Question 1

(a)

To begin, we can compute the underage (c_u) and overage cost (c_0) as follows:

$$c_u = \$10 - \$5 = \$5$$
$$c_o = \$5 - \$2 = \$3$$

To determine Q^*, the optimal number of drinks to prepare, we compute the critical ratio:

$$F(Q^*) = \frac{c_u}{c_o + c_u} = \frac{5}{3+5} = 0.625$$

Given the probability distribution table, we compute the cumulative distribution table as follows. We find that $F(Q^*) = 0.625$ lies in between 60 and 70. Following procedures outlined in the book *Production and Operations Analysis*, we choose the higher value and thus the economic number of drinks for Piney Drinks to prepare is $Q^* = 70$.

Q	$f(Q)$	$F(Q)$
30	0.05	0.05
40	0.12	0.17
50	0.2	0.37
60	0.24	0.61
70	0.17	0.78
80	0.14	0.92
90	0.08	1

Table 1: Probability and Cumulative Distribution Table of Piney Drinks Sold

(b)

Now, we assume that Piney Drink's demand changes to a normal distribution with mean 60 and variance 81 ($D \sim \mathcal{N}(60, 81)$). The critical ratio $F(Q^*)$ remains the same, but now we implement a different procedure in finding the exact value of Q^*. By definition:

$$F(Q^*) = \mathbb{P}(D \leq Q^*) = 0.625$$

and as such Q^* is the 65th percentile of the demand distribution. By this information and referencing the normal distribution table, we can find simply find Q^* as follows:

$$\begin{aligned} Q^* &= \sigma z + \mu \\ z &= 0.32 \quad (\text{since } \mathbb{P}(Z \leq 0.32) = 0.625) \\ Q^* &= \sqrt{81} \cdot 0.32 + 60 = 62.88 \\ \boxed{Q^* = 63} &\quad (\text{assuming we are not restricted to batches of 10}) \end{aligned}$$

(c)

On Sundays, Piney Drink's demand follows a uniform distribution (assumed to be continuous for simplicity's sake) between 30 and 90. As such, we can define the Piney's demand probability distribution function

as follows:

$$f(x) = \left\{ \begin{array}{ll} 1/60, & \text{if } 30 \leq x \leq 90 \\ 0, & \text{otherwise} \end{array} \right\}$$

As there are no changes to the operational costs, the critical ratio remains unchanged and thus $F(Q^*) = 0.625$. By integrating the pdf defined above, we will be able to find Q^*.

$$\begin{aligned} F(Q^*) &= \int_{30}^{Q^*} \frac{1}{60} dx \\ 0.625 &= \frac{Q^* - 30}{60} \\ Q^* &= 67.5 \\ \boxed{Q^* = 68} & \quad \text{(assuming we are not restricted to batches of 10)} \end{aligned}$$

Question 2

(a)

We are given the following set of information.

$K = \$50$
$c = \$10/\text{unit}$
$\tau = 6$ months
$h = 0.2c = \$2/\text{unit}$
$p = \$25/\text{unit}$

$D \sim U(270, 330)$ for each cycle (6 months)
$\mu = (270 + 330)/2 = 300$ units/cycle
$\lambda = \mu \cdot 2 = 600$ units/year

Plushy Toys' total average annual cost arise from three contributors: holding, setup, and penalty for unmet demand costs. This gives the following equation for $G(Q, R)$:

$$G(Q, R) = h(Q/2 + R - \lambda\tau) + K\lambda/Q + p\lambda n(R)/Q$$

where Q is the quantity of toys ordered, R is the inventory reorder level, and $n(R)$ is the expected number of shortages that occur in one cycle.

$$\begin{aligned} n(R) = \mathbb{E}(\max(D - R, 0)) &= \int_R^\infty (x - R) f(x) dx \\ f(x) &= \left\{ \begin{array}{ll} 1/60, & \text{if } 270 \leq x \leq 330 \\ 0, & \text{otherwise} \end{array} \right\} \\ \therefore n(R) &= \int_R^{330} \frac{x - R}{60} dx \end{aligned}$$

