Tutorial 1

Bay-Yuan Hsu

1. A box contains some number of white and black balls. When two balls are drawn without replacement, the probability that both are white is $\frac{1}{3}$.

   (a) Show that

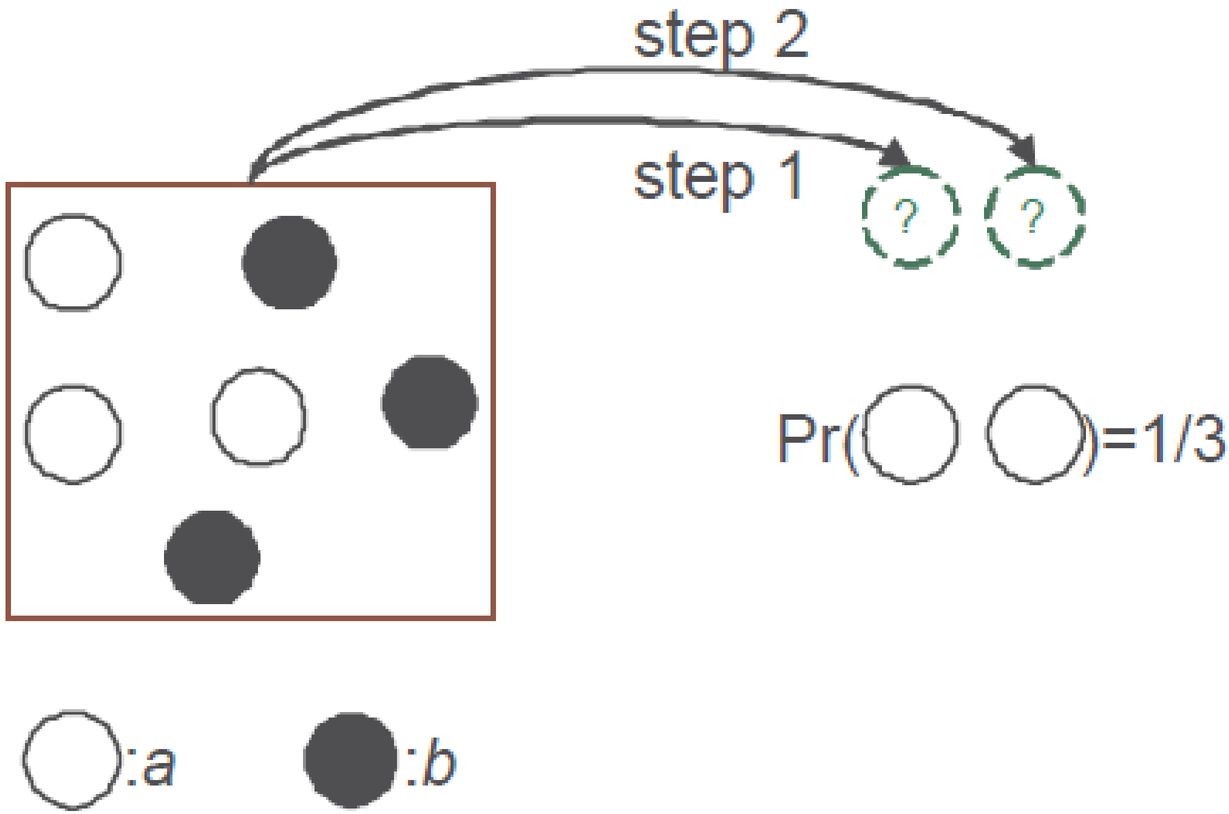
$$\left(\frac{a-1}{a+b-1}\right)^2 < \frac{1}{3} < \left(\frac{a}{a+b}\right)^2.$$

   (b) Show that

$$\frac{(\sqrt{3}+1)b}{2} < a < 1 + \frac{(\sqrt{3}+1)b}{2}.$$

   (c) Find the smallest number of balls in the box.

   (d) How small can the total number of balls be if black balls are even in number?



Hint: The probability we get a white ball at the first time is *a/a+b.*
Use conditional probability.

2. Consider the following balls-and-bin game. We start with one black ball and one white ball in a bin. We repeatedly do the following: Choose one ball from the bin uniformly at random, and then put the ball back in the bin with another ball of the same color. (That is, after each step, we have one more ball in the bin.) We repeat until there are n balls in the bin. Show that the number of white balls is equally likely to be any number between 1 and $n-1$.

