This is a take home midterm. You can use your notes, my online notes on canvas and the textbook. You are supposed to work on your own text without external help. I'll be available to answer question in person or via email. Please, write clearly and legibly and take a readable scan before uploading.

To solve the Exam problems, I have not collaborated with anyone nor sought external help and the material presented is the result of my own work.

Name (print): ______________________________

Question:	1	2	3	4	5	6	7	Total
Points:	20	10	10	20	10	20	10	100
Score:								

Question 1 . . . . *20 point*

In a drawer there are 15 spoons. Every day you pick three of them at random and use them. Afterward you clean them and put them back in the drawer.

(a) (10 points) Find the probability p that after 5 days one particular spoon has never been used.

Solution: For a given particular spoon, the probability p_1 of being selected one day is

$$p_1 = \frac{3}{15} = \frac{1}{5}$$

so that the probability that the spoon has never been used is

$$p = \left(1 - \frac{1}{5}\right)^5 = \left(\frac{4}{5}\right)^5$$

(b) (10 points) Find the probability q that after 5 days there is a spoon that has never been used.

Solution: Since in total you selected 15 spoons, the probability that they have all been used is the probability that you used each of them exactly once.

Taking into account the order in which the spoons are selected every day, the number of outcomes where all spoon are used is 15! while the total number of outcomes is $(15 \cdot 14 \cdot 13)^5$. Thus the probability is

$$q = 1 - \frac{15!}{(15 \cdot 14 \cdot 13)^5}$$

Question 2 .... *10 point*

Let X_i, $i = 1, \dots, N$ be a i.i.d family of discrete r.v. that assume only a finite number of values $x_0 < x_1 < \dots < x_k$. Let $Z = \max_{i\leq N} X_i$ and $T = \min_{i\leq N} X_i$. Show that, if $\mathbb{P}(X_i = x_0)\mathbb{P}(X_i = x_k) > 0$ then

$$\lim_{N\to\infty} \mathbb{P}(Z = x_k \,\&\, T = x_0) = 1$$

Solution: Since x_k is the maximum possible value for the X_i we have

$$\mathbb{P}(Z < x_k) = \mathbb{P}(X_1 < x_k)^N$$

where we have used independence of the X_i. Moreover we have

$$\mathbb{P}(X_1 < x_k) = 1 - \mathbb{P}(X_1 = x_k) < 1$$

from which we get

$$\lim_{N\to\infty} \mathbb{P}(Z < x_k) = 0.$$

Similarly we get

$$\lim_{N\to\infty} \mathbb{P}(T > x_0) = 0.$$

Finally we have

$$\begin{aligned}\mathbb{P}(Z = x_k \,\&\, T = x_0) =& 1 - \mathbb{P}(Z < x_k \mid T > x_0) \\ =& 1 - \mathbb{P}(Z < x_k) - \mathbb{P}(T > x_0) + \mathbb{P}(Z < x_k \,\&\, T > x_0) \\ \geq& 1 - \mathbb{P}(Z < x_k) - \mathbb{P}(T > x_0)\end{aligned}$$

so that

$$\lim_{N\to\infty} \mathbb{P}(Z = x_k \,\&\, T = x_0) \geq 1 - \lim_{N\to\infty} \mathbb{P}(Z < x_k) - \lim_{N\to\infty} \mathbb{P}(T > x_0) = 1.$$

Question 3 .... *10 point*

You want to test for Covid-19 a large population of N individuals. You collect a sample from each and then decide to proceed as follows. You randomly divide the sample in M/k groups of k samples each. For each group, you take a small part of each of the k samples and you mix this k parts together. Then you test the mixture. If the mixture is negative it means that all k samples where negative. If the mixture is positive you must test all the k samples individually since at least one is infected.

Assuming that in the population the rate of infection is 0.05, compute the expected number of test you will perform and find the optimal k to minimize the number of tests.

Solution: For each group, the number of tests performed is 1 with probability $(1-p)^k$ or $k+1$ with probability $1-(1-p)^k$. Thus the expected value of the number of test performed for each group is

$$e = q^k + (k+1)(1-q^k) = 1 + k(1-q^k)$$

and the total number of test performed is

$$E = N\left(\frac{1}{k} + (1-q^k)\right) := N\epsilon(k)$$

We thus need the minimum of $\epsilon(k)$ for k integer and positive. The value of $\epsilon(k)$ for $k = 1, \ldots, 8$ are 1.0500, 0.5975, 0.4760 0.4355, 0.4262, 0.4316, 0.4445, 0.4616. Thus it looks like the best k is 5.

To mathematically prove that $k = 5$ is optimal we need to show that $\epsilon(k) > \epsilon(4)$ for every $k \neq 5$. There are several way to show this. Below is an elementary one.

