1 Expectations and variances

1. Expectation is $\frac{1}{3}(0+1+2)=1$. Variance is $(0-1)^2/3+(1-1)^2/3+(2-1)^2/3=2/3$.

2. mean is $\frac{1}{4}(366+3*365)=365.25$. variance is $\frac{1}{4}(\frac{3}{16}+\frac{9}{16})=\frac{3}{16}$.

3. (a) $\frac{1}{6}(1+\cdots+6)=3.5$.

   (b) $4*3.5=14$.

4. If one dice, then no strategy possible, and expected payoff is 3.5. Thus we should only consider strategies which result in expected payoff larger than 3.5.

   The payoff from the second dice is also 3.5, so we should roll again if our first roll is less than 3.5. The payoff of this strategy is $\frac{1}{2}(5)+\frac{1}{2}(3.5)=\frac{17}{4}=4.25$.

   The payoff from having two dice to roll is 4.25, so we should reroll if our first roll is less than 4.25. The payoff of this strategy is $\frac{1}{3}(5.5)+\frac{2}{3}(4.25)=\frac{14}{3}$.

5. The mean is $\frac{1}{n}(1+\cdots+n)=\frac{n+1}{2}$. The variance is $\sum_{i=1}^{n}\frac{i^2}{n}-(\sum_{i=1}^{n}\frac{i}{n})^2=\frac{(n+1)(n-1)}{12}$.

6. The number of losses before four wins (or wins before four losses) is negative binomial, so we would want to compute something like the conditional expectation of a negative binomial random variable given that it is less than 4. However, since the sample space is small, it seems easier to enumerate directly:

   There are $2\binom{3+0}{0}=2$ ways for the match to end in four games, with a probability of 1/8. There are $2\binom{3+1}{1}=8$ ways for the match to end in five games, with a probability of 1/4. There are $2\binom{3+2}{2}=20$ ways for the match to end in six games, with a probability of 5/16. There are $2\binom{3+3}{3}=40$ ways for the match to end in seven games, with a probability of 5/16.

   Thus the mean of the number of games played is $8/16+20/16+30/16+35/16=93/16=5.8125$ and the variance of the number of games played is $(4-5.8125)^2/8+(5-5.8125)^2/4+(6-5.8125)^2*5/16+(7-5.8125)^2*5/16=1.02734375$.

7. (a) To get birth rank 1, you can sample any family, and then the birth rank 1 child. Thus the probability is $3/10+1/2*1/2+1/5*1/3=3/10+1/4+1/15=1/4+11/30=37/60$. To get birth rank 2, you can sample families with two or more children and then the birth rank 2 child. Thus the probability is $1/2*1/2+1/5*1/3=1/4+1/15=19/60$. To get birth rank 3, you can sample families with 3 more more children and then the birth rank 3 child. Thus the probability is $1/5*1/3=4/60$.

   The mean is $1*37/60+2*19/60+3*4/60=87/60$.

   The variance is $(-27/60)^2*37/60+(33/60)^2*19/60+(93/60)^2*4/60\approx 0.3808$

   (b) There are 190 children in the town, 100 of birth rank 1, 70 of birth rank 2, and 20 of birth rank 3. The probabilities are $10/19, 7/19, 2/19$ respectively. The mean is $1*10/19+2*7/19+3*2/19=30/19$. The variance is $(-11/19)^2*10/19+(8/19)^2*7/19+(27/19)^2*2/19\approx .4543$

8. (a) There are 20 million people in the country and 10 cities, so the average population per city is 2 million.

   (b) Keeping everything else the same, we can change the variance by redistributing the population of the cities in the south, so we cannot compute variance without knowing a city by city breakdown of population.

   (c) The expected population sizes of cities in the north, east, south, west respectively are 3/4 million, 4/3 million, 5/2 million, 8 million. This means the expected size of a city of a randomly chosen city (in this way) is 3/16 million + 1/3 million + 5/8 million + 2 million = $3\frac{7}{48}$ million.

   (d) In (c), the city with population 8 million in the west gets more weight in the corresponding sampling than any other city, since it has a 1/4 chance of being sampled, whereas e.g. a city from the north only has a 1/16 chance of being sampled.

9. (a) has expected value 16k. (b) has expected value 1/2(1k) + 1/2(1/4*64k+3/4*32k)=.5k + 8k + 12k = 20.5k. (c) has expected value 3/4(1k) + 1/4(1/2*64k + 1/2*32k) = .75k + 8k + 4k = 12.75k.

   Option (b) has the highest expected value. Option (a) has the lowest (zero) variance.

10. The expected payout is $\sum_{n\geq 1} \frac{n}{2^n} = 2$. If the pyoff is n^2, then the expected payout is 6.

11. Martin wins one dollar unless he sees five tails in a row, since he cannot continue to bet a sixth time, in which case he loses all 31 dollars. Thus Martin's average payoff is $1 * 31/32 - 31 * 1/32 = 0$.

12. The expected value is zero, since $E(X) = \sum_{i=-n}^{n} iP(X = i) = \sum_{i=-n}^{-1} iP(X = -i) + \sum_{i=1}^{n} iP(X = i) = 0$.

13. Yes. Let X be $2^{2^{2^{\cdots 2}}}$ with probability at most .1 and 0 with probability at least .99. Let Y be constant.

14. $E(X) = \sum_{k\geq 1} cp^k = \frac{cp}{1-p}$. $E(X^2) = \sum_{k\geq 1} kcp^k = cp\sum_{k\geq 1} kp^{k-1} = \frac{cp}{(1-p)^2}$. Thus $Var(X) = cp(1-cp)/(1-p)^2$

15. (a) Let q_j denote the probability with which player B chooses j and X the payoff. Then $E(X) = \sum_i iP(A = B = i) = \sum_i iP(A = i)P(B = i) = \sum_i ip_iq_i$. Player B's expected payoff is a weighted average of the numbers ip_i, so B can maximize the expected payoff by choosing to put all their weight into the i such that ip_i is maximized.

    (b) Suppose that both players set $p_j = q_j = c/j$. Then each $jp_j = c$, so B's expected payoff is a weighted sum of 100 copies of c, so B's expected payoff must be c.

    If A or B changes their strategy unilaterally, the payouts do not decrease or increase by the argument above, so there is no incentive to change. In fact, if A (or B) changes their strategy, then the other player can change their strategy to take advantage of that.

