Time Allowed: 5 hours. Total sum: 40p. Grades 3, 4 and 5 require 18p, 25p and 32p, respectively. Note:

- Motivate solutions well but no more than necessary.
- Just write on one side of the paper.
- Start new question on new page, and label the question number clearly next to your solution.
- Do not write with red-ink pen.
- Write the solutions in increasing order of question numbers.
- Hints and other information provided in a problem may also be useful for subsequent problems.
- Submit this exam paper with your solutions, i.e., you may NOT take this exam paper with you.

Permitted aids: Any books and notes.

Ex. 1 — *5p* – Suppose you observe the following five data points from some product experiment:

$$0, 2, 1, 0, 3.$$

(a)— *1p* – Report the sample mean.

(b)— *1p* – Report the sample variance.

(c)— *1p* – Report the five order statistics from minimum to maximum.

(d)— *1p* – Sketch the plot of the Empirical Mass Function showing discontinuities in the function and clearly labeling the axes.

(e)— *1p* – Sketch the plot of the Empirical Distribution Function showing discontinuities in the function and clearly labeling the axes.

Ex. 2 — *5p* –

1.— *3p* – In a survey, the two attributes of houses, whether they have a chimney and/or have a tile roof, were cross tabulated. The following table gives the resulting probabilities. Are these two attributes independent? You must justify your answer.

	Tile Roof	
Chimney	No	Yes
No	0.3	0.1
Yes	0.4	0.2

2.— *2p* – A certain river contains on average 1 trout per 100 m of length along the bank. Assuming the trout occur at random and their location is independent of other trout, what is the probability:

a)There are two or more trout in a 100 m stretch of river.

b)There are no trout in a 500 m stretch of river.

Ex. 3 — *5 p* – Assume In testing seeds for germination, batches of 20 seeds are tested at a time. If the true probability of germination for a seed is 0.9, and the germination of the seeds is independent, then:

(i)What is the expected number of seeds that will germinate?

(ii)What is the probability that exactly 19 seeds germinate.

(iii)Suppose there is a pollutant in the field and the probability of germination is 1/100. Assuming germination is independent, find an expression for the probability that none of the 1000 planted seeds in a given unit area of a field will germinate using the "Poisson approximation to the Binomial".

(iv)Justify the probability model used in the above computation in (iii). Give an example of agricultural conditions that can violate this model?

(v)Suppose there is another pollutant that reduces the probability of germination to 5/100 independently of the presence of the first pollutant. Using Poisson approximations for the second pollutant (as well as the first pollutant), what is the probability that none of the 1000 planted seeds in a unit area of a field will germinate in the presence of both pollutants?

Ex. 4 — *5p* –

1.Consider the continuous random variable, X, with the probability density function

$$f(x) = \begin{cases} Kx(1-x), & \text{if } 0 \leq x \leq 1 \\ 0, & \text{otherwise} \end{cases}.$$

a)Find the value of K.

b)Find the distribution function, $F(x)$.

c)Find the probability, $P(\,1/4 < X < 1/2\,)$.

2. Consider the discrete random variable, X, with the probability mass function given in the table below:

x	0	1	2	4
$P(X = x)$	0.4	0.3	0.2	0.1

a) Find the mean value of X, $E(X)$.

b) If $Y = 1 + X^2$, find the expected value of Y, $E(Y)$.

Ex. 5 — *5p* – Consider the continuous random variable, X, with the probability density function

$$f(x) = \begin{cases} \dfrac{2}{x^2}, & \text{if } 1 \leq x \leq 2 \\ 0, & \text{otherwise} \end{cases}$$

1. Find the mean value of X, $E(X)$.
2. Find the variance of X, $V(X)$.
3. If $Y = \sqrt{X}$, find the probability density function $f(y)$.

Ex. 6 — *5p* – Let X be Normal$(1, 3)$, Y be Normal$(-1, 1)$ and Z be Normal$(0, 1)$ RVs that are jointly independent. Obtain the following:

1. $E(3X - 3Y + 4Z)$
2. $V(2Y - 2Z)$
3. the distribution of $6 - 3Z + X - 2Y$
4. the probability that $6 - 3Z + X - 2Y > 0$
5. Covariance of X and W, where $W = X - Y$.