To find the optimal Q^*, we need to solve for Q and R iteratively using the following equations.

$$Q = \sqrt{\frac{2\lambda(K + pn(R))}{h}} \qquad\qquad 1 - F(R) = \frac{Qh}{p\lambda}$$

$1 - F(R)$ is simply the right-tail area of the demand p.d.f. given R ($\mathbb{P}(D \geq R)$ and as such the following equation can be formulated to solve for R along the way.

$$1 - F(R) = \frac{330 - R}{60}$$

To begin, set $Q_0 = \text{EOQ}$. By the formula for the basic EOQ model, we get:

$$Q_0 = \sqrt{\frac{2K\lambda}{h}} = \sqrt{\frac{2 \cdot 50 \cdot 600}{2}} = 173.205080757$$
$$1 - F(R_0) = \frac{Q_0 h}{p\lambda} = \frac{173.205080757 \cdot 2}{25 \cdot 600} = 0.0230940107676$$

By definition, $F(R_0) = \mathbb{P}(D \leq R_0)$ and thus we find R_0.

$$1 - F(R_0) = \frac{330 - R_0}{60} \longrightarrow R_0 = 328.614359354$$

Using R_0, we find $n(R_0)$, and, with that, Q_1.

$$n(R_0) = \int_{R_0}^{330} \frac{x - R_0}{60} dx = 0.016$$
$$Q_1 = \sqrt{\frac{2\lambda(K + pn(R_0))}{h}} = \sqrt{\frac{2 \cdot 600(50 + 25 \cdot 0.016)}{2}} = 173.896520954$$

We now find R_1 and $n(R_1)$.

$$1 - F(R_1) = \frac{Q_1 h}{p\lambda} = \frac{173.896520954 \cdot 2}{25 \cdot 600} = 0.0231862027939$$
$$1 - F(R_1) = \frac{330 - R_1}{60} \longrightarrow R_1 = 328.608827832$$
$$n(R_1) = \int_{R_1}^{330} \frac{x - R_1}{60} dx = 0.016128$$

We now compute Q_2 and R_2.

$$Q_2 = \sqrt{\frac{2\lambda(K + pn(R_1))}{h}} = \sqrt{\frac{2 \cdot 600(50 + 25 \cdot 0.016128)}{2}} = 173.902041391$$
$$1 - F(R_2) = \frac{Q_2 h}{p\lambda} = \frac{173.902041391 \cdot 2}{25 \cdot 600} = 0.0231869388521$$
$$1 - F(R_2) = \frac{330 - R_2}{60} \longrightarrow R_2 = 328.608783669$$
$$n(R_2) = \int_{R_2}^{330} \frac{x - R_2}{60} dx = 0.016129024$$

Here, notice that Q_1 and Q_2 are very close. As such, we have reached convergence and our optimal order quantity and reorder level, rounded to the nearest integer, are:

$$\boxed{Q^* = 174} \qquad \boxed{R^* = 329}$$

(b)

Not stocking out during the order-arrival lead time is the same as having demand not exceed the reorder level. As such, the probability of not stocking out during the lead time is simply given by $\mathbb{P}(D \leq R^*)$, where:

$$\mathbb{P}(D \leq R^*) = F(R^*) = \frac{329 - 270}{60} = 0.983333333333$$
$$\boxed{\mathbb{P}(D \leq R^*) = 98.3\bar{3}\%}$$

We conclude that there will be no stock-outs in $98.3\bar{3}$ percent of the order cycles.

(c)

The proportion of demand that is met from stock is given by:

$$\beta = 1 - n(R^*)/Q^*$$

Given $R^* = 329$ and $Q^* = 174$:

$$n(R^*) = \int_{329}^{330} \frac{x - 330}{60} dx = 0.00833333333333$$
$$\beta = 1 - 0.00833333333333/174$$
$$\boxed{\beta = 0.99995210728}$$

We conclude that nearly all of the demand is fulfilled by stock.

