**This is a take home final exam. You can use your notes, my online notes on canvas and the text book. 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.

Signature: ______________________________

Name (print): ______________________________

| Question: | 1 | 2 | 3 | 4 | 5 | 6 | 7 | Total |
|---|---|---|---|---|---|---|---|---|
| Points: | 30 | 15 | 25 | 20 | 10 | 0 | 0 | 100 |
| Score: | | | | | | | | |

| Question: | 1 | 2 | 3 | 4 | 5 | 6 | 7 | Total |
|---|---|---|---|---|---|---|---|---|
| Bonus Points: | 0 | 0 | 0 | 0 | 0 | 15 | 15 | 30 |
| Score: | | | | | | | | |

Question 1 . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . *30 point*

Let $X$ be a continuous r.v. with p.d.f. given by

$$f(x) = \begin{cases} p + 2(1-p)x & 0 < x < 1 \\ 0 & \text{otherwise} \end{cases}$$

where $p$ is a parameter.

(a) (10 points) for which value of $p$ is $f$ a valid p.d.f.? (**Hint**: remember that there are 2 conditions you should check.)

**Solution:** We have

$$\int_0^1 f(x)dx = 1$$

for every $p$ so that we just need to check that $f(x) > 0$. Since $f$ is linear in $x$ it is enough that $f(0) > 0$ and $f(1) > 0$. Thus we get

$$0 \le p \le 2\,.$$

(b) (10 points) Compute the expected value $\mathbb{E}(X)$ and the variance $\mathrm{var}(X)$ of $X$.

**Solution:** We have

$$\mathbb{E}(X) = \int_0^1 xf(x)dx = p\int_0^1 xdx + 2(1-p)\int_0^1 x^2dx = \frac{p}{2} + \frac{2}{3}(1-p) = \frac{2}{3} - \frac{p}{6}$$

and

$$\mathbb{E}(X^2) = \int_0^1 x^2f(x)dx = p\int_0^1 x^2dx + 2(1-p)\int_0^1 x^3dx = \frac{p}{3} + \frac{1}{2}(1-p) = \frac{1}{2} - \frac{p}{6}$$

so that

$$\mathrm{var}(X) = \frac{1}{36}\left(p\left(2-p\right) + 2\right)$$

(c) (10 points) Show that

$$\mathbb{E}\left(\left(X-\frac{2}{3}\right)^2\right) \geq \left(\frac{p}{6}\right)^2$$

(**Hint**: use Jensen inequality.)

**Solution:** Use Jensen inequality to get

$$\mathbb{E}\left(\left(X-\frac{2}{3}\right)^2\right) \geq \left(\mathbb{E}(X)-\frac{2}{3}\right)^2 = \left(\frac{1}{2}-\frac{p}{6}-\frac{1}{2}\right)^2 .$$

Question 2 . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . *15 point*

Let $N_k$, $k = 1, 2, 3, \ldots$, be an infinite sequence of geometric random variable with parameter $p_k = \frac{\lambda}{k}$, that is

$$\mathbb{P}(N_k = n) = (1 - p_k)^{n-1} p_k \qquad \text{for } n \geq 1\,.$$

and $\mathbb{P}(N_k = n) = 0$ for $n < 1$. Moreover let $Y$ be an exponential r.v. with parameter $\lambda$, that is

$$f_Y(y) = \lambda e^{-\lambda y} \qquad \text{for } y \geq 0$$

and $f_Y(y) = 0$ for $y < 0$.

Show that $Z_k = N_k/k$ converge in distribution to $Y$ as $k \to \infty$. (**Hint**: compute the c.d.f. of $Z_k$, that is $F_k(x) = \mathbb{P}(Z_k \leq x)$ for every real number $x$.)

