

# 复合随机需求下基于阈值策略的库存分配模型研究<sup>\*</sup>

崔利刚<sup>1,2</sup>, 谢文静<sup>1</sup>, 吴 有<sup>1</sup>, 杨京昊<sup>1</sup>

(1. 重庆交通大学 经济与管理学院; 2. 重庆市智能供应链工程技术研究中心, 重庆 400074)

**摘要:**在订单驱动的库存决策中,需求订单的到达常呈现出双重随机叠加的特点,即需求订单的随机到达和需求规模的随机呈现。为解决此背景下常见的库存问题,考虑了需求到达和规模符合复合随机条件下备件库存的最优分配问题。具体来看,在周期检查( $S, T$ )订货策略下以备件库存系统为研究对象,研究在复合随机需求条件下,基于阈值策略的积压需求清理机制建立缺货成本高和低 2 类需求的库存分配模型,实现库存成本最小化。在模型求解上,不同于之前的研究仅采用枚举法,本文应用了改进遗传算法,并设计了一种简单、高效的启发式算法。数值实验结果表明,启发式算法在求解所建立的 2 类模型都表现出了更好的求解效果,阈值策略比先到先服务和库存分割策略有明显的降本优势;敏感性分析证明阈值策略的相对成本优势随检查期、到达需求规模均值增加而减少,随提前期、单位库存持有成本增加而增加。研究结果为库存的精益管理提供决策参考。

**关键词:**复合随机需求;缺货成本;阈值策略;改进遗传算法;启发式算法

**中图分类号:**O227;F273

**文献标志码:**A

**文章编号:**1672-6693(2025)01-0001-13

库存管理者可针对不同的需求订单进行库存分类并赋予服务优先级,再通过预期分配仓储资源以满足差异化的需求订单。库存分配策略是企业为可能到来的高优先级需求预留部分库存而暂缓服务已到的低优先级需求<sup>[1]</sup>。过早或过晚开始暂停低优先级需求(订单)都会导致成本的增加。因此,在正确的时间点暂停低优先级需求正是库存分配的关键。现实经营管理中顾客订单会随机到达,且伴随出现的订单规模也存在随机性,为此,本文提出库存阈值分配策略是在考虑双重复合随机需求条件下,根据企业库存系统的存货数量降至阈值点的时间来确定商品的分配水平,具有较强的现实意义。该策略有助于企业精准把握库存水平,确保在高优先级需求出现时,能够迅速响应并满足客户需求,从而显著提升服务水平与客户满意度。同时,本文给出的库存阈值策略,可为企业有效控制库存成本,优化仓储资源分配等方面提供参考。

库存分配问题最早由 Veinott<sup>[2]</sup>在 20 世纪 60 年代提出,用于解决库存周期检查下将有限的库存分配多类需求问题。许多学者就需求特征对库存分配策略产生的影响进行了深入研究。在需求服从正态分布假设下,郑长征等人<sup>[3]</sup>研究了在需求和商品生命周期同时随机情况下的库存分配问题。Li 等人<sup>[4]</sup>假设需求分布未知,设计了“边做边学”的启发式方法对大型百货连锁商店的季节性商品库存分配问题进行研究。在单品高、低 2 类需求均服从泊松到达过程的基础上,Wan 等人<sup>[5]</sup>尝试研究订至点(order-up-to)策略下的最佳库存分配阈值,但采用的是枚举法进行遍历求解,效率不高。另外,存货的管理特征也会对库存分配造成影响。存货成本差异会影响需求呈现不同特征,科学的库存分配则会降低这种影响<sup>[6]</sup>。备件配送中,客户服务合同会根据备件单价、提前期和最小订货量进行不同优先级划分,优秀的库存分配策略组合会在维持较低系统成本的基础上满足对不同客户的服务需求<sup>[7]</sup>。在电商配送网络中,中心仓会综合考虑商品需求信息和库存信息对二级仓进行动态补货,在此条件下中心仓的库存分配方案会直接影响系统成本高低<sup>[8]</sup>。

不同的库存订货模式会影响库存分配策略的设计<sup>[9]</sup>。在周期检查策略( $S, T$ )下,库存分配策略还要考虑客户需求变化的原因或特征,如维修服务对象的停工成本特征<sup>[10]</sup>和不同客户网点需求特点<sup>[11]</sup>。在( $s, S$ )策略下,库存的分配问题更为复杂。以阿里巴巴分销系统为例,通常会有多个库存分配策略以应对不同的系统情景需

<sup>\*</sup> 收稿日期:2024-05-21 修回日期:2024-11-13 网络出版时间:2025-03-14T14:14

资助项目:国家自然科学基金面上项目(No. 72172022);教育部人文社会科学项目(No. 21YJC630016);重庆市教育委员会人文社会科学  
研究一般项目(No. 24SKGH119);重庆市研究生导师团队项目(No. JDDSTD2022005)

第一作者简介:崔利刚,男,教授,博士生导师,研究方向为补货与库存管理,E-mail: cligang@126.com

网络出版地址:https://link.cnki.net/urlid/50.1165.N.20250313.1810.014

求<sup>[12]</sup>。在 $(r, Q)$ 订货策略下,库存分配策略需要在不同类型的客户服务水平之间寻求平衡,例如在给定订单的准备率和完成率水平<sup>[1]</sup>中寻求平衡。本文重点关注周期检查 $(S, T)$ 订货策略下考虑高、低 2 类阈值需求和缺货损失的库存分配问题。

不确定性叠加对基于优先级的库存分配造成了挑战。Hung 等人<sup>[13]</sup>同时考虑了补货提前期不确定和需求随机到达的情况下,如何配置不同优先级需求的库存(高交付率)又尽可能减少库存剩余。类似地,在 base-stock 策略下以最小化库存剩余为目标,Kouki 等人<sup>[14]</sup>假设高、低优先级需求服从复合泊松分布。实践中,客户需求得不到满足时可采用计提销售损失来解决;也可以通过延迟交付的方式来解决,延迟交付当前需求订单,在后期补货到达后满足客户的最终需求<sup>[14]</sup>。相较而言,考虑需求积压条件的建模更复杂,需解决补货何时到达和到达时如何处理积压需求的问题。处理积压需求的基本思路是先满足高优先级需求,再满足低优先级需求。然而,因为存货量和需求积压量在很大程度上受补货订单到达过程的影响,这种方法无法将连续的订货周期拆分为多个独立的订货周期,这对评估系统的长期成本带来了挑战。

结合现实情况,现有文献在研究库存分配策略时存在以下特征:1) 多数研究模型都以假设连续到达的单一随机需求为前提,只有少数研究考虑构建复合随机需求条件模型和考虑离散到达的随机需求及需求规模的随机性问题。2) 仅有的一些在复合随机需求下的研究都是以服务水平为导向,缺少以成本为导向的研究。3) 针对离散到达的随机需求下的库存分配模型,目前的求解方法较为复杂、低效,缺少简便且高效的算法。

为此,本文以周期检查 $(S, T)$ 订货策略下的备件库存系统为研究对象,考虑了 2 类不同缺货成本和到达率的复合泊松需求,未满足的需求积压统一在下一补货订单到达时按阈值积压清理机制满足,以系统平均成本最小化为目标,构建基于阈值策略的库存分配模型。为应对补货订单对存货量和需求积压量的影响,在 Deshpande 等人<sup>[15]</sup>的研究基础上,考虑交付库存水平降至阈值点之前的所有需求以及降至阈值点之后的高优先级需求。同时,本文设计了改进遗传算法(improved genetic algorithm, IGA)和启发式算法(heuristics algorithm, HA)求解模型,在数值实验部分将阈值策略与先到先服务(first come first service, FCFS)、库存分割(stock-piling split, SPS)策略下的结果进行了比较,并对相关成本优势进行了敏感性分析。

1 问题描述

考虑了一个单品单级采用 $(S, T)$ 订货模式的库存系统,并以阈值策略来满足根据缺货成本高、低划分的 2 类需求,即代表高缺货成本的需求类 1 和代表低缺货成本的需求类 2。2 类需求到达过程服从泊松分布,到达的需求规模服从随机分布。考虑到指数分布只定义非负值,具有无记忆性且更能准确刻画复合泊松分布的概率密度函数,因此本文选择指数分布作为需求规模服从的分布。未满足的需求积压会产生相应的缺货成本。为方便建模,本文涉及的所有相关参数和符号说明见表 1。