(c) B's expected payoff under the situation in (b) is $c = \sum_{i=1}^{100} \frac{1}{i} \approx 5.187$. This does not depend on A's strategy since the payout is a weighted sum of the numbers $iq_i = c$.

16. (a) There are 18 courses with 360 students, so the average number of students per class is 20. On the other hand, 200 students think their course size is 100 while 160 students think their course size is 10, so the average number of students per class is $(100 * 200 + 10 * 160)/360 = 540/9 = 60$.

    This discrepancy results from the fact that there are more students in large classes, so more students experience larger classes than would be represented by taking the average per class.

    (b) Let the class sizes be $a_1, \ldots, a_n$. The dean's perspective is that the average number of students per class is $\frac{\sum_i a_i}{n}$. The student's perspective is that the average number of students per class is $\frac{\sum_i a_i^2}{\sum_i a_i}$. Then since variance are nonnegative,

$$\sum_i a_i^2/n - (\sum_i a_i/n)^2 \geq 0 \implies \frac{\frac{\sum_i a_i^2/n}{\sum_i a_i/n}}{\sum_i a_i/n} \geq 1$$

    with equality if and only if the variance is equal to 0. The variance is zero if and only if all classes have exactly the same size.

17. (a) $E(C) = m_1$.

    (b) $E(C) = m_2/m_1$.

    (c) (b) is larger by nonnegativity of variances (same as the previous problem) unless all families have the same size.

2 Named distributions

18. We have the first trial, and then we are waiting for the opposite result (+the opposite result itself), so the number of tosses is a $1 + FS(1/2)$ random variable. The expected value is 3.

19. $P(X \leq k) = \frac{1}{2} + \cdots + \frac{1}{2^k} = \frac{1}{2}(1 + \cdots + \frac{1}{2^{k-1}}) = \frac{1}{2}\frac{1-\frac{1}{2^k}}{1-\frac{1}{2}} = \frac{2^k-1}{2^k}$. This graph jumps at every integer, and increase more and more slowly (exponentially slowly) as k gets large.

20. (a) Not possible. Since Y can be larger than 100, then $X - Y$ can be negative with positive probability.

    (b) Let X represent the chance that a person is randomly selected to enter a coin flip tournament where they must flip a coin that has probability p of heads to pass each round and Y represent the chance that a person passes the first round of the tournament. The coin must have probability $5/9$.

(c) Not possible. If $P(X \leq Y) = 1$, then $E(Y - X) \geq 0$, but $E(Y - X) = E(Y) - E(X) = 50 - 90 = -40$.

21. (a) Since $\max(X, Y) + \min(X, Y) = X + Y$, $E(V) + E(W) = E(V + W) = E(X + Y) = E(X) + E(Y) = \frac{2n+1}{2}$.

    (b) $|X - Y| = \max(X, Y) - \min(X, Y) = W - V$ so applying expectation to everything gives the desired result.

    (c) $Var(n - X) = Var(-X) = Var(X) = \frac{n}{4}$. Could also compute it by knowing that $n - X \sim X \sim Bin(n, 1/2)$.

22. The most reasonable distribution that we know seems to be Poisson. The parameter λ would be 100t drops per five square inches per t minutes, so the probability of no rain drops in 3 seconds (=.05 minutes) is e^{-5}.

23. (a) Let I_j be the indicator of whether person j is friends with both Alice and Bob. $P(I_j = 1) = 1/400$ so $E(X) = E(\sum_j I_j) = \sum_j E(I_j) = \frac{1000}{400} = 2.5$.

    (b) X is hypergeometric with parameters $50, 950, 50$.

    (c) See previous part.

24. Consider an infinite sequence of Bernoulli trials where each trial results in a success with probability p. In the set of infinite bitstrings, let A denote the event that the first n trials contain fewer than r successes. Then by definition, $A = \{X < r\}$ in the sense that $P(A) = P(X < r)$ (since A depends only on the first n trials, so we can restrict to the first n and track the outcomes there).

    We now show that $A = \{Y > n - r\}$ in the sense that $P(A) = P(Y > n - r)$ as well. Suppose $a_1 a_2 \cdots \in A$. Then by definition $a_1 \ldots a_n$ contains fewer than r successes, so this is one outcome where $Y > n - r$. Conversely, if $Y > n - r$ for an outcome, then the outcome must contain strictly more than $n - r$ failures before the first r successes, so in particular the first n trials must contain strictly fewer than r successes, so the outcome is in A.

25. Let E denote the expected number of games played. Then $E = 2 + E(2pq) \implies E = \frac{2}{1-2pq}$.

26. (a) Let $Z_i = X_i$ and Y_i (in the sense that $Z_i = 1$ if $X_i = Y_i = 1$ and $Z_i = 0$ otherwise). Then the first time at which they are simultaneously successful is the first success of the Z_i's, i.e. a first success distribution with parameter $p_1 p_2$. The expected value is $\frac{1}{p_1 p_2}$.

    (b) Let $Z_i = X_i$ or Y_i (in the sense that $Z_i = 1$ if $X_i = 1$ or $Y_i = 1$ and $Z_i = 0$ otherwise). Then the first time when one of them has a success is $FS(p_1 + p_2 - 2p_1 p_2)$.

    (c) The probability of a simultaneous first success is $\sum_{i \geq 1} (1 - p)^{2i-2} p^2 = \frac{p}{2-p}$. Otherwise, the probability that Nick's first success precedes Penny's is equal to the probability that Penny's first success precedes Nick's (by symmetry) so it is equal to $\frac{1}{2}(1 - \frac{p}{2-p}) = \frac{1-p}{2-p}$.

27. T is not Poisson since it cannot achieve odd values.

28. X is t plus negative hypergeometric with parameters $t, n-t, t$, so $E(X) = t(n-t)/(t+1) + t = t(n+1)/(t+1)$.

29. $f(k) = (1-p)^k p$, so $E(f(X)) = \sum_{k\geq 0}(1-p)^{2k}p^2 = p^2/(1-(1-p)^2) = p^2/(2p-p^2) = p/2-p$, i.e. the same as 26(a) since this is the probability that two independent geometrics are equal to each other.

30. (a) $E(Xg(X)) = \sum_k kg(k)P(X=k) = \sum_k kg(k)e^{-\lambda}\frac{\lambda^k}{k!} = \lambda\sum_k g(k)e^{-\lambda}\frac{\lambda^{k-1}}{(k-1)!} = \lambda\sum_k g(k+1)e^{-\lambda}\frac{\lambda^k}{k!} = \lambda E(g(X+1))$.