Question 3

If we now assume that any unmet demand is lost, then we need to modify our equation for the total average annual cost as follows.

$$G(Q, R) = h\Big(Q/2 + R - \lambda\tau + n(R)\Big) + K\lambda/Q + c\lambda + p\lambda n(R)/Q$$

To minimize $G(Q, R)$, one must iteratively solve for a new pair of equations:

$$Q = \sqrt{\frac{2\lambda(K + pn(R))}{h}} \qquad\qquad 1 - F(R) = \frac{Qh}{Qh + p\lambda}$$

Similar to (a), we assume $Q_0 = \text{EOQ}$ to find Q_0 and R_0.

$$Q_0 = \sqrt{\frac{2K\lambda}{h}} = \sqrt{\frac{2 \cdot 50 \cdot 600}{2}} \qquad 1 - F(R_0) = \frac{Q_0 h}{Q_0 h + p\lambda} = \frac{173.205080757 \cdot 2}{173.205080757 \cdot 2 + 25 \cdot 600}$$
$$\boxed{Q_0 = 173.205080757} \qquad = 0.0225727162162$$
$$1 - F(R_0) = \frac{330 - R_0}{60} \longrightarrow \boxed{R_0 = 328.645637027}$$

Using the above, we compute $n(R_0)$ and, with it, Q_1 and R_1.

$$n(R_0) = \int_{R_0}^{330} \frac{x - R_0}{60} dx = 0.0152858255214 \qquad 1 - F(R_1) = \frac{Q_1 h}{Q_1 h + p\lambda}$$
$$Q_1 = \sqrt{\frac{2\lambda(K + pn(R_0))}{h}} \qquad = \frac{173.865716525 \cdot 2}{173.865716525 \cdot 2 + 25 \cdot 600}$$
$$= \sqrt{\frac{2 \cdot 600(50 + 25 \cdot 0.0152858255214)}{2}} \qquad = 0.0226568619973$$
$$\boxed{Q_1 = 173.865716525} \qquad 1 - F(R_1) = \frac{330 - R_1}{60} \longrightarrow \boxed{R_1 = 328.64058828}$$

We can perform another iteration to find $n(R_1)$, Q_2, and R_2.

$$n(R_1) = \int_{R_1}^{330} \frac{x - R_1}{60} dx = 0.015400001867 \qquad 1 - F(R_2) = \frac{Q_2 h}{Q_2 h + p\lambda}$$

$$Q_2 = \sqrt{\frac{2\lambda(K + pn(R_1))}{h}} \qquad = \frac{173.870641651 \cdot 2}{173.870641651 \cdot 2 + 25 \cdot 600}$$

$$= \sqrt{\frac{2 \cdot 600(50 + 25 \cdot 0.015400001867)}{2}} \qquad = 0.0226574892606$$

$$\boxed{Q_2 = 173.870641651} \qquad 1 - F(R_2) = \frac{330 - R_2}{60} \longrightarrow \boxed{R_2 = 328.640550644}$$

As Q_1 and Q_2 are incredibly close, we can conclude that convergence has been reached. Thus, our optimal order quantity and reorder level, rounded to the nearest integer, are:

$$\boxed{Q^* = 174} \qquad \boxed{R^* = 329}$$

Question 4

To achieve a 98.5% fill rate ($\beta = 0.985$) using the accurate formulation of type 2 service, we substitute the equation for the penalty cost into the total cost equation.

$$p = \frac{Qh}{(1 - F(R))\lambda}$$

It can be shown that the resulting equation can be simplified into the following equation for Q (also known as the SOQ formula). By solving this equation simultaneously with $n(R)$, we minimize our total cost.

$$Q = \frac{n(R)}{1 - F(R)} + \sqrt{\frac{2K\lambda}{h} + \left(\frac{n(R)}{1 - F(R)}\right)^2} \qquad n(R) = (1 - \beta)Q = 0.015Q$$

Now we can solve for Q_0 and $n(R_0)$. We set $Q_0 = \text{EOQ}$.

$$Q_0 = \sqrt{\frac{2K\lambda}{h}} = \sqrt{\frac{2 \cdot 50 \cdot 600}{2}} \qquad n(R_0) = 0.015Q_0$$

$$\boxed{Q_0 = 173.205080757} \qquad \boxed{n(R_0) = 2.59807621135}$$

Using the above, we solve for R_0 and $F(R_0)$.