表 1 参数及符号说明  
Tab. 1 Description of parameters and symbols

| 符号          | 含义                                           | 符号                 | 含义                                      |
|-------------|----------------------------------------------|--------------------|-----------------------------------------|
| $i$         | 需求类别, $i=1, 2$                               | $B(t)$             | $t$ 时刻对总需求的积压量                          |
| $T$         | 固定检查期即订货周期                                   | 参数 $D_i(t_1, t_2)$ | $(t_1, t_2)$ 时间段内需求类 $i$ 发生的需求          |
| $L$         | 固定补货提前期                                      | $D(t_1, t_2)$      | $(t_1, t_2)$ 时间段内产生的总需求                 |
| $\lambda_i$ | 检查期内需求类 $i$ 的平均到达率                           | 决策                 | $S$ 最大库存水平                              |
| $\lambda$   | 检查期内总需求的平均到达率, $\lambda=\lambda_1+\lambda_2$ | 变量                 | $K$ 阈值水平                                |
| 参数 $b_i$    | 单位时间单位规模下需求类 $i$ 的缺货积压成本                     | 函数                 | $p_i(x, t)$ $t$ 时段内需求类 $i$ 产生的需求的概率密度函数 |
| $h$         | 单位时间单位规模下的库存持有成本                             |                    | $p(x, t)$ $t$ 时段内产生的总需求的概率密度函数          |
| $a$         | 到达需求规模的均值                                    |                    | $g(y)$ 存货量降至阈值点的时间的概率密度函数               |
| $H(t)$      | $t$ 时刻的存货量                                   |                    | $(x)^+$ 代表 $\max\{x, 0\}$               |
| $B_i(t)$    | $t$ 时刻对需求类 $i$ 的积压量                          |                    |                                         |

该库存系统基本事件发生的顺序如下:系统在  $t_0$  时刻下达补货订单,将库存水平(存货量与在途库存之和减去需求积压量)提高至最大库存水平  $S$ ,订单将会在  $t_l = t_0 + L$  时刻到达。然后,在  $t_0 + T$  时刻下达新的补货订单,新订单将在  $t_l + T$  时刻到达。事件发生的时间顺序如图 1 所示。根据  $(S, T)$  订货模式特征,从  $t_0$  到  $t_0 + T$  时间段为一个订货周期。本文假定在整个订货周期内存在一个存货阈值水平  $K$ ,即当存货量高于  $K$  时,按 FCFS 策略来满足所有到达的需求;当存货量等于或低于  $K$  时,只满足高优先级需求类即需求类 1 的需求,低优先级需求类即需求类 2 的需求直接积压,并在下次补货订单到达时统一满足。



图 1 单个订货周期

Fig. 1 A single ordering cycle

在  $t_0$  下达的补货订单经过提前期  $L$  后到达,因此在  $t_0$  的订货决策会影响到  $t_l$  的存货量和需求积压量,包括在  $(t_l, t_l + T)$  时间段内该系统发生的成本。在采用  $(S, T)$  订货策略和考虑泊松需求的库存系统分配下形成一个重复的库存周期循环,进而只需计算 1 个库存周期循环内系统的平均成本即可得到长期平均成本。在本文考虑的复合泊松需求情况下,上述情况仍存在,因此仍可通过这种方法来计算系统的长期平均成本。本文计算系统平均成本的时间区间为  $(t_l, t_l + T)$ 。

## 2 模型构建

本文考虑的需求规模服从指数分布的概率密度函数,可表示为  $f(x) = \frac{1}{a} \exp\left(-\frac{x}{a}\right)$ ,  $x > 0$ 。令  $N$  为某一时段内需求到达总次数,  $X_i$  为第  $i$  次到达的需求规模,当  $N$  为一个固定数,即当  $N = n$  时,有总需求  $X = S_N = X_1 + X_2 + \dots + X_n$  和复合泊松分布  $p(x, t)$ ,有  $p(x, t) = \sum_{n=1}^{\infty} p_n(x) P_n(t)$ ,其中  $p_n(x)$  是关于  $X_n$  的分布函数,  $P_n(t)$  是关于  $N$  的分布函数,分别为:

$$p_n(x) = \frac{x^{n-1}}{a^n (n-1)!} e^{-x/a}, P_n(t) = \frac{(\lambda T)^n}{n!} e^{-\lambda T}。$$

由此,可以推导出总需求  $X$  的概率密度函数  $p(x, t)$  为:

$$p(x, t) = \sum_{n=1}^{\infty} \frac{x^{n-1}}{a^n (n-1)!} \frac{(\lambda t)^n}{n!} e^{-\frac{x}{a} - \lambda t} = e^{-\frac{x}{a} - \lambda t} \sqrt{\frac{\lambda t}{ax}} I_1\left(2\sqrt{\frac{\lambda tx}{a}}\right),$$

其中:  $I_1(\cdot)$  指的是一阶第一类修正贝塞尔函数。

为方便后续表达,令  $y = t_k - t_0$  来表示下达补货订单后存货量降至阈值点所经过的时间,由于需求到达过程和到达规模是不确定的,导致  $y$  也是不确定的。Kitaeva 等人<sup>[16]</sup>在定价优化研究中推导出了销售时间的概率密度函数,本文发现该研究中的销售时间本质上就等同于这里的  $y$ ,进而得到关于  $y$  的概率密度函数:

$$g(y) = \lambda e^{-\lambda y - \frac{q}{a}} + \sum_{n=2}^{\infty} \frac{\lambda^n y^{n-1}}{(n-1)!} e^{-\lambda y} \frac{q^{n-1}}{a^{n-1} (n-1)!} e^{-\frac{q}{a}} = \lambda e^{-\lambda y - \frac{q}{a}} \left[ 1 + \sum_{s=1}^{\infty} \frac{1}{(s!)^2} \left(\frac{\lambda y q}{a}\right)^s \right] = \lambda e^{-\lambda y - \frac{q}{a}} I_0\left(\sqrt{\frac{2\lambda y q}{a}}\right)。$$

由于上式较为复杂,为方便处理,在上述研究中给出一个很好近似  $y$  的分布,即一个均值为  $\frac{q}{a\lambda}$ ,方差为  $\frac{2q}{a\lambda^2}$  的正态分布,并对此进行了严谨的证明。由此,本文选用这一分布来作为  $y$  的实际分布:

$$y \sim N\left(\frac{q}{a\lambda}, 2\frac{q}{a\lambda^2}\right) = N\left(\frac{S-K}{a\lambda}, 2\frac{S-K}{a\lambda^2}\right), g(y) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(y-\mu)^2}{2\sigma^2}\right),$$

其中:  $\mu = \frac{q}{a\lambda}, \sigma^2 = \frac{2q}{a\lambda^2}$ 。

前面提到,若要计算库存系统长期平均成本就只需要计算  $(t_l, t_l + T)$  内的平均成本。由于阈值积压清理机制清理积压取决于  $t_0$  之后第  $q$  个需求量到达的时间  $y$ ,因此考虑了  $y$  所有可能的取值区间,分为以下 3 种情况:

1)  $0 < y \leq L$ , 表示第  $q$  个需求量的到达发生在  $t_0$  时刻下达的补货订单到达之前; 2)  $L < y \leq L + T$ , 表示第  $q$  个需求量的到达发生在  $t_0$  时刻下达的补货订单到达之后, 但又在  $t_0 + T$  时刻下达的补货订单到达之前; 3)  $y > L + T$ , 表示第  $q$  个需求量的到达发生在  $t_0 + T$  时刻下达的补货订单到达之后。本节后续内容分别针对上述 3 种情况, 建立了在一个库存循环即时间段  $(t_l, t_l + T)$  内的总成本模型。

