COL202: Discrete Mathematical Structures. I semester, 2020-21.
Amitabha Bagchi
Tutorial Sheet 10: Probability.
17 December 2020

Important: The question marked with a ♠ is to be submitted via gradescope by 11:59PM on the day that you have your tutorial.

Problem 1
We have m balls and n bins. Each ball is thrown into one out of the n bins chosen with some probability. Let us define some notation

- $[k] = \{1, \ldots, k\}$, for $k \geq 1$.
- $B_i \in [n]$ denotes the bin into which ball i is dropped, $1 \leq i \leq m$.
- $S_i \subset [m]$ denotes the set of balls dropped into bin i, $1 \leq i \leq n$.

Problem 1.1
In this question you do *not* have to calculate any probability. We will describe some events words and you just have to write out mathematical descriptions of these events in set builder notation using the notation introduced above.

1. The sample space Ω, i.e., the set of all possible outcomes.
2. Bin i is non-empty.
3. Bin i is empty.
4. No bin is empty.
5. No bin is non-empty.
6. Half the bins are empty. (You can assume n is even.)
7. At least half the bins are empty.
8. There is a bin with two balls in it.
9. No bin has more than k balls.
10. Bin i and Bin j have the same number of balls.
11. No two bins have the same number of balls.

Problem 1.2 ♠
Prove that $\Pr(|S_i| > 0 \cap |S_k| > 0) \leq \Pr(|S_i| > 0)\Pr(|S_k| > 0)$, for all $i, k \in [m]$.

Problem 1.3
Generalize the result of Problem 1.2 to prove that $\Pr(|S_i| > j \cap |S_k| > \ell) \leq \Pr(|S_i| > j)\Pr(|S_k| > \ell)$, for all $i, k \in [m]$ and $j, \ell \geq 0$.

Problem 2
We have two fair dice, i.e., each of the 6 numbers comes up with equal probability. Two throws of this dice gives us a uniform probability space defined on the sample space $[6]^2$. Now, suppose we change the situation slightly: throw the first dice and let's say it shows the number i. Repeatedly throw the second dice till it shows a number $\geq i$, then stop. Ignore all the numbers $< i$. Consider the following events. For each of them, first try to guess without calculation whether the probability is higher for the second kind of throw or the first kind of throw (where we just throw the second dice without applying any condition).

- The sum of the numbers is at least 7.
- Both dice show the same number.
- The sum of the numbers is at most 7.

Now, write out the sample space of the second experiment and compute all outcome probabilities. Find the probabilities of the events given above and compare them to the probability of the same event in the previous case (where the two dice are thrown normally).

Problem 3
We are given an urn (a *ghada* in Hindi) that contains r red balls and b blue balls. We are allowed to put our hand into the urn and pull out a ball but we cannot see what we are picking so let us assume that every time we pick a ball out of the urn, each ball in the urn at that time has equal probability of being picked. After taking the ball out we may put it back in (replacement) or not.

Problem 3.1
Pick k balls, replacing each ball before picking the next one. What is the sample space for this experiment? What is the probability that.

1. All balls picked are red.
2. The first ball picked and the last ball picked have the same colour.
3. The number of red balls is equal to the number of blue balls.

Problem 3.2
Solve all subparts of Problem 3.1 assuming that the ball removed at each step is *not* replaced. You may assume that $k \leq \min\{r, b\}$. Write the sample space carefully for this case.

Problem 4
If A, B are two events from an outcome space Ω such that $A \cup B = \Omega$, show that

$$\Pr(\omega \in A \cap B) = \Pr(\omega \in A)\Pr(\omega \in B) - \Pr(\omega \notin A)\Pr(\omega \notin B).$$

Problem 5
Prove or disprove: If X and Y are independent random variables then so are $f(X)$ and $g(Y)$, where f and g are any functions.

Problem 6
Prove that two events are independent iff their indicator variables and independent.