$$n(R_0) = \int_{R_0}^{330} \frac{x - R_0}{60} dx$$

$$2.59807621135 \cdot 60 = \left[\frac{1}{2}x^2 - xR_0\right]_{R_0}^{330}$$

$$2.59807621135 \cdot 60 = \frac{1}{2}(330)^2 - 330R_0 - \frac{1}{2}R_0^2 + R_0^2$$

$$R_0 = 312.343 \vee R_0 = 347.657 \quad \text{(solved using desmos)}$$

Here, we reject $R_0 = 347.657$ because this is larger than 330, the upper bound of the demand distribution. Choosing $R_0 = 347.657$ would mean we have a fill rate of 1, which differs from what is requested. Thus, $R_0 = 312.343$ and we get $F(R_0)$ through the right-tail cdf of the demand.

$$1 - F(R_0) = \frac{330 - R_0}{330 - 270} = 0.294283333333$$

We proceed to our next iteration and solve for Q_1 and $n(R_1)$.

$$Q_1 = \frac{n(R_0)}{1 - F(R_0)} + \sqrt{\frac{2\,(50)\,(600)}{(2)} + \left(\frac{n(R_0)}{1 - F(R_0)}\right)^2} \qquad n(R_1) = 0.015Q_1$$

$$\boxed{Q_1 = 182.25842018} \qquad \boxed{n(R_1) = 2.7338763027}$$

Using the above, we solve for R_1 and $F(R_1)$.

$$n(R_1) = \int_{R_1}^{330} \frac{x - R_1}{60} dx$$

$$2.7338763027 \cdot 60 = \left[\frac{1}{2}x^2 - xR_1\right]_{R_1}^{330}$$

$$2.7338763027 \cdot 60 = \frac{1}{2}(330)^2 - 330R_1 - \frac{1}{2}R_1^2 + R_1^2$$

$$R_1 = 311.887 \vee R_1 = 348.113 \quad \text{(solved using desmos)}$$

Here, we reject $R_1 = 347.657$ because this is larger than 330, the upper bound of the demand distribution. Choosing $R_1 = 348.113$ would mean we have a fill rate of 1, which differs from what is requested. Thus, $R_1 = 311.887$ and we get $F(R_1)$ through the right-tail cdf of the demand.

$$1 - F(R_1) = \frac{330 - R_1}{330 - 270} = 0.301883333333$$

We proceed to our next iteration and solve for Q_2 and $n(R_2)$.

$$Q_2 = \frac{n(R_1)}{1 - F(R_1)} + \sqrt{\frac{2\,(50)\,(600)}{(2)} + \left(\frac{n(R_1)}{1 - F(R_1)}\right)^2} \qquad n(R_2) = 0.015Q_2$$

$$\boxed{Q_2 = 182.497737561} \qquad \boxed{n(R_2) = 2.73746606341}$$

Using the above, we solve for R_2 and $F(R_2)$.

$$n(R_2) = \int_{R_2}^{330} \frac{x - R_2}{60} dx$$

$$2.73746606341 \cdot 60 = \left[\frac{1}{2}x^2 - xR_2\right]_{R_2}^{330}$$

$$2.73746606341 \cdot 60 = \frac{1}{2}(330)^2 - 330R_2 - \frac{1}{2}R_2^2 + R_2^2$$

$$R_2 = 311.876 \vee R_2 = 348.124 \quad \text{(solved using desmos)}$$

Here, we reject $R_2 = 348.124$ because this is larger than 330, the upper bound of the demand distribution. Choosing $R_2 = 348.124$ would mean we have a fill rate of 1, which differs from what is requested. Thus, $R_2 = 348.124$.

Our (Q, R) values are also within 1 unit from the previous iteration and as such we have our final optimal solution (rounded to nearest integer).

$$\boxed{Q^* = 182}$$

$$\boxed{R^* = 312}$$

Question 5

This paper [2] presents a novel method utilizing **Reinforcement Learning (RL)** to address supply chain and operations research (SCOR) problems, particularly in the area of **Multi-Echelon Inventory Management (MEIM)**. RL is a highly promising approach to solve stochastic sequential decision-making problems, but it has mostly been utilised for game and robotics development. Most RL algorithms are developed for game-based benchmarks unsuitable for SCOR problems, thus motivating the authors to develop a more appropriate approach to implement RL algorithms in SCOR.

