

清 华 大 学

# 综 合 论 文 训 练

题目：保留在制品库存对安全库存的  
影响

---

培养单位：工业工程系

学    科：工业工程

姓    名：徐驰

指导教师：吴甦教授

2014 年 6 月 15 日

# 关于学位论文使用授权的说明

本人完全了解清华大学有关保留、使用学位论文的规定，即：学校有权保留学位论文的复印件，允许该论文被查阅和借阅；学校可以公布该论文的全部或部分内容，可以采用影印、缩印或其他复制手段保存该论文。

**(涉密的学位论文在解密后应遵守此规定)**

签 名：\_\_\_\_\_ 导师签名：\_\_\_\_\_ 日 期：\_\_\_\_\_

## 中文摘要

**关键词：**在制品库存    安全库存    延迟差异化

## ABSTRACT

**Key words:** In-process inventory    Safety stock    Postponement

# 目 录

第 1 章 引言 .....	1
1.1 课题背景 .....	1
1.2 问题描述 .....	1
1.3 研究目标 .....	2
1.4 意义和价值 .....	2
第 2 章 文献综述 .....	3
2.1 安全库存和安全时间 .....	3
2.2 多阶段生产和在途库存 .....	5
2.3 风险分担 .....	7
2.4 延迟差异化 .....	8
第 3 章 研究思路 .....	9
3.1 证明优化潜力存在性 .....	9
3.2 成品需求相关性的影响 .....	10
3.3 提前期变化的影响 .....	10
3.4 最优库存策略的探索（待定） .....	11
3.5 实际生产数据的应用 .....	11
第 4 章 优化潜力存在性 .....	12
4.1 安全库存的制定标准 .....	12
4.2 两种成品需求独立正态分布 .....	15
4.3 两种以上的成品需求独立正态分布 .....	17
4.4 一些参数对改进效果的影响 .....	18
第 5 章 改进失效现象 .....	19
5.1 对称稳定分布的参数与改进失效的关系 .....	19
5.2 右偏斜的需求分布一定存在改进失效的可能 .....	21
5.3 任何服务水平都存在改进失效的可能 .....	23

5.4 泊松分布下发生改进失效的区间图示 .....	26
5.4.1 第二类服务水平与改进失效 .....	29
5.5 避免改进失效 .....	31
<b>第 6 章 需求相关性对改进效果的影响 .....</b>	<b>32</b>
6.1 需求相关性的现实意义 .....	32
6.2 需求服从二维正态分布的情况 .....	32
6.3 需求服从多维正态分布的情况 .....	35
6.4 极端情况的讨论 .....	36
<b>第 7 章 Backorder 对改进效果的影响 .....</b>	<b>38</b>
7.1 考虑 Backorder 成本的任意更新过程 .....	38
7.2 考虑 Backorder 成本的泊松过程 .....	39
7.3 服务水平约束的影响 .....	41
<b>第 8 章 提前期对改进效果的影响 .....</b>	<b>42</b>
8.1 需求正态分布的情况 .....	42
8.2 需求泊松到达的情况 .....	42
8.3 $(s,S)$ 库存策略下的情况 .....	42
<b>插图索引 .....</b>	<b>43</b>
<b>表格索引 .....</b>	<b>44</b>
<b>公式索引 .....</b>	<b>45</b>
<b>参考文献 .....</b>	<b>48</b>

# 第 1 章 引言

## 1.1 课题背景

随着 MIS、MRP 等技术在汽车制造业的应用，汽车行业对零部件供应的及时性和服务水平要求越来越高。

G 公司是全球最大的汽车制造商之一。M 公司作为汽车零部件供应商，为 G 公司某些型号的汽车生产保险杠。

保险杠的生产流程需要经过注塑、喷涂、装配三个阶段，下面对每个阶段的作用和特点作简要介绍。

**注塑** 注塑是保险杠生产的第一步，即通过注塑机器和模具，制造出特定形状的保险杠。注塑不同型号的保险杠时需要换模，换模时间相当长。为了提高生产效率，需要在注塑阶段设置合适的生产批量，尽量做到少换模。公司当前拥有注塑机器 3 台，生产的保险杠型号共十余种，并且要求每台机器每天换模不超过 2 次。这种策略下，每天能投入生产的保险杠型号是有限的，因此有必要保持一定量的安全库存。