Problem 7
In the birthday problem assume that the probability of throwing a ball in bin i is p_i, $1 \leq i \leq N$. Show that if $\max_i p_i > 1/N$ then the probability of having all balls land in different bins is strictly smaller than in the uniform case, i.e., $\max_i p_i = 1/N$.

Problem 8

Let X be a random variable that takes only non-negative integer values, then the probability generating function of X

$$g_X(z) = \sum_{k \geq 0} \Pr(X = k) z^k,$$

i.e. the generating function of the sequence $\{a_k\}_{k \geq 0}$ where $a_k = \Pr(X = k)$.

1. Prove that

   (a) $g_X(1) = 1$.

   (b) $g'_X(1) = \mathrm{E}(X)$.

   (c) $\Pr(X \leq r) \leq z^{-r} g_X(z)$, for $0 < z \leq 1$,

   (d) $\Pr(X \geq r) \leq z^{-r} g_X(z)$, for $z \geq 1$.

2. In the case where $g_X(z) = (1+z)^n/2^n$, use inequality (c) to prove the following result about binomial coefficients:

   $$\sum_{k \leq \alpha n} \binom{n}{k} \leq \frac{1}{\alpha^{\alpha n}(1-\alpha)^{(1-\alpha)n}},$$

   when $0 < \alpha < 1/2$. Check that this identity is true by testing it out at the two endpoints of α's range.

Problem 8.1

A non-negative random variable X is said to have the *Poisson distribution with mean* μ if

$$\Pr(X = k) = e^{-\mu} \frac{\mu^k}{k!}.$$

1. Write the probability generating function of X.

2. Find the mean and variance of X.

3. If X_1 is Poisson with mean μ_1 and X_2 is Poisson with mean μ_2, what is the probability that $X_1 + X_2 = n$?

Problem 9

Here's a way of choosing a random permutation of n numbers: Throw n balls into n bins. If each bin has exactly 1 ball in it then we have a permutation π where $\pi(i)$ is defined as the id of the ball that lands in bin i. If each bin does not have exactly 1 ball, we retrieve all the balls and throw them again, repeating this process till the required condition (each bin has exactly 1 ball) is achieved. Answer the questions below based on this setting.

Problem 9.1

Suppose each ball is thrown independently of all other balls, i.e., the event $\{B_i = j\}$ is independent of the event $\{B_k = \ell\}$ for all $i, j, k, l \in [n]$. Prove, using the formula for conditional probabilities that the experiment above generates a random permutation with uniform probability (i.e. each permutation is generated with equal probability.

Problem 9.2

If each time we throw all n balls is called one round of the process of generating a permuation, and if X is the number of rounds the process takes till the required condition is achieved, then clearly X is a random variable. What is the range of values X can take? What is the probability $\Pr(X = k)$? Calculate the expectation of X, $\mathrm{E}(X)$.

Problem 10
In this problem we have n bins, as before, but we have an *unlimited* supply of balls. We throw the balls one at a time, uniformly at random into one of the bins independently of all other balls, until each bin has at least one ball in it. As soon as each bin has at least one ball in it, we stop. Let X_n be the (random) number of balls thrown until we stop. This problem is known as the *Coupon Collector's problem*, and X_n is sometimes referred to as the coupon collection time. Let us solve a few problems based on this.

Problem 10.1
Compute $\mathrm{E}(X_n)$. Hint: Define Y_i to be the (random) number of steps taken to fill the i-th bin, conditioned on the fact that $i-1$ bins are already filled.

Problem 10.2
If we denote by $S_t(i) \subset [t]$ the set of balls that have fallen into bin i after exactly t balls have been thrown, compute the probability that $|S_t(i)| = 0$.

Problem 10.3
Use the calculation made in Problem 10.2 and the union bound to show that for any $c > 0$,

$$\Pr\left(X_n > \lceil n \log n + cn \rceil\right) \le e^{-c}.$$