CS442 **Practice Problems in PROB.** **Sept. 10, 2019**

Most of these problems are for your own practice and as a review of basic probability and counting. Try to work them out for your own benefit, or at least convince yourself that you know how to do them. Recitations are a good place to discuss them, and you may also do so between yourselves. Later, I might choose a few of them and ask you to write up and then submit good solutions as the first HW in the course.

1. Show $P(A \cap B) \geq 1 - P(A^c) - P(B^c)$.
2. Show that $A \cup B$ and C are independent if A, B, and C are mutually independent events.
3. Show that if A and B are independent, then also: (i) A and B^c are independent; (ii) A^c and B are indep.; (iii) A^c and B^c are indep..
4. Study the Monty Hall problem with $k > 3$ doors and show it is always better to switch, though the advantage over not switching decreases with k (Compute the probability of winning (i) when you switch and (ii) when you stick to the original choice.)
5. Describe a probability space (Ω, P) and events in it (the simpler, the better), where the following hold:
   (a) The events A_1, A_2, A_3 are not pairwise independent although A_1 and A_2 are independent and also A_1 and A_3 are independent.
   (b) In the family of events $A_1, \ldots, A_4$, A_1, A_2, A_3 are pairwise independent and A_1, A_2, A_4 are pairwise independent, but A_3 and A_4 are not independent.
   (c) Events $A_1, \ldots, A_4$ which are not 3-wise independent although A_1, A_2, A_3 are 3-wise ind., A_1, A_2, A_4 are 3-wise ind., and A_1, A_3, A_4 are 3-wise independent.
6. How many even numbers in $[100, 999]$ have distinct digits? How many palindromes (numbers that are the same when you write them backwards) are in this range? How about the range $[1000, 9999]$?
7. The numbers 1 through n are arranged in a random order (what is the sample space?). What is the probability that 1 and 2 are adjacent? That 1 and 3 are adjacent to 2? That none of these numbers are adjacent?
8. The lottery experiment is to pick 6 distinct numbers at random (the winning combination) from the first 55 integers.
   (a) What is the sample space Ω and what is $|\Omega|$, its size?
   (b) Before the experiment is performed, YOU write down 6 distinct numbers of your own on a card (your lottery ticket). What is the probability that your 6 numbers are the winning combination?
   (c) What is the probability that *none* of your numbers are in the winning combination? That exactly k of your 6 numbers are in the winning combination for each value $k = 1, \ldots, 5$.

9. In the hat experiment with $n = 6$ hats, what is the probability of the event A, that person 1 and person 2 exchange hats? of B, that person 3 and person 5 exchange hats? of C, that person 4 and person 6 exchange? What is the probability that NONE of these events occurs? That ALL occur?

10. A deck of cards is shuffled. What is the probability that no aces are adjacent? That no aces are within 2 cards of each other?

11. 4 cards from a 52 card deck are randomly dealt to each of 13 distinguishable players.

    (a) Describe the sample space and write down its size.

    (b) What is the probability that each player has one card from each suit?

    (c) What is the probability that *one* player has one card from each suit but that nobody else has cards from *more* than one suit?

12. Players A, B, C, and D toss a fair coin in order(i.e., first A then B, then C, then D, then A again, etc.) The first player to get a HEAD wins and the game is over. What are their respective chances to win? Repeat, now with only three players who toss *a pair* of coins; the first to get something different than HH wins the game

13. (*) The final step of the Floyd-Rivest randomized selection algorithm is to sort the (*random*) set $S = \{a_i \in \mathcal{A} : L \leq a_i \leq R\}$ of inputs which lie between sampled items L and R. Show that $P(|S| > 4n^{5/6}) \to 0$ as $n \to \infty$. (hints: look at the item $\sigma \in \mathcal{A}$ of appropriate rank and study the probability that $L < \sigma$; look at the item $\tau \in \mathcal{A}$ of appropriate rank and study the probability that $R > \tau$).

14. Let t_n be the (random) running time of the Floyd-Rivest selection algorithm with n inputs. We saw $P[t_n \leq 1.5n + o(n)] \to 1$ as $n \to \infty$. TRY TO GET a good bound on $E(t_n)$.

15. We do the balls and boxes experiment with r balls and n boxes. Let N denote the number of empty boxes, Y the number of boxes with exactly one ball, L_1 the load on (i.e., number of balls in) box 1, and L_2, the load on box 2.

    (a) Find the probability that $L_1 = k$, $k = 0, 1, \ldots, r$.

    (b) Find the expected value for each of these random variables, N, Y, L_1, L_2.

    (c) Are N and L_1 independent? Explain. Repeat for N and Y; for L_1 and Y; for L_1 and L_2.

    (d) If $r = n + 1$, find the probability that $N = 0$. If $r = n - 1$ find the probability that $N = 1$; that $N = 2$.

    (e) (*) What is the probability that ALL $L_i = 0$, $i = 1, \ldots, 5$? What is the probability that NO $L_i = 0$, $i = 1, \ldots, 5$?

16. The load balancing algorithm described in class randomly assigns each of r jobs (balls) to one of the n resources (boxes). Write a simple program that simulates the r balls in n boxes experiment for given values of r and n up to a million. It should keep track of the load in each box and output the maximum load. You can assume no load will exceed 15.

    (a) Perform at least 20 (and preferably 100 simulations with $r = n = 1,000,000$. Tabulate the observed max-loads.

(b) Here is another load-balancing algorithm: each ball chooses a *pair* of (distinct) boxes, and is placed in the chosen box with the smallest load; if the load is the SAME for both boxes in the chosen pair, the ball will choose one of them at random. Modify your program to implement this new scheme and again do at least 20 simulations, tabulating the max loads. Is it better? Is the result surprising?

(c) Repeat, now choosing three distinct boxes, and assigning that ball to the chosen box with the lightest load.

17. Markov's inequality states that if $X \geq 0$ is a random variable with $E(X) = \mu$, and if $t \geq 1$ is given, then

$$P(X \geq t\mu) \leq \frac{1}{t}.$$

Show that $X \geq 0$ is *necessary* for the conclusion to hold: i.e., describe a random variable (the simpler the better) with $E(X) = \mu$ but for which

$$P(X \geq t\mu) \; > \frac{1}{t},$$

for some $t \geq 1$.

18. Let N_r denote the number samples needed (*with* replacement) to collect r distinct coupons from a set T of n coupons. We showed in class that $E(N_n) = nH_n$, $H_n = 1 + 1/2 + \cdots + 1/n$, and $|H_n - \log_e n| \leq 1$.

(a) Compute the variance of N_n and show that it is equal to $V(N_n) = n^2 \sum_{i=1} \frac{1}{i^2} - n \sum_{i=1}^{n} \frac{1}{i}$

(b) Using the fact that $\sum_{i=1} \frac{1}{i^2} \to \pi^2/6$ show that

$$P(|N_n - nH_n| \geq cn) \leq \frac{\pi^2}{6c^2};$$

in other words N_n deviates from its mean by cn with probability that decreases like $1/c^2$. We will see later that in fact it decreases much more quickly.

(c) For the sake of comparison, use Markov's inequality to show that for $c > 1$,

$$P(|N_n - nH_n| \geq (c-1)nH_n) \leq \frac{1}{c}.$$

This says N_n has at most a $1/c$ chance to be $cn \log n$ above its mean. The previous statement says there is at most a $1/c^2$ chance to be <u>only</u> cn above the mean, a MUCH stronger and more precise statement.