

STAT1301 Assignment 3

Caleb Yates s49886351

September 7, 2025

1 Question 1

1.1 Part a)

Let Ω be the sample space. Therefore $P(\{\Omega\}) = 1$. Adding all the joint pmf values must sum to 1:

$$\begin{aligned}\{\Omega\} &= \bigcup_x \bigcup_y \{X = x\} \cap \{Y = y\} \\ P(\{\Omega\}) &= 1 \\ \implies 1 &= P((\{X = -1\} \cap \{Y = -1\}) \cup \dots \cup (\{X = 1\} \cap \{Y = 1\})) \\ &= P(\{X = -1\} \cap \{Y = -1\}) + \dots + P(\{X = 1\} \cap \{Y = 1\}) \\ &= (p - \frac{1}{16}) + (\frac{1}{4} - p) + (0) + (\frac{1}{8}) + (\frac{3}{16}) + (\frac{1}{8}) + (p + \frac{1}{16}) + (\frac{1}{16}) + (\frac{1}{4} - p) \\ 1 &= -\frac{1}{16} + \frac{4}{16} + \frac{7}{16} + \frac{1}{16} + \frac{1}{16} + \frac{4}{16} \\ 1 &= 1\end{aligned}$$

Unfortunately, this tells us no information about p . From the definition of probability, $P(\{c\})$ for $c \in \Omega$ must be greater or equal to 0, $P(\{c \in \Omega\}) \geq 0$. This can be used to restrict the possible values of p :

$$\begin{aligned}P(A \subseteq \Omega) &\geq 0 \\ \implies P(\{X = -1\} \cap \{Y = -1\}) &\geq 0 \\ p - \frac{1}{16} &\geq 0 \\ p &\geq \frac{1}{16} \\ \implies P(\{X = 0\} \cap \{Y = -1\}) &\geq 0 \\ \frac{1}{4} - p &\geq 0 \\ p &\leq \frac{1}{4} \\ \implies P(\{X = -1\} \cap \{Y = 1\}) &\geq 0 \\ p + \frac{1}{16} &\geq 0 \\ p &\geq -\frac{1}{16} \\ \implies p &\in [\frac{1}{16}, \frac{1}{4}] \tag{1}\end{aligned}$$

Therefore, $\frac{1}{16} \leq p \leq \frac{1}{4}$, and can be any value within this range.

1.2 Part b)

Aim is to find $P(\{X = Y\})$:

$$\begin{aligned}P(\{X = Y\}) &= \sum_a P(\{X = a\} \cap \{Y = a\}) \\&= P(\{X = -1\} \cap \{Y = -1\}) + P(\{X = 0\} \cap \{Y = 0\}) + P(\{X = 1\} \cap \{Y = 1\}) \\&= (p - \frac{1}{16}) + (\frac{3}{16}) + (\frac{1}{4} - p) \\&= \frac{6}{16} = \frac{3}{8}\end{aligned}$$

1.3 Part c)

The marginal pdf of X is $f_X(x)$, which is equal to $P(\{X = x\})$ and can be manually evaluated:

$$\begin{aligned}
 P(\{X = -1\}) &= \sum_y P(\{X = -1\} \cap \{Y = y\}) \\
 &= (p - \frac{1}{16}) + (\frac{1}{8}) + (p + \frac{1}{16}) \\
 &= 2p + \frac{1}{8} \\
 P(\{X = 0\}) &= \sum_y P(\{X = 0\} \cap \{Y = y\}) \\
 &= (\frac{1}{4} - p) + (\frac{3}{16}) + (\frac{1}{16}) \\
 &= -p + \frac{1}{2} \\
 P(\{X = 1\}) &= \sum_y P(\{X = 1\} \cap \{Y = y\}) \\
 &= (0) + (\frac{1}{8}) + (\frac{1}{4} - p) \\
 &= -p + \frac{3}{8} \\
 \Rightarrow f_X(x) = P(\{X = x\}) &= \begin{cases} 2p + \frac{1}{8} & x = -1 \\ -p + \frac{1}{2} & x = 0 \\ -p + \frac{3}{8} & x = 1 \\ 0 & \text{otherwise} \end{cases}
 \end{aligned}$$

