COL202: Discrete Math	Tutorial 6
Instructor: Venkata Koppula	Release date: 28th September, 2024

The solutions for the ($\star$) marked problems must be submitted on Gradescope by 11:59am on 4th October, 2024.

This tutorial sheet requires basic probability, conditional probability and random variables. Here is a summary of the definitions, and theorems/results that we have seen in class (Lectures 16 and 17).

- A probability distribution is defined using a sample space Ω (which is finite or countably infinite) and a function $p : \Omega \to \mathbb{R}^{\geq 0}$ such that $\sum_{x \in \Omega} p(x) = 1$. The probability of any event $\mathcal{E} \subseteq \Omega$ is $\sum_{x \in \mathcal{E}} p(x)$.

- (Union Bound) For any k events $A_1, A_2, \ldots, A_k$,

$$\Pr\left[\bigcup_{i=1}^{k} A_i\right] \leq \sum_{i=1}^{k} \Pr[A_i].$$

- (Birthday bound) Suppose we sample t numbers from $\{1, 2, \ldots, n\}$, independently and uniformly at random. Let p_{coll} denote the probability that at least two of the sampled elements are equal. Then

$$1 - e^{-t(t-1)/2n} \leq p_{\mathsf{coll}} \leq \frac{t(t-1)}{2n}$$

- Given two events A, B such that $\Pr[B] > 0$, we define the conditional probability

$$\Pr[A \mid B] = \frac{\Pr[A \cap B]}{\Pr[B]}.$$

- (Law of Total Probability) Let (Ω, p) denote a probability distribution over sample space Ω. Let $(\Omega_1, \ldots, \Omega_k)$ be any partitioning of Ω. Then for any event A,

$$\Pr[A] = \sum_{i=1}^{k} \Pr[A \cap \Omega_i] = \sum_{i=1}^{k} \Pr[A \mid \Omega_i] \cdot \Pr[\Omega_i].$$

- (Random variables and Expectation) Let (Ω, p) be a probability distribution, and $X : \Omega \to \mathbb{R}$ be any random variable. Let $Z = \{X(w) : w \in \Omega\}$, and for any $z \in Z$, let $X^{-1}(z) = \{w \in \Omega \ : \ X(w) = z\}$. The expectation of this random variable is

$$\mathbb{E}[X] = \sum_{w \in \Omega} X(w) \cdot p(w) = \sum_{z \in Z} w \cdot \Pr\left[X^{-1}(z)\right].$$

- (Linearity of Expectation) For any random variables $X_1, X_2, \ldots, X_k$ over the same probability distribution (Ω, p),

$$\mathbb{E}[X_1 + X_2 + \ldots + X_k] = \mathbb{E}[X_1] + \mathbb{E}[X_2] + \ldots + \mathbb{E}[X_k].$$

1 Tutorial Submission Problem (⋆)

The following problem has a few parts. You only need to submit the ones that are (⋆) marked.

Let $P = (\Omega, p)$ and $Q = (\Omega, q)$ be two discrete probability distributions. We define their Statistical Difference (SD) as:

$$\mathsf{SD}(P, Q) = \frac{1}{2} \sum_{x \in \Omega} |p(x) - q(x)| \, .$$

The statistical difference is between 0 and 1 and measures how "far apart" the two distributions are. Notice that $\mathsf{SD}(P, Q) = 0$ if and only if the distributions are *identical*, i.e.

$$p(x) = q(x) \;\; \forall x \in \Omega$$

Similarly, $\mathsf{SD}(P, Q) = 1$ if and only if the distributions have *disjoint supports*, i.e.

$$\{x \in \Omega \text{ s.t. } p(x) > 0\} \;\cap\; \{x \in \Omega \text{ s.t. } q(x) > 0\} = \varnothing$$

1.1. Prove that

$$\mathsf{SD}(P, Q) = \max_{S \subseteq \Omega} \Pr_P[S] - \Pr_Q[S]$$

where $\Pr_P[S] = \sum_{x \in S} p(x)$ and $\Pr_Q[S] = \sum_{x \in S} q(x)$.

1.2. (⋆) In this problem we will explore the problem of distinguishing between distributions. Let $Q_0 = (\Omega, q_0)$ and $Q_1 = (\Omega, q_1)$ be two probability distributions. Consider the following game: I sample a uniformly random bit $b \leftarrow \{0, 1\}$, then sample x from Q_b. You are given x (and you also know the two distributions Q_0, Q_1). You must guess b, and you win if your guess is correct.

- Suppose $\mathsf{SD}(Q_0, Q_1) = 1$. Give an algorithm that wins the game (that is, determines the distribution) with probability 1.
- Suppose $\mathsf{SD}(Q_0, Q_1) = \delta$. Give an algorithm that wins the game (that is, determines the distribution) with probability $(1 + \delta)/2$.
- Prove that no algorithm can determine the distribution with probability greater than $(1 + \delta)/2$.

1.3. Let S_n denote the set of all permutations of $\{1, 2, \ldots, n\}$.[1] Consider the following algorithm for sampling a permutation from S_n.

Algorithm 1 Fisher Yates Shuffling

$a = [1, 2, \ldots, n]$
for $i = 1$ to $n - 1$ **do**
    $j \leftarrow$ uniformly random integer such that $i \leq j \leq n$.
    Swap $a[i]$ and $a[j]$
end for
return a

[1] A permutation σ is a bijection from $\{1, 2, \ldots, n\}$ to itself.