**Solution:** Observe that we have

$$\mathbb{P}(Z_k \leq x) = \mathbb{P}(N_k \leq kx) = \sum_{i=1}^{\lfloor kx \rfloor} (1 - p_k)^{i-1} p_k = 1 - (1 - p_k)^{\lfloor kx \rfloor - 1} = \left(1 - \frac{\lambda}{k}\right)^{\lfloor kx \rfloor - 1}$$

while

$$\mathbb{P}(Y \leq y) = 1 - e^{-\lambda y}.$$

Observe that

$$\frac{\lfloor kx \rfloor - 1}{k} \to x$$

as $k \to \infty$ so that

$$\left(1 - \frac{\lambda}{k}\right)^{\lfloor kx \rfloor - 1} = \left(\left(1 - \frac{\lambda}{k}\right)^k\right)^{\frac{\lfloor kx \rfloor - 1}{k}} \to_{k \to \infty} e^{-\lambda x}$$

for every $x$.

Question 3 ........................................................................ *25 point*

A student is attempting a multiple choices exam. For each question there are 4 possible answers. He has a probability of 0.75 of knowing the correct answer. If he does not know the answer he chooses one answer uniformly and randomly. All questions and answers are independent.

To get a B he need to answer correctly 85% of the questions while to get an A he needs to answer correctly 95% of the questions.

To answer the questions below, you can use Matlab or R (or any other software) to compute the needed values c.d.f. of a Standard Normal r.v.. In case you do not have them available, this is an online calculator.

(a) (10 points) If the test contains 40 questions, use a normal approximation (CLT) to compute the probability $p_B$ that the student will get at least a B and the probability $p_A$ that the student will get a A.

**Solution:** Let $p$ be the probability that the student give a correct answer. We have

$$p = 0.75 + 0.25 \cdot 0.25 = 0.8125\,.$$

Let $X_i$ be 1 if he answer correctly to the $i$-th question and 0 otherwise. Thus $\mathbb{E}(X_i) = 0.8125$ and $V(X_i) = 0.1523$. We get

$$\begin{aligned} p_B =& \mathbb{P}\left(\sum_{i=1}^{40} X_i > 0.85 \cdot 40\right) = \\ =& \mathbb{P}\left(\frac{\sum_{i=1}^{40} X_i - 0.8125 \cdot 40}{0.390\sqrt{40}} > \frac{(0.85 - 0.8125) \cdot 40}{0.390\sqrt{40}}\right) = \\ =& 1 - \Phi(0.60) = 0.274 \end{aligned}$$

while

$$\begin{aligned} p_B =& \mathbb{P}\left(\sum_{i=1}^{40} X_i > 0.95 \cdot 40\right) = \\ =& \mathbb{P}\left(\frac{\sum_{i=1}^{40} X_i - 0.8125 \cdot 40}{0.390\sqrt{40}} > \frac{(0.95 - 0.8125) \cdot 40}{0.390\sqrt{40}}\right) = \\ =& 1 - \Phi(2.23) = 0.013\,. \end{aligned}$$

(b) (15 points) Let $p_B$ be the probability that a student that knows 75% of the answers will get a B or more. If the teacher wants $p_B$ to be less than 0.025, how many questions should there be on the exam.

**Solution:** He wants to find $N$ such that

$$\mathbb{P}\left(\sum_{i=1}^{N} X_i > 0.85 \cdot N\right) \leq 0.025$$

This means

$$\mathbb{P}\left(\frac{\sum_{i=1}^{N} X_i - 0.8125 \cdot N}{0.390\sqrt{N}} > \frac{(0.85 - 0.8125) \cdot \sqrt{N}}{0.390}\right) = 1-\Phi(0.096\cdot\sqrt{N}) \leq 0.025$$

From the table we

$$\Phi(1.96) = 0.975$$

so that he needs

$$N > \left(\frac{1.96}{0.096}\right)^2 = 416$$

questions.

Question 4 ........................................................................ *20 point*

Let $X$ be a continuous r.v. with uniform distribution in $[-1, 1]$ and $Y$ be such that $\mathbb{P}(Y = 1) = \mathbb{P}(Y = -1) = 0.5$, $X$ and $Y$ independent. Consider the r.v. $Z = XY$.

