1. Suppose that the individual ground up loss of an insured, denoted by X, follows a uniform distribution between 0 and 1200. The insurance policy has a deductible level of $d > 0$, such that if the loss amount is below d, the insured will not be able to claim from the insurer. If the loss amount is above d, the insurer will pay the excess amount over d denoted by $X - d$, subjected to a maximum claim payment of $M < 1200 - d$.

   (a) Derive the mean and the variance of the insurance payout under the policy in terms of d and M. (5 marks)

$$Y = (X \wedge (d+M)) - (X \wedge d) = \begin{cases} 0, & 0 < X \leq d \\ X - d, & d < X < d+M \\ M, & d+M \leq X < 1200 \end{cases}$$

$$\begin{aligned}
f_X(x) &= \frac{1}{1200} \\
E(Y) &= \frac{1}{1200}\int_d^{d+M} (x-d)dx + \frac{1}{1200}\int_{d+M}^{1200} M dx \\
&= \frac{1}{1200}\frac{M^2}{2} + \frac{M}{1200}(1200 - d - M) \\
&= M - \frac{1}{1200}\left(Md + \frac{M^2}{2}\right) \\
\text{OR } &= \int_0^{d+M} x f_X(x)dx + (d+M)(1 - F_X(d+M)) - \left(\int_0^d x f_X(x)dx + d(1 - F_X(d))\right) \\
&= \left(\frac{(d+M)^2}{2(1200)} + (d+M) \times \frac{1200 - (d+M)}{1200}\right) - \left(\frac{d^2}{2(1200)} + d \times \frac{1200 - d}{1200}\right) \\
&= (d+M) - \frac{(d+M)^2}{2(1200)} - \left(d - \frac{d^2}{2(1200)}\right) \\
&= M - \frac{1}{1200}\left(Md + \frac{M^2}{2}\right) \\
\text{Var}(Y) &= E(Y^2) - (E(Y))^2 \\
&= \frac{1}{1200}\int_d^{d+M} (x-d)^2 dx + \frac{1}{1200}\int_{d+M}^{1200} M^2 dx - \left(M - \frac{1}{1200}\left(Md + \frac{M^2}{2}\right)\right)^2 \\
&= \frac{1}{1200}\left(\frac{M^3}{3} + M^2(1200 - d - M)\right) - \left(M - \frac{1}{1200}\left(Md + \frac{M^2}{2}\right)\right)^2
\end{aligned}$$

   (b) Is the maximum random loss to be borne by the insured after taking into account the random claim payment by the insurer equal to the deductible level d? Explain your answer clearly. (1 mark)

No, the random loss is given by $X \wedge d + (X - M - d)_+$, which equals to $X - M > d$ for large losses $d + M < X < 1200$.

(c) Suppose that the insurer sold 500 such policies, each with $d = 60$ and $M = 800$. The claim frequency under each policy follows a Poisson distribution with a mean of 0.8. Using the Normal approximation, calculate the probability that the total aggregate claims from all 500 policies is greater than 125% of the pure premium received from the policies. (4 marks)

$$
\begin{aligned}
E(Y) &= 800 - \frac{1}{1200}\left(800(60) + \frac{800^2}{2}\right) \\
&= \frac{1480}{3} = 493.333 \\
\text{Var}(Y) &= \frac{1}{1200}\left(\frac{800^3}{3} + 800^2(1200 - 60 - 800)\right) - \left(\frac{1480}{3}\right)^2 \\
&= \frac{721600}{9} = 80177.778 \\
E(S) &= E(N)E(Y) = 0.8(500)E(Y) \\
&= \frac{592000}{3} = 197333.333 \\
\text{Var}(S) &= E(N)\text{Var}(Y) + (E(Y))^2\,\text{Var}(N) = 400\text{Var}(Y) + 400\,(E(Y))^2 \\
\text{OR } &= \lambda m_2 = 400E(Y^2) = 400 \times \frac{1}{1200}\left(\frac{800^3}{3} + 800^2(1200 - 60 - 800)\right) \\
&= \frac{1164800000}{9} = 129422222.222 \\
\Pr\left(S > 1.25E(S)\right) &= \Pr\left(Z > \frac{1.25E(S) - E(S)}{SD(S)}\right) \\
&= \Pr(Z > 4.336467) = \Pr(Z > 4.34) \\
&\approx 0
\end{aligned}
$$

Overall, the performance in this question was poorer than expected.