Many researchers have proposed optimisation techniques in the area of multi-echelon inventory management. However, traditional research methods still face many difficulties in practical application, namely: (1) current methods involve a **large solution space** entailing many nodes with complex network relationships, which causes **large solution times**, and (2) **large operational uncertainty** arise, both exogenous and endogenous, due to fluctuations in demand, price, and production among other factors. As such, deterministic supply chain models are largely erroneous or inefficient, prompting the authors to develop a new approach based on RL.

Reinforcement learning (RL) is a subfield of machine learning that computes an optimal decision-making policy by interacting with the underlying system via an iterative approach [2]. The RL problem is formalized as a **Markov Decision Process (MDP)**, which provides an effective framework for modelling uncertain, sequential decision making problems [1]. RL can transcend traditional SCOR research methods for two reasons: (1) it enables **flexible use** of different classes of system model and forms of uncertainty, and (2) it identifies operational decisions from the predictions of optimal policy function approximations, enabling **faster solution times**. Effectively, the two problems in traditional SCOR methods mentioned earlier can be solved using RL.

Currently, RL algorithms for SCOR face certain shortcomings, particularly due to their dependence on Q-learning and policy gradient algorithms to identify an optimal solution. When these algorithms are combined with nonlinear function approximation, the algorithms are **generally unstable** in training (vanishing or exploding gradient problem) and get stuck easily in **local optimal solutions**. The paper thus proposes a new approach in RL which does not involve derivatives, thereby solving these challenges.

A key challenged faced is that RL algorithms often show great differences in performance across different use cases. Thus, the new method proposed has to be safe, robust, and performance consistent. For this reason, the paper introduces a **hybrid derivative-free algorithm** for policy optimization and risk sensitive formulations for distributional RL. The proposed method combines four algorithms: evolution strategy (ES), artificial bee colony (ABC), particle swarm optimization (PSO), and simulated annealing (SA). By combining these algorithms, the method balances **exploitation** and **exploration** in searching for the optimal solution in a given SCOR problem. Furthermore, the algorithm is proposed through a distributional RL formulation, which accounts for low-probability, worse-case events in MEIM—something which ordinary RL formulations fail to account for.

To evaluate the proposed method, the paper uses a multi-echelon supply chain inventory management problem and two common operations research studies as comparison. The hybrid method proposed outperforms the state-of-the-art **proximal policy optimization (PPO)** algorithm by approximately **11%**, the **MILP** formulation by **20%**, and the **LP relaxation** of the problem by **6%**. Furthermore, the authors investigated the hybrid algorithm's performance for optimization of the distributional **CVaR** (conditional value at risk) objective in order to minimise worst-case performances. Through 1000 **Monte Carlo** samples, it is found that the proposed distributional RL algorithm yields a return value which is **7.2%** better compared to expectation-based RL algorithms. Essentially, the distributional RL successfully learns a **risk-sensitive policy** and is greatly improved compared to non-distributional RLs.

In conclusion, whereas most literature advocates for the traditional policy gradient approach to tackling SCOR problems, this paper introduces a hybrid derivative-free algorithm based on distributional RL to significantly improve solution times and risk-sensitivity compared to pre-existing state-of-the-art techniques. Their algorithm is more efficient and flexible and demonstrate promising value to the industry.

References

[1] Martin L. Puterman. *Markov Decision Processes: Discrete Stochastic Dynamic Programming*. Hoboken, New Jersey: John Wiley Sons, Inc., 2005.

[2] Guoquan Wu, Miguel Ángel de Carvalho Servia, and Max Mowbray. "Distributional reinforcement learning for inventory management in multi-echelon supply chains". In: *Digital Chemical Engineering* 6.100073 (Mar. 2023). DOI: https://doi.org/10.1016/j.dche.2022.100073.