Sequential Decision-Making for Inventory Control using Bayesian Decision Theory

Jonas Petersen
Ander Runge Walther

November 21, 2024

Abstract

Inventory control is a core aspect of supply chain management, balancing customer demand fulfillment against the costs of holding and ordering inventory. This paper addresses inventory management through the lens of Bayesian decision theory, enabling probabilistic decision-making under uncertainty. By establishing a framework for sequential decision-making over an infinite time horizon, an optimal solution tailored to a specific asymmetric cost function is derived, offering insights for practical implementations in inventory management.

1 Introduction

Inventory management is a critical component of supply chain management that involves overseeing the ordering, storage, and use of a company's inventory, including raw materials, components, and finished goods. Effective inventory control ensures that a company has adequate stock to meet demand while minimizing costs associated with holding inventory. In practice, this often requires a balance between minimizing stockouts, which disrupt service levels, and limiting overstocking, which incurs unnecessary holding costs.

The coordination of inventory and transportation decisions has gained significant attention over recent decades, driven by the need for specific practices like vendor-managed inventory (VMI), third-party logistics (3PL), and time-definite delivery (TDD) (see, e.g., (Çetinkaya and Lee, 2000; Alumur, 2012; Gürler, 2014)). These programs aim to optimize the balance between inventory holding costs and transportation costs. Traditional inventory models often assume immediate delivery to meet demand, yet this can be inefficient due to the fixed costs of transportation, prompting companies to adopt shipment consolidation policies that merge smaller demands into larger, less frequent shipments (Çetinkaya and Bookbinder, 2003; Higginson and Bookbinder, 1994).

Three commonly implemented shipment consolidation policies—quantity-based, time-based, and hybrid—are particularly useful in improving cost efficiency by regulating shipment size or timing. Each policy type has distinct impacts on performance metrics, such as delay penalties, average inventory, and annual costs, guiding companies toward designing more efficient inventory-transportation systems (Çetinkaya, 2005; Wei, 2020; Çetinkaya and Lee, 2006).

While these approaches provide valuable frameworks, they do not explicitly account for the uncertainties inherent in demand fluctuations or the associated asymmetric costs of overstocking and understocking. In settings where inventory decisions must be made sequentially, uncertainty can be more effectively managed through Bayesian decision theory, which offers a probabilistic approach to decision-making under uncertainty. This framework allows for more refined, sequential adjustments to inventory based on demand forecasting, which dynamically incorporates new data over time.

In this study, Bayesian decision theory is utilized to address a sequential inventory control problem. Specifically, an optimal decision-making framework is derived based on a discounted asymmetric cost function, considering the impact of decisions made infinitely far into the future. Our model involves a decision-maker (henceforth referred to as the Robot) who manages stock levels over time in an interactive environment shaped by a random demand process. Through this setup, a mathematically rigorous and computationally feasible solution for optimal inventory decisions that account for uncertainty and asymmetric costs is derived.

2 Sequential Decision-Making Framework

To formalize the sequential decision-making problem, consider a stock management scenario in which the Robot is set in a game against an entity termed Nature, which represents environmental randomness (Lavalle, 2006). In this game the Robot and Nature each have to make a decision, $u \in \Omega_U \subset \mathbb{Z}_{\geq 0}$ and $s \in \Omega_S \subset \mathbb{Z}_{\geq 0}$ respectively, for each time step $t \in [1, T]$ into the future.

Definition 1 (Stock level)**.** *Given an initial stock N_0, the stock level at time* *t evolves according to*

$$N_t \equiv N_0 + \sum_{t'=1}^{t}(u_{t'-L} - s_{t'}) \\ = N_0 + \upsilon_t - \zeta_t \quad , \tag{1}$$

where $u_{t'-L}$ represents the decision made at an earlier time due to a potential lag L and $\zeta_t \equiv \sum_{t'=1}^{t} s_{t'}$ and $\upsilon_t \equiv \sum_{t'=1}^{t} u_{t'-L}$. To support probabilistic decision-making, the Robot uses a probabilistic forecast based on data D

$$p(s_1, s_2, \dots s_T | D, I), \tag{2}$$

where I denotes any additional background information (Sivia and Skilling, 2006).

Definition 2 (Discounted Cost). *The Robot receives a numerical penalty, assigned by a cost function, depending on the decisions $\{u\}, \{s\}$ made over the forecast horizon. The cost C is assumed to be discounted and is defined as*

$$C = \sum_{t=1}^{T} \gamma_{disc}^{t-1} \left(h_t 1_{N_t>0} + c_t(1_{N_t>0} - 1)\right) N_t, \tag{3}$$

where $\gamma_{disc} \in [0, 1]$ is the discount factor, h_t and c_t represent storage (holding) and understocking costs at time t, respectively, and $1_{N_t>0}$ is the indicator function that equals 1 when $N_t > 0$ (see definition 1) and 0 otherwise.