### 2.1 $0 < y \leq L$ 的情况

此情况下第  $q$  个需求量的到达是在收到  $t_0$  时刻下达的补货订单即  $t_l$  之前, 根据阈值积压清理机制,  $t_k$  之前发生的需求即前  $q$  个需求量包括第  $q$  个需求量均会得到满足, 剩余的存货量  $K$  只能用于满足在  $(t_k, t_l + T)$  内需求类 1 发生的需求。此外, 在  $t_k$  之后需求类 2 发生的所有需求均会被直接积压。因此,  $\forall t \in (t_l, t_l + T)$  时, 有  $H(t)$  为  $[K - D_1(t_k, t)]^+$ ,  $B_1(t)$  为  $[D_1(t_k, t) - K]^+$ ,  $B_2(t)$  为  $D_2(t_k, t)$ 。令  $C_1(y)$  表示在情况 1) 下且给定  $y$  时的库存系统总成本, 在  $(t_l, t_l + T)$  的系统总成本为:

$$\begin{aligned} C_1(y) = & \int_{t_l}^{t_l+T} hE[(K - D_1(t_k, t))^+]dt + \int_{t_l}^{t_l+T} b_1E[(D_1(t_k, t) - K)^+]dt + \int_{t_l}^{t_l+T} b_2E[D_2(t_k, t)]dt = \\ & h \int_L^{L+T} \int_{-\infty}^{+\infty} (K - x)^+ p_1(x, t - t_k) dx dt + b_1 \int_L^{L+T} \int_{-\infty}^{+\infty} (x - K)^+ p_1(x, t - t_k) dx dt + b_2 \int_L^{L+T} a\lambda_2(t - t_k) dt = \\ & h \int_L^{L+T} \int_0^K (K - x) p_1(x, t - y) dx dt + b_1 \int_L^{L+T} \int_K^{+\infty} (x - K) p_1(x, t - y) dx dt + b_2 \int_L^{L+T} a\lambda_2(t - y) dt. \end{aligned}$$

上式中第 1 部分代表库存系统的库存持有成本, 第 2 和第 3 部分分别代表需求类 1、需求类 2 的缺货积压成本。

### 2.2 $L < y \leq L + T$ 的情况

已知第  $q$  个需求量是在  $t_k$  时到达, 则在  $(t_0, t_k)$  内到达的总需求为  $q$ ,  $D_L \in [0, q]$ 。令  $X(t)$  是长度为  $t$  的任意时间段内到达的总需求, 有  $X(t_k - t_0) = X(y)$ 。根据 Ross<sup>[17]</sup> 对随机过程的研究, 在本文已知时间段  $(t_0, t_k)$  内到达的总需求为  $q$  的条件下, 发生在  $(t_0, t_l)$  内的总需求  $D_L$  服从二项分布, 那么有:

$$\Pr(D_L = i | X(y) = q) = C_q^i \left(\frac{L}{y}\right)^i \left[1 - \left(\frac{L}{y}\right)\right]^{q-i}, i \in [0, q].$$

知道总需求  $D_L$  后, 可以得出在  $(t_l, t_k)$  发生的总需求为  $q - D_L$ , 令  $u = q - D_L$ , 则  $u \in [0, q]$ 。在  $t_k$  之前发生的 2 类需求均会得到满足, 所以在  $(t_l, t_k)$  内不会产生缺货成本。因此, 在给定  $(t_l, t_k)$  内发生的总需求为  $u$  的情况下, 只需要计算在该时间段内的库存持有成本。设  $X_{l+j}$  是提前期后到达的第  $j$  批需求,  $v_j$  是  $t_l$  与  $X_{l+j}$  到达时间之差,  $j = 1, 2, \dots, n$  ( $n = \lceil (q - i)/a \rceil$ ), 令  $\mathbf{v} = (v_1, v_2, \dots, v_n)$ , 基于在  $t_l$  之后存货量  $S - D_L$  的已知信息, 在给定  $y$  的条件下可以计算出在  $(t_l, t_k)$  内的库存持有成本, 用  $H(\mathbf{v})$  表示, 具体如下:

$$\begin{aligned} H(\mathbf{v}) = & h(S - D_L)v_1 + h(S - D_L - X_{l+1})(v_2 - v_1) + \dots + h(S - D_L - X_{l+n})(v_n - v_{n-1}) = \\ & h(S - D_L - \sum_{j=1}^{n-1} X_{l+j})v_n + h \sum_{j=1}^{n-1} X_{l+j}v_j = h(S - D_L - \sum_{i=1}^{n-1} X_{l+i})(t_k - t_l) + h \sum_{j=1}^{n-1} X_{l+j}v_j. \end{aligned}$$

当  $(t_l, t_k)$  内有  $n$  个泊松事件到达时, 各事件在到达时间上的分布同  $(t_l, t_k)$  内均匀分布的  $n$  个独立随机变量的顺序统计量的分布一样<sup>[17]</sup>。本文中的  $v_1, v_2, \dots, v_n$  的分布同  $(0, y - L)$  内均匀分布的  $n$  个独立随机变量的顺序统计量的分布  $V_1, V_2, \dots, V_n$  一样。于是, 在给定  $D_L$  条件下发生在  $(t_l, t_k)$  内的库存持有成本  $C_h(D_L, y)$  可表示为:

$$\begin{aligned} C_h(D_L, y) = & E[H(\mathbf{v})] = h(S - D_L - E[\sum_{j=1}^{n-1} X_{l+j}])(t_k - t_l) + hE[\sum_{j=1}^{n-1} X_{l+j}v_j] = \\ & h(S - D_L - E[\sum_{j=1}^{n-1} X_{l+j}])(y - L) + hE[\sum_{j=1}^{n-1} X_{l+j}V_j] = \\ & h(S - D_L - (n - 1)a)(y - L) + (n - 1)h \sum_{j=1}^{n-1} E[X_{l+j}]E[V_j] = \\ & h(S - D_L - (n - 1)a)(y - L) + (n - 1)ha \frac{y - L}{2} = h(y - L) \left(S - D_L - \frac{(n - 1)a}{2}\right). \end{aligned}$$

基于上述结果, 很容易得到在  $(t_l, t_k)$  内的库存持有成本  $C_{21}(y)$  为:

$$C_{21}(y) = \sum_{i=0}^q C_h(i, y) \Pr(D_L = i | X(y) = q) = \sum_{i=0}^q h(y - L) \left(S - i - \frac{(n - 1)a}{2}\right) C_q^i \left(\frac{L}{y}\right)^i \left[1 - \left(\frac{L}{y}\right)\right]^{q-i}.$$



前面提到,在  $t_k$  时的存货量为  $K$ ,仅用于满足之后需求类 1 发生的需求,而需求类 2 的需求直接积压到  $t_l + T$  时。因此,与 2.1 节中计算成本的思想相同,那么在给定  $y$  的条件下发生在  $(t_k, t_l + T)$  的总成本  $C_{22}(y)$  为:

$$\begin{aligned} C_{22}(y) = & \int_{t_k}^{t_l+T} hE[(K - D_1(t_k, t))^+]dt + \int_{t_k}^{t_l+T} b_1E[(D_1(t_k, t) - K)^+]dt + \int_{t_k}^{t_l+T} b_2E[D_2(t_k, t)]dt = \\ & h \int_y^{L+T} \int_{-\infty}^{+\infty} (K - x)^+ p_1(x, t - t_k) dx dt + b_1 \int_y^{L+T} \int_{-\infty}^{+\infty} (x - K)^+ p_1(x, t - t_k) dx dt + b_2 \int_y^{L+T} a\lambda_2(t - t_k) dt = \\ & h \int_y^{L+T} \int_0^K (K - x) p_1(x, t - y) dx dt + b_1 \int_y^{L+T} \int_K^{+\infty} (x - K) p_1(x, t - y) dx dt + b_2 \int_y^{L+T} a\lambda_2(t - y) dt. \end{aligned}$$

由于上式与 2.1 节中类似,此处便不再说明。

综上,就可得到在该情况下发生在  $(t_l, t_l + T)$  内的库存系统总成本  $C_2(y) = C_{21}(y) + C_{22}(y)$ 。

### 2.3 $y > L + T$ 的情况

易知在  $(t_0, t_l + T)$  内到达的总需求至多为  $q$  个,因此可以得到  $y > L + T$  发生的概率为:

$$\Pr(y > L + T) = \sum_{i=0}^q \Pr[X(L + T) = i] = \sum_{i=0}^q p(i, L + T).$$

显然在这种情况下,库存系统在  $(t_l, t_l + T)$  内的存货量始终高于  $K$ ,在此期间不会出现任何需求积压的情况,所以也不会产生任何的缺货积压成本,那么在  $(t_l, t_l + T)$  内的总成本就只包含库存持有成本。令  $z$  表示在  $(t_0, t_l + T)$  内发生的总需求,则  $z \in [0, q]$ 。为了方便地计算库存持有成本,本文计算了  $z$  取每个可能值时的成本。在给定  $z$  的条件下,采用 2.2 节中的方法,计算出了在  $(t_l, t_l + T)$  内库存系统的总成本  $C_3(z)$  为:

$$\begin{aligned} C_3(z) = & \sum_{i=0}^z C_h(i, L + T) \Pr(D_L = i | X(L + T) = z) dz = \\ & \sum_{i=0}^z hT \left( S - i - \frac{(n-1)a}{2} \right) C_z^i \left( \frac{L}{L+T} \right)^i \left( 1 - \frac{L}{L+T} \right)^{z-i} = \\ & \sum_{i=0}^z hT \left( S - i - \frac{(n-1)a}{2} \right) C_z^i \left( \frac{L}{L+T} \right)^i \left( \frac{T}{L+T} \right)^{z-i}. \end{aligned}$$

