Independence, Hashing and Bin Packing.

1. **Independence.** For each of the following examples, decide whether the listed events are mutually independent, pairwise independent, or neither (if the events are mutually independent, there is no need to also select pairwise independent).

   (a) The event of drawing a jack of hearts from the deck and the event of drawing a jack of clubs from the same deck.

   Recall the definition of independence–is the probability of both events equal to the product of the probabilities of each event? Should one event influence the probability of the other?

   - Mutually Independent
   - Only Pairwise Independent
   - Neither

   (b) The event of drawing a jack of hearts from the deck and the event of drawing an ace of diamonds from the same deck.

   Recall the definition independence–is the probability of both events equal to the product of the probabilities of each event? Should one event influence the probability of the other?

   - Mutually Independent
   - Only Pairwise Independent
   - Neither

   (c) The outcomes of three consecutive coinflips.

   Recall the definition of independence–is the probability of all three events equal to the product of the probabilities of each event? Should one event influence the probability of the others?

   - Mutually Independent
   - Only Pairwise Independent
   - Neither

   (d) Given 2 random integers x, y, the event that $x = 5 \mod n$, the event that $y = 7 \mod n$, and the event that $x + y = 20 \mod n$.

   Recall the definition of independence–is the probability of all events equal to the product of the probabilities of each event? Should one event influence the probability of the others?

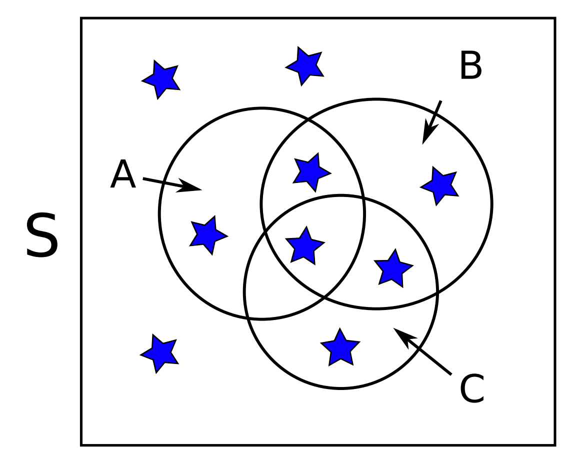
   - Mutually Independent
   - Only Pairwise Independent
   - Neither

2. **The Principle of Inclusion-Exclusion.** The Principle of Inclusion-Exclusion states that for events $A_1, \ldots, A_n$ in probability space S,

$$\mathbb{P}[\cup_{i=1}^n A_i] = \sum_{i=1}^n \mathbb{P}[A_i] - \sum_{\{i,j\}} \mathbb{P}[A_i \cap A_j] + \sum_{\{i,j,k\}} \mathbb{P}[A_i \cap A_j \cap A_k] - \cdots + (-1)^{n-1}\mathbb{P}[\cap_{i=1}^n A_i].$$

   That is, the probability of the union of events is the sum of the probabilities, minus the sum of the pairwise intersections (which were counted twice), plus the sum of the 3-way intersections (which were subtracted with the pairwise intersections), etc.

   Consider the picture below–this depicts a probability space S, where each of the small stars is one outcome.



   (a) How many outcomes are there in S in total?

   \# outcomes =

   Each outcome is represented as a star–just count the number of stars.

   (b) How many outcomes are in the union $A \cup B \cup C$?

   \# outcomes =

   Each outcome is represented as a star–just count the number of stars in A, B, and C.

   (c) What is $\mathbb{P}[A \cup B \cup C]$? Please enter your answer as a fully reduced fraction (i.e. in the form x/y where x, y are the smallest possible integers they can be).

   $\mathbb{P}[A \cup B \cup C] =$

   Review the definition of probability in a discrete space.

   (d) Now, we will calculate this probability using the Principle of Inclusion-Exclusion. First, what is $\mathbb{P}[A] + \mathbb{P}[B] + \mathbb{P}[C]$? Please enter your answer as a fully reduced fraction (i.e. in the form x/y where x, y are the smallest possible integers they can be).