Definition 3 (Optimal Policy). *Given the data D, the Robot's objective is to formulate a sequence of decision rules, called a policy, $\pi = \{U_0(D) = u_0, U_1(D) = u_1, \dots\}$, where each $U_j(D) = u_j$ minimizes the expected cost (definition 2)*

$$\pi^* = \arg\min_{\pi} \mathbb{E}[C|D, I] \tag{4}$$

over the probability distribution of definition 1. The optimal decisions rules satisfy the first- and second-order conditions

$$\frac{d}{dU_m}\mathbb{E}[C|D, I]\Big|_{U_m=U_m^*} = 0 \quad \forall m, \\ \frac{d^2}{dU_m^2}\mathbb{E}[C|D, I]\Big|_{U_m=U_m^*} > 0 \quad \forall m. \tag{5}$$

Theorem 1 (Optimal Policy Rule for Inventory Control). *Given the sequential decision-making framework and the asymmetric cost function of definition 2, the optimal policy for the Robot at each time step t is defined by (see* *appendix A for derivation)*

$$p(N_t^* > 0|D, I) = \frac{c_t}{c_t + h_t}, \tag{6}$$

where

$$N_t^* = N_0 + \upsilon_t^* - \zeta_t, \tag{7}$$

$\zeta_t - \zeta_{t-1} \geq 0$ and $\upsilon_t - \upsilon_{t-1} \geq 0$ by definition. However, if $N_0 + \upsilon_{t-1} > \zeta_t$, the optimal decision is $\upsilon_t^* = 0$. In this case, the condition in equation (6) is not strictly satisfied, as $p(N_t^* > 0|D, I)$ becomes irrelevant due to the sufficient inventory at time t.

3 Optimal Policy under Conditional Independence

Equation (6) can be written

$$\begin{aligned} p(N_t^* > 0|D, I) &= \sum_{s_1,\dots s_t} 1_{N_t \geq 0} p(s_1, \dots s_t|D, I) \sum_{s_{t+1},\dots s_T} p(s_{t+1}, \dots s_T|s_1, \dots s_t, D, I) \\ &= \sum_{\zeta_t=0}^{N_0+\upsilon_t} p(\zeta_t|D, I) \sum_{s_{t+1},\dots s_T} p(s_{t+1}, \dots s_T|\zeta_t, D, I) \end{aligned} \tag{8}$$

Assuming conditional independence $p(s_{t+1}, \dots s_T|\zeta_t, D, I) = p(s_{t+1}, \dots s_T|D, I)$

$$p(N_t^* > 0|D, I) = \sum_{\zeta_t=0}^{N_0+\upsilon_t} p(\zeta_t|D, I). \tag{9}$$

Combining eqation (9) with theorem 1 implies that the optimal policy is related to quantiles of ζ_t viz

$$\upsilon_t^* = \max(\mathrm{round}(\mathcal{Q}_{\frac{c_t}{c_t+h_t}} \zeta_t - N_0), 0), \tag{10}$$

where $\mathcal{Q}_q(X)$ denotes the q-quantile of the random variable X and the rounding ensures the decisions belong to $\Omega_U \subset \mathbb{Z}_{\geq 0}$.

3.1 Policy Efficiency Ratio

In order to gauge the effect of using the optimal policy (equation (10)), the associated expected cost is compared to the expected cost associated to a baseline policy via the policy efficiency ratio (PER)

$$\text{PER} \equiv \frac{\mathbb{E}[C|D,I]|_{\pi=\pi^*}}{\mathbb{E}[C|D,I]|_{\pi=\pi'}}, \tag{11}$$

where π^* is the optimal policy (theorem 1) and π' is the baseline policy (definition 4).

Definition 4 (Baseline Policy)**.** *The baseline policy is an (R,Q) inventory policy (Bartmann and Beckmann, 1992; Axsater, 2006), where the reorder point R determines when to reorder, and the batch quantity, Q, is based on the expected demand. Under this policy, the order quantity is determined by the expected demand, $\mathbb{E}[s_t|D,I]$, adjusted to bring the stock level up to the reorder point R*

$$\upsilon'_t = \max(\mathit{round}(\mathbb{E}[\zeta_t|D,I] + R - N_0), 0). \tag{12}$$

The rounding ensures the order quantity is an integer, and the maximum function prevents negative orders.

