CS 70 Discrete Mathematics and Probability Theory
Summer 2019 James Hulett and Elizabeth Yang

HW 5

1 Combinatorial Proof!

Prove that for $0 < k < n$, $\binom{n}{k} = \sum_{i=0}^{k} \binom{n-i-1}{k-i}$.

Solution:

The left hand side of the equation is just the typical way of counting the number of bit strings of length n that have exactly k ones.

For the right hand side, we first look at a single element of the summation. This counts the number of bit strings of length n that have exactly k ones and that have their first zero at position $i+1$. To see why this is, consider that in order for a bit string's first zero to come at position $i+1$, the first i digits have to all be ones. Thus, there are $i+1$ digits that need to be fixed. Once we've dealt with those, there are $n-i-1$ positions left that are not fixed, of which $k-i$ need to be ones (since we already used up i ones in the first $i+1$ digits). Hence, there are $\binom{n-i-1}{k-i}$ ways to complete the string, as claimed.

Since $n > k$, we know that every bit string must have at least one zero–and since there can only be k ones, we know that the first zero can appear no later that position $k+1$. Thus, by summing over i from 0 to k, we cover each of the bit strings covered by the left hand side exactly once, and so the two sides must be equal.

2 Cliques in Random Graphs

In last week's homework you worked on a graph $G = (V,E)$ on n vertices which is generated by the following random process: for each pair of vertices u and v, we flip a fair coin and place an (undirected) edge between u and v if and only if the coin comes up heads. Now consider:

(a) Prove that $\binom{n}{k} \leq n^k$. *Optional:* Can you come up with a combinatorial proof? Of course, an algebraic proof would also get full credit.

(b) Prove that the probability that the graph contains a k-clique, for $k \geq 4\log n + 1$, is at most $1/n$. (The log is taken base 2). *Hint:* Apply the union bound and part (a).

Solution:

(a) The algebraic solution is an application of the definition of $\binom{n}{k}$:

$$\binom{n}{k} = \frac{n!}{(n-k)!k!} = \frac{n\cdot(n-1)\cdots(n-k+1)}{k!} \tag{1}$$

$$\tag{2}$$

We can also translate the proof above into a combinatorial proof. The number of ways there are to pick k books out of a book shelf with n books to read is smaller than the number of ways you can read k books in n days, where each day you can read any of these k books. (1) is less than or equal to the number of ways to place k books from n on the shelf without placement (**??**) (as if you have n books but only k spaces on your bookshelf) is less than or equal to the number of ways to place k books from n on the shelf with replacement (**??**) (as if you own a bookstore and you have k spaces to place n different book titles).

(b) Let A_S denote the event that S is a k-clique, where $S \subseteq V$ is of size k. Then, the event that the graph contains a k-clique can be described as the union of A_S's over all $S \subseteq V$ of size k. Using the union bound,

$$\mathbb{P}\left[\bigcup_{S\subseteq V,|S|=k} A_S\right] \leq \sum_{S\subseteq V,|S|=k} \mathbb{P}[A_S] = \sum_{S\subseteq V,|S|=k} \frac{1}{2^{\binom{k}{2}}}.$$

Now, since there are $\binom{n}{k}$ ways of choosing a subset $S \subseteq V$ of size k, the right-hand side of the above equality is

$$\frac{\binom{n}{k}}{2^{\binom{k}{2}}} = \frac{\binom{n}{k}}{2^{k(k-1)/2}} \leq \frac{n^k}{\left(2^{(k-1)/2}\right)^k} \leq \frac{n^k}{\left(2^{(4\log n+1-1)/2}\right)^k} = \frac{n^k}{\left(2^{2\log n}\right)^k} = \frac{n^k}{n^{2k}} = \frac{1}{n^k} \leq \frac{1}{n}.$$

3 Balls and Bins, All Day Every Day

You throw n balls into n bins uniformly at random, where n is a positive *even* integer.

(a) What is the probability that exactly k balls land in the first bin, where k is an integer $0 \leq k \leq n$?

(b) What is the probability p that at least half of the balls land in the first bin? (You may leave your answer as a summation.)

(c) Using the union bound, give a simple upper bound, in terms of p, on the probability that some bin contains at least half of the balls.