**喷涂** 喷涂的作用是对注塑成型的保险杠上色。每种型号的保险杠有几种不同的颜色可选。喷涂工序是在连续运转的环形传送带上完成的，工人在某个位置将未喷涂的保险杠注塑件挂到传送带上的支架上，传送一段距离后进入喷涂机器，完成喷涂后在某个位置取下保险杠，然后空支架又运转到上货处。喷涂机器更换喷涂颜料需要一定的时间和人力，因此也要尽量设置合理的喷涂批量，减少颜料更换次数。

**装配** 装配是指给保险杠安装必要的配件。经过前两道工序，保险杠的主体已经基本完成。在打包发往 G 公司之前，还需要在保险杠上安装一些配件。这道工序耗时不多，且相当稳定，一般能保证在规定时间内完成装配并发货。

## 1.2 问题描述

如前所述，受限于注塑机器的换模次数，M 公司无法在一天内生产每一种型号的保险杠塑件，因此需要保有足量的安全库存，以应对可能发生的需求波动。不仅如此，由于每种型号的保险杠有不同颜色之分，进一步增大了需求的

不确定性。如果每种型号每种颜色的保险杠都保有足量安全库存，无疑使得安全库存量变得更大。

公司希望能在不影响服务水平、不降低生产效率的前提下，通过采用合理的生产和库存策略，减少安全库存总量。

### **1.3 研究目标**

通过数学模型分析将分化的成品安全库存合并前置到在制品库存是否有助于减少安全库存，根据研究结果制定合适的库存策略，并进一步分析此改进方案的适用范围和优化效果。最后将企业实际生产线数据代入到方案中，理论分析此方案在企业可能产生的改善效果。

### **1.4 意义和价值**

对于一般企业而言，在不降低服务水平和生产效率的基础上，通过库存合并前移减少安全库存，能够减少生产线的库存成本，节省场地和设备，解放更多的流动资金，对生产线的生产柔性有一定的提升。

除此之外，库存的前移对于汽车保险杠塑件生产线还有更多的价值。保险杠塑件喷漆后存放过程中，可能因为磕碰、受热等各种原因导致油漆脱落、变脏等缺陷。如果把库存移到喷涂工序之前，就能在一定程度上避免这些风险，提高成品质量。



## 第 2 章 文献综述

为了更深入地研究上述问题，我们首先要明确安全库存、在途库存等概念的准确定义和计算方法，然后了解多阶段生产的特点以及用于研究多阶段生产的现有数学模型。

### 2.1 安全库存和安全时间

Fetter 和 Dalleck (1961)<sup>[1]</sup> 认为安全库存是为了满足一定服务水平而保留的库存。作者根据日销售量和补货周期等因素，提出了计算一定服务水平下的安全库存的方法。

Lambrecht 等人 (1984)<sup>[2]</sup> 在研究物料需求计划系统 (MRP) 时指出，安全库存是为了对抗需求预测、生产批量、生产运输提前期等因素中存在的不确定性。作者根据供应链中的级库存模型，将多阶段生产计划建模成动态规划问题和马尔科夫决策问题，改进了 Clark 和 Scarf (1962)<sup>[3]</sup> 的启发式算法，并指出了启发式算法的上下界和获得最优解条件。

Baker 等人 (1986)<sup>[4]</sup> 研究了多阶段生产线中，提高零部件通用性对安全库存的影响。作者结合风险分担思想在供应链等领域中的应用，用数学模型证明，在给定最低服务水平的前提下，引入通用件能够降低两阶段、两产品生产线的安全库存，并总结了引入通用件产生的影响：所有零件的安全库存总数降低；通用件的安全库存低于它所替代的两种零件安全库存之和；没有被通用件替代的其他零件安全库存增加。