3.1.1 Scaling relations

In general the PER is a function of R, c_t, h_t, N_0 as well as the distribution (and its details) of ζ_t. Assuming i) $p(\zeta_t|D,I)$ follows a Poisson distribution and ii) conditional independence, the expected cost can be written (see appendix B)

$$\begin{aligned}\mathbb{E}[C|D,I] = \sum_{t=1}^{T}\gamma_{\text{disc}}^{t-1}\bigg((h_t+c_t)(N_0+\upsilon_t)&\frac{\Gamma(N_0+\upsilon_t+1,\lambda_t)}{\Gamma(N_0+\upsilon_t+1)}\\ -(h_h+c_t)\lambda_t&\frac{\Gamma(N_0+\upsilon_t,\lambda_t)}{\Gamma(N_0+\upsilon_t)}\\ -c_t(N_0+\upsilon_t-\lambda_t)\bigg)&.\end{aligned} \tag{13}$$

where Γ is the gamma function and $\lambda_t = \mathbb{E}[\zeta_t|D,I]$ is the expected value of the Poisson distribution. Under further loose approximations specified in appendix **??**, the PER can be approximated viz

$$\begin{aligned}\text{PER}|_{R\sim 0} &\sim \frac{4}{2+\frac{h}{c}+\frac{c}{h}},\\ \text{PER}|_{R\gg 0} &\sim \frac{L-1}{T-1} + \frac{1}{R(1+\frac{h}{c})}.\end{aligned} \tag{14}$$

Several things can be noted from (14), including: i) the PER is a function of $r \equiv \frac{c}{h}$, ii) the scaling according to r is inverted for $R \sim 0$ and $R \gg 0$ and iii) the PER approach a constant value for $R \to \infty$.

3.1.2 Example: Numerical Policy Efficiency Ratio

To provide an example of implementation (see source code on Github) and test equation (14) on an anecdotal level, consider the case where $N_0 = 37$, $L = 6$ and $U_j \in \mathbb{Z}$, $0 \leq U_j \leq 0$ for any time step. Data are mocked from a Poisson distribution with rate parameter $\lambda = 3$. 300 random samples are drawn to represent a mock time series, with the last 52 reserved as test data and the remaining as training data. Based on the training data, the λ is estimated using MLE and based on a Poisson distribution with this estimated rate parameter, a forecast of dimensions (number_of_samples, forecast_horizon), with number_of_samples=5 000 and forecast_horizon=52, is drawn. The optimal and baseline policies applied to yield decision over the forecast horizon. Using these decisions, the actual costs are computed using the test data. This is done 5 000 times to yield an estimate of the PER for a particular h, c. This process is repeated for $c \in \mathbb{Z}$, $1 \leq c \leq 100$ and $h \in \mathbb{Z}$, $1 \leq c \leq 20$, yielding the PERs shown in figure 1. Figure 1 clearly show the same pattern as figure **??**, however, with slight differences (as expected for an approximation); i) the maximum slightly exceeds 1 and ii) the minimum significantly deviates from 0. The former is likely due to random fluctuations from approximating the expected cost from 5 000 samples. Increasing the sample size gradually brings the maximum closer to 1.

========= TO THIS POINT ========== The latter? Is it an artefact, or due to numerical limitations? In the upper limit inaccuracy seems to be ~ 0.04. In the lower limit we have ~ 0.2... It can be inaccuracies of the approximation. I have defined the scale of the approximation based on deduction. Perhaps this is not perfectly accurate?

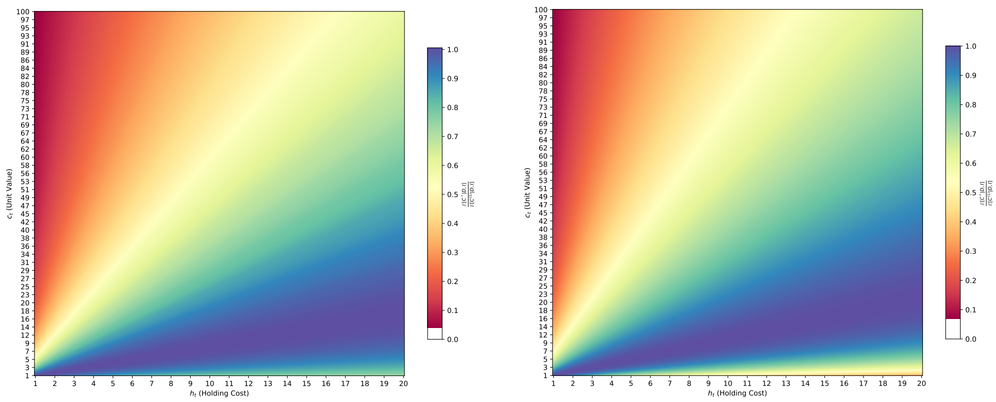

Figure 1: Heatmap of the numerical policy efficiency ratio calculated via simulation.

