MIDTERM 2
MATH 3215-C (PROBABILITY AND STATISTICS)

TUESDAY, OCTOBER 20

INSTRUCTOR: ARMENAK PETROSYAN

IMPORTANT: Please read carefully (1pt)

- You have a **12 hour** window to take and submit your exam (**7 am - 7 pm**).
- Be warned: **exam ends at 7 pm** (e.g. if you start at 6 pm, you only have 1 hour).
- After you opened this file you have 100 minutes to finish the work and 20 minutes to submit it **(120 minutes in total)**.
- If you run into difficulties submitting on GradeScope, email the files to the instructor before the 120 minutes expire and before 7 pm. **Late submissions will not be accepted**.
- If you encounter technical problems, email the instructor as soon as possible.
- You **CAN** use the course textbook and the lecture notes/slides for reference.
- You **CAN** use any fact we presented in class without proving them; anything else used must be proved.
- You **CANNOT** get any help from or collaborate with anyone.
- Posting the problems online to get help or to let others know what the problems are will be a violation; it will be reported and result in a penalty.
- **To get full credit you need to write complete answers.**
- Numerical answers must be up to 4 decimal precision.
- The total amount of points for this exam is 75. Different problems have different weights.
- Be wise with your time. You can handwrite your answers on a different paper, and submit a photocopy. Make sure it is readable. No need to print the problem sheet or copy the problems.

Calculator and software use

- You can use a calculator for arithmetic computations.
- To find the values of standard distributions, you **MUST** use the tables in the appendix of the textbook (e.g. if it is asking to find a cdf value for the normal distribution, you need to reduce to the standard normal and find the value from the table in the back of the book).
- You may verify your answers for yourself with a calculator.
- You can plot the graphs by hand or you may use any graphical software.

Problem 1 (15pt). *For the following functions, check if there exists a number c for which $f(x)$ is a pdf.*
(a) $f(x) = c(1-x^2)$, $x \in [-1,1]$.
(b) $f(x) = c(2-x^2)$, $x \in [-2,2]$.
If such c exists, compute the expected value and the variance of a random variable X with pdf $f(x)$.

Solution. *(a)* $1-x^2 \geq 0$ *for* $x \in [0,1]$ *so for* $c \geq 0$, $f(x) \geq 0$.

$$1 = \int_{-1}^{1} (1-x^2)\,\mathrm{d}x = c(x - \frac{x^3}{3})\Big|_{-1}^{1} = c(1 - \frac{1}{3} + 1 - \frac{1}{3}) = c\frac{4}{3}.$$

Hence $\boxed{c = \frac{3}{4}}$.

$$E[X] = \int_{-1}^{1} \frac{3}{4}x(1-x^2)\,\mathrm{d}x = \frac{3}{4}(\frac{x^2}{2} - \frac{x^4}{4})\Big|_{-1}^{1} = \frac{3}{4}(\frac{1}{2} - \frac{1}{4} - \frac{1}{2} + \frac{1}{4})\boxed{= 0}.$$

$$E[X^2] = \int_{-1}^{1} x^2\frac{3}{4}(1-x^2)\,\mathrm{d}x = \frac{3}{4}(\frac{x^3}{3} - \frac{x^5}{5})\Big|_{-1}^{1} = \frac{3}{4}(\frac{1}{3} - \frac{1}{5} + \frac{1}{3} - \frac{1}{5}) = \frac{3}{4}\,\frac{4}{15}.$$

Hence

$$\operatorname{Var}(X) = E[X^2] - E[X]^2 \boxed{= \frac{1}{5}}.$$

(b) Notice that $f(0) = 2c$ *and* $f(2) = -2c$ *. The only c for which* $f(x) \geq 0$ *is when* $c = 0$ *but in that case the integral will be 0 so there is no c for which it is a pdf.*

Problem 2 (10pt). *Assume X has $N(\mu, \sigma^2)$ distribution (normal distribution with mean μ and variance σ). Find the probability that the value of X is 1.55σ distance away from the mean.*