### 2.4 成本整合

前面推导出了  $y$  所有取值情况下库存系统在  $(t_l, t_l + T)$  内的总成本。同时,也提到  $y$  是服从参数为  $\lambda$  和  $q$  的正态分布,通过整合各种情况下的总成本可以得到库存系统的长期平均成本为:

$$C(S, K) = \frac{1}{T} \left[ \int_0^L C_1(y) g(y) dy + \int_L^{L+T} C_2(y) g(y) dy + \sum_{z=0}^q C_3(z) \Pr(X(L + T) = z) \right]. \quad (1)$$

## 3 求解算法设计

### 3.1 IGA

由于平均成本式(1)内含贝塞尔函数,而该函数在数学上无法解析,只能进行数值计算且耗时较长,所以库存分配模型无法通过解析的方式求解。求解该类库存分配模型的已有研究中通常采用枚举法(enumeration method, EM)<sup>[5]</sup>,显然该方法运用到本文的问题中费时且低效。而遗传算法(genetic algorithm, GA)操作简单、求解效果好且其中的交叉和变异操作适用于解决离散问题,因此本文选择了 GA 进行模型求解。但传统 GA 在搜索后期易陷入局部最优进而出现搜索停滞等状态,为避免于此,本文对 GA 的交叉和变异操作进行了适当改进。

1) 染色体设置:染色体中只包含代表决策变量  $S$  和  $K$  的基因,基于  $S$  和  $K$  都是已知范围内的整数,故采用二进制编码,方便后续的交叉、变异操作。以图 2 为例,前 4 个基因对应一组  $(S, K)$  中的  $S$ ,后 6 个基因对应这组  $(S, K)$  中的  $K$ 。

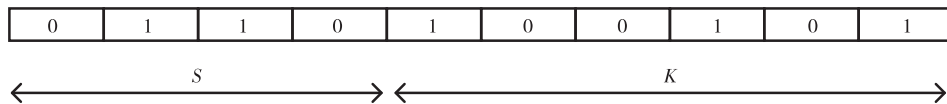


图 2 一组  $(S, K)$  的染色体设置示意图

Fig. 2 Schematic diagram of a piece of chromosome  $(S, K)$

2) 初始化:  $S$  和  $K$  初始化均采用在对应基因位上进行 0-1 随机赋值的方式。

3) 选择: 计算初始化得到染色体的适应值, 记为  $U_i$ 。采用轮盘赌的方式来选择进入交叉操作的染色体。由于适应值越小代表染色体越好, 因此先将所有染色体的适应值求和, 用总适应值减去每一个染色体的适应值, 最后计算每个染色体被选中的概率  $p_i$  和累计概率  $P_n$  分别为:

$$p_i = \frac{(\sum_{i=1}^n U_i) - U_i}{\sum_{i=1}^n [(\sum_{i=1}^n U_i) - U_i]}, P_n = \sum_{i=1}^n p_i。$$

4) 最优交叉:  $\beta$  先产生一个随机数, 当交叉概率  $p_c$  大于  $\beta$  时, 从种群中随机选择 2 个染色体并在最优染色体上随机产生一个交叉的基因位置, 如图 3 所示随机产生的交叉位置为 3, 将最优染色体前 3 个基因替换到染色体 1 上, 将后 7 个基因替换到染色体 2 上。

| 最优交叉前  |   |   |   |   |   |   |   |   |   |   |
|--------|---|---|---|---|---|---|---|---|---|---|
| 随机染色体1 | 0 | 1 | 1 | 0 | 0 | 0 | 1 | 0 | 1 | 1 |
| 随机染色体2 | 1 | 0 | 0 | 1 | 1 | 0 | 0 | 1 | 0 | 0 |
| 最优染色体  | 0 | 1 | 0 | 1 | 0 | 0 | 1 | 1 | 1 | 1 |
| 最优交叉后  |   |   |   |   |   |   |   |   |   |   |
| 随机染色体1 | 0 | 1 | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 1 |
| 随机染色体2 | 1 | 0 | 0 | 1 | 0 | 0 | 1 | 1 | 1 | 1 |

图 3 交叉操作示意图

Fig. 3 Crossover operation diagram

5) 自适应变异: 在进化初期采用较大的变异概率来保持种群的多样性, 而在后期变异概率将逐步缩小以提高局部微调能力。对于每一次种群进化, 都会产生一个相应的变异概率  $p_m$ , 每一个染色体都会先产生一个随机数  $\theta$ , 当  $p_m$  大于  $\theta$  时, 随机产生一个变异的基因位置, 并在此位置上发生基因的替换(0→1, 1→0)。改进后的自适应变异算子为:

$$p_m(g) = p_{\min} + (p_{\max} - p_{\min})e^{\frac{1-G}{G-g+1}}$$

其中:  $p_{\max}$ 、 $p_{\min}$  分别表示最大和最小变异率,  $g$  表示当前迭代次数,  $G$  表示最大迭代次数。

### 3.2 HA

本文构建的是离散情况下的库存分配模型, 决策变量  $S$ 、 $K$  均为非负整数, 由于  $K \leq S$  且通过预实验可以确定  $S$  的范围, 故存在有限个可行解, 则一定存在一个最优解。因此, 本文根据所建立模型的结构特点设计了一种简单高效的 HA, 具体搜索步骤如下:

步骤 1, 在  $S$  的取值范围内随机确定一个  $S$ , 作为初始的最优  $S$ , 通过式(1)搜索出成本最小时的  $K$ , 即为此时的最优  $K$ 。

步骤 2, 固定新确定的最优  $K$ , 通过式(1)搜索出成本最小时的  $S$ , 即为此时的最优  $S$ 。

步骤 3, 判断前后 2 个最优  $S$  是否相等。若相等, 则得到最优解, 算法到此结束; 若不等, 则更新解的空间即在  $S$  的取值范围内去掉前一个最优  $S$ , 然后继续进行下一步。

步骤 4, 固定新确定的最优  $S$ , 通过式(1)搜索出成本最小时的  $K$ , 即为此时的最优  $K$ 。

步骤 5, 判断前后 2 个最优  $K$  是否相等。若相等, 则得到最优解, 算法到此结束; 若不等, 则更新解的空间即在  $K$  的取值范围内去掉前一个最优  $K$ , 然后返回步骤 2, 重复上述步骤, 直至满足结束条件得到最优解。

## 4 数值实验与参数分析

下面先对 IGA 的求解效果进行验证, 再在 EM、IGA、HA 之间进行比较。随后, 进行大量的数值实验, 通过与 FCFS 策略和 SPS 策略进行比较, 展示了阈值策略在降低库存系统成本上的优势。最后, 分析提前期  $L$ 、检查期  $T$ 、到达需求规模均值  $a$ 、单位库存持有成本  $h$  等系统参数对阈值策略相对优势的影响。

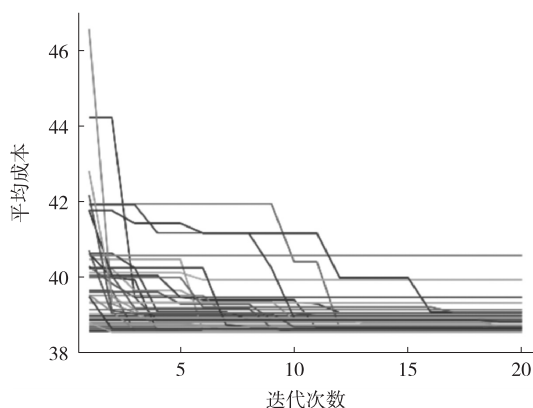
### 4.1 算法效果比较与分析

本文设计了一个基准算例, 其中  $T=1$ ,  $L=0.25$ ,  $\lambda=20$ ,  $b_1=100$ ,  $h=1$  且  $a=2$ , 并基于该算例需求差异的