4 Conclusion

The derived decision rule offers a structured approach to managing inventory in a stochastic environment using Bayesian decision theory. The probabilistic framework accounts for demand uncertainty, guiding optimal actions based on expected costs. This work contributes to the field of inventory management by presenting an analytically derived solution, supporting practical applications in real-world inventory systems.

A Minimization of Expected Cost

To determine the policy, ξ^*, that minimize the expected cost $\mathbb{E}[C|D,I]$, the derivative of the cost function with respect to ξ is needed

$$\frac{dC}{dU_m} = \sum_{t=1}^{\infty} \gamma_{\text{disc}}^{t-1} \left(h_t 1_{N_t>0} + c_t(1_{N_t>0} - 1)\right) \frac{dN_t}{dU_m} \tag{15}$$

where

$$\begin{aligned} \frac{dN_q}{dU_m} &= \sum_{t'=1}^{t} \frac{dU_{t'-L}}{dU_m} \\ &= \sum_{t'=1}^{t} \delta_{t'-L,m} \end{aligned} \tag{16}$$

Using equation (16) in equation (15)

$$\frac{dC}{dU_m} = \sum_{t=1}^{\infty} \gamma_{\text{disc}}^{t-1} \Big(h_t 1_{N_t \geq 0} + c_t(1_{N_t \geq 0} - 1) \Big) \sum_{t'=1}^{t} \delta_{t'-L,m}. \tag{17}$$

For some generic function g_t

$$\begin{aligned} \sum_{t=1}^{\infty} g_t \sum_{t'=1}^{t} \delta_{t'-L,m} &= g_1 \delta_{1-L,m} + g_2(\delta_{1-L,m} + \delta_{2-L,m}) + \dots \\ &= \sum_{t=L+m}^{\infty} g_t \end{aligned} \tag{18}$$

meaning

$$\frac{dC}{dU_m} = \sum_{t=L+m}^{\infty} \gamma_{\text{disc}}^{t-1} (h_t 1_{N_t \geq 0} + c_t(1_{N_t \geq 0} - 1)). \tag{19}$$

Combining equations (5) and (19)

$$\sum_{s_1, s_2, \dots} \sum_{t=L+m}^{\infty} \gamma_{\text{disc}}^{t-1} (h_t 1_{N_t \geq 0} + c_t(1_{N_t \geq 0} - 1)) p(s_1, s_2, \dots | D, I) \stackrel{!}{=} 0 \quad \forall m \tag{20}$$

The sums can be evaluated viz

$$\begin{aligned} \sum_{s_1, s_2, \dots} 1_{N_t \geq 0} p(s_1, s_2, \dots | D, I) &= p(N_t \geq 0 | D, I), \\ \sum_{s_1, s_2, \dots} p(s_1, s_2, \dots | D, I) &= 1. \end{aligned} \tag{21}$$

Let

$$\psi_t \equiv (h_t + c_t)p(N_t \geq 0|D, I) - c_t, \tag{22}$$

then

$$\begin{aligned}\frac{d}{dU_m}\mathbb{E}[C|D, I] &= \sum_{t=L+m}^{\infty} \gamma_{\text{disc}}^{t-1}\psi_t \\ &\stackrel{!}{=} 0 \quad \forall m\end{aligned} \tag{23}$$

A recursion relation can be derived viz

$$\begin{aligned}\frac{d}{dU_0}\mathbb{E}[C|D, I] &= \sum_{t=L}^{\infty} \gamma_{\text{disc}}^{t-1}\psi_t \\ &= \gamma_{\text{disc}}^{L-1}\psi_L + \sum_{t=L+1}^{\infty} \gamma_{\text{disc}}^{t-1}\psi_t \\ &= \gamma_{\text{disc}}^{L-1}\psi_L + \frac{d}{dU_1}\mathbb{E}[C|D, I] \\ &= \gamma_{\text{disc}}^{L-1}\psi_L + \gamma_{\text{disc}}^{L}\psi_{L+1} + \frac{d}{dU_2}\mathbb{E}[C|D, I] \\ &= \dots \\ &\stackrel{!}{=} 0\end{aligned} \tag{24}$$

meaning

$$\frac{d}{dU_m}\mathbb{E}[C|D, I] = \gamma_{\text{disc}}^{L+m-1}\psi_{L+m} + \frac{d}{dU_{m+1}}\mathbb{E}[C|D, I]. \tag{25}$$

Since all derivatives are required to be simultaneously zero,

$$\gamma_{\text{disc}}^{j-1}\psi_j \stackrel{!}{=} 0 \quad \forall j \Rightarrow \psi_j = 0 \tag{26}$$

meaning

$$p(N_t^* \geq 0|D, I) = \frac{c_t}{c_t + h_t}, \tag{27}$$

where

$$N_t^* \equiv N_0 + \upsilon_t^* - \zeta_t \tag{28}$$

denote the units on stock given optimal decisions.

