Sequential Decision-Making for Inventory Control using Bayesian Decision Theory

Jonas Petersen
Ander Runge Walther

January 20, 2025

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 $t \in [1, T]$ $(T \geq L)$ into the future.

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

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

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

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

Theorem 1 (Optimal Policy Rule for Inventory Control)**.** *Given the asymmetric cost function of definition 2 the optimal policy (definition 3) 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. Since negative decisions are not allowed,* $\upsilon_t^* = 0$ *is the optimal decision in case of* $N_t \leq 0$*. 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

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

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

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

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 can be 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. Note that PER$\in [0,1]$ by definition since the optimal policy by definition yield the lowest possible expected cost. Assuming i) $p(\zeta_t|D,I)$ follows a Poisson distribution (with time varying rate parameter) and ii) conditional independence, the expected cost can be written (see appendix B)

$$\begin{aligned}\mathbb{E}[C|D,I] = \sum_{t=1}^{T} \gamma_{\text{disc}}^{t-1} \Bigg(& (h_t + c_t)(N_0 + \upsilon_t) \frac{\Gamma(N_0 + \upsilon_t + 1, \lambda_t)}{\Gamma(N_0 + \upsilon_t + 1)} \\ & - (h_h + c_t)\lambda_t \frac{\Gamma(N_0 + \upsilon_t, \lambda_t)}{\Gamma(N_0 + \upsilon_t)} \\ & - c_t(N_0 + \upsilon_t - \lambda_t) \Bigg).\end{aligned} \tag{12}$$

Equation (12) can be used to calculate the exact PER for any baseline policy.

3.1.1 Example: Numerical Policy Efficiency Ratio

To provide an example of usage of both theorem 1 and the PER, the PER is considered with $N_0 = 37$, $L = 6$, constant unit value $c_t = c$, constant holding cost $h_t = h$, Natures decisions shown in figure 1 and the baseline policy of 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 expectation of Natures decisions*

$$\upsilon_t' = \max(\mathit{round}(\mathbb{E}[\zeta_t|D,I] + R - N_0), 0). \tag{13}$$

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

Given the above setup, figure 2 presents a heatmap illustrating the Policy Efficiency Ratio (PER) as a function of the reorder point R and the cost-to-holding ratio $\frac{c}{h}$. From the figure, several patterns emerge. In general

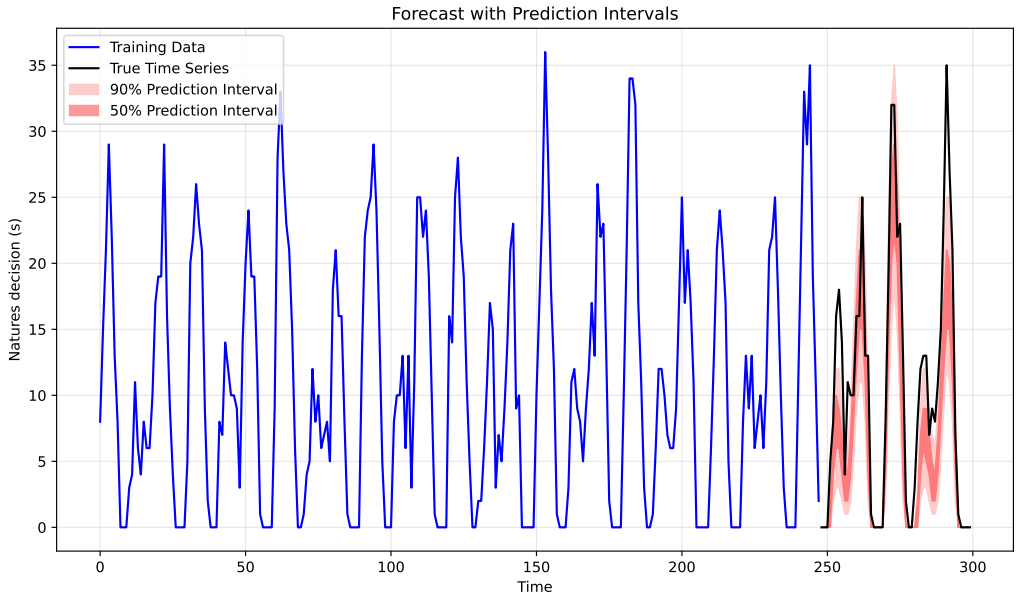

Figure 1: Natures decisions as a function of time. Blue denotes the historical data available for the Robot, red denote the forecast made by the Robot (based on the historical data) and the black denotes the true future values for reference (not used).