    (b) $E(X^3) = E(XX^2) = \lambda E((X+1)^2) = \lambda(E(X^2)+2\lambda+1) = \lambda((\lambda+\lambda^2)+2\lambda+1) = \lambda(\lambda^2+3\lambda+1)$.

31. (a) $P(X=0) = p+(1-p)e^{-\lambda}$ and $P(X=k>0) = (1-p)e^{-\lambda}\frac{\lambda^k}{k!}$.

    (b) $P((1-I)Y=k) = P(1-I=1)P(Y=k)$ if $k>0$ and $P((1-I)Y=0) = P(1-I=0)+P(1-I=1)P(Y=0)$ so X and $(1-I)Y$ are identically distributed.

    (c) $E(X) = \sum_k x(1-p)e^{-\lambda}\frac{\lambda^k}{k!} = (1-p)\lambda$. Alternatively, $E(X) = E(1-I)E(Y) = (1-p)\lambda$.

    (d) $Var(X) = E(X^2)-E(X)^2 = (1-p)E(Y^2)-(1-p)^2E(Y)^2 = (1-p)(Var(Y)+E(Y)^2-(1-p)E(Y)^2) = (1-p)(\lambda+p\lambda^2)$.

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

    (b) The only discrete distributions we have which can take on arbitrary integer values are geometric, first success, negative binomial, and Poisson. Among these, the geometric and first success seem to have the memoryless property.

    For geometric/first success, it is clear: if we have not yet had a success, knowing that we have already experienced n trials does not matter since we are no "closer" to a success.

33. $w=1, b=n, r=1$. Since all sequences are equally likely, we are equally likely to draw any number of black balls before the one white ball. $P(X=k) = \frac{1}{n+1}$, since the white ball is equally likely to be in between any two black balls (or on the far left or right).

3 Indicator r.v.s

34. Let I_j be the indicator of the j^{th} box being empty. Then $P(I_j=1) = (\frac{n-1}{n})^k$, so the expected number of empty boxes is $n(\frac{n-1}{n})^k$.

35. Let $I_{j,k}$ be the indicator of whether people j and k share a birthday. Then $P(I_{j,k} = 1) = \frac{1}{365}$. Let I_l be the indicator of whether day l has at least two people born on that day. Then $P(I_l=0)$ is the probability that at most one person was born on

that day, which is $(364/365)^{50} + 50(1/365)(364/365)^{49}$ and $P(I_l = 1) = 1 - P(I_l = 0)$, so the expected number of days on which at least two people were born is $365(1 - (364/365)^{50} - (1/365)(364/365)^{49})$.

36. I_{12} is independent of I_{34} (e.g. the first two and last two coin tosses in four independent coin tosses are independent). I_{12} is independent of I_{13}. They are not independent. Knowing 12 and 23 have the same birthday implies 13 have the same birthday as well.

37. Let I_j be the indicator of whether the j^{th} bag goes to one of the first three students. Then $P(I_j = 1) = \frac{3}{20}$ so the expected total number of bags going to the first three students is 3.

    Let I_j be the indicator of whether student j receives at least one bag. Then $P(I_j = 0)$ is the probability that student j receives 0 bags, which has probability $(19/20)^{20}$. The expected total number of students who get at least one bag is $n(1 - (19/20)^{20})$.

38. Let I_j be the indicator variable of whether person j receives their own name. $P(I_J = 1) = \frac{1}{n}$ since each person (a priori) is equally likely to receive any one of the n slips of paper. Thus the average number of people who draw their own names is 1.

39. Let I_j be the indicator of whether person j is in both samples. Then $P(I_j = 1) = \frac{m}{N}\frac{n}{N}$, so the expected size of overlap is $\frac{mn}{N}$.

40. Let I_j be the indicator of whether the indices $j, j+1, j+2$ are HTH. Then $P(I_j = 1) = \frac{1}{8}$ and there are $n-2$ possible indices j, so the expected number of occurrences is $(n-2)/8$.

41. Let I_j be the indicator of whether cards j and $j+1$ are both red. Then $P(I_j = 1) = \frac{26}{52}\frac{25}{51}$. There are 51 such pairs, so the expected number of occurrences is $\frac{25}{2}$.

42. Let I_j be the indicator of whether you have collected toy type j by t toys. Then $P(I_j = 0) = ((n-1)/n)^t$ so $E(I_j) = 1 - ((n-1)/n)^t$ and thus the expected number of toys is $n(1 - ((n-1)/n)^t)$.

43. (a) Let I_j be the indicator of whether someone presses the floor j button. Then $P(I_j = 0) = ((n-2)/(n-1))^k$, so $E(I_j) = 1 - ((n-2)/(n-1))^k$ and thus the expected number of stops is $(n-1)(1 - ((n-2)/(n-1))^k)$.

    (b) In this case, $P(I_j = 0) = (p_2 + \cdots + \widehat{p_j} + \cdots + p_n)/(p_2 + \cdots + p_n)$, but the rest is the same as before.

44. At the beginning, there are 200 ends and each step reduces the number of ends by 2, so there are 100 steps always.

    Let I_j be the indicator variable of whether a loop was created at the j^{th} stage. We can consider each stage as removing a shoelace (whether two fuse to become a longer one or one gets tied into a loop) so at each stage there are $100 - (j-1)$ shoelaces and we create a loop if we connect two ends of the same shoelace. Since we choose the ends randomly, then the probability of choosing the ends of one shoelace is $\frac{1}{200-2(j-1)-1}$. Thus the expected number of loops is $1/199 + 1/197 + \cdots + 1/1$.

45. $P(A_1 \cap \cdots \cap A_n) = 1 - P(A_1^c \cup \cdots \cup A_n^c) \geq 1 - \sum_i P(A_i^c) = 1 - \sum_i 1 - P(A_i) = 1 - n + \sum_i P(A_i)$.

46. This is negative hypergeometric $w = 4, b = 48, r = 1$. The expected number before the first ace is the same as the expected numbed between the first and second ace. The expected number before the first ace is $\frac{48}{5}$.

47. (a) Let I_j be the indicator of getting the j^{th} card correct. No matter your guess, you have a $\frac{1}{52}$ chance of guessing correctly, so $P(I_j = 1) = 1/52$ so the expected number of cards correct is 1. (If you guess a permutation, this is equivalent to the fact that an average permutation has one fixed point.)