B Expected Cost

In this appendix, the policy efficiency ratio (PER, equation (11)) is calculated under the assumption of i) $p(\zeta_t|D, I)$ follows a Poisson distribution and ii) conditional independence between ζ_t. The first step consist of re-writing the expected cost

$$
\begin{aligned}
\mathbb{E}[C|D,I] &= \sum_{t=1}^{\infty}\sum_{\zeta_t=0}^{\infty}\gamma_{\text{disc}}^{t-1}\left(h_t 1_{N_t>0} + c_t(1_{N_t>0}-1)\right)N_t p(\zeta_t|D,I)\\
&= \sum_{t=1}^{\infty}\sum_{\zeta_t=0}^{\infty}\gamma_{\text{disc}}^{t-1}\Big((h_t+c_t)(N_0+\upsilon_t)1_{N_t>0}\\
&\qquad\qquad - (h_t+c_t)1_{N_t>0}\zeta_t - c_t(N_0+\upsilon_t) + c_t\zeta_t\Big)p(\zeta_t|D,I).
\end{aligned} \tag{29}
$$

The sums over ζ_t can be expressed as follows (see appendix D)

$$
\begin{aligned}
\sum_{\zeta_t=0}^{\infty} 1_{N_t>0}p(\zeta_t|D,I) &= \sum_{\zeta_t=0}^{N_0+\upsilon_t} p(\zeta_t|D,I)\\
&= \frac{\Gamma(N_0+\upsilon_t+1,\lambda_t)}{\Gamma(N_0+\upsilon_t+1)},\\
\sum_{\zeta_t=0}^{\infty} 1_{N_t>0}\zeta_t p(\zeta_t|D,I) &= \sum_{\zeta_t=0}^{N_0+\upsilon_t} \zeta_t p(\zeta_t|D,I)\\
&= \lambda_t\frac{\Gamma(N_0+\upsilon_t,\lambda_t)}{\Gamma(N_0+\upsilon_t)},\\
\sum_{\zeta_t=0}^{\infty}\zeta_t p(\zeta_t|D,I) &= \lambda_t,\\
\sum_{\zeta_t=0}^{\infty} p(\zeta_t|D,I) &= 1.
\end{aligned} \tag{30}
$$

Collecting the results means

$$
\begin{aligned}
\mathbb{E}[C|D,I] = \sum_{t=1}^{\infty}\gamma_{\text{disc}}^{t-1}\Big(&(h_t+c_t)(N_0+\upsilon_t)\frac{\Gamma(N_0+\upsilon_t+1,\lambda_t)}{\Gamma(N_0+\upsilon_t+1)}\\
&- (h_h+c_t)\lambda_t\frac{\Gamma(N_0+\upsilon_t,\lambda_t)}{\Gamma(N_0+\upsilon_t)}\\
&- c_t(N_0+\upsilon_t-\lambda_t)\Big).
\end{aligned} \tag{31}
$$

C Analytical Policy Efficiency Ratio

Using an approximation from Bartmann and Beckmann (1992)

$$\frac{\Gamma(x,y)}{\Gamma(x)} \simeq \frac{1}{1+e^{-\frac{m}{\sqrt{y}}(x-y-\frac{1}{2})}} \tag{32}$$

with $m \simeq 1.8$. Applying this yields

$$\begin{aligned}\sum_{\zeta_t=0}^{\infty} 1_{N_t>0}\, p(\zeta_t|D,I) &\simeq \frac{1}{1+e^{-\frac{m}{\sqrt{\lambda_t}}(N_0+\upsilon_t+\frac{1}{2}-\lambda_t)}},\\ \sum_{\zeta_t=0}^{\infty} 1_{N_t>0}\, \zeta_t p(\zeta_t|D,I) &\simeq \lambda_t \frac{1}{1+e^{-\frac{m}{\sqrt{\lambda_t}}(N_0+\upsilon_t-\frac{1}{2}-\lambda_t)}}.\end{aligned} \tag{33}$$

The expected cost can thus be written

$$\begin{aligned}\mathbb{E}[C|D,I] \simeq \sum_{t=1}^{\infty} \gamma_{\text{disc}}^{t-1}\Bigg(&\frac{(h_t+c_t)(N_0+\upsilon_t)}{1+e^{-\frac{m}{\sqrt{\lambda_t}}(N_0+\upsilon_t+\frac{1}{2}-\lambda_t)}} - \frac{(h_t+c_t)\lambda_t}{1+e^{-\frac{m}{\sqrt{\lambda_t}}(N_0+\upsilon_t-\frac{1}{2}-\lambda_t)}}\\ &- c_t(N_0+\upsilon_t-\lambda_t)\Bigg).\end{aligned} \tag{34}$$

