COL202: Discrete Mathematical Structures. I semester, 2017-18.
Amitabha Bagchi
Tutorial Sheet 8: Conditional probability, independence, random variables
23 September 2017

Important: The boxed question is to be submitted at the beginning of class on a plain sheet of paper with your name, entry number and the tutorial sheet number clearly written at the top of the sheet.

Problem 1
First, let's solve a selection of questions from [1].

Problem 1.1
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 1.2
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 1.3
Construct a random variable that has finite mean and infinite variance.

Problem 1.4
Let X be a random variable that takes only non-negative integer values and let the probability generating function of X be

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

1. Prove that
   (a) $\Pr(X \leq r) \leq z^{-r} g_X(z)$, for $0 < z \leq 1$,
   (b) $\Pr(X \geq r) \leq z^{-r} g_X(z)$, for $z \geq 1$.
2. In the case where $g_X(z) = (1+z)^n/2^n$, use the first of the inequalities above 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 1.5
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 2

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 2.1

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 2.2

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 2.3 *

Recall that we discussed in class that to generate a random number between 0 and $n - 1$ we need to create $\log n$ random bits. Another way of saying this is that it takes $\log n$ random bits to select a single element at random out of a set of cardinality n. How many random bits does it take to generate a permutation using the method described in Problem 2.1? Independent of this method, give a lower bound on how many random bits are required to generate a random permutation. How do the two numbers compare?

Problem 3

In this problem we study the relationship between k-wise independence and $k - 1$-wise independence.

Problem 3.1

Give an example of a collection of events $\{A_i : 1 \leq i \leq n\}$ that are k-wise independent for some $3 \leq k \leq n$ but are not $k - 1$-wise independent. Under what condition would we be able to say that they are also $k - 1$-wise independent?

Problem 3.2

Recall the balls and bins experiment, i.e. there are m balls to be thrown into n bins, and we use the notation B_j to denote the bin in which ball j, $j \in [m]$ lands. Suppose we throw the balls in such a way that the collection of events

$$\{B_j = i : i \in [n], j \in [m]\}$$

is k-wise independent for some $k \geq 3$. Prove that this implies that this collection is k'-wise independent for all $2 \leq k' \leq k$.

Problem 4

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, 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 4.1
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, write a mathematical description of X_n.

Problem 4.2
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 4.3
Show that for any $c > 0$,

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

Try to figure out (qualitatively if not rigorously) what this result means for the variance of X_n? Hint: The approach of Problem 4.2 is *not* useful here. But a simple application of inclusion-exclusion works well. A quick look at the solution of Problem 4.1 may help.

Problem 5 *
Given the set integers S, we construct a random binary search tree as follows:

- Pick an element x uniformly at random from S. Put x in a node and make it the root of the tree.
- Recursively apply this procedure to the sets $\{y \in S : y \leq x\}$ and $\{y \in S : y > x\}$ and make the binary search trees thus created the left and right subtrees (respectively) of the node containing x.

You may assume that all random choices are independent of each other. Recall that the *height* of a binary tree is the length of the longest path from the root to a leaf in the tree. Now answer the following questions:

Problem 5.1
If h_n is the (random) height of the random binary search tree built on the set $[n]$, find $\mathrm{E}(h_n)$.

Problem 5.2
Show that there is constant c (not depending on n) such that

$$\lim_{n \to \infty} \Pr(h_n > c \log n) = 0.$$

Note: This problem (both parts) is approached using recurrences in Probs 13, 14, 15 of the exercise set following Ch 5.6 of BSD05. Have a look at that as well but try to avoid using the recurrence approach: it may work well for Problem 5.1 but inclusion-exclusion will give you an easier path to Problem 5.2.

References

[1] R. L. Graham, D. E. Knuth, O. Patashnik. Concrete Mathematics: A foundation for computer science, 2nd ed. Pearson, 1994.