(d) What is the probability, in terms of p, that at least half of the balls land in the first bin, or at least half of the balls land in the second bin?

(e) After you throw the balls into the bins, you walk over to the bin which contains the first ball you threw, and you randomly pick a ball from this bin. What is the probability that you pick up the first ball you threw? (Again, leave your answer as a summation.)

Solution:

(a) The probability that a particular ball lands in the first bin is $1/n$. We need exactly k balls to land in the first bin, which occurs with probability $(1/n)^k$, and we need exactly $n-k$ balls to land in a different bin, which occurs with probability $(1-1/n)^{n-k}$, and there are $\binom{n}{k}$ ways to choose which of the n balls land in first bin. Thus, the probability is $\binom{n}{k}(1/n)^k(1-1/n)^{n-k}$.

(b) This is the summation over $k = n/2, \ldots, n$ of the probabilities computed in the first part, i.e., $\sum_{k=n/2}^{n}\binom{n}{k}(1/n)^k(1-1/n)^{n-k}$.

(c) The event that some bin has at least half of the bins is the union of the events A_k, $k = 1, \ldots, n$, where A_k is the event that bin k has at least half of the balls. By the union bound, $\mathbb{P}(\bigcup_{i=1}^{n} A_k) \le \sum_{i=1}^{n}\mathbb{P}(A_k) = np$.

(d) The probability that the first bin has at least half of the balls is p; similarly, the probability that the second bin has at least half of the balls is also p. There is overlap between these two events, however: the first bin has half of the balls and the second bin has the second half of the balls. The probability of this event is $\binom{n}{n/2}n^{-n}$: there are n^n total possible configurations for the n balls to land in the bins, but if we require exactly $n/2$ of the balls to land in the first bin and the remaining balls to land in the second bin, there are $\binom{n}{n/2}$ ways to choose which balls land in the first bin. By the principle of inclusion-exclusion, our desired probability is $p + p - \binom{n}{n/2}n^{-n} = 2p - \binom{n}{n/2}n^{-n}$.

(e) Condition on the number of balls in the bin. First we calculate the probability $\mathbb{P}(A_k)$, where A_k is the event that the bin contains k balls and $k \in \{1, \ldots, n\}$ (note that $k \neq 0$ since we know at least one ball has landed in this bin). A_k is the event that, in addition to the first ball you threw, an additional $k-1$ of the other $n-1$ balls landed in this bin, which by the reasoning in Part (a) has probability

$$\mathbb{P}(A_k) = \binom{n-1}{k-1}(1/n)^{k-1}(1-1/n)^{n-k} \quad .$$

If we let B be the event that we pick up the first ball we threw, then

$$\mathbb{P}(B \mid A_k) = 1/k$$

since we are equally likely to pick any of the k balls in the bin. Thus the overall probability we are looking for is, by an application of the law of total probability,

$$\mathbb{P}(B) = \sum_{k=1}^{n}\mathbb{P}(A_k \cap B) = \sum_{k=1}^{n}\mathbb{P}(A_k)\mathbb{P}(B \mid A_k) = \sum_{k=1}^{n}\frac{1}{k}\binom{n-1}{k-1}\left(\frac{1}{n}\right)^{k-1}\left(1-\frac{1}{n}\right)^{n-k}.$$

4 Indicator Variables

(a) After throwing n balls into m bins at random, what is the expected number of bins that contains exactly k balls?

(b) Alice and Bob each draw k cards out of a deck of 52 distinct cards with replacement. Find k such that the expected number of common cards that both Alice and Bob draw is at least 1.

(c) How many people do you need in a room so that you expect that there is going to be a shared birthday on a Monday of the year (assume 52 Mondays in a year and 365 days in a year)? By "expect" we mean on expectation, there should be 1 pair of people that have a shared birthday.