Due to the lead time, no decisions can be implemented until $t \geq L$, leading to

$$\begin{aligned}\mathbb{E}[C|D,I] \simeq E_L + \sum_{t=L}^{\infty} \gamma_{\text{disc}}^{t-1}\Bigg((h_t+c_t)\Bigg[&\frac{N_0+\upsilon_t}{1+e^{-\frac{m}{\sqrt{\lambda_t}}(N_0+\upsilon_t+\frac{1}{2}-\lambda_t)}}\\ -\frac{\lambda_t}{1+e^{-\frac{m}{\sqrt{\lambda_t}}(N_0+\upsilon_t-\frac{1}{2}-\lambda_t)}}\Bigg] &- c_t(N_0+\upsilon_t-\lambda_t)\Bigg),\end{aligned} \tag{35}$$

where E_L is defined as

$$E_L \equiv \sum_{t=1}^{L-1} \gamma_{\text{disc}}^{t-1}\Bigg(\frac{N_0(h_t+c_t)}{1+e^{-\frac{m}{\sqrt{\lambda_t}}(N_0+\frac{1}{2}-\lambda_t)}} - \frac{\lambda_t(h_t+c_t)}{1+e^{-\frac{m}{\sqrt{\lambda_t}}(N_0-\frac{1}{2}-\lambda_t)}} - c_t(N_0-\lambda_t)\Bigg). \tag{36}$$

C.1 Optimal Expected Cost

For the optimal cost, use that

$$\frac{c_t}{c_t+h_t} 1_{t\geq L} \simeq \frac{1}{1+e^{-\frac{m}{\sqrt{\lambda_t}}(N_0+\upsilon_t+\frac{1}{2}-\lambda_t)}} \tag{37}$$

with the implicit criteria $\upsilon_t \geq 0$ and assuming $\upsilon_t \in$

$$\upsilon_t \simeq \max(\text{round}(\lambda_t + \frac{\sqrt{\lambda_t}}{m} \ln \frac{c_t}{h_t} - N_0 - \frac{1}{2}), 0) \tag{38}$$

yielding the optimal expected cost

$$\begin{aligned} \mathbb{E}[C^*|D,I] &\simeq E_L + \sum_{t=L}^{\infty} \gamma_{\text{disc}}^{t-1} \left(c_t(N_0 + \upsilon_t) - \lambda_t c_t \frac{c_t + h_t}{c_t + e^{\frac{m}{\sqrt{\lambda_t}}} h_t} - c_t(N_0 + \upsilon_t - \lambda_t) \right) \\ &= E_L + \sum_{t=L}^{\infty} \gamma_{\text{disc}}^{t-1} \lambda_t h_t \frac{e^{\frac{m}{\sqrt{\lambda_t}}} - 1}{e^{\frac{m}{\sqrt{\lambda_t}}} \frac{h_t}{c_t} + 1}. \end{aligned} \tag{39}$$

C.2 Baseline Expected Cost

Finally, for the baseline cost, with $\upsilon_t = 1_{t \geq L}(\lambda_t - N_0 + R)$ with R the constant reorder point, the baseline expected cost is

$$\mathbb{E}[C'|D,I] \simeq E_L + \sum_{t=L}^{\infty} \gamma_{\text{disc}}^{t-1} \left(\frac{(h_t + c_t)(\lambda_t + R)}{1 + e^{-\frac{m}{\sqrt{\lambda_t}}(R+1)}} - \frac{(h_t + c_t)\lambda_t}{1 + e^{-\frac{m}{\sqrt{\lambda_t}} R}} - c_t R \right) \tag{40}$$

with

$$\begin{aligned} \lim_{R \to 0} \mathbb{E}[C'|D,I] &\simeq E_L + \sum_{t=L}^{\infty} \gamma_{\text{disc}}^{t-1} \frac{h_t \lambda_t}{2} \left(1 + \frac{c_t}{h_t} \right) \frac{e^{\frac{m}{\sqrt{\lambda_t}}} - 1}{e^{\frac{m}{\sqrt{\lambda_t}}} + 1}, \\ \lim_{R \to \infty} \mathbb{E}[C'|D,I] &\simeq E_L + \sum_{t=L}^{\infty} \gamma_{\text{disc}}^{t-1} h_t R. \end{aligned} \tag{41}$$