    (b) Let the cards you would guess in order be $c_1, \ldots, c_{52}$ and let I_j be the indicator of that you correctly guess at least j cards in your order. Then $P(I_j = 1)$ is the probability that the cards $c_1, \ldots, c_j$ appear in the deck in that order, which is $\frac{\binom{52}{j}(52-j)!}{52!} = \frac{1}{j!}$. (In other words, we only have to focus on the cards c_1 through c_j.)

    The total number of cards you guess correctly is $\sum_j I_j$, so the expected number of cards guessed correctly is $\sum_{j=1}^{52} \frac{1}{j!} \approx e - 1$ since this is the (truncation of the) Taylor expansion of e^x evaluated at $x = 1$.

    (c) Let I_j be the indicator that you guess card j correctly. At card j, there are $j - 1$ cards revealed, so there is a $\frac{1}{52-(j-1)}$ chance that this strategy guesses card j correctly. Thus the expected number of cards guessed correctly is $\sum_j \frac{1}{52-j} = 1 + \frac{1}{2} + \cdots + \frac{1}{52}$.

48. (a) $\binom{X}{2}$ is the number of pairs of white balls in the sampled n balls. Let $I_{j,k}$ by the indicator of whether the pair j, k of white balls is chosen in the sample. There are $\binom{w+b-2}{n-2}$ samples which contain this pair of white balls and $\binom{w+b}{n}$ samples total, so the probability that a given sample includes the pair j, k of white balls is $\binom{w+b-2}{n-2}/\binom{w+b}{n} = \binom{n}{2}/\binom{w+b}{2}$. Then the expected number of pairs of white balls is $\binom{w}{2}\binom{n}{2}/\binom{w+b}{2}$.

    (b) $2E\binom{X}{2} = E(X^2 - X) = Var(X) + E(X)^2 - E(X)$. The mean of a hypergeometric distribution is $\frac{wn}{w+b}$, so we find

$$Var(X) = 2\binom{w}{2}\binom{n}{2}/\binom{w+b}{2} - (wn/(w+b))^2 + wn/(w+b)$$

$$= np\left((w-1)(n-1)/(N-1) - pn + 1\right) = np\left((nw - n - w + N)/(N-1) - pn\right)$$

$$np\frac{(nw - n - w + N)N - nw(N-1))}{N(N-1)} = np\frac{N^2 - Nw - Nn + nw}{N(N-1)}$$

$$= np\frac{(N-w)(N-n)}{N(N-1)} = npq\frac{N-n}{N-1}$$

49. Let I_j be the indicator of whether j dollar prize is chosen. Then total value is $X = \sum_j I_j$, so the expected total value is $\frac{k}{n}(1 + \cdots + n) = k\frac{n-1}{2}$.

50. The dumb way to find the probability of two chords intersecting is to integrate $2x(1-x)$ from $x = 0$ to $x = 1$ since we can view a pair of chords as one chord with a fixed endpoint and a randomly chosen point x along the perimeter, then the probability of an intersection is $2x(1-x)$ (i.e. just saying that the endpoints of the second chord need to be on opposite sides of the circle). The integral is $2(x^2/2 - x^3/3)|_0^1 = 1/3$.

    The hint gives a cleverer way to do this: two chords correspond to four randomly chosen points, and among those four points there are three ways to connect them with chords and one of the three has an intersection.

    The expected number of intersections is then $15 = \binom{10}{2} * 1/3$.

51. Let E_j be the indicator of a location being empty. $P(E_j = 1) = (\frac{n-1}{n})^k$, so the expected number of empty locations is $n(\frac{n-1}{n})^k$.

    Let O_j be the indicator of a location having exactly one phone number. $P(O_j = 1) = \frac{k}{n}(\frac{n-1}{n})^{k-1}$, so the expected number of locations with exactly one phone number is $k(\frac{n-1}{n})^{k-1}$.

    Let M_j be the indicator of more than one phone number. $P(M_j = 0) = 1 - \frac{1}{n}(\frac{n-1}{n})^{k-1} - k(\frac{n-1}{n})^k$, so the expected number of locations with more than one phone number is n times this probability.

    These quantities should add up to n by linearity of expectation, since these cases are exhaustive.

52. Let I_j be the indicator variable of whether flip j begins a new run (i.e. is different from flip $j-1$). Then $P(I_j = 1) = p(1-p) + (1-p)p = 2p(1-p)$. $I_1 = 1$ always, so the expected number of runs is $1 + (n-1)2p(1-p)$.

53. Let I_j be the indicator of whether the sequence starting at position j is HH. Then $P(I_j = 1) = p^2$ and $1 \leq j \leq 3$. Then $E(X) = 3p^2$.

    $Var(X) = E(X^2) - E(X)^2 = 3p^2 + 2(p^3 + p^4 + p^3) - 9p^4 = 3p^2 + 4p^3 - 7p^4$.

54. (a) The j^{th} person in the sample is equally likely to be any member of the population, so $E(W_j) = \sum_{k=1}^{N} y_k \frac{1}{N}$. Then $E(\overline{W}) = \frac{1}{n}\sum_{j=1}^{n} E(W_j) = \frac{1}{n}\sum_{j=1}^{n} \overline{y} = \overline{y}$.

    (b) Let Y_j be the random variable whose value is y_j if the j^{th} person is included in the sample and 0 otherwise. Then $\overline{W} = \frac{1}{n}\sum_{i=1}^{n} W_j = \frac{1}{n}\sum_{i=1}^{N} Y_j$. Then using properties of expectation, $E(\overline{W}) = \frac{1}{n}\sum_{i=1}^{N} E(Y_j) = \frac{1}{n}\sum_{i=1}^{N} \frac{n}{N} y_j = \overline{y}$.

55. (a) Let $I_{j,k}$ be the indicator for whether numbers j and k are out of order. Then $P(I_{j,k} = 1) = \frac{1}{2}$, so the expected number of inversions is $\frac{\binom{n}{2}}{2}$.

    (b) There are at least as many comparisons as swaps, so that means the expected number of comparisons is at least $\binom{n}{2}/2$. There are maximum $n-1$ rounds and round j involves $n-j$ comparisons, so there are at most $\binom{n}{2}$ comparisons if every sweep round occurs.

56. (a) Let I_j be the indicator of whether the 7 shots starting from shot j are successful. Then $P(I_j = 1) = p^7$. There are $n - 7 + 1$ such indices, so the expected number of streks of 7 shots is $(n - 6)p^7$.