情况,在  $\lambda_1/\lambda=0.3, b_2/b_1=0.05$  条件下进行了算法测试,验证求解效果。通过预实验事先确定了传统 GA 和 IGA 较好的参数设置,分别为  $p_c=0.7, p_m=0.5$  和  $p_c^1=1, p_{\min}^1=0.2, p_{\max}^1=0.8$  以及种群规模 pop 均为 10、最大迭代次数  $G$  均为 20,并采用这些参数进行后续实验。

#### 4.1.1 IGA 求解效果

为更直观地观察 IGA 相较于 GA 在求解效果上的提升,绘制了重复运行 50 次的算法进化过程图,如图 4、5 所示。从图中可以看出,GA 在 20 代左右勉强收敛且所得解的质量不好也不稳定,而 IGA 大部分都在 10 代左右收敛,14 代以内全部收敛,所得解的质量较好且十分稳定。由此可知,本文设计的 IGA 缩短了求解时间,提高了求解的质量,在求解效果上远胜于 GA。



注:平均成本为相对量,无单位。下同。

图 4 GA 运行 50 次目标函数变化情况

Fig. 4 Changes of the optimal solution after  
GA algorithm 50 times runs

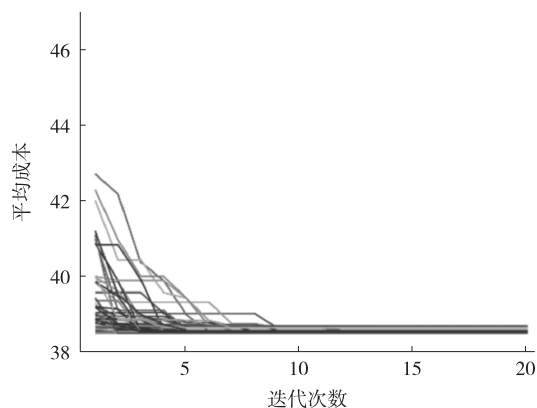


图 5 IGA 运行 50 次目标函数变化情况

Fig. 5 Changes of the optimal cost after  
IGA 50 times runs

#### 4.1.2 EM、IGA、HA 比较结果

为更好地比较 IGA 与 HA 的求解效果,将已有研究中采用的 EM 作为基本比较对象。为保证算法实验结果的准确性,基于上面 3 种方法重复求解了 50 次。从表 2 中可以看出:在平均求解时间上,EM 花费的时远远大于 IGA 和 HA,HA 最小且与 IGA 的差距明显;在解的质量上,由于 EM 求得的是真实最优解,所以重点比较 IGA、HA 与它的差距,显然 HA 较 IGA 在解的均值、标准差、平均最优偏差、最优偏差的标准差上与 EM 的差距更小,因此解的质量更高。总的来说,本文设计的 HA 在求解效果上优于 IGA。

表 2 基于 EM、IGA、HA 的求解结果

Tab. 2 Results obtained by EM, IGA and HA

| 算法  | 平均求解时间/s | 解的均值     | 解的标准差 | 平均最优偏差 | 最优偏差的标准差 |
|-----|----------|----------|-------|--------|----------|
| EM  | 5 956.37 | 38.546 0 |       |        |          |
| IGA | 675.58   | 38.630 7 | 0.067 | 0.220% | 0.174    |
| HA  | 375.58   | 38.557 1 | 0.038 | 0.029% | 0.099    |

#### 4.1.3 不同算例下 IGA 与 HA 的差值结果

为进一步证明本文设计的 HA 在求解效果上优于 IGA,在不同算例下进行了实验,根据实验结果得到了比较后的差值,见表 3。从表 3 可以看出,所有数据均为正,由此可得在各种参数设置发生变动的算例下,本文设计的 HA 在求解效果上普遍优于 IGA。

#### 4.2 不同参数设置下的求解结果

首先,在给定其他参数情况下( $T=1, L=0.25, \lambda_1=12, \lambda_2=8, a=2, b_1=100, b_2=5, h=1$ ),分析库存系统长期平均成本随阈值水平  $K$  变化而变化的情况,结果见图 6。从图 6 可以看出,平均成本先是随  $K$  增加而减少,减少到最小点后又随  $K$  增加而增加,这就意味着存在一个最优阈值水平  $K^*$ 。

本文后续实验采用与现实背景较为接近的参数设置,旨在探究阈值策略在成本上的表现。因此,基于备件

库存系统选择了一个最接近现实情况的参数设定： $T=1, L=0.25, \lambda=20, a=2, b_1=100$  和  $h=1$  来作为基准例子。前面多次提到, 2 类需求的差异主要在于缺货成本和到达率, 因此  $b_2/b_1$  和  $\lambda_1/\lambda$  可以用来表示 2 类需求的差异大小。此外,  $(S^*, K^*)$  表示阈值策略下的最优分配方案, 涉及订货和库存分配 2 个方面。表 4 展示了在基准例子下 2 类需求不同差异大小时的最优分配方案, 表 5 展示了基准例子的 4 个对照例子的最优分配方案。

表 3 不同算例下 IGA 与 HA 的求解结果差值  
Tab. 3 Result differences between IGA and HA under different numerical cases

| 参数                  | 设置变化     | 平均求解时间差值/s | 解的均值差值  | 解的标准差差值 | 平均最优偏差差值 | 最优偏差的标准差差值 |
|---------------------|----------|------------|---------|---------|----------|------------|
| $b_2/b_1$           | 0.05→0.1 | 301.90     | 0.082 5 | 0.082   | 0.192%   | 0.117      |
| $\lambda_1/\lambda$ | 0.3→0.5  | 559.74     | 0.101 8 | 0.179   | 0.237%   | 0.418      |
| $L$                 | 0.25→0.5 | 138.05     | 0.179 2 | 0.189   | 0.439%   | 0.509      |
| $T$                 | 1→1.5    | 159.05     | 0.191 0 | 0.079   | 0.393%   | 0.561      |
| $a$                 | 2→3      | 204.57     | 0.069 3 | 0.191   | 0.120%   | 0.138      |
| $h$                 | 1→1.5    | 307.07     | 0.081 1 | 0.136   | 0.157%   | 0.263      |

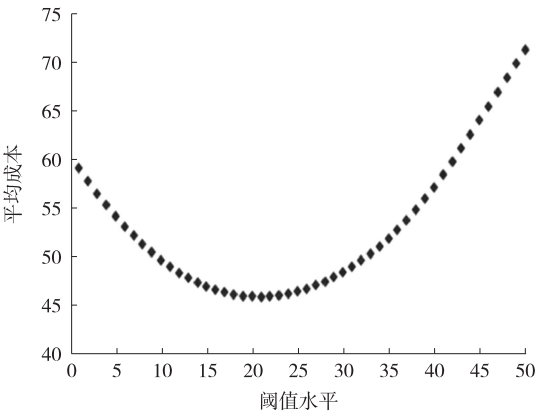


图 6 给定下参数条件下平均成本随阈值水平变化的情况  
Fig. 6 Variation of the average total cost with  $K$  given parameters

表 4 基准例子下的最优分配方案  $(S^*, K^*)$   
Tab. 4 Optimal allocation schemes under the base case  $(S^*, K^*)$

| $b_2/b_1$ | $\lambda_1/\lambda$ |        |         |         |         |
|-----------|---------------------|--------|---------|---------|---------|
|           | 0.1                 | 0.2    | 0.3     | 0.5     | 0.7     |
| 0.05      | (52,5)              | (55,8) | (57,11) | (63,17) | (66,23) |
| 0.1       | (57,3)              | (60,6) | (62,8)  | (66,12) | (69,18) |
| 0.2       | (63,1)              | (64,3) | (66,5)  | (68,8)  | (71,11) |
| 0.3       | (66,0)              | (67,2) | (68,3)  | (70,5)  | (72,8)  |
| 0.5       | (70,0)              | (71,0) | (71,1)  | (73,2)  | (73,4)  |