Solution. *We want to compute the probability $|X - \mu| < 1.55\sigma$. Denote by Z the z-score of X: $Z = \frac{X-\mu}{\sigma}$. Then*

$$\begin{aligned} P(|X-\mu| < 1.55\sigma) &= P(|Z| < 1.55\sigma) = P(-1.55 < Z < 1.55) \\ &= F(1.55) - F(-1.55) = F(1.55) - (1 - F(1.55)) \\ &= 2F(1.55) - 1 = 2 \cdot 0.9394 - 1 = 0.8788. \end{aligned}$$

Problem 3 (12pt)**. *Fixed:*** *A web page gets 24 visits on average per hour. Assuming visits are governed by a Poisson process, what is the probability that the first 30 visits will happen within the first 50 minutes?*

(Hint: let X be the time of the 30th visit. Observe that the distribution of $Z = 2 \cdot \frac{60}{24} X$ is one of the distributions with a table in the Appendix of the textbook).

Solution. *Let X be the time the 30th visit happens. We want to compute*

$$P(X \leq 50).$$

X has a Gamma distribution with $\theta = \frac{60}{24} = \frac{5}{2}$ and $\alpha = 30$. Notice that $Z = \frac{2}{\theta} X = \frac{4X}{5}$ is also a Gamma distribution with $\theta = 2$ and $\alpha = 30$. This can be proved by computing the cdf of Z or we can make the following observation. We have a Poisson process with waiting time $\theta = 60/24 = 5/2$ (the average time for the next visit to happen). X is the time when the 30th visit happens. $Z = cX$ will be the time when the 30th visit happens if the waiting time was $c\theta$ for any $c > 0$. Take $c = 2/\theta$ then $Z = 2X/\theta$ will be the time when the 30th visit happens for the Poisson process with waiting time $c\theta = 2$.

Z having Gamma distribution with $\theta = 2$ and $\alpha = 30$ is the same as having χ^2 distribution with $r = 2\alpha = 60$ degrees of freedom. In that case,

$$P(X \leq 50) = P(Z \leq 40) \approx 0.025$$

from the table in the back of the book.

Problem 4 (12pt)**.** *Assume the length of a side of a cube is a random variable that has exponential distribution with parameter $\theta = 2$. Compute the expected volume of the cube.*

Solution. *We want to compute $E[X^3]$. This is the third moment of the Gaussian distribution and so $E[X^3] = M'''(0)$ where $M(t)$ is the moment generating function of the exponential distribution. We have shown that*

$$M(t) = \frac{1}{1-\theta t}.$$

And therefore

$$M'(t) = \frac{\theta}{(1-\theta t)^2}$$
$$M''(t) = \frac{2\theta^2}{(1-\theta t)^3}$$
$$M'''(t) = \frac{6\theta^3}{(1-\theta t)^4}.$$

And, therefore,

$$M'''(0) = 6\theta^3 \boxed{= 48}.$$

Problem 5 (20pt). *The following table contains values of the joint pmf of two discrete random variables. The top row and the leftmost column are the corresponding ranges of the random variables.*

(a) Are the random variables independent. Explain your answer.

(b) Compute the least squares line:

$$y = \frac{\mathrm{Cov}(X,Y)}{\mathrm{Var}(X)}(x-\mu_X)+\mu_Y.$$

(c) Compute the conditional mean $m(x) = \mu_{Y|x}$.

Y \ X	-1	0	1	2
-1	0.01	0.04	0.03	0.1
0	0.08	0.22	0.24	0.06
1	0.03	0.12	0.03	0.04

TABLE 1.

Solution. *(a) The random variables in Table 1 are dependent.*

$$f_X(-1) = 0.12, f_X(0) = 0.38, f_X(1) = 0.30, f_X(2) = 0.20$$
$$f_Y(-1) = 0.18, f_Y(0) = 0.60, f_Y(1) = 0.22.$$