Let $\mathcal{D}$ be the distribution obtained via Algorithm 1. Let $\mathcal{U}(S_n)$ be the uniform distribution on S_n. Show that $\mathsf{SD}(\mathcal{D},\mathcal{U}(S_n)) = 0$.

2 Problems - General Probability

2.1. A fair coin is flipped n times. What's the probability that all the heads occur at the end of the sequence? (If no heads occur, then "all the heads are at the end of the sequence" is vacuously true.)

2.2. First one digit is chosen uniformly at random from $\{1, 2, 3, 4, 5\}$ and is removed from the set; then a second digit is chosen uniformly at random from the remaining set. What is the probability that an odd digit is picked the second time?

2.3. Consider a population where the probability of being born on any given day is given by the following (non-uniform) distribution: the probability of being born on day i, $p_i = \frac{C}{i^2}$ where $i \in [365]$ and C is a constant chosen such that $\sum_{i=1}^{365} p_i = 1$. What is the probability that, in a room of 23 people, at least two people share the same birthday under this distribution? How does this compare to the uniform case?

2.4. (♦) Let S_n denote the set of all permutations of $\{1, 2, \ldots, n\}$. Given a permutation σ, let $\mathsf{inv}(\sigma)$ denote the number of pairs of indices (i, j) such that $i < j$ and $\sigma(i) > \sigma(j)$.

Find the expected value of $\mathsf{inv}(\sigma)$, when σ sampled from the following distributions:

1. (Easy) Uniform distribution over S_n
2. A distribution $\mathcal{D}$ defined as follows: let $T_1 = \{1, 2, \ldots, n\}$ and $s_1 = \sum_{x \in T_1} x$. For $i = 1$ to n do the following:
   (a) for each $x \in T_i$, let $p_i(x) = x/s_i$. Sample an element y from (T_i, p_i).
   (b) Set $\sigma(i) = y$.
   (c) Let $T_{i+1} = T_i \setminus \{y\}$ and $s_{i+1} = s_i - y$.

Hint: in each of these two cases, decompose the random variable appropriately.

2.5. (♦) There are n seats in an aeroplane, and n passengers each having a unique ticket to a seat. Passengers come one at a time to occupy seats. The first passenger loses his/her ticket and thereby chooses a seat uniformly at random. For each successive passenger, if the seat corresponding to his/her ticket is vacant, he/she occupies that seat, else he/she chooses a vacant seat uniformly at random. What is the probability that the n^{th} passenger occupies the seat corresponding to his/her ticket?

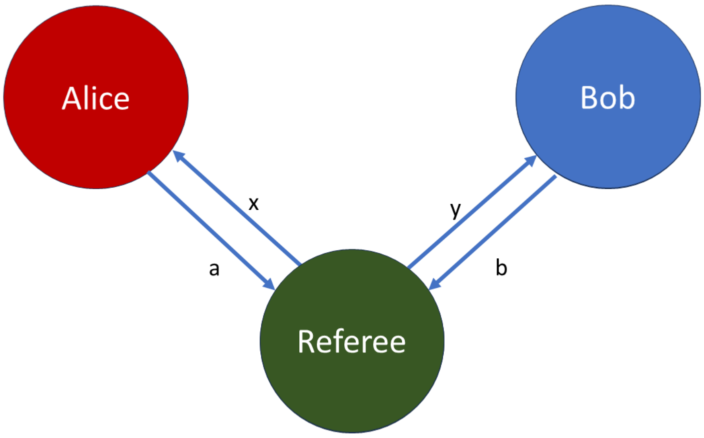
2.6. We play a game with a deck of 52 regular playing cards, of which 26 are red and 26 are black. I randomly shuffle the cards and place the deck face down on a table. You have the option of "taking" or "skipping" the top card. If you skip the top card, then that card is revealed and we continue playing with the remaining deck. If you take the top card, then the game ends; you win if the card you took was revealed to be black, and you lose if it was red. If we get to a point where there is only one card left in the deck, you must take it. Prove that you have no better strategy than to take the top card—which means your probability of winning is 1/2.

2.7. (♦) Consider a game played between Alice, Bob and the Referee where the Refree asks questions x to Alice and y to Bob. Alice and Bob reply with their respective answers a, b. Here $x, y, a, b \in \{0, 1\}$. Note that the questions may be sampled from any distribution by the Referee.

They win the game if

$$x \wedge y = a \oplus b$$

Alice and Bob are allowed to agree on a strategy before the game begins, but once it starts, they are *spatially separated* and cannot communicate.



Determine an optimal strategy for Alice and Bob to win the game. Provide a proof showing that this strategy is indeed optimal.

This is version 1.0 of the tutorial sheet. Let me know if something is unclear. In case of any doubt or for help regarding writing proofs, feel free to contact me or TAs.

Venkata Koppula - kvenkata@iitd.ac.in
Ananya Mathur - cs5200416@iitd.ac.in
Anish Banerjee - cs1210134@cse.iitd.ac.in
Eshan Jain - cs5200424@cse.iitd.ac.in
Mihir Kaskhedikar - cs1210551@iitd.ac.in
Naman Nirwan - Naman.Nirwan.cs521@cse.iitd.ac.in
Pravar Kataria - Pravar.Kataria.cs121@cse.iitd.ac.in
Shashwat Agrawal - csz248012@cse.iitd.ac.in
Subhankar Jana - csz248009@iitd.ac.in