$$\begin{aligned}
P(\{Y = -1\}) &= \sum_x P(\{X = x\} \cap \{Y = -1\}) \\
&= \left(p - \frac{1}{16}\right) + \left(\frac{1}{4} - p\right) + (0) \\
&= \frac{3}{16} \\
P(\{Y = 0\}) &= \sum_x P(\{X = x\} \cap \{Y = 0\}) \\
&= \left(\frac{1}{8}\right) + \left(\frac{3}{16}\right) + \left(\frac{1}{8}\right) \\
&= \frac{7}{16} \\
P(\{Y = 1\}) &= \sum_x P(\{X = x\} \cap \{Y = 1\}) \\
&= \left(p + \frac{1}{16}\right) + \left(\frac{1}{16}\right) + \left(\frac{1}{4} - p\right) \\
&= \frac{6}{18} = \frac{3}{8} \\
\Rightarrow f_Y(x) = P(\{Y = y\}) &= \begin{cases} \frac{3}{16} & y = -1 \\ \frac{7}{16} & y = 0 \\ \frac{3}{8} & y = 1 \\ 0 & \text{otherwise} \end{cases}
\end{aligned}$$

1.4 Part d)

X and Y are independent if

$$P(\{X = x\} \cap \{Y = y\}) = P(\{X = x\}) \cdot P(\{Y = y\}) = f_X(x) \cdot f_Y(y) \quad (2)$$

for all possible values x and y . Therefore, this must be true for $x = -1$ and $y = 1$:

$$\begin{aligned} \text{LHS} &= P(\{X = -1\} \cap \{Y = 1\}) \\ &= p + \frac{1}{16} \end{aligned} \qquad \begin{aligned} \text{RHS} &= P(\{X = -1\})P(\{Y = 1\}) \\ &= (2p + \frac{1}{8})(\frac{3}{8}) \\ &= \frac{3}{4}p + \frac{3}{64} \end{aligned}$$

As shown above, LHS and RHS are only equal for zero or one values of p . Letting $\text{LHS} = \text{RHS}$, we can find this exact value (or lack thereof):

$$\begin{aligned} p + \frac{1}{16} &= \frac{3}{4}p + \frac{3}{64} \\ \frac{1}{4}p &= \frac{3}{64} - \frac{1}{16} \\ p &= -\frac{1}{64} \cdot 4 = -\frac{1}{16} \end{aligned}$$

Therefore $\text{LHS} = \text{RHS}$ only when $p = -\frac{1}{16}$, however from (1) this is not within the potential domain of p . Therefore $\text{LHS} \neq \text{RHS}$, showing one counterexample to (2), hence X and Y are not independent.

1.5 Part e)

$$\begin{aligned}
E(X) &= \sum_x xP(\{X = x\}) \\
&= -1(2p + \frac{1}{8}) + 0(-p + \frac{1}{2}) + 1(-p + \frac{3}{8}) \\
&= -2p - \frac{1}{8} - p + \frac{3}{8} \\
\therefore E(X) &= -3p + \frac{1}{4} \\
E(Y) &= \sum_y yP(\{Y = y\}) \\
&= -1(\frac{3}{16}) + 0(\frac{7}{16}) + 1(\frac{3}{8}) \\
\therefore E(Y) &= -\frac{3}{16} + \frac{3}{8} = \frac{3}{16} \\
\text{Cov}(X, Y) &= E[(X - E(X))(Y - E(Y))] \\
\text{Cov}(X, Y) &= \sum_{c \in \Omega} (X(c) - E(X))(Y(c) - E(Y))P(\{c\}) \\
&= \sum_{x,y} (x - (-3p + \frac{1}{4}))(y - (\frac{3}{16}))P(\{X = x\} \cap \{Y = y\})
\end{aligned}$$

Expanding this sum is tedious and results in nine trinomials. The following sum for the $\text{Cov}(X, Y)$ expansion significantly reduces the algebra necessary by computing $E(XY)$ instead:

$$\begin{aligned}
\text{Cov}(X, Y) &= E(XY) - E(X)E(Y) \\
E(XY) &= \sum_{c \in \Omega} X(c)Y(c)P(\{c\}) \\
&= \sum_{x,y} xyP(\{X = x\} \cap \{Y = y\}) \\
&= (-1)(-1)(p - \frac{1}{16}) + (-1)(0)(\frac{1}{4} - p) + (-1)(1)(0) \\
&\quad + (0)(-1)(\frac{1}{8}) + (0)(0)(\frac{3}{16}) + (0)(1)(\frac{1}{8}) \\
&\quad + (1)(-1)(p + \frac{1}{16}) + (1)(0)(\frac{1}{16}) + (1)(1)(\frac{1}{4} - p) \\
&= (p - \frac{1}{16}) - (p + \frac{1}{16}) + (\frac{1}{4} - p) \\
\therefore E(XY) &= -p + \frac{1}{8} \\
\implies \text{Cov}(X, Y) &= (-p + \frac{1}{8}) - (-3p + \frac{1}{4})(\frac{3}{16}) \\
&= -p + \frac{1}{8} + \frac{9}{16}p - \frac{3}{64} \\
\therefore \text{Cov}(X, Y) &= -\frac{7}{16}p - \frac{5}{64}
\end{aligned}$$

2 Question 2

Ω is continuous, which implies X is continuous and Y is continuous. Let D be the set of all $\langle x, y \rangle$ that is inside (or on the boundary) of the triangle given.

2.1 Part a)

We are told that the joint pdf of X and Y is uniform over D , and assume it is 0 everywhere else. The area of the triangle D on a cartesian plane is $A = \frac{1}{2}bh = 1$, and we know

$$\int_{d \in D} f_{X,Y}(d) dd = 1$$

Since $f_{X,Y}$ is uniform, this integral can be interpreted as the geometric volume of a triangular prism, extruded from D by $f_{X,Y}$

$$\begin{aligned} A \cdot f_{X,Y} &= 1 \\ f_{X,Y} &= 1 \end{aligned}$$

Therefore the joint pdf of (X, Y) is $f_{X,Y} = 1$.

2.2 Part b)

The set of vectors D_2 containing all vectors $\langle x, y \rangle$ satisfying $x > y$ in D forms a triangle on a cartesian plane with vertices at $(0, 0)$ $(0.5, 0.5)$ and $(1, 0)$. The area of this triangle is exactly $\frac{1}{4}$ the total area of D , since $A = \frac{1}{2}bh = \frac{1}{2} \cdot 1 \cdot 0.5 = \frac{1}{4}$. Therefore:

$$\begin{aligned} P(\{X \geq Y\}) &= \int_{d \in D_2} f_{X,Y}(d) dd \\ &= f_{X,Y} \cdot A \\ &= \frac{1}{4} \\ \therefore P(\{X \geq Y\}) &= \frac{1}{4} \end{aligned}$$

2.3 Part c)

Since

$$F_X = \int_{-\infty}^x f_X(x) dx$$

$$f_X(x) = \frac{d}{dx} F_X(x)$$

2.3.1 PDF of X

Figure 1 illustrates the geometric cases involved with evaluating f_X :