    (b) Consider each 7 shots (e.g. shots 1-7, 7-13, etc) as a Bernoulli trial with probability of success p^7. Then $X \leq FS(p^7)$ since crudely thinking of shots $7i$ to $7i + 6$ as a Bernoulli trial ignores the possibility that the streak of 7 starts at some index not divisible by 7. However, this is enough to get the bound $E(X) \leq \frac{7}{p^7}$.

57. (a) The probability of obtaining a red ball is r, so the expected number of balls until a red one has distribution $Geom(r)$, so the expected number of balls chosen before the first red ball is $\frac{1-r}{r}$.

    (b) Let I_g and I_b denote the indicators of drawing a red ball after a green or blue ball respectively. Then by previous computations, $P(I_g = 1) = \frac{g}{r+g}$ and $P(I_b = 1) = \frac{b}{r+b}$. $X = I_g + I_b$ is the number of distinct colors before the first red ball, so $E(X) = \frac{g}{r+g} + \frac{b}{r+b}$.

    (c) This is the complement of the event that exactly one ball is red given that at least one ball is red. The probability that exactly one ball is red is $\binom{n}{1}r(g+b)^{n-1}$ while the probability that at least one ball is red is $1 - (g + b)^n$, so we have $nr(g + b)^{n-1}/(1 - (g + b)^n)$.

58. Let I_j be the indicator of whether candidate 1 is the best of the first j candidates. Then $X = \sum_j I_j$ has expected value $\sum_j \frac{1}{j} = \infty$.

59. (a) The event $X \leq j$ is equivalent to the event that among j people, there is a birthday match. (Note that the event $X = j$ is more restrictive: there must be a match among the first j people, but not among the first $j - 1$ people.) The event $X \geq j$ is equivalent to the event that there is no birthday match among $j - 1$ people.

    For any $m < 23$, $P(X \leq m) \leq P(X \leq 22)$ which is less than .5 as computed previously, so m cannot be a median. For any $m > 23$, $P(X \geq m) \geq P(X \geq 24)$ which is also less than .5 as computed previously. Thus there are no possible medians other than 23. Finally, $P(X \geq 23) = 1 - P(X < 23) = 1 - P(X \leq 22) \geq$ .5, so 23 is a median of X.

    (b) These are the tail probabilities, so we can use the same theorem as the previous problem. But explicitly, suppose $X = j$. Then $X \geq k$ for every $k = 1, \ldots, j$, so the indicators $I_1, \ldots, I_j$ are 1 and the remainder are zero, so $\sum_j I_j = j$.

    As in the previous problem, the event $X \geq j$ is equivalent ot the event that there is no birthday match among $j - 1$ people. This event imposes no restrictions on the first person, but from the second person onward, they cannot match birthdays with any of the previous people. Thus $P(I_j = 1) = p_j$, so $E(X) = \sum_{j \geq 1} p_j$

    (c) $E(X) = 24.61659$.

    (d) $Var(X) = 149.6403$ (note: this seems large, though I suppose the sd is the square root which is around 12, which seems okay since I would expect a decent amount of variance)

In terms of the $p_j's$, $Var(X) = E(X^2) - E(X)^2 = E((\sum_j I_j)^2) - (\sum_j p_j)^2$. $E(I_j^2) = E(I_j) = p_j$. $E(I_j I_{j+k}) = E(I_{j+k}) = p_{j+k}$, so $E((\sum_j I_j)^2) = \sum_j p_j + \sum_j (j-1)p_j$.

60. (a) X is the number of negative samples until finding m positive samples, so this is negative hypergeometric with parameters $n, N-n, m$ and $Y = X + m$ since Y just needs to include the successes as well.

    (b) $E(Y) = m + E(X)$. The expectation of the negative hypergeometric X is $\frac{m(N-n)}{n+1}$. We can define an indicator variable I_j of which two tagged elks the j^{th} untagged elk is sampled between.

    (c) The number of tagged elk is now hypergeometric with parameters $n, N-n, E(Y)$, so the expected value is $E(Y) * n/N = m(N+1)/(n+1) * n/N = m(N+1)/N * n/(n+1) < m$. (Since $n < N$, then $n/n+1 > N/N+1$.)

4 LOTUS

61. $E(X!) = \sum_k k! e^{-\lambda} \frac{\lambda^k}{k!} = e^{-\lambda} \sum_k \lambda^k = \frac{1}{e^{\lambda}(1-\lambda)}$ if $\lambda < 1$.

62. $E(2^X) = \sum_k 2^k e^{-\lambda} \frac{\lambda^k}{k!} = e^{-\lambda} \sum_k (2\lambda)^k / k! = e^{\lambda}$

63. $E(2^X) = \sum_k 2^k q^k p = p/(1-2q)$ if $2q < 1$. $E(2^{-X}) = \sum_k 2^{-k} q^k p = p/(1-q/2)$.

64. $E(e^{tx}) = \sum_k e^{tk} q^k p = p/(1-qe^t)$.

65. (a) $E(1 + 2X + X^2) = 1 + 2\lambda + \lambda + \lambda^2 = \lambda^2 + 3\lambda + 1$.

    (b) $E(1/Y) = \sum_k \frac{1}{k+1} e^{-\lambda} \frac{\lambda^k}{k!} = \frac{e^{-\lambda}}{\lambda} \sum_{k \geq 1} \frac{\lambda^k}{k!} = \frac{e^{-\lambda}}{\lambda}(e^{\lambda} - 1)$.

66. (a) $E(e^{-3X}) = \sum_k e^{-3k} e^{-\lambda} \frac{\lambda^k}{k!} = e^{-\lambda} \sum_k \frac{(\lambda e^{-3})^k}{k!} = e^{-\lambda} e^{\lambda e^{-3}}$.

    (b) $E((-2)^X) = \sum_k (-2)^k e^{-\lambda} \frac{\lambda^k}{k!} = e^{-\lambda} e^{-2\lambda} = e^{-3\lambda}$.

    (c) $g(X)$ is a reasonable estimate if $\lambda = 0$. Otherwise, the absolute value of $(-2)^X$ is larger than 2 and $(-2)^X$ can be negative, so even though it is unbiased, no particular value of $(-2)^X$ gives a good approximation since $0 < e^{-3\lambda} < 1$ and we'd only make good use of the unbiased estimator if we repeated a lot of trials and averaged them.

    As a trivial improvement, we could take h to be the constant estimator 1.