Ex. 7 — *5p* – Suppose the collection of RVs $X_1, X_2, \ldots, X_n$ model the number of errors in n computer programs named $1, 2, \ldots, n$, respectively. Suppose that the RV X_i modeling the number of errors in the i-th program is the Poisson$(\lambda = 3)$ for any $i = 1, 2, \ldots, n$. Further suppose that they are independently distributed. Succinctly, we suppose that

$$X_1, X_2, \ldots, X_n \overset{IID}{\sim} Poisson(\lambda = 3) \ .$$

Suppose we have $n = 123$ programs and want to make a probability statement about the sample mean from these 123 samples which is the average error per program out of these 123 programs. Since $E(X_i) = \lambda = 3$ and $V(X_i) = \lambda = 3$, we want to know how often our sample mean differs from the expectation of 3 errors per program. Using the CLT find the probability that the sample mean from the 123 samples is

1. less than 4
2. less that or equal to 0

Ex. 8 — *5p* –

1. Show that the minimum of k IID Exponential(λ) RVs is another Exponential($k\lambda$) RV. Explain your argument step-by-step.
2. Show that the Exponential(λ) RV satisfies the following:

$$P(X > s + t | X > s) = P(X > t).$$

   Explain your argument step-by-step.

Lycka till!

Answer (Ex. 1) —
This was done in the last day. Note that you will get 0 points if the points of discontinuity are not perfectly plotted with open and closed circles and solid and dashed lines with axes labelled and marked properly.

(a)sample mean =

(b)sample variance =

(c)order statistics is:

(d)Empirical Mass Function is given by:

(e)Empirical Distribution Function is given by:

Answer (Ex. 2) —

1.For independence $P(A \cap B) = P(A)\,P(B)$ or equivalently $P(A|B) = P(A)$.

$$\begin{aligned} P(\text{ Chimney }) &= 0.4 + 0.2 = 0.6 \\ P(\text{ Tile roof }) &= 0.1 + 0.2 = 0.3 \\ P(\text{ Chimney} \cap \text{Tile roof}) &= 0.2 \neq 0.6 \times 0.3 = 0.18 \end{aligned}$$

So the attributes are not independent.

2.

$$(a) \qquad \begin{aligned} P(\, X \geq 2\,) &= 1 - P(0) - P(1) \\ &= 1 - e^{-1} - e^{-1} = 1 - 2e^{-1} \\ &\approx 1 - 2 \times 0.368 = 1 - 0.736 = 0.264 \end{aligned}$$

$$(b) \qquad \lambda = 5, \qquad \begin{aligned} P(\, X = 0\,) &= e^{-5}\frac{5^0}{0!} = e^{-5} \\ &= 0.007 \text{ from the tables. OR } (0.368)^5 \end{aligned}$$

Answer (Ex. 3) — (i) Expected number $= np = 20 \times 0.9 = 18.$

(ii) $P(\, X = 19\,) = \binom{20}{19}(0.9)^{19}(0.1)^1 = 2\,(0.9)^{19}$

(iii) [only partial answer and hints given]: ... [see Lacunae problem on steel specimens to figure out how to get the expression here...]

(iv) Heterogeneity in other field conditions like soil fertility, moisture, competing species, could violate the Poisson assumption.

(v) [only partial answer and hints given]: Get Poisson approxmation under second pollutant and recall Exercise 3.54 that the sum of independent Poissons is Poisson with paramater given by the sum of the individual Poisson parameters.

Answer (Ex. 4) — 1. a)To be a probability distribution

$$\begin{aligned}1 = \int_0^1 f(x)\,dx &= K\int_0^1 x(1-x)\,dx \\ &= K\int_0^1 x - x^2\,dx \\ &= K\left[\frac{x^2}{2} - \frac{x^3}{3}\right]_0^1 \\ &= K\frac{1}{6}.\end{aligned}$$

So that $1 = K/6$, implying $K = 6$.

b)For $0 < x < 1$,

$$F(x) = 6\left[\frac{x^2}{2} - \frac{x^3}{3}\right] = 3x^2 - 2x^3.$$

Therefore,

$$F(x) = \begin{cases} 0, & x \le 0 \\ 3x^2 - 2x^3, & 0 < x < 1 \\ 1, & x \ge 1. \end{cases}$$

c)

$$\begin{aligned}P(\frac{1}{4} < x < \frac{1}{2}) &= F\left(\frac{1}{2}\right) - F\left(\frac{1}{4}\right) \\ &= \left[3\left(\frac{1}{2}\right)^2 - 2\left(\frac{1}{2}\right)^3\right] - \left[3\left(\frac{1}{4}\right)^2 - 2\left(\frac{1}{4}\right)^3\right] \\ &= \left[\frac{3}{4} - \frac{1}{4}\right] - \left[\frac{3}{16} - \frac{1}{32}\right] \\ &= \frac{1}{2} - \frac{5}{32} = \frac{11}{32}\end{aligned}$$

2. a)– *p —

$$\begin{aligned}E(X) &= \sum_i x_i\,p_i = 0\times 0.4 + 1\times 0.3 + 2\times 0.2 + 4\times 0.1 \\ &= 0 + 0.3 + 0.4 + 0.4 \\ &= 1.1\end{aligned}$$

b)

$$E(Y) = \sum_i y_i\, p_i = (1+0^2)\times 0.4 + (1+1^2)\times 0.3 + (1+2^2)\times 0.2 + (1+4^2)\times 0.1$$
$$= 1\times 0.4 + 2\times 0.3 + 5\times 0.2 + 17\times 0.1$$
$$= 0.4 + 0.6 + 1.0 + 1.7$$
$$= 3.7$$