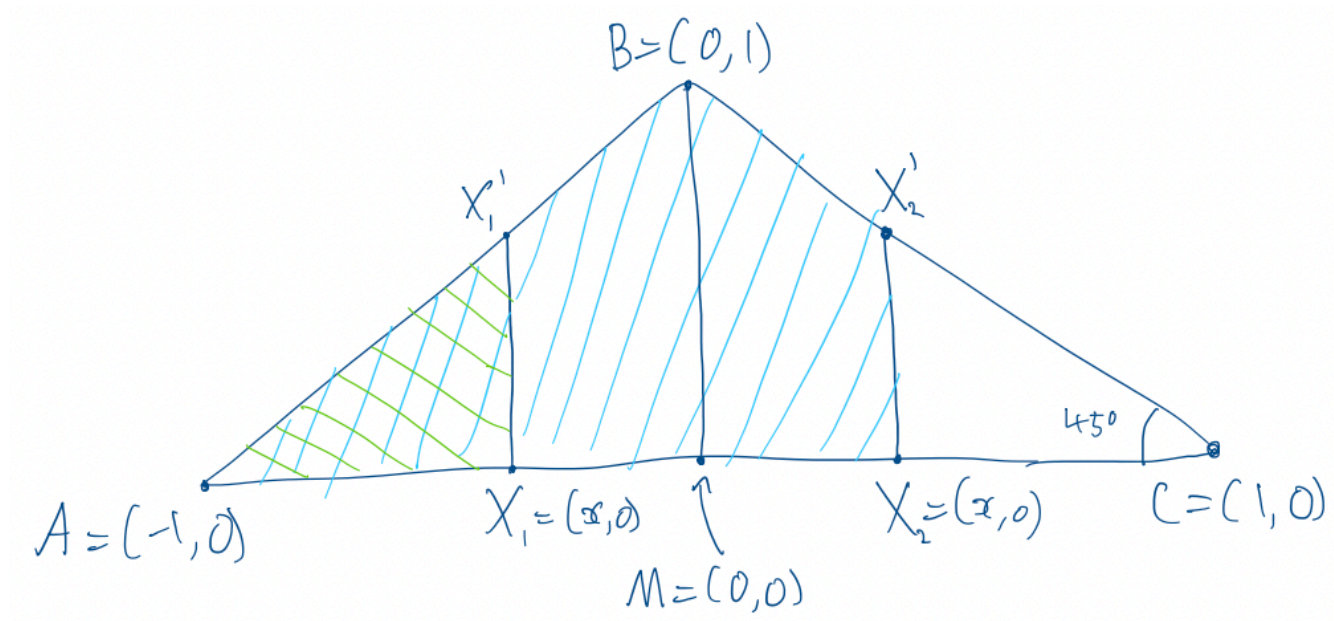


Figure 1: Geometric rendering of D showing the cases of F_X

We know

$$F_X(-1) = 0$$

$$F_X(1) = 1$$

$$F_X(x < -1) = 0$$

$$F_X(x > 1) = 1$$

Case 1: $-1 \leq x \leq 0$. This implies F_X is the area of $\triangle AX_1'X_1$. Let $AX_1 = x_1 = X_1X_1'$, which implies $x_1 = x + 1$:

$$\begin{aligned} \text{Area}(AX_1'X_1) &= \frac{1}{2}bh \\ &= \frac{1}{2}x_1^2 \end{aligned}$$

Therefore

$$F_X(-1 \leq x \leq 0) = \frac{1}{2}(x+1)^2$$

Case 2: $0 < x \leq 1$. This implies $F_X(x)$ is the area of ABX'_2X_2 . Note $x = MX_2$, and that $X_2X'_2 = X_2C$

$$\begin{aligned}
 \text{Area}(ABX'_2X_2) &= \text{Area}(\triangle ABM) + \text{Area}(MBX'_2X_2) \\
 \text{Area}(\triangle ABM) &= \frac{1}{2} \\
 \text{Area}(MBX'_2X_2) &= \text{Area}(MBC) - \text{Area}(X_2X'_2C) \\
 \text{Area}(MBC) &= \text{Area}(ABM) = \frac{1}{2} \\
 \text{Area}(X_2X'_2C) &= \frac{1}{2}bh \\
 &= \frac{1}{2} \cdot (1-x) \cdot (1-x) \\
 &= \frac{1}{2}(1-x)^2
 \end{aligned}$$

This is enough information to express F_X :

$$\begin{aligned}
 \implies F_X(0 < x \leq 1) &= \left(\frac{1}{2}\right) + \left(\frac{1}{2} - \frac{1}{2}(1-x)^2\right) \\
 &= 1 - \frac{1}{2}(1 - 2x + x^2) \\
 &= 1 - \frac{1}{2} + x - \frac{1}{2}x^2 \\
 &= -\frac{1}{2}x^2 + x + \frac{1}{2}
 \end{aligned}$$

Combining cases:

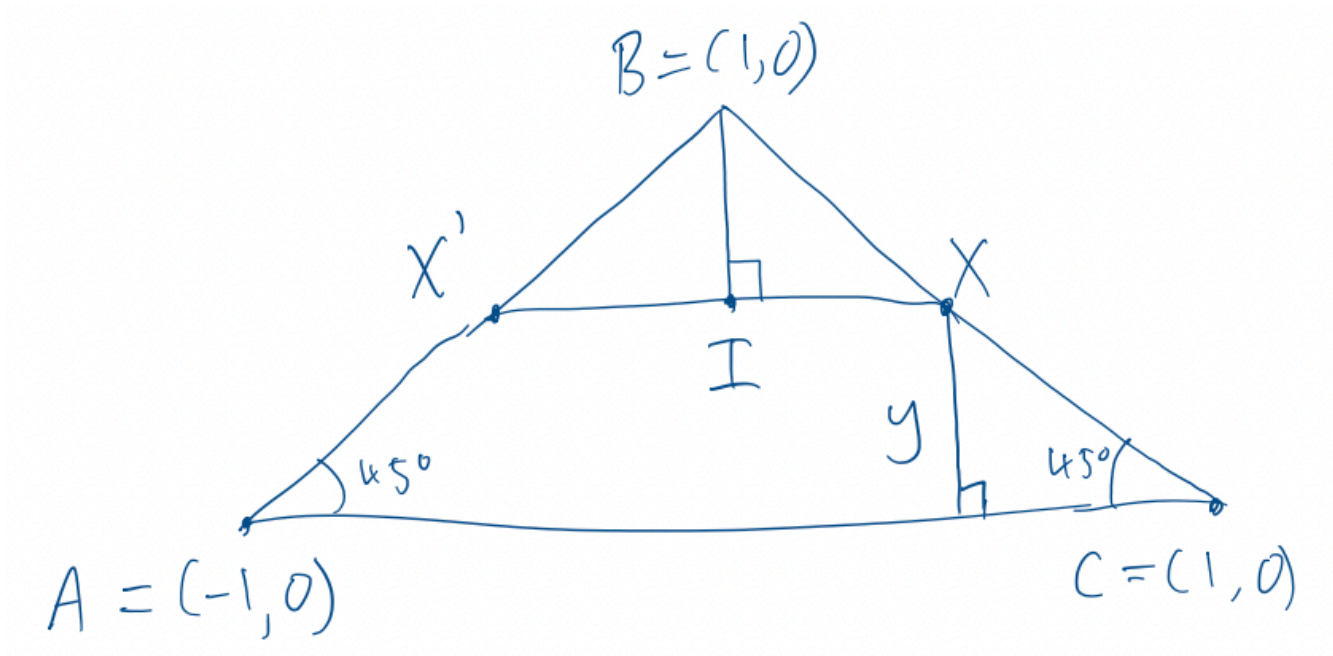
$$\begin{aligned}
 \implies F_X &= \begin{cases} 0 & : x < -1 \\ \frac{1}{2}(1+x)^2 & : -1 \leq x \leq 0 \\ -\frac{1}{2}x^2 + x + \frac{1}{2} & : 0 < x \leq 1 \\ 1 & : x > 1 \end{cases} \\
 \therefore f_X(x) &= \begin{cases} x+1 & : -1 \leq x \leq 0 \\ -x+1 & : 0 < x \leq 1 \\ 0 & : \text{otherwise} \end{cases}
 \end{aligned}$$

This so happens to be the geometric shape of D on a cartesian plane, ABC.

2.3.2 PDF Y

$$F_Y(y \leq 0) = 0$$

$$F_Y(y \geq 1) = 1$$

Figure 2: Geometric rendering of D showing the cases of F_Y

Case $0 < y < 1$

$$\begin{aligned}
 F_Y(y) &= \text{Area}(AX'XC) \\
 &= \text{Area}(\triangle ABC) - \text{Area}(\triangle XBX') \\
 &= 1 - 2 \cdot \text{Area}(\triangle XIB)
 \end{aligned}$$

Note $IB = 1 - y = IX$

$$\begin{aligned}
 \text{Area}(\triangle XIB) &= \frac{1}{2}bh \\
 &= \frac{1}{2}(1 - y)^2 \\
 &= \frac{1}{2}(y^2 - 2y + 1) \\
 \implies F_Y &= 1 - (y^2 - 2y + 1) \\
 &= -y^2 + 2y
 \end{aligned}$$

Combining cases

$$\begin{aligned}
 F_Y(y) &= \begin{cases} 1 & : y \geq 1 \\ -y^2 + 2y & : 0 < y < 1 \\ 0 & : y < 0 \end{cases} \\
 f_Y(y) &= \begin{cases} -2y + 2 & : 0 < y < 1 \\ 0 & : \text{otherwise} \end{cases}
 \end{aligned}$$

The geometric interpretation of this is not as intuitive to realise. Morph A to $(0, 2)$, B to $(0, 1)$ and C to $(0, 0)$, and the initial value of $f_X(0) = AC$ is now placed on the y-axis.