C.3 Policy Efficiency Ratio

Given the expected cost of the optimal and baseline policies, the PER can be written

$$\frac{\mathbb{E}[C^*|D,I]}{\mathbb{E}[C'|D,I]} \simeq \frac{E_L + \sum_{t=L}^{\infty} \gamma_{\text{disc}}^{t-1} \lambda_t h_t \frac{e^{\frac{m}{\sqrt{\lambda_t}}} - 1}{e^{\frac{m}{\sqrt{\lambda_t}}} \frac{h_t}{c_t} + 1}}{E_L + \sum_{t=L}^{\infty} \gamma_{\text{disc}}^{t-1} h_t \left(\frac{(1 + \frac{c_t}{h_t})(\lambda_t + R)}{1 + e^{-\frac{m}{\sqrt{\lambda_t}}(R+1)}} - \frac{(1 + \frac{c_t}{h_t})\lambda_t}{1 + e^{-\frac{m}{\sqrt{\lambda_t}} R}} - \frac{c_t}{h_t} R \right)} \tag{42}$$

Assuming constant h, c, that E_L can be neglected in relative magnitude to the sums over infinity and using that

$$1 \leq e^{\frac{m}{\sqrt{\lambda_t}}} \lesssim 1.5 \Rightarrow e^{\frac{m}{\sqrt{\lambda_t}}} \approx k \tag{43}$$

the limit of $R \sim 0$ yields

$$\left.\frac{\mathbb{E}[C^*|D,I]}{\mathbb{E}[C'|D,I]}\right|_{R\sim 0} \approx 2\frac{\frac{k-1}{k\frac{h}{c}+1}}{\left(1+\frac{1}{\frac{h}{c}}\right)\frac{k-1}{1+k}}\frac{\sum_{t=L}^{\infty}\gamma_{\text{disc}}^{t-1}\lambda_t}{\sum_{t=L}^{\infty}\gamma_{\text{disc}}^{t-1}\lambda_t}$$
$$= \frac{2(1+k)}{\left(1+\frac{1}{\frac{h}{c}}\right)\left(k\frac{h}{c}+1\right)} \tag{44}$$
$$\simeq \frac{4}{2+\frac{h}{c}+\frac{c}{h}}$$

where for the last equality it was used that $k \simeq 1$. In the limit of $R \gg 0$

$$\left.\frac{\mathbb{E}[C^*|D,I]}{\mathbb{E}[C'|D,I]}\right|_{R\gg 0} \simeq \frac{\sum_{t=L}^{\infty}\gamma_{\text{disc}}^{t-1}\lambda_t h_t \frac{e^{\frac{m}{\sqrt{\lambda_t}}}-1}{e^{\frac{m}{\sqrt{\lambda_t}}}\frac{h_t}{c_t}+1}}{\sum_{t=L}^{\infty}\gamma_{\text{disc}}^{t-1}h_t R}$$
$$= \frac{1}{\frac{h}{c}+1}\frac{m}{R}\frac{\sum_{t=L}^{\infty}\gamma_{\text{disc}}^{t-1}\sqrt{\lambda_t}}{\sum_{t=L}^{\infty}\gamma_{\text{disc}}^{t-1}} \tag{45}$$

By definition

$$\frac{m}{R}\frac{\sum_{t=L}^{\infty}\gamma_{\text{disc}}^{t-1}\sqrt{\lambda_t}}{\sum_{t=L}^{\infty}\gamma_{\text{disc}}^{t-1}} = 1 \tag{46}$$

since $0 \leq \text{PER} \leq 1$. Therefore

$$\lim_{R\to\infty}\frac{\mathbb{E}[C^*|D,I]}{\mathbb{E}[C'|D,I]} \simeq \frac{1}{1+\frac{h}{c}}. \tag{47}$$

D Identities Related to the Poisson Distribution

The Poisson distribution with rate parameter λ describes the probability of a discrete random variable ζ taking integer values. Here, some key identities relevant for the PER (equation (11)) are explored. The cumulative probability of observing $\zeta \leq k$ given a Poisson rate λ is

$$p(\zeta \leq k|\lambda) = e^{-\lambda}\sum_{j=0}^{k}\frac{\lambda^j}{j!}$$
$$= \frac{\Gamma(k+1,\lambda)}{\Gamma(k+1)}, \tag{48}$$

where $\Gamma(k+1, \lambda)$ is the incomplete gamma function and $\Gamma(s)$ denotes the complete gamma function

$$\Gamma(k+1) = k!. \tag{49}$$

Similarly, an expression for the conditional expectation of ζ, given that $\zeta \leq k$, can be derived viz

$$\begin{aligned}
\mathbb{E}[\zeta|\zeta \leq k, \lambda] &= e^{-\lambda} \sum_{j=0}^{k} j \frac{\lambda^j}{j!} \\
&= e^{-\lambda} \sum_{j=0}^{k} \frac{\lambda^j}{(j-1)!} \\
&= \lambda e^{-\lambda} \sum_{j=0}^{k-1} \frac{\lambda^j}{j!} \\
&= \lambda \frac{\Gamma(k, \lambda)}{\Gamma(k)}.
\end{aligned} \tag{50}$$