(a) (10 points) Find the p.d.f. of $Z$.

**Solution:** Observe that

$$\mathbb{P}(Z < z) = \mathbb{P}(X < z)\mathbb{P}(Y = 1) + \mathbb{P}(X > -z)\mathbb{P}(Y = -1)\,.$$

Observing that $\mathbb{P}(X < z) = \mathbb{P}(X > -z)$ we get $\mathbb{P}(Z < z) = \mathbb{P}(X < z)$, that is $Z$ is uniform in $[-1, 1]$ and

$$f_Z(z) = \begin{cases} \frac{1}{2} & -1 \leq z \leq 1 \\ 0 & \text{otherwise} \end{cases}$$

(b) (10 points) Are $Z$ and $Y$ independent?

**Solution:** For any $A \subset [-1, 1]$ we have

$$\mathbb{P}(Z \in A \,\&\, Y = 1) = \mathbb{P}(X \in A)\mathbb{P}(Y = 1) = \mathbb{P}(Z \in A)\mathbb{P}(Y = 1)$$

while

$$\mathbb{P}(Z \in A \,\&\, Y = -1) = \mathbb{P}(X \in -A)\mathbb{P}(Y = 1) = \mathbb{P}(X \in A)\mathbb{P}(Y = -1) = \mathbb{P}(Z \in A)\mathbb{P}(Y = -1)$$

so that $Z$ and $Y$ are independent.

Question 5 ........................................................................ *10 point*

Let $X_1$ and $X_2$ be two independent exponential r.v. with expected value $\lambda$. Find the joint p.d.f. of $Y_1 = X_1 + X_2$ and $Y_2 = X_1 - X_2$ and the marginal p.d.f. of $Y_2$. (**Hint**: pay attention to the possible values of $Y_1$ and $Y_2$)

**Solution:** The p.d.f. of $X_1$ and $X_2$ is

$$f(x) = \lambda^{-1} e^{-\frac{x}{\lambda}} .$$

Note: if students used $\lambda$ as the parameter instead of $\lambda^{-1}$, do not penalize them.

Clearly we have

$$X_1 = (Y_1 + Y_2)/2 \qquad X_1 = (Y_1 - Y_2)/2$$

and the Jacobian of the change of variables id

$$|J(y_1, y_2)| = \frac{1}{2}$$

and we need $y_1 + y_2 > 0$ and $y_1 - y_2 > 0$, so that we get

$$f_{Y_1,Y_2}(y_1, y_2) = \frac{1}{2}\lambda^{-2} e^{-\frac{y_1}{\lambda}} \qquad \text{for } y_1 > 0,\ |y_2| < y_1$$

To find the marginal p.d.f. of $Y_1$ we compute

$$f_{Y_2}(y_2) = \frac{1}{2}\int_{|y_2|}^{\infty} \lambda^{-2} e^{-\frac{y_1}{\lambda}}\, dy_1 = \frac{1}{2}\lambda^{-1} e^{-\frac{|y_2|}{\lambda}} .$$

6. (15 points (bonus)) Let $N_1$, $N_2$ and $N_3$ be discrete random variables with joint probability mass function

$$p(n_1, n_2, n_3) = \mathbb{P}(N_1 = n_1 \,\&\, N_2 = n_2 \,\&\, N_3 = n_3) = \frac{3^{-N} N!}{n_1! n_2! n_3!}$$

if $n_1 + n_2 + n_3 = N$ and 0 otherwise.

Compute the marginal mass function $p_{N_1}$ of $N_1$, that is

$$p_{N_1}(n_1) = \mathbb{P}(N_1 = n_1)$$

and the conditional mass function $p_{N_2,N_3|N_1}$ of $N_2$ and $N_3$ given $N_1$, that is

$$p_{N_2,N_3|N_1}(n_2, n_3|n_1) = \mathbb{P}(N_2 = n_2 \,\&\, N_3 = n_3 \mid N_1 = n_1)\,.$$

(**Hint**: you can answer the question without doing any computation. Think what situation is described by $N_1$, $N_2$ and $N_3$.)