2.4 Part d)

The continuous independence rule can be stated like so:

$$f_{X,Y}(< x, y >) = f_X(x) \cdot f_Y(y) \quad \forall x, y \in \mathbb{R}$$

Suppose $< x, y > \in D$.

$$\begin{aligned} \text{LHS} &= f_{X,Y}(< x, y >) \\ &= 1 \\ \text{RHS} &= f_Y(y) \cdot f_X(x) \\ &= (-2y + 2) \cdot \begin{cases} x + 1 & : -1 \leq x \leq 0 \\ -x + 1 & ; 0 < x \leq 1 \end{cases} \end{aligned}$$

Suppose further $< x, y > = < 0, 0 >$

$$\begin{aligned} \text{RHS} &= (-2 \cdot 0 + 2)(0 + 1) \\ &= 2 \end{aligned}$$

Therefore there exists an $< x, y > \in D$ such that the independence rule fails, $\text{LHS} \neq \text{RHS}$. Therefore X and Y are not independent by counterexample.

2.5 Part e)

Due to the symmetry across the $x = 0$ "line" and the fact that $\text{Cov}(X, Y) = \int_{x,y} xyP(\{X = x\} \cap \{Y = y\})$ is negative (or zero) for $x < 0$ and positive (or zero) for $x > 0$, this demonstrates intuitively that $\text{Cov}(X, Y) = 0$.

Somehow this seems dissatisfying to me. $\text{Cov}(X, Y) = E(XY) - E(X)E(Y)$ is another way to evaluate $\text{Cov}(X, Y)$. This method requires evaluating a double integral or equivalent.

$$\begin{aligned} E(X) &= \int_{-\infty}^{+\infty} x f_X(x) dx \\ &= \int_{-1}^0 x(x+1) dx + \int_0^1 x(-x+1) dx \\ &= \int_{-1}^0 x^2 + x dx + \int_0^1 -x^2 + x dx \\ &= \left. \frac{1}{3}x^3 + \frac{1}{2}x^2 \right|_{x=-1}^{x=0} + \left. -\frac{1}{3}x^3 + \frac{1}{2}x^2 \right|_{x=0}^{x=1} \\ &= (0+0) - \left(\frac{1}{3}(-1)^3 + \frac{1}{2}(-1)^2 \right) + \left(-\frac{1}{3}(1)^3 + \frac{1}{2}(1)^2 \right) - (0+0) \\ &= +\frac{1}{3} + \frac{1}{2} - \frac{1}{3} + \frac{1}{2} \\ &= 1 \end{aligned}$$

$$\begin{aligned} E(Y) &= \int_{-\infty}^{+\infty} y f_Y(y) \, dy \\ &= \int_0^1 -2y + 2 \, dy \\ &= -y^2 + 2y \Big|_{y=0}^{y=1} \\ &= (-(1)^2 + 2(1)) - (0 + 0) \\ &= -1 + 2 \\ &= 1 \end{aligned}$$

$$\begin{aligned} E(XY) &= \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} xy f_X(x) f_Y(y) \, dx \, dy \\ &= \int_{-\infty}^{+\infty} x \left(\int_0^1 y f_Y(y) \, dy \right) f_X(x) \, dx \\ &= E(Y) \int_{-1}^{+1} x f_X(x) \, dx \\ &= E(Y) \cdot E(X) \\ &= 1 \end{aligned}$$

Therefore $\text{Cov}(X, Y) = 1 - 1 \cdot 1 = 0$

3 Question 3

The sample space Ω is parameterized by the number of orders in the month.

$$\Omega_n = \{(x_1, x_2, \dots, x_n) : \forall n \in \mathbb{Z} \text{ and } x_n \in \{1, 2, 3, 4\}\}$$

Let X_i be the random variable for the number of items in the i 'th order. Notice X_i are all independent and identically distributed (iid).

$$X_i((x_1, \dots, x_i, \dots, x_n) \in \Omega_n) = x_i$$

From the question statement, for all $i \in \mathbb{Z}$:

$$P(\{X_i = 1\}) = 0.54$$

$$P(\{X_i = 2\}) = 0.22$$

$$P(\{X_i = 3\}) = 0.15$$

$$P(\{X_i = 4\}) = 0.09$$

Let T be the random variable for the total number of items:

$$T((x_1, \dots, x_n) \in \Omega_n) = \sum_{i=1}^n x_i$$

3.1 Part a)

Note $n = 500$. The question is asking to evaluate $P(\{T < 900\}) \geq 0.95$, equivalently $P(\{T > 900\}) < 0.05$.