Hint: Use induction

3. Suppose that a fair coin is flipped n times. For $k > 0$, show that the probability of obtaining a sequence of $\log_2 n + k$ consecutive heads is at most $1/2^k$.

Hint : You don’t need to give an exact probability

4. Let n be the number of vertices of a graph. Show that the number of distinct min-cut sets in the graph is at most $n(n-1)/2$.

Hint: [Textbook, Theorem 1.8]
The algorithm outputs a min-cut set with probability at least $2/n(n-1)$.

5. The following problem is known as the Monty Hall problem, after the host of the game show "Let's Make a Deal". There are three curtains. Behind one curtain is a new car, and behind the other two are goats. The game is played as follows. The contestant chooses the curtain that she thinks the car is behind. Monty then opens one of the other curtains to show a goat. (Monty may have more than one goat to choose from; in this case, assume he chooses which goat to show uniformly at random.) The contestant can then stay with the curtain she originally chose, or switch to the other unopened curtain. After that, the location of the car is revealed, and the contestant wins the car or the remaining goat. Should the contestant switch curtains or not, or does it make no difference?

Hint : 1. You can list all the cases and count the probability.

2. Try to think what if there are 100 curtains and Monty opens the rest 98 curtains?

6. (a) Show that if E_1 and E_2 are mutually independent, then so are $\overline{E_1}$ and $\overline{E_2}$.

(b) Show that if $E_1, E_2, \ldots, E_n$ are mutually independent, then so are $\overline{E_1}, \overline{E_2}, \ldots, \overline{E_n}$.

Hint : Think about the basic definition.

. Let X and Y be independent geometric random variables, where X has parameter p and Y has parameter q.

(a) What is the probability that $X = Y$?

(b) What is $\mathrm{E}[\max(X, Y)]$?

(c) What is $\Pr(\min(X, Y) = k)$?

(d) What is $\mathrm{E}[X \mid X \leq Y]$?

Hint: 1. Solve it by basic definition

2. Use memory-less property

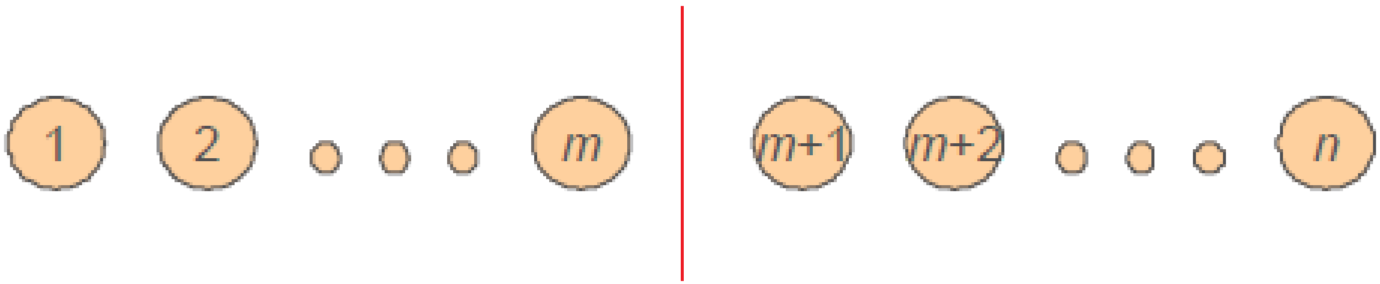
Advanced Question

You need a new staff assistant, and you have n people to interview. You want to hire the best candidate for this position. When you interview the candidates, you can give each of them a score, with the highest score will be the best and no ties being possible.

You interview the candidates one by one. Because of your company's hiring policy, after you interview the kth candidate, you either offer the candidate the job immediately, or you will forever lose the chance to hire that candidate.

We suppose that the candidates are interviewed in a random order, chosen uniformly at random from all $n!$ possible orderings.

1. First, we give everyone a number card. The number card means the order of interview.



2. Interview first m people and choose the best grade from them as A.
3. Interview follow the ordering.
   - If someone got a better score than A, accept him.
4. After interviewing all remaining candidates but no one get better score, we choose the last one.