5 Poisson approximation

67. (a) Let A_j be the event that student j has the same seat in both courses. Then by inclusion exclusion, the probability that no one has the same seat is $\sum_{k=0}^{100} (-1)^k \binom{100}{k} \frac{(100-k)!}{100!} = \sum_{k=0}^{100} (-1)^k \frac{1}{k!}$.

(b) $\sum_{k=0}^{100}(-1)^k\frac{1}{k!} \approx e^{-1}$. Alternatively, we could also realize this by finding that the expected number of people with the same seat is 1, so this number is distributed approximately as $Pois(1)$.

(c) Using the $Pois(1)$ approximation, we find $1-2e^{-1}$.

68. (a) $E(X) = 1$ by letting I_j be the indicator of person j receiving their own name. $P(I_j = 1) = \frac{1}{n}$ so $X = \sum_j I_j \implies E(X) = 1$.

(b) Let $I_{j,k}$ be the indicator of people j,k swapping names. Then $P(I_j) = (n-2)!/n! = \frac{1}{n(n-1)} = \frac{1}{2\binom{n}{2}}$.

(c) If n is large, then X is approximately $Pois(1)$. $P(X=0) \xrightarrow{n\to\infty} e^{-1}$.

69. The probability that a pair of people of the 1000 coincide is $1/10^6$. There are $\binom{1000}{2}$ pairs, so the expected number of coincidences is $\binom{1000}{2}/10^6 \approx \frac{1}{2}$, so we may approximate this number as $Pois(1/2)$. Thus, our approximation is $1-e^{-.5}$.

70. (a) The number of winners is distributed as $Bin(10^7, 10^{-7}) \approx Pois(1)$.

(b) $P(M) = \sum_{k\geq 0} P(M|W=k)P(W=k) = \sum_k \frac{1}{k+1}e^{-1}/k! = \sum_k e^{-1}/((k+1)!) = e^{-1}(e-1) = 1-e^{-1}$.

71. The probability of two people having the same birthday and mother's birthdays is $\approx \frac{1}{360^2}$. There are $\binom{90}{2} \approx \frac{90^2}{2}$ pairs of people, so the expected number of matches is $\approx \frac{1}{32}$, meaning that we can approximate this event using a $Pois(\frac{1}{32})$ distribution. Under this approximation, the probability that at least one pair shares a birthday and mother's birthday is $1-e^{-\frac{1}{32}} \approx 1-(1-\frac{1}{32}) \approx \frac{1}{32}$.

72. (a) If there are n people, the probability that nobody shares a birthday with me is $(364/365)^n$, so the n for which the probability is at least 50% is 253. As a Poisson approximation, the probability that someone shares a birthday with me is $\frac{1}{365}$, so the expected number of people sharing a birthday with me is $\frac{n}{365}$, and the Poisson approximation is $Pois(n/365)$. We want to find the n for which $P(X=0) < .5$, so $e^{-n/365} = .5 \implies n = \ln(2)*365 \approx 252.9987$ so this is a very good approximation.

(b) There are $N = 365*24$ day-hours. The probability that nobody in n people were born on the same day-hour is $(\frac{N}{N})(\frac{N-1}{N})\cdots(\frac{N-n+1}{N})$, so we want to find when this probability is first smaller than .5, which seems to happen around $n = 110$.

As a Poisson approximation, the probability that two people share the same birthday-hour is $1/N$, so the expected number of birthday-hour matches is $\binom{n}{2}/N$, and the Poisson approximation is $Pois(\binom{n}{2}/N)$. We want to find when $P(X = 0) < .5$, so $e^{-\binom{n}{2}/N} = .5 \implies \binom{n}{2} = ln(2)*N \implies n \approx 110.700$.

(c) This is the same reason that the original birthday problem isn't 365. This answer assumes that the number of birthday matches grows linearly with n since it says that after 24 birthday matches, we can expect to see one pair where the hours also match, but since the number of pairs grows quadratically, then we should not expected to need all 24*23 people.

(d) The indicator for the days of the year are probably more accurate since they are less dependent. Let $I_{j,k,l}$ be the indicator for people j,k,l sharing the same birthday. This event has probability $\frac{1}{365^2}$. There are $\binom{100}{3}$ triplets, so the expected number of birthday triples is $\binom{100}{3}/365^2$. Thus we have a $Pois(\binom{100}{3}/365^2)$ approximation, yielding $1 - P(X=0) \approx .7029$.

If we use days of the year, then the probability of a triple birthday match on day j is $X = 1-(364/365)^{100} - 100(1/365)*(364/365)^{99} - \binom{100}{2}(1/365^2)(\binom{364}{365})^{98}$. There are 365 days, so the expected number of triple birthday matches is $365X$. Thus we have a $Pois(365X)$ approximation, yielding $1 - P(X=0) \approx .6305$.

73. (a) Let $I_{j,k}$ be the indicator of whether players j,k play each other twice. The opponent of j is uniformly one of the other 99 players, so the probability is $1/99^2$. Then the expected value of X is $2\binom{100}{2}\frac{1}{99^2} = \frac{100}{99}$.

(b) X is not approximately Poisson because whether one pair played each other twice rules out other people playing with anyone in this pair twice.

(c) $X = 0$ occurs if each pair did not play each other in the previous round. This is 1 minus the probability that at least one pair played each other last round. This is $1 - \sum_{i=1}^{50}(-1)^{i+1}\binom{50}{i}\frac{1}{99*\cdots*(2*(50-i)+1)}$. As a crude approximation, we can use $Pois(100/99)$ giving $\approx .3642$. However, $1 - e^{-.5}$ seems to be a much better approximation...

$X = 2$ if exactly one pair played each other in the previous round. The probability of this is approximately $1-\sum_{i=1}^{50}(-1)^{i+1}\binom{50}{i}\frac{1}{99*\cdots*(2*(50-i)+1)}+\sum_{i=2}^{50}(-1)^{i}\binom{50}{i}\frac{1}{99*\cdots*(2*(50-i)+1)}$. Using the same crude approximation as before, ≈ 0.3679. However, $1-ln(2)$ seems to be a much better approximation...

6 Existence

74. (a) This is hypergeometric with parameters 5,6,5. However, using indicator variables, we find $X = \sum_j I_j$ where I_j is the indicator that both Alice and Bob select movie j. $P(I_j = 1) = \frac{5}{11}^2$. Since there are 11 movies, the expected number is $\frac{25}{11}$.