Solution:

(a) Let $X_i = 1$ if bin i contains exactly k balls and $X_i = 0$ otherwise.

$$\mathbb{E}[X_i] = \binom{n}{k}\left(\frac{1}{m}\right)^k\left(\frac{m-1}{m}\right)^{n-k} = \binom{n}{k}\frac{(m-1)^{n-k}}{m^n}$$
$$\mathbb{E}[X] = \sum_{i=1}^{m}\binom{n}{k}\frac{(m-1)^{n-k}}{m^n} = \binom{n}{k}\frac{(m-1)^{n-k}}{m^{n-1}}$$

(b) Let $X_i = 1$ if card i is chosen by both Alice and Bob and $X_i = 0$ otherwise.
After drawing k cards, the probability that any given card appears at least once $1-(51/52)^k$ so

$$\mathbb{E}[X_i] = \left(1-\left(\frac{51}{52}\right)^k\right)\cdot\left(1-\left(\frac{51}{52}\right)^k\right)$$
$$\mathbb{E}[X] = \sum_{i=1}^{52}\left(1-\left(\frac{51}{52}\right)^k\right)^2 = 52\cdot\left(1-\left(\frac{51}{52}\right)^k\right)^2.$$

Setting $\mathbb{E}[X] = 1$, we have $k = 7.69 \approx 8$.

(c) Two interpretations:
Interpretation 1 (number of pairs who share a birthday on a specific day):
For $i < j$, let $X_{i,j} = 1$ if i, j share a birthday and $X_{i,j} = 0$ otherwise. Then, the total number of shared birthdays is $X = \sum_{i=1}^{k-1}\sum_{j=i+1}^{k} X_{i,j}$, where k is the total number of people in the room. There is $52/365$ chance that person i has a birthday on a Monday and $1/365$ chance that person j has same birthday as person i so

$$\mathbb{E}[X] = \sum_{i=1}^{k-1}\sum_{j=i+1}^{k}\frac{52}{365}\cdot\frac{1}{365} = \binom{k}{2}\frac{52}{365^2} = \frac{k(k-1)}{2}\cdot\frac{52}{365^2}.$$

We want $\mathbb{E}[X] = 1$ so $k \geq 73$.

Interpretation 2 (number of days on which there are multiple people with that birthday):
Let X_i represent whether or not there are shared birthdays on the i-th Monday of the year. Then $X = X_1 + X_2 + \cdots + X_{52}$. By linearity of expectation, $E[X] = E[X_1] + E[X_2] + \cdots + E[X_{52}] =$

$52 * E[X_i]$, and $E[X_i] = 1 - P$[there are no shared birthdays on the i-th Monday] $= 1 - (P[0$ people are born on the i-th Monday] $+ P$[exactly 1 person is born on the i-th Monday])

$$\mathbb{E}[X] = 52 \cdot E[X_i] = 52 \cdot \left(1 - \left(\frac{364}{365}\right)^n + n \cdot \frac{1}{365}\left(\frac{364}{365}\right)^{n-1}\right)$$

We want $\mathbb{E}[X] = 1$ so $n \geq 77.1$.

5 Poisoned Smarties

Supposed there are 3 men who are all owners of their own Smarties factories. Burr Kelly, being the brightest and most innovative of the men, produces considerably more Smarties than his competitors and has a commanding 45% of the market share. Yousef See, who inherited his riches, lags behind Burr and produces 35% of the world's Smarties. Finally Stan Furd, brings up the rear with a measly 20%. However, a recent string of Smarties related food poisoning has forced the FDA investigate these factories to find the root of the problem. Through his investigations, the inspector found that one Smarty out of every 100 at Kelly's factory was poisonous. At See's factory, 1.5% of Smarties produced were poisonous. And at Furd's factory, the probability a Smarty was poisonous was 0.02.

(a) What is the probability that a randomly selected Smarty will be safe to eat?

(b) If we know that a certain Smarty didn't come from Burr Kelly's factory, what is the probability that this Smarty is poisonous?

(c) Given this information, if a randomly selected Smarty is poisonous, what is the probability it came from Stan Furd's Smarties Factory?

Solution:

(a) Let S be the event that a smarty is safe to eat.
Let BK be the event that a smarty is from Burr Kelly's factory.
Let YS be the event that a smarty is from Yousef See's factory.
Finally, let SF be the event that a smarty is from Stan Furd's factory.

$$\begin{aligned}\mathbb{P}(S) &= \mathbb{P}(BK)P(S \mid BK) + \mathbb{P}(YS)\mathbb{P}(S \mid YS) + \mathbb{P}(SF)\mathbb{P}(S \mid SF) \\ &= (0.45)(0.99) + (0.35)(0.985) + (0.2)(0.98) = 0.98625.\end{aligned}$$

Therefore the probability that a Smarty is safe to eat is about 0.98625.