为更好地说明阈值策略在该库存系统下的表现, 将阈值策略同常见的 FCFS、SPS 策略进行成本比较。通过定义一个相对优势  $R_F$  来说明阈值策略相对于 FCFS 策略的优势大小, 具体如下:

$$R_F = \frac{C(\text{FCFS}) - C(S, K)}{C(\text{FCFS})}。$$

同理, 定义另一个相对优势  $R_S$  来说明阈值策略相对于 SPS 策略的优势大小, 具体如下:

$$R_S = \frac{C(\text{SPS}) - C(S, K)}{C(\text{SPS})},$$

其中:  $C(\text{FCFS})$ 、 $C(\text{SPS})$  分别代表 FCFS、SPS 策略下的系统成本。



下面表 6 和表 7 分别展示了在 2 类需求不同差异大小时,相对优势  $R_F$  和  $R_S$  的大小及变化情况,侧面反映了 2 类需求差异大小对这 2 个相对优势的影响。

首先,从表 6 可以看出,相对优势  $R_F$  在上述情况下基本为正,最大达到 15.94%,但也存在少数为 0 的情况,这表明在上述情况下阈值策略基本优于 FCFS 策略。同样在表 7 中,可以看到相对优势  $R_S$  均为正,且普遍在 10%左右,这表明在上述情况下阈值策略明显优于 SPS 策略。其次,在表 6 中可以发现,相对优势  $R_F$  明显随  $b_2/b_1$  增加而减少,这是因为  $b_2/b_1$  增加意味着 2 类需求差异越来越小,降低了需求分类的必要性,阈值策略也就越来越接近 FCFS 策略。在表 7 中,在  $\lambda_1/\lambda$  小于 0.5 情况下,相对优势  $R_S$  随  $b_2/b_1$  增加而减少;在  $\lambda_1/\lambda$  大于等于 0.5 情况下,相对优势  $R_S$  随  $b_2/b_1$  增加而增加,但这 2 种情况下发生的变化都不太明显,变化量都较小。

表 5 4 个基准对照例子下的最优分配方案( $S^*, K^*$ )Tab. 5 Optimal allocation schemes under four benchmark cases ( $S^*, K^*$ )

| 基准对照                      | $b_2/b_1$ | $\lambda_1/\lambda$ |         |         |          |          |
|---------------------------|-----------|---------------------|---------|---------|----------|----------|
|                           |           | 0.1                 | 0.2     | 0.3     | 0.5      | 0.7      |
| $L: 0.25 \rightarrow 0.5$ | 0.05      | (63,6)              | (66,9)  | (69,12) | (75,18)  | (79,24)  |
|                           | 0.1       | (69,3)              | (72,6)  | (75,8)  | (78,13)  | (81,18)  |
|                           | 0.2       | (74,1)              | (77,3)  | (79,4)  | (81,8)   | (84,12)  |
|                           | 0.3       | (78,0)              | (80,2)  | (81,3)  | (83,5)   | (84,9)   |
|                           | 0.5       | (82,0)              | (83,0)  | (84,1)  | (85,2)   | (86,4)   |
| $T: 1 \rightarrow 1.5$    | 0.05      | (69,6)              | (73,9)  | (75,13) | (81,19)  | (86,27)  |
|                           | 0.1       | (76,4)              | (78,6)  | (81,9)  | (85,15)  | (89,20)  |
|                           | 0.2       | (82,2)              | (84,4)  | (86,6)  | (89,9)   | (92,13)  |
|                           | 0.3       | (86,1)              | (87,2)  | (89,3)  | (91,6)   | (93,8)   |
|                           | 0.5       | (91,1)              | (91,1)  | (93,1)  | (94,2)   | (95,4)   |
| $a: 2 \rightarrow 3$      | 0.05      | (78,8)              | (82,13) | (86,17) | (94,25)  | (100,35) |
|                           | 0.1       | (86,6)              | (89,9)  | (92,12) | (98,19)  | (101,27) |
|                           | 0.2       | (94,2)              | (97,5)  | (99,7)  | (103,12) | (106,18) |
|                           | 0.3       | (99,0)              | (101,2) | (102,4) | (105,8)  | (107,13) |
|                           | 0.5       | (105,0)             | (106,0) | (107,1) | (109,3)  | (109,7)  |
| $h: 1 \rightarrow 1.5$    | 0.05      | (47,6)              | (50,9)  | (53,11) | (58,17)  | (62,25)  |
|                           | 0.1       | (53,4)              | (55,6)  | (57,8)  | (62,13)  | (64,19)  |
|                           | 0.2       | (58,1)              | (60,3)  | (62,5)  | (64,8)   | (66,12)  |
|                           | 0.3       | (62,0)              | (63,2)  | (64,3)  | (66,6)   | (68,9)   |
|                           | 0.5       | (66,0)              | (67,0)  | (67,1)  | (69,2)   | (70,4)   |

表 6 2 类需求不同差异大小下的  $R_F$ Tab. 6  $R_F$  under different differences between two types of demands

| $b_2/b_1$ | $\lambda_1/\lambda$ |        |        |        |        |
|-----------|---------------------|--------|--------|--------|--------|
|           | 0.1                 | 0.2    | 0.3    | 0.5    | 0.7    |
| 0.05      | 10.90%              | 15.05% | 15.94% | 14.27% | 11.23% |
| 0.1       | 3.68%               | 7.05%  | 8.18%  | 8.50%  | 7.12%  |
| 0.2       | 0.38%               | 1.93%  | 2.95%  | 3.75%  | 3.45%  |
| 0.3       | 0                   | 0.42%  | 1.04%  | 1.76%  | 1.84%  |
| 0.5       | 0                   | 0      | 0.01%  | 0.28%  | 0.55%  |

4.3 参数敏感性分析

在本小节中,依次针对提前期  $L$ 、检查期  $T$ 、到达需求规模均值  $a$ 、单位库存持有成本  $h$  等系统参数对阈值策略相对优势  $R_F$  和  $R_S$  的影响进行了敏感性分析。不同提前期下的相对优势  $R_F$  和  $R_S$  的大小及变化情况如表 8、表 9 所示。

表 7 2 类需求不同差异大小下的  $R_S$   
Tab. 7  $R_S$  under different differences between two types of demands

| $b_2/b_1$ | $\lambda_1/\lambda$ |        |        |        |        |
|-----------|---------------------|--------|--------|--------|--------|
|           | 0.1                 | 0.2    | 0.3    | 0.5    | 0.7    |
| 0.05      | 6.48%               | 12.22% | 13.56% | 11.85% | 7.79%  |
| 0.1       | 5.99%               | 11.57% | 12.97% | 12.30% | 8.65%  |
| 0.2       | 5.55%               | 11.02% | 12.91% | 12.80% | 9.51%  |
| 0.3       | 5.39%               | 10.78% | 12.89% | 13.12% | 10.16% |
| 0.5       | 4.68%               | 10.63% | 12.88% | 13.63% | 11.04% |

表 8 不同提前期下的  $R_F$   
Tab. 8  $R_F$  under different lead times

| $L$  | 当 $\lambda_1/\lambda=0.3$ 时,不同 $b_2/b_1$ 取值的 $R_F$ |       |       | 当 $b_2/b_1=0.1$ 时,不同 $\lambda_1/\lambda$ 取值的 $R_F$ |       |       |
|------|----------------------------------------------------|-------|-------|----------------------------------------------------|-------|-------|
|      | 0.05                                               | 0.1   | 0.2   | 0.2                                                | 0.3   | 0.5   |
| 0.25 | 15.94%                                             | 8.47% | 2.95% | 7.05%                                              | 8.47% | 8.50% |
| 0.5  | 15.98%                                             | 8.50% | 2.90% | 7.19%                                              | 8.50% | 8.51% |
| 0.75 | 16.04%                                             | 8.54% | 3.08% | 7.30%                                              | 8.54% | 8.53% |
| 1    | 16.10%                                             | 8.59% | 3.10% | 7.31%                                              | 8.59% | 8.56% |

