Solutions 5

1. Let X be a random day of the week, coded so that Monday is 1, Tuesday is 2, etc. (so X takes values 1, 2,..., 7, with equal probabilities). Let Y be the next day after X (again represented as an integer between 1 and 7). Do X and Y have the same distribution? What is P(X< Y)?

   Solution: They have the same distribution since Y is also equally likely to represent any day of the week, but P(X<Y)=P(X≠7)=6/7.

2. Are there discrete random variables X and Y such that E(X) > 100E(Y) but Y is greater than X with probability at least 0.99?

   Solution: Yes: consider what happens if we make X usually 0 but on rare occasions, X is extremely large (like the outcome of a lottery); Y , on the other hand, can be more moderate. For example, let X be 10^6 with probability 1/100 and 0 with probability 99/100, and let Y be the constant 1.

3. A group of 50 people are comparing their birthdays (as usual, assume their birthdays are independent, are not February 29, etc.). Find the expected number of pairs of people with the same birthday, and the expected number of days in the year on which at least two of these people were born.

   Solution: Creating an indicator r.v. for each pair of people, we have that the expected number of pairs of people with the same birthday is $C_{50}^{2}\frac{1}{365}$ by linearity. Now create an indicator r.v. for each day of the year, taking the value 1 if at least two of the people were born that day (and 0 otherwise). Then the expected number of days on which at least two people were born is $365(1-\left(\frac{364}{365}\right)^{50}-50*\frac{1}{365}*\left(\frac{364}{365}\right)^{49})$.

4. A total of 20 bags of Haribo gummi bears are randomly distributed to the 20 students in a certain Stat 110 section. Each bag is obtained by a random student, and the outcomes of who gets which bag are independent. Find the average number of bags of gummi bears that the first three students get in total, and find the average number of students who get at least one bag.

   Solution: Let X_j be the number of bags of gummi bears that the jth student gets, and let I_j be the indicator of X_j ≥1. Then X_j ~ Bin(20, 1/20), so E(X_j) = 1. So E(X_1 + X_2 + X_3) = 3 by linearity. The average number of students who get at least one bag is E(I_1 + ⋯ + I_{20}) = 20E(I_1) = 20 P(I_1 = 1) = 20 (1-$(\frac{19}{20})^{20}$).

5. There are 100 shoelaces in a box. At each stage, you pick two random ends and tie them together. Either this results in a longer shoelace (if the two ends came from different pieces), or it results in a loop (if the two ends came from the same piece). What are the expected number of steps until everything is in loops, and the expected number of loops after everything is in loops? (This is a famous interview problem; leave the latter answer as a sum.)

Hint: for each step, create an indicator r.v. for whether a loop was created then, and note that the number of free ends goes down by 2 after each step.

Solution: Initially there are 200 free ends. The number of free ends decreases by 2 each time since either two separate pieces are tied together, or a new loop is formed. So exactly 100 steps are always needed. Let I_J be the indicator r.v. for whether a new loop is formed at the jth step. At the time when there are n unlooped pieces (so 2n ends), the probability of forming a new loop is $\frac{n}{C_{2n}^{2}}$ = 1/2n-1 since any 2 ends are equally likely to be chosen, and there are n ways to pick both ends of 1 of the n pieces. By linearity, the expected number of loops is $\sum_{n=1}^{100} \frac{1}{2n-1}$.

6. Randomly, k distinguishable balls are placed into n distinguishable boxes, with all possibilities equally likely. Find the expected number of empty boxes.

   Let I_j be the indicator random variable for the jth box being empty, so $I_1 + \cdots + I_n$ is the number of empty boxes (the above picture illustrates a possible outcome with n = 9, k = 13). Then E(I_j) = P(I_j = 1) =$\left(1-\frac{1}{n}\right)^k$. By linearity, E($\sum_{j=1}^{n} I_j$) = $\sum_{j=1}^{n} E(I_j) = n\left(1-\frac{1}{n}\right)^k$.
   Miracle check: for any k ≥ 1, there can be at most n - 1 empty boxes, so the expected number of empty boxes must be at most n - 1. Here, we do have $n\left(1-\frac{1}{n}\right)^k \leq n\left(1-\frac{1}{n}\right) = n-1$. And for k = 0 the answer should reduce to n, while for n = 1, k ≥ 1 it should reduce to 0. Also, it makes sense that the expected number of empty boxes converges to 0 (without ever reaching 0, if n ≥ 2) as k → ∞.

7. Athletes compete one at a time at the high jump. Let X_j be how high the jth jumper jumped, with X_1, X_2,... i.i.d. with a continuous distribution. We say that the jth jumper set a record if X_j is greater than all of X_{j-1},..., X_1.

   (a) Is the event "the 110th jumper sets a record" independent of the event "the 111th jumper sets a record"? Justify your answer by finding the relevant probabilities in the definition of independence and with an intuitive explanation.

   Solution: Let I_j be the indicator r.v. for the jth jumper setting a record. By symmetry, P(I_j = 1) = 1/j (as all of the first j jumps are equally likely to be the highest of those jumps). Also, P(I_{110} = 1, I_{111} = 1) = 109!/ 111! =$\frac{1}{110*111}$, since having the 110th and 111th jumps both being records is the same thing as having the highest of the first 111 jumps being in position 111, the second highest being in position 110, and the remaining 109 being in any order. So P(I_{110} = 1, I_{111} = 1) = P(I_{110} = 1)P(I_{111} = 1), which shows that the 110th jumper setting a record is independent of the 111th jumper setting a record. Intuitively, this makes sense since learning that the 111th jumper sets a record gives us no information about the "internal" matter of how the first 110 jumps are arranged amongst themselves.

   (b) Find the mean number of records among the first n jumpers (as a sum). What happens to the mean as n → ∞?

   Solution: By linearity, the expected number of records among the first n jumpers is $\sum_{j=1}^{n} \frac{1}{j}$, which goes to ∞ as k → ∞(as this is the harmonic series).