- Part (a) was poorly answered. Common mistakes include not accounting for the upper limit M, incorrectly using M as the upper bound for $Y = X-d$, incorrect subsequent steps in deriving the mean, incorrect (part of) initial integral such as using $\int_d^{d+M} \frac{x}{1200}dx$, did not realise that $E\left(((X \wedge (d+M)) - (X \wedge d))^2\right) \neq E\left((X \wedge (d+M))^2\right) + E\left((X \wedge d)^2\right)$, probably due to time constraint / pressure.

- The performance in part (b) was mixed. Common mistakes include not explaining how / why the loss is more than d, incorrectly using the deductible d as the reason, or simply did not provide any explanation.
- Part (c) was well answered, considering the performance in part (a), which **emphasizes the importance of not ignoring subsequent parts even if you know that you did not manage to solve the earlier parts correctly**. The formula for the mean and the variance of the aggregate sum is straightforward and the same goes to the probability calculation. Common mistakes include not using 500 policies into the calculation, mis-interpretation of pure premium and not calculating the numerical values etc.

2. You are given the following joint distribution function for bivariate random variables X_1 and X_2

$$\Pr(X_1 \leq x_1, X_2 \leq x_2) = 1 - \left(1 - \left(1 - e^{-3x_1\theta}\right)\left(1 - (1 - x_2)^\theta\right)\right)^{\frac{1}{\theta}}, \quad x_1 > 0, 0 < x_2 < 1,$$

for $\theta \geq 1$.

(a) Determine and identify the underlying marginal distributions of X_1 and X_2 respectively and their associated parameters. Note that simply writing down the marginal cumulative distribution function is not sufficient. (3 marks)

$$\begin{aligned}
F_{X_1}(x_1) &= \Pr(X_1 \leq x_1, X_2 \leq 1) = 1 - e^{-3x_1} \\
&\Rightarrow X_1 \sim \text{Exp}(3), \quad X_1 > 0 \\
F_{X_2}(x_2) &= \Pr(X_1 \leq \infty, X_2 \leq x_2) = x_2 \\
&\Rightarrow X_2 \sim U(0,1), \quad 0 < X_2 < 1
\end{aligned}$$

(b) Determine the copula function $C(u_1, u_2)$ in terms of u_1 and u_2. (3 marks)

$$
\begin{aligned}
F_{X_1}(x_1) &= 1 - e^{-3x_1} = u_1 \\
\Rightarrow x_1 &= -\frac{\log(1-u_1)}{3} \\
F_{X_2}(x_2) &= x_2 = u_2 \\
C(u_1, u_2) &= 1 - \left(1 - \left(1 - e^{-3\theta \times -\frac{\log(1-u_1)}{3}}\right)\left(1 - (1-u_2)^\theta\right)\right)^{\frac{1}{\theta}} \\
&= 1 - \left(1 - \left(1 - (1-u_1)^\theta\right)\left(1 - (1-u_2)^\theta\right)\right)^{\frac{1}{\theta}}
\end{aligned}
$$

(c) The copula function in part (b) can be used to link any two marginal distributions (i.e., not necessarily of the same form as those obtained in part (a)). True or false? Explain. (1 mark)

True. Copula is flexible in that it can take in any marginal cumulative distribution functions and express the joint cumulative distribution function via its explicit form in terms of both marginals u_1 and u_2.

(d) What is the property of the copula if $\theta = 1$? (2 marks)

We obtain $C(u_1, u_2) = u_1 u_2$, which is an independence copula.

(e) Following from the expression in part (b), explain **in words** whether the copula is an Archimedean copula. You do not need to work out the mathematical formulation or details. Instead, describe your thought process and be as specific as possible. (2 marks)

The product of $(1 - (1-u_1)^\theta)$ and $(1 - (1-u_2)^\theta)$ can be expressed in terms of the addition of two generator functions so long as we take logarithm before we perform the addition and we invert the operation by taking exp(.) through its inverse function (i.e., the inverse of the generator function) afterwards.

Overall, the performance in this question was in line with expectation.

- Part (a) was well answered in general. Common mistakes include not simplifying the expression further (to cancel out the impact of θ), stating incorrect parameters for the distributions and in some cases identify the wrong distributions.