(b) Let P be the event that a smarty is poisonous.

$$\begin{aligned}\mathbb{P}(P \mid \neg BK) &= \mathbb{P}(YS \mid \neg BK)\mathbb{P}(P \mid YS) + \mathbb{P}(SF \mid \neg BK)\mathbb{P}(P \mid SF) \\ &= \frac{\mathbb{P}(YS)}{\mathbb{P}(\neg BK)}\mathbb{P}(P \mid YS) + \frac{\mathbb{P}(SF)}{\mathbb{P}(\neg BK)}\mathbb{P}(P \mid SF) \\ &= \frac{0.35}{0.55} \cdot 0.015 + \frac{0.2}{0.55} \cdot 0.02 = 0.0168.\end{aligned}$$

(c)

$$\mathbb{P}(SF \mid P) = \frac{\mathbb{P}(P \mid SF)\mathbb{P}(SF)}{\mathbb{P}(P)}$$

In the first part we calculate the probability that any random Smarty was safe to eat. We can use that since $\mathbb{P}(P) = 1 - \mathbb{P}(S)$. Therefore the solution becomes:

$$\begin{aligned}\mathbb{P}(SF \mid P) &= \frac{\mathbb{P}(P \mid SF)\mathbb{P}(SF)}{1 - \mathbb{P}(S)} \\ &= \frac{(0.02)(0.2)}{(1 - 0.98625)} = 0.29.\end{aligned}$$

6 Testing Model Planes

Dennis is testing model airplanes. He starts with n model planes which each independently have probability p of flying successfully each time they are flown, where $0 < p < 1$. Each day, he flies every single plane and keeps the ones that fly successfully (i.e. don't crash), throwing away all other models. He repeats this process for many days, where each "day" consists of Dennis flying any remaining model planes and throwing away any that crash. Let X_i be the random variable representing how many model planes remain after i days. Note that $X_0 = n$. Justify your answers for each part.

(a) What is the distribution of X_1? That is, what is $\mathbb{P}[X_1 = k]$?

(b) What is the distribution of X_2? That is, what is $\mathbb{P}[X_2 = k]$? Show that X_2 follows a binomial distribution by finding some n' and p' such that $X_2 \sim \text{Binom}(n', p')$.

(c) Repeat the previous part for X_t for arbitrary $t \geq 1$.

(d) What is the probability that at least one model plane still remains (has not crashed yet) after t days? Do not have any summations in your answer.

(e) Considering only the first day of flights, is the event A_1 that the first and second model planes crash independent from the event B_1 that the second and third model planes crash? Recall that two events A and B are independent if $\mathbb{P}[A \cap B] = \mathbb{P}[A]\mathbb{P}[B]$. Prove your answer using this definition.

(f) Considering only the first day of flights, let A_2 be the event that the first model plane crashes *and* exactly two model planes crash in total. Let B_2 be the event that the second plane crashes on the first day. What must n be equal to in terms of p such that A_2 is independent from B_2? Prove your answer using the definition of independence stated in the previous part.

(g) Are the random variables X_i and X_j, where $i < j$, independent? Recall that two random variables X and Y are independent if $\mathbb{P}[X = k_1 \cap Y = k_2] = \mathbb{P}[X = k_1]\mathbb{P}[Y = k_2]$ for all k_1 and k_2. Prove your answer using this definition.

Solution:

(a) Since Dennis is performing n trials (flying a plane), each with an independent probability of "success" (not crashing), we have $X_1 \sim \text{Binom}(n, p)$, or $\mathbb{P}[X = k] = \binom{n}{k} p^k (1-p)^{n-k}$, for $0 \le k \le n$.

(b) Each model plane independently has probability p^2 of surviving both days. Whether a model plane survives both days is still independent from whether any other model plane survives both days, so we can say $X_2 \sim \text{Binom}(n, p^2)$, or $\mathbb{P}[X = k] = \binom{n}{k} p^{2k} (1-p^2)^{n-k}$, for $0 \le k \le n$.