**Solution:** Observe that $N_1$, $N_2$ and $N_3$ are the result of repeating an experiment with 3 possible equiprobable outcomes (say 1,2,3) $N$ times. $\mathbb{P}(N_1 = n_1)$ represents the probability of obtaining $n_1$ 1s when the probability of a 1 in 1/3. Thus $\mathbb{P}(N_1 = n_1)$ is a binomial with $p = 1/3$ that is

$$\mathbb{P}(N_1 = n_1) = \frac{N!}{(N-n_1)!n_1!}\left(\frac{1}{3}\right)^{n_1}\left(\frac{2}{3}\right)^{N-n_1}\,.$$

On the other hand if you know you had exactly $n_1$ 1's, the remaining outcomes are 2 or 3, with equal probability. Thus

$$p_{N_2,N_3|N_1}(n_2, n_3|n_1) = \frac{2^{-(N-n_1)}(N-n_1)!}{n_2! n_3!}$$

if $n_2 + n_3 = N - n_1$ and 0 otherwise.

In formulas we have

$$\begin{aligned} p_{N_1}(n_1) = \sum_{n_2,n_3} p(n_1, n_2, n_3) &= \sum_{n_2+n_3=N-n_1} \frac{3^{-N} N!}{n_1! n_2! n_3!} = \\ &= \frac{3^{-N} 2^{N-n_1} N!}{(N-n_1)!n_1!} \sum_{n_2+n_3=N-n_1} \frac{2^{-(N-n_1)}(N-n_1)!}{n_2! n_3!} = \\ &= \frac{N!}{(N-n_1)!n_1!}\left(\frac{1}{3}\right)^{n_1}\left(\frac{2}{3}\right)^{N-n_1} \end{aligned}$$

so that $N_1$ is a binomial r.v. with $N$ trials and $p = 1/3$.

Moreover we have

$$
\begin{aligned}
p_{N_2,N_3|N_1}(n_2,n_3|n_1) =& \frac{3^{-N}N!}{n_1!n_2!n_3!}\left(\frac{N!}{(N-n_1)!n_1!}\left(\frac{1}{3}\right)^{n_1}\left(\frac{2}{3}\right)^{N-n_1}\right)^{-1} = \\
=& \frac{2^{-(N-n_1)}(N-n_1)!}{n_2!n_3!}
\end{aligned}
$$

if $n_2 + n_3 = N - n_1$ and 0 otherwise.

Thus $N_2$ is a binomial r.v with $N - n_1$ trials and $p = 1/2$.

7. (15 points (bonus)) Let $X_i,\ i = 1, \ldots, N$ be independent and identically distributed continuous r.v. with median $m$, that is

$$\mathbb{P}(X_i \leq m) = \frac{1}{2}\,.$$

Show that

$$\lim_{N\to\infty} \mathbb{P}\left(\min_{1\leq i\leq N}(X_i) \leq m < \max_{1\leq i\leq N}(X_i)\right) = 1\,.$$

**Solution:** Observe that due to independence

$$\mathbb{P}\left(\min_{1\leq i\leq N}(X_i) > m\right) = \prod_{i=1}^{N} \mathbb{P}(X_i > m) = 2^{-N}$$

and similarly

$$\mathbb{P}\left(\max_{1\leq i\leq N}(X_i) \leq m\right) = \prod_{i=1}^{N} \mathbb{P}(X_i \leq m) = 2^{-N}$$

so that

$$\begin{aligned}&\mathbb{P}\left(\min_{1\leq i\leq N}(X_i) \leq m < \max_{1\leq i\leq N}(X_i)\right) = \\ &1 - \mathbb{P}\left(\min_{1\leq i\leq N}(X_i) > m\right) - \mathbb{P}\left(\max_{1\leq i\leq N}(X_i) \leq m\right) = \\ &1 - 2^{N-1}\,.\end{aligned}$$