Eppen 和 Martin (1988)<sup>[5]</sup> 总结了各种情况下安全库存的计算方式。当需求和提前期分布已知时，令  $W$  代表提前期内需求， $X$  代表单位时间需求， $Y$  代表提前期长度， $R$  代表补货点。则有以下公式。

$$\mu_W = \mu_X \mu_Y \quad (2-1)$$

$$\sigma_W^2 = \sigma_X^2 \mu_Y + \sigma_Y^2 \mu_X^2 \quad (2-2)$$

$$R = \mu_W + Z_\delta \sigma_W \quad (2-3)$$

当需求和提前期分布未知时，需要视情况建立合适的模型估算它们的分布。

Molinder (1997)<sup>[6]</sup> 认为生产中的不确定性主要来自需求、制造、运输等环节, 并进一步指出制造环节的不确定性是由产能负载、机器故障、排队效应、返工等多种因素引起的。作者认为这些不确定性对多阶段生产的影响尤其显著, 因为某一工序的延迟可能导致物料清单 (BOM) 上所有更高级别的物料都受到影响。作者通过模拟退火优化了多种参数情形下的订货批量、安全库存和提前期, 并比较了每种情形下用安全库存或安全时间作为对抗风险准则的效果。实验结果表明:

1. 需求不确定性高、提前期不确定性低时, 宜采用安全库存作为优化准则; 需求和提前期不确定性都高时, 宜采用安全时间作为准则。
2. 产品结构 (或 BOM) 基本不影响采用哪种准则的抉择。
3. 当缺货成本与库存成本之比较低时, 安全库存准则比其他准则都好。
4. 当上述比值较高时, 准则的抉择往往取决于需求不确定性。需求不确定性低, 宜采用安全库存; 反则宜采用安全时间。

Chopra 等人 (2004)<sup>[7]</sup> 认为降低供应商的供货提前期和降低供货提前期的不确定性是两种降低安全库存的有效方法。一般认为, 在正态模型下, 降低供货提前期的不确定性比降低供货提前期更为有效, 尤其是在供货提前期不确定性较大的情况下。本文作者以 Eppen 和 Martin (1988)<sup>[5]</sup> 的工作为基础, 证明了该结论不是完全正确的, 而是存在一个不小于 50% 的服务水平临界点, 当服务水平低于这个临界点时, 降低提前期的不确定性反而会使安全库存增大。若模型是正态的, 则这个临界点为 50%; 若不是正态的, 临界点一般在 50% 到 70%, 正好是一般企业的服务水平范围。因此企业如果希望降低安全库存, 必须谨慎检查自己的服务水平和提前期分布模型, 盲目降低提前期的不确定性反而可能增大安全库存。文章的最后, 作者还提出了通过累积分布函数 (CDF) 估算临界点的方法。

Fang 等人 (2013)<sup>[8]</sup> 考虑需求不确定性、提前期、提前期不确定性的边际价值, 建立了能同时处理需求和供应两边不确定性的模型, 并且不依赖于最优库存策略。相反, 非最优的库存策略下, 从这几个方面获得的边际价值更大。因此, 即使面临糟糕的库存管理策略, 管理人员也能够根据此模型很快找出最有效率的改进方向。

## 2.2 多阶段生产和在途库存

Gershwin (1994)<sup>[9]</sup> 研究了如图 2.1 所示的单产品多阶段生产线。其中  $M_i$  代表机器， $B_i$  代表缓存。

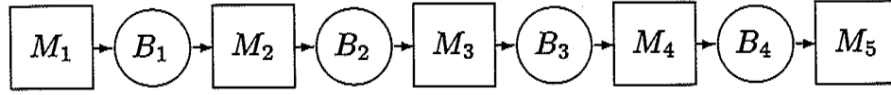


图 2.1 单产品多阶段生产

多阶段生产中的重要制约因素是机器故障。机器故障有两种模式：基于时间的故障模式和基于运作的故障模式。两种模式都是随机的。基于时间的故障模式下，机器故障的发生取决于使用的时间长短，比如材料的强度失效；基于运作的故障模式下，机器故障的发生取决于使用次数或产量，比如刀具失效。

机器即使没有出故障，也不一定运转，因为还受到上下游库存的限制。如果上游库存不足，导致下游机器停工，我们称下游机器为缺料状态；如果下游库存已满，导致上游机器停工，我们称上游机器为阻塞状态。

缓存区的库存被称为在途库存。在多阶段生产中，在途库存往往是必不可少的，因为它能减少机器故障带来的损失。当一台机器出现故障时，上游的机器可以继续运转，生产的产品进入在途库存；下游机器也能依靠故障机器的在途库存提供原料继续运转。但在途库存也会带来一些问题，Gershwin 在书中总结了在途库存的主要缺点：