Observe that for $\epsilon(90) = 1.0012 > 1$ while $\epsilon(k) > \epsilon(5)$ for $6 \leq k \leq 100$.

If $k/(k+1) > q$ and $\epsilon(k) > 1$, that is $1/k - q^k > 0$, we have

$$\epsilon(k+1) = \frac{1}{k+1} + 1 - q^{k+1} > \frac{k}{k+1}\frac{1}{k} - qq^k + 1 > q\left(\frac{1}{k} - q^k\right) + 1 > 1$$

Since $99/100 > 0.95$ we know that $\epsilon(k) > 1 > \epsilon(4)$ for $k > 36$.

In grading I will take out 1pt if you did not realize that you need an argument to show that 4 is a global minimum. Any reasonable attempt to prove that the minimum is global will be considered as a bonus.

Question 4 . . . . *20 point*

A factory produces 1000 computers. Each computer has a probability $p = 0.05$ of having a defect. The factory has a quality control department. If a computer is defective it will be detected and discarded with probability 1. If a computer is not defetective it will be discarded with a probability $s = 0.03$.

(a) (10 points) Compute the probability that a randomly selected computer will be discarded by the quality control department and the probability that a discarded computer is actually defective. (**Hint**: Call A the event $\{$*computer is defective*$\}$ and B the event $\{$*computer is discarded*$\}$. You are asked to find $\mathbb{P}(B)$ and $\mathbb{P}(A|B)$.)

Solution: Call A the event $\{$*computer is defective*$\}$ and B the event $\{$*computer is discarded*$\}$. Then we want to find $\mathbb{P}(B)$ and $\mathbb{P}(A|B)$. We have

$$\mathbb{P}(B) = \mathbb{P}(B|A)\mathbb{P}(A) + \mathbb{P}(B|A')\mathbb{P}(A') = 1 \cdot 0.05 + 0.03 \cdot 0.95 = 0.079$$

while

$$\mathbb{P}(A|B) = \mathbb{P}(B|A)\frac{\mathbb{P}(A)}{\mathbb{P}(B)} = 1\frac{0.05}{0.079} = 0.64$$

(b) (10 points) Call Y the r.v. that describes the number of discarded computers that are working. Write the p.m.f. of Y. Using the Poisson approximation, compute the probability that $Y = 20$ and the expected value of Y, $E(Y)$.

Solution:

Clearly Y is a Binomial r.v.. with parameters 1000 and p, where p is the probability that a randomly selected computer is both working and discarded. This is given by

$$\mathbb{P}(A' \cap B) = \mathbb{P}(A')\mathbb{P}(B|A') = 0.95 \cdot 0.03 = 0.0285$$

It follows that

$$p_Y(y) = \binom{1000}{y} 0.9715^{1000-y} 0.0285^y.$$

Since $1000 * 0.0285 = 28.5$ we have

$$\mathbb{P}(Y = 20) \simeq e^{-28.5}\frac{28.5^{20}}{20!} = 0.0215$$

while

$$\mathbb{E}(Y) = 28.5.$$

Question 5 .... *10 point*

(a) Let X_1 and X_2 be two Bernoulli r.v. with $\mathbb{E}(X_1) = p_1$ and $\mathbb{E}(X_2) = p_2$. Show that if $\text{Cov}(X_1, X_2) = 0$ then X_1 and X_2 are independent.

Solution: Observe that $\mathbb{E}(X_i) = \mathbb{P}(X_i = 1)$ so that

$$\text{Cov}(X_1, X_2) = \mathbb{P}(X_1 = 1 \,\&\, X_2 = 1) - \mathbb{P}(X_1 = 1)\mathbb{P}(X_2 = 1)$$

from which we have

$$\mathbb{P}(X_1 = 1 \,\&\, X_2 = 1) = \mathbb{P}(X_1 = 1)\mathbb{P}(X_2 = 1)$$

Since

$$\mathbb{P}(X_1 = 1 \,\&\, X_2 = 0) = \mathbb{P}(X_1 = 1) - \mathbb{P}(X_1 = 1 \,\&\, X_2 = 1)$$

we get

$$\mathbb{P}(X_1 = 1 \,\&\, X_2 = 0) = \mathbb{P}(X_1 = 1)\mathbb{P}(X_2 = 0)$$

and similarly

$$\mathbb{P}(X_1 = 0 \,\&\, X_2 = 1) = \mathbb{P}(X_1 = 0)\mathbb{P}(X_2 = 1)$$
$$\mathbb{P}(X_1 = 0 \,\&\, X_2 = 0) = \mathbb{P}(X_1 = 0)\mathbb{P}(X_2 = 0)$$

so that X_1 and X_2 are independent.