注:其他参数为  $b_1=100, \lambda=20$ 。下同。

表 9 不同提前期下的  $R_S$   
Tab. 9  $R_S$  under different lead times

| $L$  | 当 $\lambda_1/\lambda=0.3$ 时,不同 $b_2/b_1$ 取值的 $R_S$ |        |        | 当 $b_2/b_1=0.1$ 时,不同 $\lambda_1/\lambda$ 取值的 $R_S$ |        |        |
|------|----------------------------------------------------|--------|--------|----------------------------------------------------|--------|--------|
|      | 0.05                                               | 0.1    | 0.2    | 0.2                                                | 0.3    | 0.5    |
| 0.25 | 13.56%                                             | 13.24% | 12.91% | 11.57%                                             | 13.24% | 12.30% |
| 0.5  | 16.71%                                             | 16.16% | 15.67% | 15.34%                                             | 16.16% | 14.57% |
| 0.75 | 18.25%                                             | 17.60% | 17.15% | 17.25%                                             | 17.60% | 15.75% |
| 1    | 19.26%                                             | 18.56% | 18.05% | 18.35%                                             | 18.56% | 16.68% |

表 8 说明相对优势  $R_F$  随着提前期  $L$  增加而微妙增加,这是因为提前期  $L$  增加意味着需求不确定性增加,而阈值策略是为应对不确定需求而设计出来的,所以当面对的需求不确定性越大时,阈值策略就会更加有效。表 9 说明相对优势  $R_S$  随提前期  $L$  增加而增加,原因同前面一样。

不同检查期下相对优势  $R_F$  和  $R_S$  的大小及变化情况如表 10 和表 11 所示。

表 10 不同检查期下的  $R_F$   
Tab. 10  $R_F$  under different inspection periods

| $T$ | 当 $\lambda_1/\lambda=0.3$ 时,不同 $b_2/b_1$ 取值的 $R_F$ |       |       | 当 $b_2/b_1=0.1$ 时,不同 $\lambda_1/\lambda$ 取值的 $R_F$ |       |       |
|-----|----------------------------------------------------|-------|-------|----------------------------------------------------|-------|-------|
|     | 0.05                                               | 0.1   | 0.2   | 0.2                                                | 0.3   | 0.5   |
| 1   | 15.94%                                             | 8.47% | 2.99% | 7.05%                                              | 8.47% | 8.50% |
| 1.5 | 15.10%                                             | 8.15% | 2.97% | 6.97%                                              | 8.15% | 7.87% |
| 2   | 14.59%                                             | 7.94% | 2.95% | 6.91%                                              | 7.94% | 7.53% |
| 3   | 13.92%                                             | 7.60% | 2.92% | 6.72%                                              | 7.60% | 7.04% |

表 10 说明了相对优势  $R_F$  随检查期  $T$  增加而微妙减少。表 11 说明了相对优势  $R_S$  随检查期  $T$  增加而减少。其一是因为阈值策略需要为高优先级需求预留库存,一旦检查期  $T$  变长,在库存持有成本和缺货成本上带来的潜在风险会明显增加;其二是需求的不确定性可以用协方差来表示,随着检查期  $T$  的增加,需求协方差  $\frac{1}{a}\sqrt{\lambda T}$  会变小,意味着需求的不确定性减少,因此阈值策略相对优势就会减少。

不同需求均值下相对优势  $R_F$  和  $R_S$  的大小及变化情况如表 12 和表 13 所示。

表 12 说明了相对优势  $R_F$  随到达需求规模均值  $a$  增加而微妙增加。表 13 说明了相对优势  $R_S$  随到达需求规模均值  $a$  增加而微妙增加。随着到达需求规模均值  $a$  的增加,需求协方差  $\frac{1}{a}\sqrt{\lambda T}$  也会变小,意味着需求的不确定性减少,因此阈值策略相对优势也就会减少。

不同库存持有成本下相对优势  $R_F$  和  $R_S$  的大小及变化情况如下表 14 和表 15 所示。

表 11 不同检查期下的  $R_S$ Tab. 11  $R_S$  under different inspection periods

| $T$ | 当 $\lambda_1/\lambda=0.3$ 时,不同 $b_2/b_1$ 取值的 $R_S$ |        |        | 当 $b_2/b_1=0.1$ 时,不同 $\lambda_1/\lambda$ 取值的 $R_F$ |        |        |
|-----|----------------------------------------------------|--------|--------|----------------------------------------------------|--------|--------|
|     | 0.05                                               | 0.1    | 0.2    | 0.2                                                | 0.3    | 0.5    |
| 1   | 13.56%                                             | 13.24% | 13.05% | 12.07%                                             | 13.24% | 12.30% |
| 1.5 | 12.85%                                             | 12.99% | 13.00% | 11.86%                                             | 12.99% | 11.94% |
| 2   | 12.14%                                             | 12.51% | 12.65% | 11.66%                                             | 12.51% | 11.49% |
| 3   | 10.95%                                             | 11.47% | 11.79% | 10.84%                                             | 11.47% | 10.53% |

表 12 不同需求均值下的  $R_F$ Tab. 12  $R_F$  under different demand averages

| $a$ | 当 $\lambda_1/\lambda=0.3$ 时,不同 $b_2/b_1$ 取值的 $R_F$ |       |       | 当 $b_2/b_1=0.1$ 时,不同 $\lambda_1/\lambda$ 取值的 $R_F$ |       |       |
|-----|----------------------------------------------------|-------|-------|----------------------------------------------------|-------|-------|
|     | 0.05                                               | 0.1   | 0.2   | 0.2                                                | 0.3   | 0.5   |
| 1.5 | 15.74%                                             | 8.35% | 2.84% | 6.86%                                              | 8.35% | 8.39% |
| 2   | 15.94%                                             | 8.47% | 2.95% | 7.05%                                              | 8.47% | 8.50% |
| 3   | 16.14%                                             | 8.65% | 3.07% | 7.20%                                              | 8.65% | 8.79% |
| 4   | 16.24%                                             | 8.74% | 3.10% | 7.27%                                              | 8.74% | 8.90% |

表 13 不同需求均值下的  $R_S$ Tab. 13  $R_S$  under different demand averages

| $a$ | 当 $\lambda_1/\lambda=0.3$ 时,不同 $b_2/b_1$ 取值的 $R_S$ |        |        | 当 $b_2/b_1=0.1$ 时,不同 $\lambda_1/\lambda$ 取值的 $R_S$ |        |        |
|-----|----------------------------------------------------|--------|--------|----------------------------------------------------|--------|--------|
|     | 0.05                                               | 0.1    | 0.2    | 0.2                                                | 0.3    | 0.5    |
| 1.5 | 13.40%                                             | 13.15% | 12.88% | 11.44%                                             | 13.15% | 12.26% |
| 2   | 13.56%                                             | 13.24% | 12.91% | 11.57%                                             | 13.24% | 12.30% |
| 3   | 14.06%                                             | 13.66% | 13.28% | 12.00%                                             | 13.66% | 12.81% |
| 4   | 14.27%                                             | 13.86% | 13.42% | 12.21%                                             | 13.86% | 13.02% |

表 14 不同持有成本下的  $R_F$ Tab. 14  $R_F$  under different inventory holding cost

| $h$  | 当 $\lambda_1/\lambda=0.3$ 时,不同 $b_2/b_1$ 取值的 $R_F$ |       |       | 当 $b_2/b_1=0.1$ 时,不同 $\lambda_1/\lambda$ 取值的 $R_F$ |       |       |
|------|----------------------------------------------------|-------|-------|----------------------------------------------------|-------|-------|
|      | 0.05                                               | 0.1   | 0.2   | 0.2                                                | 0.3   | 0.5   |
| 0.75 | 14.45%                                             | 7.69% | 2.64% | 6.36%                                              | 7.69% | 7.72% |
| 1    | 15.94%                                             | 8.18% | 2.95% | 7.05%                                              | 8.18% | 8.50% |
| 1.25 | 16.87%                                             | 9.12% | 3.19% | 7.63%                                              | 9.12% | 9.18% |
| 1.5  | 17.63%                                             | 9.60% | 3.53% | 8.06%                                              | 9.60% | 9.57% |