1. 在途库存是消耗资金制造出来的，占用了资金却不产生任何效益。
2. 根据 Little's Law<sup>[10]</sup>，在途库存对提前期有影响。在途库存越多，提前期越长。提前期过长可能导致顾客不满，同时也使得生产中的故障更难以被观察到。在观察到故障之前，故障机器可能已经制造了很多残次品。
3. 零部件在库存期间可能遭受破损、偷窃等损失。库存量越大、时间越长，遭受的损失越大。
4. 储存用的场地和设备会消耗空间和资金。

Gershwin 还用马尔科夫模型分析了多阶段生产线在缓存容量上限  $N = 0$  和  $N = \infty$  的极端情形下的表现，得出了两个重要的结论：

1. 在  $N = 0$  的情况下，多阶段生产线的运作时间与总时间之比总是随机器数目增加而减小，并且收敛于 0。也就是说，如果没有在途库存，当产品需要的生产阶段很多时，生产线的效率将趋近于 0。

2. 在  $N = \infty$  的情况下，多阶段生产线的产能总是取决于最慢的一台机器。也就是说，总体产能只有在瓶颈阶段可能得到优化。

Gershwin 指出，用马尔科夫模型研究多阶段生产还存在一些难点。最突出的难点就是状态空间过大。每个机器有 2 种状态，每个缓存有  $N + 1$  种状态，马尔科夫模型的状态空间就是它们的乘积，即

$$M = 2^k \prod_{i=0}^{k-1} (N_i + 1)$$

随着阶段数目的增加，状态空间  $M$  很快就变得非常庞大。

Goyal (1978)<sup>[11]</sup> 考虑需求分布均匀、不允许缺货的多阶段生产，建立了以减少总成本为优化准则的数学模型，用于分析最优的生产批量以及各阶段之间最优的运输批量。

Kimura 和 Terada (1981)<sup>[12]</sup> 分析了多阶段拉动式生产的需求和库存波动是如何逐级传递的，证明了拉动式生产系统比推动式生产系统更能适应需求的波动，并能够防止下游的生产和库存波动向上游逐级扩大。Kimura 和 Terada 所定义的拉动式生产系统需要满足以下运行方式：建立标准的再订货点和订货批量大小；随时了解库存水平和延期订单；连续检查库存水平，一旦某些零部件数量低于再订货点，就立即向上游订货。拉动式生产系统中使用看板来维持运作，流程如图 2.2 所示。