   $\mathbb{P}[A] + \mathbb{P}[B] + \mathbb{P}[C] =$

   Review the definition of probability in a discrete space–how many outcomes fall in each of the events A, B, C? Take a sum of each of the individual probabilities.

   (e) Notice that the above was a severe overestimate of $\mathbb{P}[A \cup B \cup C]$. This is because all of the two-way intersections were counted twice, once for each individual event.

   Now, calculate the sum of probabilities of the intersections, $\mathbb{P}[A \cap B] + \mathbb{P}[B \cap C] + \mathbb{P}[C \cap A]$. Please enter your answer as a fully reduced fraction (i.e. in the form x/y where x, y are the smallest possible integers they can be).

   $\mathbb{P}[A \cap B] + \mathbb{P}[B \cap C] + \mathbb{P}[C \cap A] =$

   Review the definition of probability in a discrete space–how many outcomes fall in each of the two-way intersections of the events? Take a sum of each of the individual probabilities.

   (f) Now, subtract the two-way intersections from the sum of the individual probabilities. What is $\mathbb{P}[A] + \mathbb{P}[B] + \mathbb{P}[C] - (\mathbb{P}[A \cap B] + \mathbb{P}[B \cap C] + \mathbb{P}[C \cap A])$?. Please enter your answer as a fully reduced fraction (i.e. in the form x/y where x, y are the smallest possible integers they can be).

   $\mathbb{P}[A] + \mathbb{P}[B] + \mathbb{P}[C] - (\mathbb{P}[A \cap B] + \mathbb{P}[B \cap C] + \mathbb{P}[C \cap A]) =$

   Simply subtract your answer from part (e) from your answer in part (d), and ensure that your input is of the proper form.

   (g) Finally, notice that this probability is slightly less than $\mathbb{P}[A \cup B \cup C]$. This is because the 3-way intersection was originally added 3 times, but then subtracted 3 times, which is once too many.

   What is $\mathbb{P}[A \cap B \cap C]$? Please enter your answer as a fully reduced fraction (i.e. in the form x/y where x, y are the smallest possible integers they can be).

   $\mathbb{P}[A \cap B \cap C] =$

   Review the definition of probability in a discrete space–how many outcomes fall in $A \cap B \cap C$?

   (h) Now, add back $\mathbb{P}[A \cap B \cap C]$ to your answer in the part (f). Does this probability match the probability you calculated in part (c)?

   - Yes
   - No

3. **Balls and Bins.** Suppose you have m labeled balls $a_1, \ldots, a_m$ that you have thrown one by one, uniformly at random into n labeled bins $b_1, \ldots, b_n$. For each of the probabilities below, select all answers that apply from the list of choices.

   (a) What is the probability that b_j contains a_i? Select all answers that apply.

   When a_i is being thrown, what is the probability that it lands in b_j? How many choices are there?

   i. $\left(1 - \frac{1}{n}\right)^m$
   ii. $\frac{1}{n}$
   iii. $\frac{1}{m}$
   iv. $\frac{m}{n}$

   (b) What is the probability that b_j is empty? Select all answers that apply.

   For each ball, what is the probability that it does not end up in b_j? Are these events independent?

   i. $\left(1 - \frac{1}{n}\right)^m$
   ii. $\binom{m}{0} \cdot \left(1 - \frac{1}{n}\right)^m$
   iii. $\frac{1}{m!}$
   iv. $\frac{1}{n!}$

   (c) What is the probability that b_j contains all of the balls? Select all answers that apply.

   For each ball, what is the probability that it ends up in b_j? Are these events independent?

   i. $\left(1 - \frac{1}{n}\right)^m$
   ii. $\left(\frac{1}{n}\right)^m$
   iii. $\binom{m}{m} \cdot \left(\frac{1}{n}\right)^m$
   iv. $\frac{1}{m!}$

   (d) What is the probability that b_j contains exactly k balls? Select all answers that apply.