表 15 不同持有成本下的  $R_s$   
Tab. 15  $R_s$  under different inventory holding cost

| $h$  | 当 $\lambda_1/\lambda=0.3$ 时,不同 $b_2/b_1$ 取值的 $R_s$ |        |        | 当 $b_2/b_1=0.1$ 时,不同 $\lambda_1/\lambda$ 取值的 $R_s$ |        |        |
|------|----------------------------------------------------|--------|--------|----------------------------------------------------|--------|--------|
|      | 0.05                                               | 0.1    | 0.2    | 0.2                                                | 0.3    | 0.5    |
| 0.75 | 12.93%                                             | 12.87% | 12.67% | 11.10%                                             | 12.87% | 12.07% |
| 1    | 13.56%                                             | 12.97% | 12.91% | 11.57%                                             | 12.97% | 12.30% |
| 1.25 | 13.80%                                             | 13.52% | 13.12% | 11.93%                                             | 13.52% | 12.58% |
| 1.5  | 14.05%                                             | 13.76% | 13.44% | 12.18%                                             | 13.76% | 12.62% |

表 14 说明了相对优势  $R_F$  随单位库存持有成本  $h$  的增加而增加。表 15 说明相对优势  $R_s$  随单位库存持有成本  $h$  的增加而微妙增加。发生上述变化情况主要是由于阈值策略的最大优势是为高优先级需求预留库存,但随着库存持有成本的增加,库存系统往往会采取持有更少库存的做法,这让库存分配过程显得尤为重要,更能体现出阈值策略的优势所在。

5 结论

本文在复合泊松需求下,研究了在周期检查  $(S, T)$  订货策略下备件库存系统的最优分配问题,在阈值策略下成功建立了考虑需求积压的库存分配模型,并分别设计了 IGA 和 HA 顺利进行求解,最终实现了库存系统长期平均成本的最小化。数值实验结果证明本文提出的 2 个算法相较 EM 明显更高效。最后,对数值实验结果进行比较分析得到以下结论:

- 1) 阈值策略在多类需求库存分配问题上总是最优的,阈值策略相比 FCFS 和 SPS 策略有明显的降本优势,尤其是在缺货成本差异大、需求到达比例恰当、检查期短、提前期长的情况下。
- 2) 阈值策略相对优势随检查期  $T$ 、需求批量均值  $a$  的增加而减少,随提前期  $L$ 、单位库存持有成本  $h$  的增加而增加。

在未来,本文的研究可以从以下几个方向进行适当拓展。首先,本文仅考虑了 2 类需求的情况,这也是大多数相关文献考虑的情况,如果将这一研究拓展到 2 类以上甚至  $n(n \gg 2)$  类需求的情况将会是非常有意义但又极具挑战性的。其次,由于本文考虑的阈值水平  $K$  是静态的,如果  $K$  可以随着时间的推移而动态调整,那么无疑会给阈值策略带来更大的优势。最后,可以将本研究拓展到采用其他订货策略的库存系统中去,例如连续检查下的  $(r, Q)$  订货策略,这也是本文后续工作的重点。

参考文献:

[1] ESCALONA P, ANGULO A, WESTON J, et al. On the effect of two popular service-level measures on the design of a critical level policy for fast-moving items[J]. Computers & Operations Research, 2019, 107: 107-126.

[2] VEINOTT A F. Optimal policy in a dynamic, single product, nonstationary inventory model with several demand classes[J]. Operations Research, 1965, 13(5): 761-778.

[3] 郑长征, 刘志学, 左晓露. 基于购买行为的随机生命周期易逝品库存策略[J]. 控制与决策, 2012, 27(1): 28-34.

ZHENG C Z, LIU Z X, ZUO X L. Optimal policies for deteriorating inventory problems based on purchasing behavior[J]. Control and Decision, 2012, 27(1): 28-34.

[4] LI J X, TORIELLO A, WANG H, et al. Dynamic inventory allocation for seasonal merchandise at Dillard's[J]. INFORMS Journal on Applied Analytics, 2021, 51(4): 297-311.

[5] WAN G Y, CAO Y. A continuous cost evaluation approach for periodic review inventory systems with threshold rationing policy[J]. Computers & Industrial Engineering, 2018, 126: 75-87.

[6] BAO L N, LIU Z Y, YU Y M, et al. On the decomposition property for a dynamic inventory rationing problem with multiple demand classes and backorder[J]. European Journal of Operational Research, 2018, 265(1): 99-106.

[7] ALFIERI A, PASTORE E, ZOTTERI G. Dynamic inventory rationing: how to allocate stock according to managerial priorities. An empirical study[J]. International Journal of Production Economics, 2017, 189: 14-29.

[8] 熊浩, 鄢慧丽. B2C 电子商务二级仓储网络订单履行的在线动态库存分配优化研究[J]. 中国管理科学, 2024, 32(7): 248-257.



- XIONG H, YAN H L. The optimization of online dynamic inventory allocation for order fulfillment in two echelon warehouse network of B2C e-commerce[J]. Chinese Journal of Management Science, 2024, 32(7): 248-257.
- [9] 崔利刚, 刘孝林, 邓洁, 等. 考虑贸易信贷周期的中欧班列联合补货决策[J]. 中国管理科学, 2024, 32(4): 76-88.
- CUI L G, LIU X L, DENG J, et al. Stochastic multi-item joint replenishment problem considering trade credit under China-Europe trade[J]. Chinese Journal of Management Science, 2024, 32(4): 76-88.
- [10] WANG Y, ZHANG S H, ZHOU S X, et al. Anticipated rationing policy for inventory systems with two demand classes and backlogging costs[J]. International Journal of Production Research, 2020, 58(20): 6300-6314.
- [11] MIAO S T, JASIN S, CHAO X L. Asymptotically optimal Lagrangian policies for multi-warehouse, multi-store systems with lost sales[J]. Operations Research, 2022, 70(1): 141-159.
- [12] CHEN H X, DAI B, LI Y, et al. Stock allocation in a two-echelon distribution system controlled by  $(s, S)$  policies[J]. International Journal of Production Research, 2022, 60(3): 894-911.
- [13] HUNG Y F, HSIAO J Y. Inventory rationing decision models during replenishment lead time[J]. International Journal of Production Economics, 2013, 144(1): 290-300.
- [14] KOUKI C, LARSEN C. Rationing policies in a spare parts inventory system with customers differentiation[J]. International Journal of Production Research, 2021, 59(20): 6270-6290.
- [15] DESHPANDE V, COHEN M A, DONOHUE K. A threshold inventory rationing policy for service-differentiated demand classes[J]. Management Science, 2003, 49(6): 683-703.
- [16] KITAEVA A V, ZHUKOVSKAYA A O, ZMEEV O A. Compound Poisson demand with price-dependent intensity for fast moving items: price optimization and parameters estimation[J]. International Journal of Production Research, 2017, 55(14): 4153-4163.
- [17] ROSS S M. Stochastic processes[M]. New York: John Wiley, 1996.

## Operations Research and Cybernetics

### An Investigation of Threshold-Based Inventory Rationing Strategy under Compound Stochastic Demand

CUI Ligang<sup>1,2</sup>, XIE Wenjing<sup>1</sup>, WU You<sup>1</sup>, YANG Jinghao<sup>1</sup>

(1. School of Economics and Management, Chongqing Jiaotong University;

2. Intelligent Supply Chain Engineering & Technology Research Center of Chongqing, Chongqing 400074, China)

**Abstract:** In order-driven inventory decision-making, the arrival of demand orders is often characterized by multiple stochastic superposition, i. e., random arrival of demand orders and random presentation of demand sizes. To solve the common inventory problem in this context, it considers the optimal allocation problem of spare parts inventory under the demand arrival and size conforming to the compound stochastic conditions. Specifically, with the spare parts inventory system as the research object under the cycle check  $(S, T)$  ordering strategy, it studies the inventory allocation model for two types of demands with high and low shortage costs based on the backlog demand clearance mechanism of the threshold strategy under the composite stochastic demand conditions to achieve the minimization of inventory costs. For model solving, compared with the enumeration method only used in previous studies, an improved genetic algorithm is applied for the first time, and a simple and efficient heuristic algorithm is designed. The results of numerical experiments show that both the heuristic algorithms are better than the first-come-first-served (FCFS) and stock-piling split (SPS) strategies and indicate that the threshold strategy has a significant advantage over the first-come-first-served (FCFS) and stock-piling split (SPS) strategies. The sensitivity analysis proves that the relative cost advantage of the threshold strategy decreases with the increase of the checking period, the mean value of the arriving demand size, and increases with the increase of the lead time and the cost of the unit stock-piling holding. The results of the project provide decision-making reference for lean inventory management.

**Keywords:** compound poisson demand; out-of-stock cost; threshold strategy; heuristic algorithms; improved genetic algorithm

(责任编辑 黄 颖)