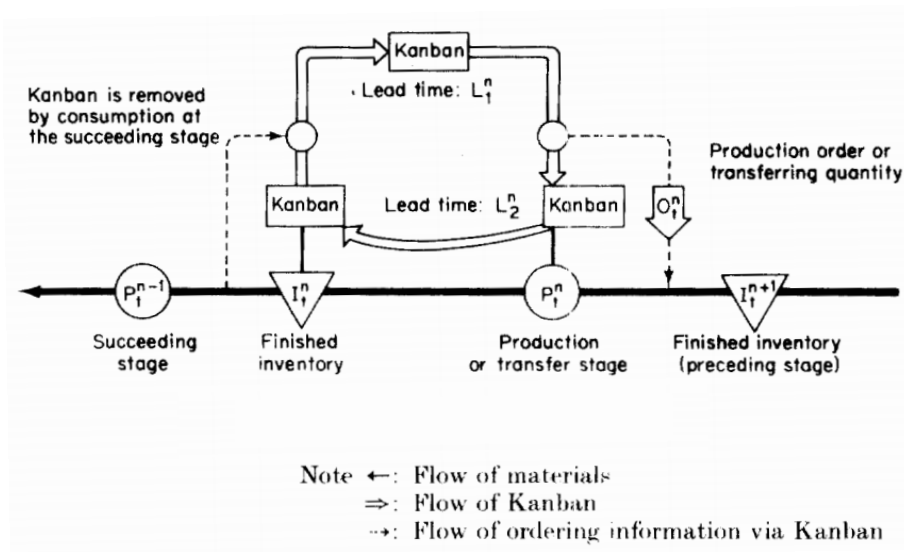


图 2.2 拉动式生产系统流程图

研究中使用的推动式生产系统是 Tabe 等人 (1980)<sup>[13]</sup> 建立的模型, 如图 2.3 所示。

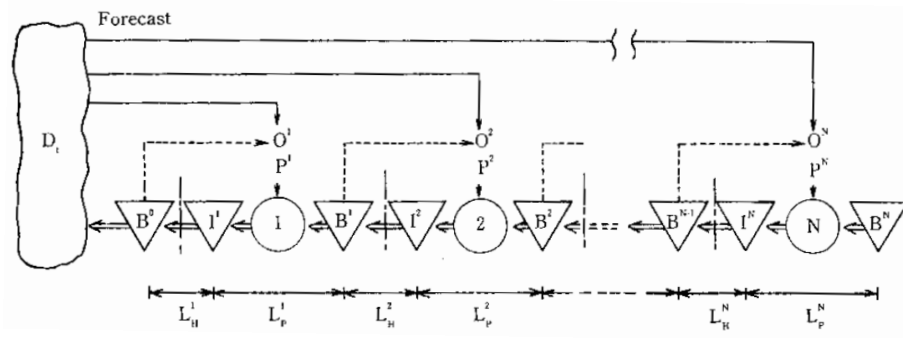


图 2.3 推动式生产系统流程图

通过数学模型分析和仿真模型实验, Kimura 和 Terada 得到如下结论:

1. 拉动式生产系统中, 单位订货量的大小非常重要。在单位订货量与产量水平相比很小的情况下, 拉动式生产中的下游生产波动不会在向上游传递的过程中放大。
2. 推动式生产系统中, 生产和库存波动的放大率取决于预测误差。以波动放大率为准则时, 推动式系统和拉动式系统之间的抉择取决于预测误差的大小。
3. 拉动式生产系统中另一个影响波动放大率的参数是从看板离开容器到该阶段完成生产的时间差。这个时间差越长, 波动放大率越高。

Conway 等人 (1988)<sup>[14]</sup> 反对一味追求降低在途库存的做法, 全面总结了在途库存在多阶段生产中的意义和作用, 并分析各种情况下如何真正改善多阶段生产的绩效。该研究中有以下一些重要的结论: 靠前的几个阶段对整条生产线的产能影响最大, 长生产线的绩效比短生产线差得不多; 添加缓存对绩效的提升较大, 但增加缓存容量的效果不显著; 在生产线靠中间的位置添加缓存, 效果比靠两端的位置好得多; 在未平衡的生产线中添加 buffer 效果不显著。

## 2.3 风险分担

已读 17 篇 risk pooling 相关文献, 待补充文献综述。

## 2.4 延迟差异化

已读 10 篇 postponement 相关文献，待补充文献综述。

## 第3章 研究思路

### 3.1 证明优化潜力存在性

将成品库存合并前移到在途库存，会造成安全库存的变动。影响这个变动的因素是多方面的：风险分担（risk pooling）使安全库存减少；但在发货之前必须先把在途库存制成成品，相当于增加了提前期，使安全库存增大。

为了证明生产线确实存在优化潜力，我们从极端情况入手，先假设喷涂工序不消耗时间，即忽略库存前移对提前期的影响。如果这种情况下安全库存减少了，那么至少可以认为：新的库存点与成品阶段之间的工序耗时不多的情况下，库存合并前移能减少安全库存。如此即证明了生产线确实存在可优化的潜力。

研究这个问题时，可以首先从最简单的情况入手：假设某注塑件在喷涂阶段分化成两种不同颜色的保险杠成品，并且两种保险杠成品的需求服从独立同分布的正态分布。我们需要证明的是，在这种情况下，如果不保留两种保险杠成品的库存，而是保留喷涂前的注塑件在制品库存，是否能在保持原有服务水平的同时，使得安全库存总量降低。

在此基础上，可以把结论推广到两种不同颜色的保险杠需求独立，但服从不同的正态分布的情况。最后可以推广到多种不同颜色的保险杠需求独立，且服从不同的正态分布的情况。这样，我们就能够证明，当保险杠需求服从独立的正态分布时，将在制品库存保留在分化前的阶段，能够使安全库存总量降低。