   For each ball, what is the probability that it ends up in b_j? Are these events independent? Does order matter?

   i. $\binom{m}{k} \cdot \left(\frac{1}{n}\right)^{m-k} \cdot \left(1 - \frac{1}{n}\right)^k$
   ii. $\left(\frac{1}{n}\right)^k \cdot \left(1 - \frac{1}{n}\right)^{m-k}$
   iii. $\binom{m}{k} \cdot \left(\frac{1}{n}\right)^k \cdot \left(1 - \frac{1}{n}\right)^{m-k}$
   iv. $\left(\frac{1}{n}\right)^k$

   (e) What is the probability that b_j contains at most k balls?

   What is the probability that b_j contains exactly i balls? Review the probability of a union of disjoint events in note 10.

   i. $\sum_{i=0}^k \binom{m}{i} \cdot \left(\frac{1}{n}\right)^{m-i} \cdot \left(1 - \frac{1}{n}\right)^i$
   ii. $\sum_{i=0}^k \binom{m}{i} \cdot \left(\frac{1}{n}\right)^i \cdot \left(1 - \frac{1}{n}\right)^{m-i}$
   iii. $1 - \sum_{i=k+1}^m \binom{m}{i} \cdot \left(\frac{1}{n}\right)^i \cdot \left(1 - \frac{1}{n}\right)^{m-i}$
   iv. $1 - \sum_{i=0}^{m-k} \binom{m}{i} \cdot \left(\frac{1}{n}\right)^{m-i} \cdot \left(1 - \frac{1}{n}\right)^i$

4. **Processes, servers and overloading.** I have M processes (jobs) and N servers that I can assign the jobs to. Any job may be assigned to any server. Suppose I assign each job to a randomly chosen server, with all servers being equally likely. We say that a server is overloaded if it is assigned greater than or equal to K jobs, where $K \leq M$. What is the probability that the first server is overloaded?

   (a) $\sum_{i=0}^{M-K-1} \binom{M}{K+i} \frac{(N-1)^{M-K-i}}{N^M}$

   (b) $\sum_{i=0}^{N-K} \binom{N}{K+i} \frac{(M-1)^{N-K-i}}{M^N}$

   (c) $\sum_{i=0}^{M-K} \binom{M}{K+i} \frac{(N-1)^{M-K-i}}{N^M}$

   (d) $\binom{M}{K} \frac{(N-1)^{M-K}}{N^M}$

   (e) $\left(\frac{N-1}{N}\right)^{M-K}$

5. **Jobs and servers, without immediate repetition.** I have $M \geq 2$ jobs and $N \geq M$ servers to assign the jobs to.

   I use the following system to assign the jobs: the first job is randomly assigned to one of the N servers (with all servers being equally likely). The second job is again assigned randomly to a server, except that the server that got the first job is excluded from the selection (any of the other $N - 1$ servers are equally likely to get the job). Similarly, the third job is assigned randomly, except this time, the server that got the second job is excluded (any of the other $N - 1$ servers have equal likelihood of being chosen for this job). And so on until all M jobs have been assigned.

   What is the probability that each of the M jobs goes to a different server?

   (a) $\sum_{i=0}^{N-M} \frac{(N-i-2)!}{(N-M-i)!(N-1)^{M-2}}$

   (b) $\frac{(M-2)!}{(N-M)!(M-2)^{N-1}}$

   (c) $\frac{(N-1)!}{(N-M)!N^{M-1}}$

   (d) $\frac{(N-2)!}{(N-M)!(N-2)^{M-3}}$

   (e) $\frac{(N-2)!}{(N-M)!(N-1)^{M-2}}$

6. **Bounding overload probabilities.** I have a load balancing set up with N servers, where I can guarantee that the probability of the i^{th} server being overloaded (where $1 \leq i \leq N$) is at most p, which is independent of i. Which of the following is true? Check all that apply.

   (a) Pr (no server is overloaded) $\leq 1 - Np$

   (b) Pr (no server is overloaded) $\geq 1 - Np$

   (c) Pr (at least one server is overloaded) $\geq p$

   (d) Pr (no server is overloaded) $\leq 1 - N(1-p)$

   (e) Pr (no server is overloaded) $\geq 1 - (1-p)^{2N}$

   (f) Pr (at least one server is overloaded) $\leq Np$