Answer (Ex. 5) — 1.

$$\begin{aligned} E(X) &= \int_1^2 x\, f(x)\, dx \\ &= \int_1^2 x \frac{2}{x^2}\, dx = \int_1^2 2x^{-1}\, dx = [\, 2\ln(x)\,]_1^2 \\ &= 2\ln 2 - 2\ln 1 = 2\ln 2 \approx 1.386 \end{aligned}$$

2.

$$\mathrm{Var}(x) = E(X^2) - E(X)^2.$$
$$E(X^2) = \int_1^2 x^2 \frac{2}{x^2}\, dx = \int_1^2 2\, dx = 2.$$

Therefore,

$$\mathrm{Var}(X) = 2 - [2\ln(2)]^2 \approx 0.0782.$$

3.

$$y = \sqrt{x}, \qquad g(x) = \sqrt{x} = x^{1/2} \implies g'(x) = \frac{1}{2}x^{-1/2} = \frac{1}{2\sqrt{x}}.$$

Hence, $\sqrt{x}$ is strictly increasing and therefore one to one. Hence g is invertible. $x = g^{-1}(y) = y^2$.

$$\left|\frac{d}{dy}g^{-1}(y)\right| = \frac{d}{dy}y^2 = 2y.$$

Therefore, for $1 < y < 2$,

$$\begin{aligned} f_Y(y) &= f_X\Big(g^{-1}(y)\Big)\cdot\left|\frac{d}{dy}g^{-1}(y)\right| \\ &= \frac{2}{(y^2)^2}\cdot 2y = \frac{4y}{y^4} = \frac{4}{y^3}. \end{aligned}$$

Thus

$$f_Y(y) = \begin{cases} \frac{4}{y^3}, & 1 < y < \sqrt{2}, \\ 0, & \text{otherwise.} \end{cases}$$

Answer (Ex. 6) — This is just a modification of the exact solution for Example 96 given below (you should find the answer for *this problem*! on your own):

1.

$$E(3X-2Y+4Z) = 3E(X)-2E(Y)+4(Z) = (3\times2)+(-2\times(-1))+4\times0 = 6+2+0 = 8$$

2.

$$V(2Y-3Z) = 2^2V(Y)+(-3)^2V(Z) = (4\times2)+(9\times1) = 8+9 = 17$$

3. From the special property of normal RVs, the distribution of $6-2Z+X-Y$ is

$$\begin{aligned}
&Normal\Big(6+(-2\times0)+(1\times2)+(-1\times-1), ((-2)^2\times1)+(1^2\times4)+((-1)^2\times2)\Big) \\
&= Normal(6+0+2+1, 4+4+2) \\
&= Normal(9,10)
\end{aligned}$$

4. Let $U = 6-2Z+X-Y$ and we know U is $Normal(9,10)$ RV.

$$\begin{aligned}
P(6-2Z+X-Y>0) = P(U>0) = P(U-9>0-9) &= P\Big(\frac{U-9}{\sqrt{10}} > \frac{-9}{\sqrt{10}}\Big) \\
&= P\Big(Z > \frac{-9}{\sqrt{10}}\Big) \\
&= P\Big(Z < \frac{9}{\sqrt{10}}\Big) \\
&\approxeq P(Z < 2.85) = 0.9978
\end{aligned}$$

5.

$$\begin{aligned}
\mathbf{Cov}(X,W) &= E(XW)-E(X)E(W) = E(X(X-Y))-E(X)E(X-Y) \\
&= E(X^2-XY)-E(X)(E(X)-E(Y)) = E(X^2)-E(XY)-2\times(2-(-1)) \\
&= E(X^2)-E(X)E(Y)-6 = E(X^2)-(2\times(-1))-6 \\
&= (V(X)+(E(X))^2)+2-6 = (4+2^2)-4 = 4
\end{aligned}$$

Answer (Ex. 7) — See solution to Exercise 5.3 in the draft book for the course and modify it with the numbers specific to *this problem*!.

Answer (Ex. 8) —

1. This was done in lectures on the last day (so only partial answer and hints on how to proceed are given). Think about what happens when $k = 2$ first. Let $Y = \min(X_1, X_2)$, then try to derive $F(Y)$ using the fact that $P(Y > y) = P(X_1 > y)P(X_2 > y)$. Use the complementary DF and proceed. You will have to show your reasoning step-by-step.

2. This was done in lectures. It is just the proof of Proposition 26 on memorylessness of Exponential RV but with x and y replaced by s and t.