(a) Let E be the event that we hire the best assistant, and let E_i be the event that ith candidate is the best and we hire him. Determine $\Pr(E_i)$, and show that

$$\Pr(E) = \frac{m}{n} \sum_{j=m+1}^{n} \frac{1}{j-1}.$$

(b) Bound $\sum_{j=m+1}^{n} \frac{1}{j-1}$ to obtain

$$\frac{m}{n}(\log_e n - \log_e m) \leq \Pr(E) \leq \frac{m}{n}(\log_e(n-1) - \log_e(m-1)).$$

(c) Show that $m(\log_e n - \log_e m)/n$ is maximized when $m = n/e$. Explain why this means $\Pr(E) \geq 1/e$ for this choice of m.

(a) **Ans.** If the ith candidate is the best of all candidates, we will choose him if and only if (1) $i > m$ and (2) the best of the first $i-1$ candidates (say y) is among the first m candidates (otherwise, if y is not among the first m candidates, we will choose y by our interview strategy).

Let B_i denote the event that ith candidate is the best, and Y_i denote the event that the best of first $i-1$ candidates is among the first m candidates. Then, we have

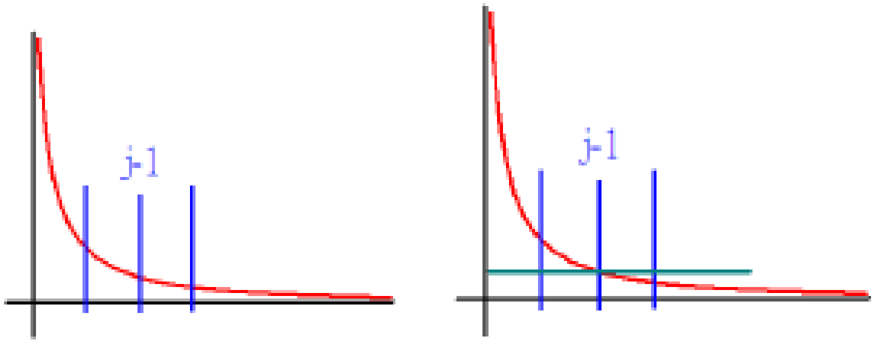
$$\Pr(E_i) = \begin{cases} 0 & \text{for } i \leq m \\ \Pr(B_i \cap Y_i) & \text{for } i > m \end{cases}$$

$$\Pr(B_i)\Pr(Y_i \mid B_i) = \frac{1}{n} \cdot \frac{m}{i-1} = \frac{m}{n} \cdot \frac{1}{i-1}.$$

From our definition, we can see that $\Pr(E) = \sum_{i=1}^{n} \Pr(E_i)$. Thus, we have

$$\Pr(E) = \sum_{i=1}^{n} \Pr(E_i) = \sum_{i=m+1}^{n} \Pr(E_i) = \frac{m}{n} \sum_{i=m+1}^{n} \frac{1}{i-1}.$$

Consider the curve $f(x) = 1/x$. The area under the curve from $x = j - 1$ to $x = j$ is less than $1/(j-1)$.



Thus,

$$\sum_{j=m+1}^{n} \frac{1}{j-1} \geq \int_{m}^{n} f(x)dx = \log_e n - \log_e m$$

$$\frac{m}{n}(\log_e n - \log_e m) \leq \Pr(E) \leq \frac{m}{n}(\log_e(n-1) - \log_e(m-1)).$$

Similarly, the area under the curve from $x = j - 2$ to $x = j - 1$ is greater than $1/(j-1)$. Thus,

$$\sum_{j=m+1}^{n} \frac{1}{j-1} \leq \int_{m-1}^{n-1} f(x)dx = \log_e(n-1) - \log_e(m-1)$$

Combining these two inequalities with part (a).

(c) **Ans.** Let $g(m) = m(\ln n - \ln m)/n$. By differentiating $g(m)$, we get

$$g'(m) = \frac{\ln n - \ln m}{n} - \frac{1}{n},$$

which is 0 when $m = n/e$. Also, if we differentiate $g'(m)$, we get

$$g''(m) = \frac{-1}{mn} < 0,$$

which indicates that $g(m)$ attains maximum when $m = n/e$.
By substituting $m = n/e$ in the inequality of part(b), we get

$$\Pr(E) \geq \frac{m(\ln n - \ln m)}{n} = \frac{n\,(\ln n - \ln(n/e))}{ne} = \frac{n\,(\log_e e)}{ne} = \frac{1}{e}.$$