(c) By extending the previous part we see each model plane has probability p^t of surviving t days, so $X_t \sim \text{Binom}(n, p^t)$, or $\mathbb{P}[X = k] = \binom{n}{k} p^{tk} (1-p^t)^{n-k}$, for $0 \le k \le n$.

(d) We consider the complement, the probability that no model planes remain after t days. By the previous part we know this to be $\mathbb{P}[X_t = 0] = \binom{n}{0} p^{t(0)} (1-p^t)^{n-0} = (1-p^t)^n$. So the probability of at least model plane remaining after t days is $1 - (1-p^t)^n$.

(e) No. $\mathbb{P}[A_1 \cap B_1]$ is the probability that the first three model planes crash, which is $(1-p)^3$. But $\mathbb{P}[A_1]\mathbb{P}[B_1] = (1-p)^2(1-p)^2 = (1-p)^4$. So $\mathbb{P}[A_1 \cap B_1] \ne \mathbb{P}[A_1]\mathbb{P}[B_1]$ and A_1 and B_1 are not independent.

(f) $\mathbb{P}[A_1 \cap B_1]$ is the probability that only the first model plane and second model plane crash, which is $(1-p)^2 p^{n-2}$. $\mathbb{P}[A_1]$ is the probability that the first model plane crashes, and exactly one of the remaining $n-1$ model planes crashes, so $\mathbb{P}[A_2] = (1-p) \cdot \binom{n-1}{1} (1-p) p^{n-1-1} = (n-1)(1-p)^2 p^{n-2}$. Trivially, we have $\mathbb{P}[B_2] = 1-p$, so $\mathbb{P}[A_2]\mathbb{P}[B_2] = (n-1)(1-p)^3 p^{n-2}$ which is equal to $\mathbb{P}[A_2 \cap B_2] = (1-p)^2 p^{n-2}$ only when $(n-1)(1-p) = 1$, or when $n = \frac{1}{1-p} + 1$.

(g) No. Let $k_1 = 0$ and $k_2 = 1$. Then $\mathbb{P}[X_i = k_1 \cap X_j = k_2] = 0$ because the sequence of X_i is by definition non-increasing. But $\mathbb{P}[X_i = k_1] > 0$ and $\mathbb{P}[X_j = k_2] > 0$ so $\mathbb{P}[X_i = k_1]\mathbb{P}[X_j = k_2] > 0$.

7 Geometric Distribution

Two faulty machines, M_1 and M_2, are repeatedly run synchronously in parallel (i.e., both machines execute one run, then both execute a second run, and so on). On each run, M_1 fails with probability p_1 and M_2 fails with probability p_2, all failure events being independent. Let the random variables X_1, X_2 denote the number of runs until the first failure of M_1, M_2 respectively; thus X_1, X_2 have geometric distributions with parameters p_1, p_2 respectively. Let X denote the number of runs until the first failure of *either* machine.

(a) Show that X also has a geometric distribution, with parameter $p_1 + p_2 - p_1 p_2$.

(b) Now, two technicians are hired to check on the machines every run. They decide to take turns checking on the machines every run. What is the probability that the first technician is the first one to find a faulty machine?

Solution:

(a) We have that $X_1 \sim \text{Geom}(p_1)$ and $X_2 \sim \text{Geom}(p_2)$. Also, X_1, X_2 are independent r.v.'s. We also use the following definition of the minimum:

$$\min(x,y) = \begin{cases} x & \text{if } x \le y; \\ y & \text{if } x > y. \end{cases}$$

Now, for all $k \in \{1,2,\dots\}$, $\min(X_1,X_2) = k$ is equivalent to $(X_1 = k) \cap (X_2 \ge k)$ or $(X_2 = k) \cap (X_1 > k)$. Hence,

$$\begin{aligned} \mathbb{P}[X = k] &= \mathbb{P}[\min(X_1,X_2) = k] \\ &= \mathbb{P}[(X_1 = k) \cap (X_2 \ge k)] + \mathbb{P}[(X_2 = k) \cap (X_1 > k)] \\ &= \mathbb{P}[X_1 = k] \cdot \mathbb{P}[X_2 \ge k] + \mathbb{P}[X_2 = k] \cdot \mathbb{P}[X_1 > k] \end{aligned}$$

