COL202: Discrete Mathematical Structures
Tutorial/Homework: 14

1. Discuss Quiz-11 questions.

2. Complete discussion of Tutorial-13 problems in case needed.

3. We will use the following notion of independence of random variables:

   Definition 14.0.1 (Independent random variables) *Random variables X, Y on a sample space S are independent iff*

   $$\mathbf{Pr}[X = r_1 \quad and \quad Y = r_2] = \mathbf{Pr}[X = r_1] \cdot \mathbf{Pr}[Y = r_2].$$

   Use the above definition in the problem below.

   k objects are picked independently at random with replacement from a set of n distinct objects. For $1 \leq i < j \leq k$, let X_{ij} denote the indicator random variable that is 1 if the i^{th} and j^{th} objects are the same otherwise 0. Show that for any $i < j$ and $p < q$ such that $(i, j) \neq (p, q)$, the random variables X_{ij} and X_{pq} are independent.

4. (**Coupon-collector problem**) Every time you go to the superstore, you get a random coupon out of n distinct coupons. What is the expected number of times you have to visit the store to be able to collect all distinct coupons?

5. (**Balls and bins**) n balls are thrown randomly into n bins. Let E be the event that no bin has more than $\frac{3 \ln n}{\ln \ln n}$ balls. Show that $\mathbf{Pr}[E] \geq (1 - 1/n)$.

6. (**Universal Hashing**) Hashing is a technique used to store elements from a large universe $U = \{0, ..., m-1\}$ using a small table $T = \{0, ..., n-1\}$ using a hash function $h : U \to T$ such that the number of collisions are minimized [1].

   Using a fixed hash function might does not work. So, we use a *family* of hash functions H and then pick a hash function randomly from this family. A hash function family H is called 2-universal if

   $$\forall x, y \in U, x \neq y, \mathbf{Pr}_{h \leftarrow H}[h(x) = h(y)] \leq 1/n.$$

   Show how a 2-universal hash function family is useful in hashing and give an example of such a family.

[1] Assume that collisions are resolved using auxiliary data structure