there seem to be bands of approximately constant PER following $R \sim \ln \frac{c}{h}$ -lines. This band represent the optimal balancing point for R for a given $\frac{c}{h}$. At $\frac{c}{h} = 1, R = 0$ the PER achieves its maximum at $\simeq 1$, which makes intuitive sense since if $h = c$, the cost function is balanced and the optimal balancing point would be at $R = 0$. The maximum value of the band slightly decrease for increasing $\frac{c}{h}$, with PER~ 0.9 as a mean value. Hence, even in case where the reorder point R is optimized perfectly, there is a significantly higher expected cost (around 10%) for the baseline (R,Q) - policy relative to the optimal policy (theorem 1).

4 Conclusion

Assuming a company has a cost function on the form of definition 2, they can use the PER in comparison with their current decision policy to i) gauge the expected improvement by implementing the optimal policy or alternatively tune the value of R for their values of $\frac{c}{h}$. The PER shown in figure 2 indicate that for $\frac{c}{h} \gg 1$ the optimal policy yields a reduction of $\sim 10\%$ in expected

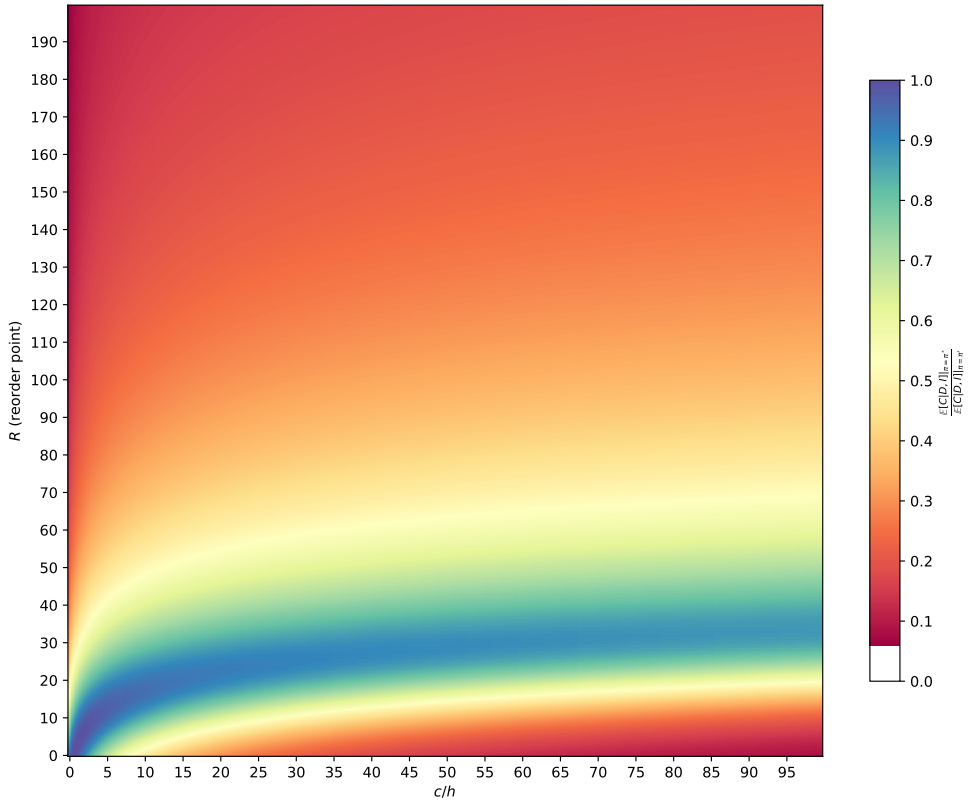

Figure 2: Heatmap of the policy efficiency ratio (PER) calculated using equations (12), (11), definition 4 and theorem 1 with $N_0 = 37$, $L = 6$ and Natures decisions shown in figure 1.

cost. In case a company is not using decision theory in a formal way, the expected cost reduction may be significantly higher.