The random variables are dependent because

$$f(-1,1) = 0.01 \neq f_X(-1)f_Y(-1) = 0.0216.$$

(b) To compute the least squares line we need to compute

$$\mu_X, \mu_Y, \mathrm{Var}(X), \mathrm{Cov}(X,Y).$$

$$\begin{aligned}
\mu_X &= -1\cdot 0.12 + 0\cdot 0.38 + 1\cdot 0.30 + 2\cdot 0.20 \boxed{= 0.58}.\\
E[X^2] &= (-1)^2\cdot 0.12 + 0^2\cdot 0.38 + 1^2\cdot 0.30 + 2^2\cdot 0.20 = 1.22\\
\mathrm{Var}(X) &= 1.22 - 0.58^2 \boxed{= 0.8836}\\
\mu_Y &= -1\cdot 0.18 + 0\cdot 0.6 + 1\cdot 0.22 \boxed{= 0.04}.\\
E[XY] &= 0.01 - 0.03 - 2\cdot 0.1 - 0.03 + 0.03 + 2\cdot 0.04 = -0.14\\
\mathrm{Cov}(X,Y) = E[XY] - \mu_X\mu_Y &= -0.14 - 0.58\cdot 0.04 \boxed{= -0.1632}.
\end{aligned}$$

Thus, the least squares line is given by

$$y = \frac{\mathrm{Cov}(C,Y)}{\mathrm{Var}(X)}(x-\mu_X)+\mu_Y = -\frac{0.1632}{0.8836}(x-0.58)+0.04.$$

(c)

$$\begin{aligned}
m(-1) &= (-1\cdot 0.01 + 0\cdot 0.08 + 1\cdot 0.03)/0.12 = 0.02/0.12 = \frac{1}{6}\\
m(0) &= (-1\cdot 0.04 + 0\cdot 0.22 + 1\cdot 0.12)/0.38 = 0.08/0.38 = \frac{4}{19}\\
m(1) &= (-1\cdot 0.03 + 0\cdot 0.24 + 1\cdot 0.03)/0.3 = 0/0.3 = 0\\
m(2) &= (-1\cdot 0.1 + 0\cdot 0.06 + 1\cdot 0.04)/0.2 = -0.06/0.2 = -\frac{3}{10}
\end{aligned}$$

Problem 6 (5pt)**.** *Let X and Y be two discrete random variables on the same space of outcomes S. We proved the following two facts:*

(1) If X and Y are independent then $\mathrm{Cov}(X, Y) = 0$ *(in class).*
(2) If X and Y are independent then, for any two functions $g, h : (-\infty, \infty) \to (-\infty, \infty)$, the random variables $g(X)$ and $h(Y)$ are also independent (as a homework problem).

Consequently, if X and Y are two independent random variables then, for any two functions $g, h : (-\infty, \infty) \to (-\infty, \infty)$,

$$\mathrm{Cov}(g(X), h(Y)) = 0.$$

Show that the opposite of the above statement is true as well: if, for any two functions $g, h : (-\infty, \infty) \to (-\infty, \infty)$,

$$\mathrm{Cov}(g(X), h(Y)) = 0$$

then X and Y are independent.

(Hint: fix any $(x, y) \in \mathrm{Range}(X, Y)$ and select appropriate functions f, g such that $\mathrm{Cov}(X, Y) = 0$ becomes $f(x, y) = f_X(x)f_Y(y)$.)

Solution. *Using the fact that*

$$\mathrm{Cov}(g(X), h(Y)) = E[g(X)h(Y)] - E[g(X)] \cdot E[h(Y)] = 0$$

we have

$$E[g(X)h(Y)] = E[g(X)] \cdot E[h(Y)]$$

or, in expanded form,

$$\sum_{(x,y)\in\mathrm{Range}(X,Y)} g(x)h(y)f(x, y) = \sum_{x\in\mathrm{Range}(X)} g(x)f_X(x) \cdot \sum_{y\in\mathrm{Range}(Y)} h(y)f_Y(y). \tag{1}$$

For fixed $(x_0, y_0) \in \mathrm{Range}(X, Y)$ define the functions

$$g(x) = \begin{cases} 1 & x = x_0 \\ 0 & x \neq x_0 \end{cases}, \quad h(y) = \begin{cases} 1 & y = y_0 \\ 0 & y \neq y_0 \end{cases}.$$

Then, if we compute the sums in (1), we see that it becomes

$$f(x_0, y_0) = f_X(x_0) \cdot f_Y(y_0).$$

Since (x_0, y_0) were any pair in $\mathrm{Range}(X, Y)$, from the definition of independence, we conclude that X and Y are independent.