(since X_1 and X_2 are independent)

$$= [(1-p_1)^{k-1} p_1](1-p_2)^{k-1} + [(1-p_2)^{k-1} p_2](1-p_1)^k$$

(since X_1 and X_2 are geometric)

$$\begin{aligned} &= ((1-p_1)(1-p_2))^{k-1} (p_1 + p_2(1-p_1)) \\ &= (1 - p_1 - p_2 + p_1 p_2)^{k-1} (p_1 + p_2 - p_1 p_2). \end{aligned}$$

But this final expression is precisely the probability that a geometric r.v. with parameter $p_1 + p_2 - p_1 p_2$ takes the value k. Hence $X \sim \text{Geom}(p_1 + p_2 - p_1 p_2)$, and $\mathbb{E}[X] = (p_1 + p_2 - p_1 p_2)^{-1}$.

An alternative, slightly cleaner approach is to work with the *tail probabilities* of the geometric distribution, rather than with the usual point probabilities as above. In other words, we can work with $\mathbb{P}[X \ge k]$ rather than with $\mathbb{P}[X = k]$; clearly the values $\mathbb{P}[X \ge k]$ specify the values $\mathbb{P}[X = k]$ since $\mathbb{P}[X = k] = \mathbb{P}[X \ge k] - \mathbb{P}[X \ge (k+1)]$, so it suffices to calculate them instead.

We then get the following argument:

$$\begin{aligned}
\mathbb{P}[X \geq k] &= \mathbb{P}[\min(X_1, X_2) \geq k] \\
&= \mathbb{P}[(X_1 \geq k) \cap (X_2 \geq k)] \\
&= \mathbb{P}[X_1 \geq k] \cdot \mathbb{P}[X_2 \geq k] && \text{since } X_1, X_2 \text{ are independent} \\
&= (1-p_1)^{k-1}(1-p_2)^{k-1} && \text{since } X_1, X_2 \text{ are geometric} \\
&= ((1-p_1)(1-p_2))^{k-1} \\
&= (1-p_1-p_2+p_1p_2)^{k-1}.
\end{aligned}$$

This is the tail probability of a geometric distribution with parameter $p_1 + p_2 - p_1p_2$, so we are done.

(b) Let the required probability be denoted by p (the probability that the first technician is the first to find a faulty machine) and X denote the first point of failure of either of the machines. Now, we have:

$$p = \mathbb{P}[X = 1] + \mathbb{P}[X = 3] + \mathbb{P}[X = 5] + \dots$$

as the first technician will be the first to find a faulty machine if and only if the first failure occurs in an odd run. Now, let us decompose the sum as follows:

$$p = \mathbb{P}[X = 1] + \sum_{i=1}^{\infty} \mathbb{P}[\{X = 2i+1\} \cap \{X \neq 1\}] = \mathbb{P}[X = 1] + \sum_{i=1}^{\infty} \mathbb{P}[X \neq 1]\mathbb{P}[X = 2i+1 | X \neq 1]$$

Using the memoryless property of the geometric distribution to further simplify the sum, we get:

$$\begin{aligned}
p &= \mathbb{P}[X = 1] + \sum_{i=1}^{\infty} \mathbb{P}[X \neq 1]\mathbb{P}[X = 2i] = \mathbb{P}[X = 1] + \mathbb{P}[X \neq 1] \sum_{i=1}^{\infty} \mathbb{P}[X = 2i] \\
&= \mathbb{P}[X = 1] + \mathbb{P}[X \neq 1] \left(1 - \sum_{i=1}^{\infty} \mathbb{P}[X = 2i-1]\right) = \mathbb{P}[X = 1] + \mathbb{P}[X \neq 1](1-p)
\end{aligned}$$

where the second equality follows because the probability that X is even is the complement of the event that X is odd. The final equation is intuitive as in the event that the first technician doesn't find either of the machines broken in the first run, the memoryless property of the geometric distribution ensures that the probability that the second technician finds the machines broken first is the same as the probability that the first technician does when we have no knowledge of the first run. That is, we have:

$$p = \mathbb{P}[\text{Second technician finds the machines broken first} \mid \text{No machine fails in run 1}]$$

.

By solving the above equation, we get that:

$$p = \frac{1}{2 - p_1 - p_2 + p_1 p_2}$$