- The performance in part (b) was a little below expectation. Common mistakes include careless mistakes due to the use of many brackets in the original joint cdf that result in final incorrect expression.
- Part (c) was poorly answered given its simplicity. The keyword here is the **copula function in part (b) is in terms of u and v** and not in terms of x_1 and x_2. Common mistakes include incorrectly arguing the function is only valid for the given distribution in the question, the range of values for x_1 and x_2 etc, the existence of unique copula for each joint cdf etc.
- Part (d) was well answered and errors carried forward receive full credit in many cases. However, many students state the answer as independence copula even when their correct simplification on an incorrect earlier part did not suggest so. Such attempt should be avoided in general, especially in the workplace. It is perfectly fine to follow the process and get an incorrect result based on earlier errors, than say arguing for the correct final outcome, because "two wrongs don't make a right".
- Part (e) was in line with expectation. This is a tricky question. Common mistakes include not realising we can use generator to turn the product into summation that satisfies the definition of Archimedean copula. Many students simply stated the conditions without analysing the copula function.

3. Suppose that the random variable X is uniformly distributed between 0 and 1.

   (a) Derive the cumulative distribution function of random variable Y defined as

$$Y = -2025 \log X.$$

(2 marks)

$$\begin{aligned} F_X(x) &= x \\ \Pr(Y \leq y) &= \Pr(-2025 \log X \leq y) = \Pr(X \geq e^{-\frac{y}{2025}}) \\ &= 1 - e^{-\frac{y}{2025}}, \quad y > 0 \end{aligned}$$

(b) Following from part (a), we denote the complement of the threshold exceedance probability as

$$\overline{F}_u(x) = 1 - F_u(x) = \Pr\left(Y - u > x | Y > u\right)$$

Show that $\overline{F}_u(x) = 1 - G_{\gamma,\sigma}(x)$ for suitably chosen parameters γ and σ of a Generalised Pareto distribution (GPD) G. (3 marks)

$$\begin{aligned}
\overline{F}_u(x) &= 1 - \frac{F(x+u) - F(u)}{1 - F(u)} = \frac{1 - F(x+u)}{1 - F(u)} \\
&= \frac{1 - \left(1 - e^{-\frac{x+u}{2025}}\right)}{1 - \left(1 - e^{-\frac{u}{2025}}\right)} \\
&= e^{-\frac{x}{2025}} = 1 - \left(1 - e^{-\frac{x}{2025}}\right) = 1 - G_{\gamma,\sigma}(x)
\end{aligned}$$

for $\gamma = 0$ and $\sigma = 2025$.

(c) Consider independently and identically distributed (iid) random variables $Y_i, i = 1, 2, \ldots, n$ with distribution function obtained in part (a). Show that the limiting distribution of the normalized n-block maximum is from the Generalised Extreme Value (GEV) distribution. Show your steps clearly and specify the range of parameter values for γ. (4 marks)

$$\begin{aligned}
\lim_{n\to\infty} \Pr\left(\frac{Y_{1,n} - b_n}{a_n} \le y\right) &= \lim_{n\to\infty} \Pr\left(Y_{1,n} \le a_n y + b_n\right) \\
&= \lim_{n\to\infty} F^n\left(a_n y + b_n\right) \\
&= \lim_{n\to\infty} \left(1 - e^{-\left(\frac{a_n y + b_n}{2025}\right)}\right)^n \\
&= \lim_{n\to\infty} \left(1 - \frac{e^{-y}}{n}\right)^n = e^{-e^{-y}}
\end{aligned}$$

by choosing normalizing constants $a_n = 2025, b_n = 2025 \log n, \mu = 0, \sigma = 1$, where $x \in \mathbb{R}$. We have $\gamma = 0$, i.e., the GEV of the Gumbel type.

Overall, the performance in this question was better than expected.

- Part (a) was well answered. Common mistakes include forgetting to switch the sign of the inequality when they divided both sides by -2025.

- Part (b) was very well answered. A number of students incorrectly used the cdf of uniform distribution instead of their cdf obtained in part (a).
- Part (c) was well answered in general. Common mistakes include not able to determine a_n and b_n due to some mistakes in algebra or struggle to understand the question. Partial marks were awarded to students who did not show full working but able to provide relevant information to GEV and correctly stated $\gamma = 0$.