(b) There are $\binom{11}{2} = 55$ pairs of movies. Each of 111 people selects 5 movies out of 11 for a total of 10 pairs. Thus, 1110 pairs of movies are named, so by pigeonhole there must be at least one pair selected by at least $\lceil 1110/55 \rceil = 21$ people. (For an expected value phrasing, let X be the random variable on the sample space of pairs of movies which says how many people have that pair in their favorites.)

75. Let X be the number of red vertices and let I_j be the indicator of whether vertex j is red or blue. Then $P(I_j) = 1$ is $\frac{2}{3}$ (note that the I_j are very non-independent), so $E(X) = \frac{8}{3}$. Since X can only take integer values and $E(X) > 2$, then there must be some configuration where at least three vertices are red.

76. There were at least $8*65 = 520$ questions answered correctly, so at least one student got at least six questions right. Now for every pair of questions, there must be at least

30 students who got both correct since for each question at least 65 students answered it correctly, so the collective of the student who got at least six questions and any other student who got the remaining questions correct will suffice.

77. For any tiling of the plane, each point will be inside exactly one cell. Thus in particular for any random hexagonal tiling where the side lengths are chosen in order to perfectly inscribe the circular coins, there are at most 10 cells containing points, and we can use the 10 coins to cover these (at most) 10 cells. Let X be the number of covered points in a random hexagonal tiling, and let I_j be the indicator of whether the j^{th} point is covered by a coin.

    Since the ratio of the area of the circle to the area of its circumscribed hexagon is larger than .9, then $P(I_j = 1) > .9$, so $E(X) > .9 * 10 = 9$. Thus, there must be at least one tiling which we can use to cover all 10 points by placing coins as the inscribed coins of hexagons.

78. For a particular binary string b and sequence of indices $\mathbf{i} = (i_1, \ldots, i_k)$, the probability that a randomly chosen set of size m is not b-complete in indices $\mathbf{i}$ is $(1-2^{-k})^m$. Let X be the number of pairs $(b, \mathbf{i})$ for which a randomly chosen set of size m is not b-complete in indices $\mathbf{i}$ and let $I_{b,\mathbf{i}}$ be the indicator of whether S is b-complete in indices $\mathbf{i}$. $E(X) = \sum_{b,\mathbf{i}} E(I_{b,\mathbf{i}}) = \binom{n}{k} 2^k (1-2^{-k})^m < 1$ so there must be some set S where the number of failures to be k-complete is zero.

7 Mixed practice

79. (a) The probability of success is 1/m, so this distribution is first success with parameter 1/m, which have expected value m.

    (b) This is a negative hypergeometric distribution with parameters 1,m-1,1, also known as the discrete uniform distribution on $1, \ldots, m$, which has average $(m+1)/2$.

    (c) $m > (m+1)/2$ if $m > 1$. This makes sense intuitively since sampling with replacement causes the chance of resampling previously sampled passwords, thus dumbly increasing the number of attempts.

    (d) If X is the number of guesses, then we want to compute the PMF of $Y = \min(n, X)$. For $k < n$, the pmf of Y agrees with that of X, and then $P(Y = n) = P(X \geq n)$.

80. (a) The probability of success is $(21-m)/20$ and this is a first success distribution with parameter $(21-m)/20$ of success, so the expected number of rolls is $20/(21-m)$.

    (b) Using LOTUS, this is $\sum_{k \geq 1} \sqrt{k} q^{k-1} p$, where $q = 1-p$ and $p = (21-m)/20$.

81. (a) $N = \frac{360!}{(3!)^{120} 120!}$

(b) Let I_j be the indicator of whether team j contains a married couple. Then $P(I_j = 1) = \frac{180*358}{\binom{360}{3}}$, so the expected number of teams containing married couples is $120 * 180 * 358/\binom{360}{3}$.

82. (a) Let I_j be the indicator of whether the j^{th} roll was all sixes. Then $P(I_j = 1) = 1/6^n$ so $E(\sum_j I_j) = \frac{4}{6}$

(b) The probability of no sixes is $(1 - 1/6^n)^{4*6^{n-1}} = ((1 - (1/6)/6^{n-1})^{6^{n-1}})^4 \approx e^{-4/6}$, so the probability of at least one all sixes is $1 - e^{-4/6}$. (We could also have gotten the result directly from Poisson approximation.)

(c) Let I_j be the same as before. Note that $P(I_1 = 1) = 1/6^n$ has not changed. Then by induction, $P(I_j = 1) = P(I_{j-1} = 1)6/7 + (1/6^n)(1/7) = 1/6^n$ so the expected number of times all sixes is achieved does not change.

83. (a) Each person gets 10, and the remaining 50 can be distributed arbitrarily. There are $\binom{54}{4}$ ways to distribute the remaining.

(b) Let I_j be the indicator of whether dollar $50 + j$ goes to the first person. Then $P(I_j = 1) = 1/5$ so the expected amount of money person 1 receives is $10+50/5 = 20$.

(c) They are not independent since knowing person 2 received more money than person 1 makes it somewhat easier for person 3 to also receive more money than person 1.

84. (a) The number of songs from his favorite album has a binomial distribution with parameters $11, 10/500$.

(b) This is 1 minus the probability that every song is from a different album, which has probability $(50/50)(49/50)\cdots(50 - i + 1)/50$.

(c) Let $I_{j,k}$ be the indicator of sings j, k matching. Then $P(I_{j,k} = 1) = 1/50$, so the expected number of matches is $\binom{11}{2}/50$.

85. (a) The probability of guessing no numbers correctly is $p = \binom{30}{5}/\binom{35}{5} = (30*29*28*27*26)/(35*34*33*32*31)$, so the probability of guessing at least one is $1-p$. The probability of guessing exactly three numbers right is $\binom{5}{3}\binom{30}{2}/\binom{35}{5}$. Thus the conditional probability is $\binom{5}{3}\binom{30}{2}/(1-p)$.

(b) Let X_j be the number of days after seeing $j-1$ distinct outcomes needed to see j of the possible lottery outcomes. Then each X_j is distributed as a first success distribution with parameter $p_j = 1-(j-1)/\binom{35}{5}$ and thus $E(\sum_j X_j) = \sum_j 1/p_j$.

(c) The probability that a particular number has never been picked is $p = (\binom{34}{5}/\binom{35}{5})^{50} = (30/35)^{50}$. If X is the number of numbers that have never been chosen in the first 50 days, we can approximate X as Poisson with parameter $35p$, so the desired probability is $P(X = 0) = e^{-35(30/35)^{50}}$.

86. (a) $\binom{n_A}{x}\binom{n_B}{y}\binom{n_C}{z}/\binom{n}{m}$.

