomly chosen child; we want $P(G_1 \cap G_2 | G_3 \cap W_3)$. Conditioning on the event A that the randomly chosen child is the first-born,

$$P(G_1 \cap G_2 | G_3 \cap W_3) = P(G_1 \cap G_2 | G_3, W_3, A) P(A | G_3, W_3) + P(G_1 \cap G_2 | G_3, W_3, A^c) P(A^c | G_3, W_3).$$

But

$$P(G_1 \cap G_2 | G_3, W_3, A) = P(G_1 \cap G_2 | G_1, W_1, A) = P(G_2 | G_1, W_1, A) = 1/2,$$

and similarly $P(G_1 \cap G_2 | G_3, W_3, A^c) = 1/2$, so the desired probability is 1/2.

(b) Find the probability that both children are girls, given that at least one of the two is a girl who was born in winter (simplify).

Since the probability that a specific child is a winter-born girl is 1/8,

$$\begin{aligned}
P(\text{both girls} | \text{at least one winter girl}) &= \frac{P(\text{both girls, at least one born in winter})}{P(\text{at least one winter girl})} \\
&= \frac{(1/4)(1 - (3/4)^2)}{1 - (7/8)^2} = \frac{7/64}{15/64} \\
&= 7/15.
\end{aligned}$$

Surprisingly, the seemingly irrelevant information about the season of birth matters, unlike in the previous part!

2. In each day that the "Mass Cash" lottery is run in Massachusetts, 5 of the integers from 1 to 35 are chosen (randomly and without replacement).

(a) Suppose you guess 5 numbers for the lottery. Find the probability of guessing exactly 3 numbers right, given that you guess at least 1 of the numbers right (leave your answer in terms of binomial coefficients).

The distribution is Hypergeometric (think of capture-recapture, "tagging" the numbers you choose). So

$$\begin{aligned} P(\text{exactly 3 right}|\text{at least 1 right}) &= \frac{P(\text{exactly 3 right})}{1 - P(\text{none right})} \\ &= \frac{\binom{5}{3}\binom{30}{2}/\binom{35}{5}}{1 - \binom{5}{0}\binom{30}{5}/\binom{35}{5}}. \end{aligned}$$

(b) Find an exact expression for the expected number of days needed so that all of the $\binom{35}{5}$ possible lottery outcomes will have occurred (leave your answer as a sum, which can involve binomial coefficients).

Let $n = \binom{35}{5}$. By the coupon collector problem (or directly by linearity, writing the expected number of days as a sum of T_j's with $T_j - 1$ a Geometric), the expected value is

$$n\left(\frac{1}{n} + \frac{1}{n-1} + \cdots + \frac{1}{2} + 1\right).$$

(c) Approximate the probability that after 50 days of the lottery, every number from 1 to 35 has been picked at least once (don't simplify, but your answer shouldn't involve a sum).

Let A_j be the event that j doesn't get picked, so

$$P(A_j) = (30/35)^{50} = (6/7)^{50}.$$

Let X be the number of A_j that occur. A Poisson approximation for X is reasonable since these events are rare and weakly dependent. This gives

$$P(X = 0) \approx e^{-35 \cdot (6/7)^{50}}.$$

3. Let $U \sim \text{Unif}(0,1)$, and $X = \ln\left(\frac{U}{1-U}\right)$.
(a) Write down (but do not compute) an integral giving $E(X^2)$.

By LOTUS,

$$E(X^2) = \int_0^1 \left(\ln\left(\frac{u}{1-u}\right)\right)^2 du.$$

(b) Find the CDF of X (simplify).

This can be done directly or by Universality of the Uniform. For the latter, solve $x = \ln(\frac{u}{1-u})$ for u, to get $u = \frac{e^x}{1+e^x}$. So $X = F^{-1}(U)$ where

$$F(x) = \frac{e^x}{1+e^x}.$$

This F is a CDF (by the properties of a CDF, as discussed in class). So by Universality of the Uniform, $X \sim F$.

(c) Find $E(X)$ *without using calculus* (simplify).
Hint: $1-U$ has the same distribution as U.

By symmetry, $1-U$ has the same distribution as U, so by linearity,

$$E(X) = E(\ln U - \ln(1-U)) = E(\ln U) - E(\ln(1-U)) = 0.$$

4. Let $X_1, X_2, X_3, \ldots, X_{10}$ be the total number of inches of rain in Boston in October of 2011, 2012, 2013, ..., 2020, with these r.v.s independent $\mathcal{N}(\mu, \sigma^2)$. (Of course, rainfall can't be negative, but μ and σ are such that it is extremely likely that all the X_j's are positive.) We say that a *record value* is set in a certain year if the rainfall is greater than all the previous years (going back to 2011; so by definition, a record is always set in the first year, 2011).

(a) On average, how many of these 10 years will set record values? (Your answer can be a sum but the terms should be simple.)

Let I_j be the indicator r.v. of the jth year setting a record. Then $P(I_j = 1)$ is $1/j$, since by symmetry all orderings of $X_1, \ldots, X_j$ are equally likely (so the largest of these values is equally likely to be anywhere among them). By linearity, the expected number of record values is

$$1 + \frac{1}{2} + \frac{1}{3} + \cdots + \frac{1}{10}.$$

(b) Is the indicator of whether the year 2013 sets a record independent of the indicator of whether the year 2014 sets a record? (Justify briefly.)

Yes, they are independent (somewhat surprisingly). Determining whether there is a record in 2014 is not related to the "internal squabble" of which of X_1, X_2, X_3 is the biggest. Let J be the index of whichever of X_1, X_2, X_3 is largest (so J takes values 1,2,3). By symmetry, the probability that X_4 is larger than all of X_1, X_2, X_3 is not affected by conditioning on J; note though that saying X_3 is a record is the same as saying that $J = 3$.

(c) Later in the course, we will show that if $Y_1 \sim \mathcal{N}(\mu_1, \sigma_1^2)$ and $Y_2 \sim \mathcal{N}(\mu_2, \sigma_2^2)$ are independent, then $Y_1 - Y_2 \sim \mathcal{N}(\mu_1 - \mu_2, \sigma_1^2 + \sigma_2^2)$. Using this fact, find the probability that the October 2014 rainfall will be more than double the October 2013 rainfall in Boston, in terms of Φ.

We have

$$P(X_4 > 2X_3) = P(2X_3 - X_4 < 0) = P(Y < 0) = P\left(\frac{Y - \mu}{\sqrt{5}\sigma} < \frac{-\mu}{\sqrt{5}\sigma}\right),$$

for $Y \sim \mathcal{N}(\mu, 5\sigma^2)$. This is $\Phi\left(\frac{-\mu}{\sqrt{5}\sigma}\right)$, which can also be written as $1 - \Phi\left(\frac{\mu}{\sqrt{5}\sigma}\right)$.