$$\begin{aligned}
E(X_i) &= \sum_x xP(\{X_i = x\}) \\
&= 1 \cdot 0.54 + 2 \cdot 0.22 + 3 \cdot 0.15 + 4 \cdot 0.09 \\
&= 1.79 \\
\text{Var}(X_i) &= E[(X_i - E(X_i))^2] \\
&= \sum_{c \in \Omega} (X_i(c) - 1.79)^2 P(\{c\}) \\
&= \sum_x (x - 1.79)^2 P(\{X_i = x\}) \\
&= (1 - 1.79)^2 \cdot 0.55 + (2 - 1.79)^2 \cdot 0.22 + (3 - 1.79)^2 \cdot 0.15 + (4 - 1.79)^2 \cdot 0.09 \\
&= 1.0059 \\
E(T) &= \sum_{i=1}^{500} E(X_i) \\
&= 500 \cdot 1.79 \\
&= 895 \\
\Rightarrow \mu_T &= 895 \\
\text{Var}(T) &= \sum_{i=1}^{500} \text{Var}(X_i) \\
&= 500 \cdot 1.0059 \\
&= 502.95 \\
\Rightarrow \sigma_T &= \sqrt{502.95} \approx 22.4265
\end{aligned}$$

From the Central Limit Theorem (CLT), we can assume T follows a normal distribution. $n = 500$ is a reasonably large n for this to be a good approximation.

$$T \sim \mathcal{N}(\mu = 895, \sigma = 22.4265)$$

$$\begin{aligned}
P(\{T > 900\}) &= P\left(\left\{\frac{T - \mu}{\sigma} > \frac{900 - 895}{22.4265}\right\}\right) \\
&= P(\{Z > 0.2230\}) \\
&= 1 - P(\{Z < 0.2230\}) \text{ Using stats table} \\
&= 1 - 0.5871 \\
&= 0.4129
\end{aligned}$$

\therefore No, there is approximately a 41% chance of exceeding the 900-item limit, which is significantly higher than the threshold 5%, therefore the company cannot process $n = 500$ orders.

3.2 Part b)

With an unknown n , the aim is to find the greatest n with the constraint that $P(\{T < 900\}) \geq 0.95$. Note $T \sim \mathcal{N}(\mu_T, \sigma_T)$

$$\begin{aligned}
E(T) &= 1.79n = \mu_T \\
\text{Var}(T) &= 1.0059n \\
\implies \sigma_T &= \sqrt{1.0059n} \\
P(\{T < 900\}) &= P\left(\left\{\frac{T - \mu}{\sigma} < \frac{900 - \mu}{\sigma}\right\}\right) \\
&= P\left(\left\{Z < \frac{900 - \mu}{\sigma}\right\}\right) \\
\text{let } z &= \frac{900 - \mu}{\sigma} \\
\implies P(\{Z < z\}) &\geq 0.95
\end{aligned}$$

From stats table $z \geq 1.65$. Ignoring the inequality until the end, and skipping algebra steps:

$$\begin{aligned}
z = 1.65 &= \frac{900 - \mu}{\sigma} \\
1.65\sqrt{1.0059n} &= 900 - 1.79n \\
0 &= 1.79^2n^2 + (-1.65^2 \cdot 1.0059 - 2 \cdot 900)n + 900^2
\end{aligned}$$

Using the quadratic equation

$$\implies n = 482.5 \text{ and } n = 523.96$$

Since we know $z = \frac{900 - \mu}{\sigma} \geq 1.65$, we can reject $n = 523.96$:

$$\begin{aligned}
1.65 &\leq \frac{9001.79 \cdot 523.96}{\sqrt{1.0059 \cdot 523.96}} \\
1.65 &\leq \approx -1.577
\end{aligned}$$

And $n = 482.5$ is reasonable, considering rounding imprecision.

$$\begin{aligned}
1.65 &\leq \frac{9001.79 \cdot 482.5}{\sqrt{1.0059 \cdot 482.5}} \\
1.65 &\leq \approx 1.649
\end{aligned}$$

$\therefore n \approx 482.5$, which erring on the side of $P(\{T < 900\}) > 0.95$ requires us to round down to $n = 482$. Therefore the largest number of orders the company can process in the month is 482, before it exceeds the limit with more than a 5% probability.