(b) Let I_j be the indicator of whether the j^{th} sampled individual is party A. Then $P(I_j = 1) = \frac{n_A}{n}$ so $E(X) = E(\sum_j I_j) = \frac{mn_A}{n}$.

(c) $E(X^2) = E(\sum_j I_j + \sum_{j<k} 2I_jI_k) = \frac{mn_A}{n} + m(m-1)\frac{n_A(n_A-1)}{n(n-1)}$. If $m = 1$, then $X = I_1$ and $\binom{m}{2} = 0$, so $Var(X) = \frac{mn_A}{n} - (\frac{mn_A}{n})^2 = (n_A/n)(1-(n_A/n))$ which is just the variance for an indicator variable.

If $m = n$, then $Var(X) = n_A + n_A(n_A-1) - n_A^2 = 0$ since $X = n_A$ with probability 1 since we are sampling the entire population without replacement.

87. (a) The number of democrats on the committee is hypergeometric with parameters d, 100-d, c. The expected value is $\frac{cd}{100}$.

(b) Let I_j be the indicator of whether state j is represented. Then $P(I_j = 1) = 1 - \binom{98}{c}/\binom{100}{c}$, so $E(\sum_j I_j) = 50(1 - \binom{98}{c}/\binom{100}{c})$.

(c) Let I_j be the indicator of both of state j's senators being on the committee. Then $P(I_j = 1) = \binom{98}{c-2}/\binom{100}{c}$, so $E(\sum_j I_j) = 50\binom{98}{c-2}/\binom{100}{c}$.

(d) The distribution is hypergeometric with parameters $50, 50, 20$. The expected number of junior senators on the committee is $10 = 20 * 50/100$.

(e) Let I_j be the indicator of state j having both senators on the committee of 20. Then $P(I_j = 1) = \binom{98}{18}/\binom{100}{20}$ so $E(\sum_j I_j) = 50\binom{98}{18}/\binom{100}{20}$.

88. (a) This is a negative hypergeometric distribution with parameters g, b, 1. The expected number of courses is $b/(g+1)$.

(b) This should be less than b/g, since b/g is the expected value of the geometric distribution which waits for the first success when sampling with replacement.

89. (a) Let I_j be the indicator of whether position j is a C. Then $P(I_j = 1) = p_C$ so $E(X = \sum_j I_j) = 115p_C$. $E(X^2) = \sum_j E(I_j) + \sum_{j<k} 2E(I_jI_k) = 115p_C + (115)(114)p_C^2$, so $Var(X) = 115p_C + 115(114)p_C^2 - 115^2p_C^2 = 115p_C - 115p_C^2 = 115p_C(1-p_C)$, which we could also have computed directly by recognizing X as $Bin(115, p_C)$ distribution.

(b) Let I_j be the indicator of whether the six letters beginning at index j are CATCAT. Then $P(I_j = 1) = p_C^2p_A^2p_T^2$, so $E(\sum_j I_j) = 110p_C^2p_A^2p_T^2$.

(c) There are four ways for CAT to appear at least once and one way for CAT to appear twice. Thus the probability is $4p_Cp_Ap_T - p_C^2p_A^2P_T^2$.

90. (a) This has a hypergeometric distribution with parameters 100, 1000, 20 (or equivalently 20, 980, 100). The expected value is $20 * 100/1000 = 100 * 20/1000 = 2$.

(b) Let I_j be the indicator of whether person j was surveyed by Bob or Alice. Then $P(I_j = 1) = 1 - (9/10)(98/100)$, so $E(\sum_j I_j) = 1000 - 9 * 98 = 118$.

(c) Let I_j be the indicator of whether couple j is in Bob's sample. Then $P(I_j = 1) = \binom{998}{18}/\binom{1000}{20} = (20*19)/(1000*999)$ so $E(\sum_j I_j) = 500\binom{998}{18}/\binom{1000}{20}$.

91. (a) If $F = G$, then the X's and Y's are iid so the rank of X_j is equally likely to be any number between 1 and $m+n$, so R_j has the discrete uniform distribution on $1, \dots, m+n$ and $E(\sum_j R_j) = \sum_j (m+n+1)/2 = (m+n+1)m/2$.

(b) Let I_k be the indicator of X_j being greater than Y_k and J_l be the indicator of whether X_j is greater than X_l. Then $R_j = 1 + \sum_k I_k + \sum_l J_l$ and $E(R_j) = 1 + \sum_k E(I_k) + \sum_l E(J_l)$. $P(J_l = 1) = 1/2$ so $\sum_l E(J_l) = (m-1)/2$ and $P(I_k = 1) = p$ so $\sum_k E(I_k) = np$. Thus $E(R_j) = (m+1)/2 + np$ and $E(R) = m(m+1)/2 + mnp$.

If $p = 1/2$, then this becomes $m(m+1)/2 + mn/2 = m(m+1+n)/2$ which reduces to the previous part, as expected.

92. (a) If X is the rank of the best dish found in the exploration phase, then the sum of the ranks for exploitation is $(m-k)X$. For the exploration, the ranks are any uniformly selected subset of k elements between $1, \ldots, n$. Let R_j be the rank of the j^{th} dish tried during the exploration phase. Then R_j is discrete uniform on $1, \ldots, n$, so $E(R_j) = (n+1)/2$. The sum of all ranks is $(m-k)X + \sum_j R_j$ and the expected value is $(m-k)E(X) + k(n+1)/2$. (Note that $X = \max(R_1, \ldots, R_K)$.)

(b) For $k \leq j \leq n$, $P(X = j) = \binom{j-1}{k-1} / \binom{n}{k}$.

(c) $E(X) = \sum_j j\binom{j-1}{k-1}/\binom{n}{k} = 1/\binom{n}{k}\sum_j k\binom{j}{k} = k/\binom{n}{k}\sum_j \binom{j}{k} = k\binom{n+1}{k+1}/\binom{n}{k} = k(n+1)/(k+1)$.

(d) Taking a derivative of $(m-k)k(n+1)/(k+1) + k(n+1)/2$ yields $-k(n+1)/(k+1) + (m-k)(n+1)/(k+1) - (m-k)k(n+1)/(k+1)^2 + (n+1)/2$ and setting equal to zero, we find

$$k^2 + 2k - (2m+1) = 0 \implies k = \frac{-2 \pm \sqrt{4 + 4(2m+1)}}{2} = \sqrt{2m+2} - 1.$$