A Optimal Policy

In accordance with theorem 3, the optimal policy π^* is defined by the first and second order conditions of equation (5).

A.1 First order condition

The first order condition can be written

$$\begin{aligned}\frac{d}{dU_m}\mathbb{E}[C|D,I]\Big|_{U_m=U_m^*} &= \frac{d}{dU_m}\sum_{s_1,s_2,\dots s_T} p(s_1,s_2,\dots s_T|D,I)\frac{dC}{dU_m}\Big|_{U_m=U_m^*} \\ &= 0\end{aligned} \tag{14}$$

with

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

and

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

Using equation (16) in equation (15)

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

For some generic function g_t

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

meaning

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

Combining equations (14) and (19)

$$\sum_{s_1,s_2,\ldots s_T} \sum_{t=L+m}^{T} \gamma_{\rm disc}^{t-1}(h_t 1_{N_t>0} + c_t(1_{N_t>0} - 1))p(s_1, s_2, \ldots s_T|D, I)\Big|_{U_m=U_m^*} = 0 \tag{20}$$

The sums can be evaluated viz

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

Let

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

then

$$\begin{aligned} \frac{d}{dU_m}\mathbb{E}[C|D, I]\Big|_{U_m=U_m^*} &= \sum_{t=L+m}^{T} \gamma_{\rm disc}^{t-1}\psi_t \\ &= 0 \end{aligned} \tag{23}$$

A recursion relation can be derived viz

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

meaning

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

For the optimal policy, all first order derivative vanish, meaning

$$\forall t \geq L, \quad \gamma_{\rm disc}^{t-1}\psi_t|_{\pi=\pi^*} = 0 \Rightarrow \psi_t|_{\pi=\pi^*} = 0. \tag{26}$$

Combining equations (26) and (22) yields

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

for $t \geq L$ and $\upsilon_t^* \geq 0$, where

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

denote the units on stock given optimal decisions.

A.2 Second order condition

For $N_t = 0$, the indicator function $1_{N_t>0}$ transitions from 1 to 0, leading to a discontinuity in the derivative of the cost function (equation 19). For this reason, the second order derivative is not well defined at this point. However, since $\frac{dC}{dU_m} > 0$ for $N_t > 0$ and $\frac{dC}{dU_m} < 0$ for $N_t < 0$ it can be inferred that the cost function achieves a local minimum in the neighborhood of $N_t = 0$.

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

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

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

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

Collecting the results means

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

C 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

$$\begin{aligned}
p(\zeta \leq k|\lambda) &= e^{-\lambda} \sum_{j=0}^{k} \frac{\lambda^j}{j!} \\
&= \frac{\Gamma(k+1,\lambda)}{\Gamma(k+1)},
\end{aligned} \tag{32}$$

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

$$\Gamma(k+1) = k!. \tag{33}$$

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

$$\begin{aligned}
\mathbb{E}[\zeta|\zeta \leq k, \lambda] &= e^{-\lambda} \sum_{j=0}^{k} j \frac{\lambda^j}{j!} \\
&= e^{-\lambda} \sum_{j=0}^{k} \frac{\lambda^j}{(j-1)!} \\
&= \lambda e^{-\lambda} \sum_{j=0}^{k-1} \frac{\lambda^j}{j!} \\
&= \lambda \frac{\Gamma(k,\lambda)}{\Gamma(k)}.
\end{aligned} \tag{34}$$

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.