(b) (10 points) Let now X_1, X_2 and X_3 be three Bernoulli r.v. with $\mathbb{E}(X_i) = p_i$. Assume that $\text{Cov}(X_i, X_j) = 0$ for $i < j = 1, 2, 3$. Is it true that X_1, X_2 and X_3 are independent? Proof it or find a counterexample.

Solution: Let Y_1 and Y_2 be two independent r.v. that take value 1 with probability 0.5 and -1 with probability 0.5. Moreover let $Y_3 = Y_1 Y_2$. It is easy to see that $\mathbb{P}(Y_3 = 0) = 0.5$ and $\text{Cov}(Y_i, Y_j) = 0$ for $i < j = 1, 2, 3$. Clearly $\mathbb{P}(Y_3 = -1 \,|\, Y_2 = 1 \& Y_2 = 1) = 0$ so that Y_1, Y_2, and Y_3 are not independent.

Call $X_i = (Y_i + 1)/2$. Then the X_i are Bernoulli r.v. with $\text{Cov}(X_i, X_j) = 0$ for $i < j = 1, 2, 3$ but they are not independent.

Question 6 .... *20 point*

In a bowl there are 10 balls. Each ball has a probability $p = 0.7$ of being red and a probability $q = 0.3$ of being blue. Let N_r be the number of red balls and N_b be the number of blue balls.

(a) (10 points) Compute the probability that N_b is bigger than N_r that is $\mathbb{P}(N_b > N_r)$.

Solution:

$$\mathbb{P}(N_b > N_r) = \mathbb{P}(N_b > 5) = \sum_{i=6}^{10} \binom{10}{i} 0.3^i 0.7^{10-i} = 0.047$$

(b) (10 points) You cannot see the content of the bowl. Someone extract a ball and show it to you and then reinsert it back in the bowl. Compute the conditional probability that N_b is grater that N_r given that the extracted ball was blue.(**Hint**: find first the conditional probability that there are k blue balls given that the extracted ball was blue.)

Solution: We can compute the conditional probability of $N_b = k$ given that the extracted ball was blue using Bayes theorem:

$$\mathbb{P}(N_b = k \,|\, \text{blue}) = \frac{\mathbb{P}(\text{blue} \,|\, N_b = k)\mathbb{P}(N_b = k)}{\mathbb{P}(\text{blue})}$$

where

$$\mathbb{P}(\text{blue}) = \sum_{i=0}^{10} \mathbb{P}(\text{blue} \,|\, N_b = i)\mathbb{P}(N_b = i) = \sum_{i=0}^{10} \frac{i}{10} \binom{10}{i} 0.3^i 0.7^{10-i} = 0.3$$

so that for $k > 0$

$$\mathbb{P}(N_b = k \,|\, \text{blue}) = \frac{1}{0.3} \frac{k}{10} \binom{10}{k} 0.3^k 0.7^{10-k} = \binom{9}{k-1} 0.3^{k-1} 0.7^{9-(k-1)}$$

while for $k = 0$ we have $\mathbb{P}(N_b = 0 \,|\, \text{blue}) = 0$. Finally

$$\mathbb{P}(N_b > N_r) = \mathbb{P}(N_b > 5) = \sum_{i=6}^{10} \binom{9}{i-1} 0.3^{i-1} 0.6^{9-(i-1)} = 0.099\,.$$

Alternatively it was enough to observe that once you saw a blue ball, you know that the number M_b of blue ball left in the bowl after the extraction, but before reinsertion, is a binomial with parameter 9 and 0.3. Thus we have

$$\mathbb{P}(N_b = k \,|\, \text{blue}) = \mathbb{P}(M_b = k - 1) = \binom{9}{k-1} 0.3^{k-1} 0.7^{9-(k-1)}\,.$$

Question 7 .... *10 point*

Let X_i be independent geometric r.v. with parameter p and N a further geometric r.v. with parameter P independent from the X_i. Compute the p.m.f. of

$$Z = \sum_{i=1}^{N} X_i.$$

Solution: The p.g.f.of the X_i and N are

$$G_{X_i}(s) = \frac{ps}{1 - qs} \qquad\qquad G_N(s) = \frac{Ps}{1 - Qs}$$

so that

$$G_Z(s) = \frac{Pps}{1 - qs}\frac{1 - qs}{1 - qs - Qps} = \frac{pPs}{1 - (1 - Pp)s}$$

so that Z is a geometric r.v. with parameter Pp and

$$\mathbb{P}(Z = z) = (1 - Pp)^{z-1}Pp$$

Equivalently it was enough to observe that each X_i may be thought of as flipping a coin with probability of giving H equal to p and waiting for the first H. If I have a second coin with probability of H equal to P, I can think of Z as flipping the first coin till I get a H and, in that moment, flipping the second coin. If the second coin give H I stop, if not I go on and repeat the procedure. This is in turn equivalent to flipping both coins together till I get H on both coins. That is a geometric r.v. with parameter Pp.