E Notes

1. In order to eliminate υ_t from the optimal policy, the the equation

$$\begin{aligned}
\sum_{\zeta_t=0}^{\infty} 1_{N_t>0} p(\zeta_t|D, I) &= \sum_{\zeta_t=0}^{N_0+\upsilon_t} p(\zeta_t|\upsilon_t \geq 0, D, I) \\
&= \frac{c_t}{c_t + h_t} 1_{\upsilon_t \geq 0}
\end{aligned} \tag{51}$$

   must be solvable for υ_t.

2. There are differences in the implementation of the policy using the approximation of the Poisson distribution. This implementation is extremely sensitive, because when you roll over time, you slice the distribution in many places and at some point, your slice will end up between a round down and a round up. From that point, the decisions coming after are affected. In terms of the cost function, it can therefore have a relatively significant effect and an approximation that seems excellent on the face, can have unintended effects. This just means the implementation of the policy should always follow the quantile rather than the approximation.

1. A few words on the intuitive interpretation of the cost function? The number of units overstocked multiplied with the cost of overstocking

per unit. The number of units understocked multiplied with the value of each unit. The latter represents the lost value.

2. Mention the benefit of the analytical solution; the computation speed relative to a numerical optimization is highly beneficial at scale.

3. Compress the recursive relation to m-notation.

4. The baseline policy is an (R,Q) policy with $R = 0$ and $Q = \mathbb{E}[s_t|D, I]$.

5. Is our result in any of the inventory control books?

6. will the relationship between baseline and optimal policy depend on forecasting method? I would say yes. How do we handle this?

7. if there is around the same cost for over/under stocking, there is a $30 - 40\%$ reduction in costs with the optimal policy compared to the baseline. In the limit of $c \gg h$, the reduction in cost approach 0% and the gains are minor. This is the relevant limit in most cases, where the value of the unit significantly outweigh the holding cost. This is relevant, however, it is underlined, that the baseline policy is always equal or worse (statistically, meaning the expected cost is always lower. Expected cost is over a distribution. draws from that distribution can fall either way) than the optimal policy.

8. PER is highly dependent on the policy. If the reference level is moved away from $N_t = 0$, for example, the plot completely changes and the blue area shifts to the top left corner.

1. S. Axsäter. Inventory control

2. Gasthaus paper

3. inspiration for introduction

4. probabilistic trucks

5. Amazon paper (inspiration)

References

et al. Alumur, S. A. Optimizing inventory and transportation decisions in supply chain management. *European Journal of Operational Research*, 216 (3):705–716, 2012.

Sven Axsater. *Inventory Control*. Springer, New York, 2006. ISBN 9780387243096.

Dieter Bartmann and Martin J. Beckmann. *Inventory Control: Models and Methods*. Springer-Verlag, Berlin; New York, 1992. ISBN 9783540542216.

et al. Gürler, Ü. Inventory and transportation optimization in supply chains. *Journal of Supply Chain Management*, 50(1):34–42, 2014.

J. Higginson and J. H. Bookbinder. The transport-inventory problem. *Journal of Operational Research Society*, 45:969–978, 1994.

I. Lavalle. *Fundamentals of Decision Theory*. Cambridge University Press, Cambridge, UK, 2nd edition, 2006.

D. S. Sivia and J. Skilling. *Data Analysis - A Bayesian Tutorial*. Oxford Science Publications. Oxford University Press, 2nd edition, 2006.

et al. Wei, Y. Hybrid shipment policies for inventory management. *Journal of Supply Chain Management*, 56:123–134, 2020.

S. Çetinkaya. Consolidation policies in inventory management. *Production and Operations Management*, 14(2):233–243, 2005.

S. Çetinkaya and J. H. Bookbinder. Optimal integrated policies for managing inventory and outbound shipments. *Transportation Science*, 37:39–55, 2003.

S. Çetinkaya and C.-Y. Lee. Stock replenishment and shipment scheduling for vendor-managed inventory systems. *Management Science*, 46(2):217–232, 2000. doi: 10.1287/mnsc.46.2.217.11925.

S. Çetinkaya and C.-Y. Lee. Inventory and transportation cost coordination in vmi systems. *Operations Research*, 54:123–137, 2006.