正态分布在实际生活中是普遍存在的，并且它拥有很多良好的数学性质，因此我们常常通过对正态分布的研究来作为问题的切入点。但是在讨论生产方面的问题时，我们也常常用泊松过程来模拟一些过程，比如服务的时间、设备的失效或者需求的到达。如果需求的到达是一个泊松过程，则一段时间内的需求就是服从泊松分布的。

可以证明，当泊松分布的参数  $\lambda$  较大时，可将泊松分布近似处理为正态分布，这样就能把问题转化为刚刚已解决的正态分布的情况。然而在企业的实际生产中，有一些产品的需求量不大，不适合正态近似，因此对泊松分布本身的研究也是必要的。由于泊松分布并不具有正态分布一样良好的数学性质，在对

其进行纯数学上的证明遇到困难时，可以考虑用数值模拟的方式进行一些简单的实验。

### 3.2 成品需求相关性的影响

将不同颜色相同型号的成品库存合并前移，成品之间的相关性必然会影响到方案的效果。因此需要分类讨论成品需求相互独立、需求正相关、需求负相关等情况。

**独立** 成品需求相互独立，意味着一种颜色的保险杠需求量对相同型号另一种颜色的保险杠需求量没有影响。这是最简单的假设，但有时候可能不符合实际情况。

**正相关** 某种颜色的保险杠需求变化，可能是由于该车型的市场需求变化。此时该车型的各种颜色需求可能同时变化。也就是说，相同型号不同颜色的保险杠需求可能倾向于同时增大或减小。这是生产实际中可能发生的情况。

**负相关** 如果某种车型的市场需求相对稳定，则意味着，一种颜色的需求增长会抑制另一种颜色的需求，表现为相同型号不同颜色的保险杠需求负相关。这也是生产实际中可能发生的情况。

不同的需求相关性会影响库存改进方案的效果。显然，成品需求负相关的情况下，此改进方案的效果将会更加突出。分析成品需求完全正相关和完全负相关的极端情况，能帮助我们明确改进方案的绩效范围，找到改进效果的上下界，从而更好地作出决策。

研究的最后，需要对企业生产线的实际数据进行统计分析，找出不同颜色相同型号的成品需求之间的实际相关性，并针对该结果做具体分析。

### 3.3 提前期变化的影响

在实际生产中，库存前移会导致供货提前期发生变化。以汽车保险杠的生产为例，假设原定的供货提前期是  $T$ ，喷涂工序所需时间是  $t$ ，在新的库存方案下，由于需要先喷涂再发货，相当于把留给企业的反应时间缩减到了  $T - t$ ，这就对企业的敏捷性和柔性提出了挑战。

显然，当  $t$  较小时，我们可以忽略它，把问题转化为上面讨论过的情况。但是当  $t$  值大到不能忽略时，必然会影响到方案的效果，最糟糕的情况下甚至可



能使安全库存增大。因此，研究中需要考虑提前期变化对改进方案的影响，希望能找到一定条件下，方案能产生改进效果的提前期临界点，作为决策支持。

研究提前期的影响时仍然可以以正态分布和泊松分布来研究。

正态分布的情况下，可以考虑相同库存总量情况下，改进前后的缺货概率大小。对于改进前后的系统来说，提前期内能够供货的最大量都等于成品库存与提前期内产量之和。以课题背景中提到的汽车保险杠生产为例：

**改进前** 持有各颜色的保险杠成品库存。一旦发现需求过高，成品库存无法满足需求，就需要立即开工生产，先注塑，再喷涂，争取在提前期内生产出足够的数量。

**改进后** 不保留成品库存。当需求到来时，从在制品库存中选取相应数量的塑件，经过喷涂后发货；如果发现塑件总量不足以满足需求，则同时启动注塑，以补充塑件在制品库存。

注塑、喷涂各自的生产时间和提前持有的库存总量都会影响两种方案的绩效。此处需要进行更为细致的讨论。

泊松分布的情况下，由于需求是泊松到达，我们可以把问题建模成连续时间的马尔科夫过程，通过求解稳态概率来比较两种方案的缺货概率。由于生产过程不止一个阶段，可能还需要用到 Jackson 网络。

### 3.4 最优库存策略的探索（待定）

考虑  $(S, s)$  库存策略下如何确定最优的  $S$  和  $s$ 。也可以考虑如何是否只把一部分库存前移能达到最优。

### 3.5 实际生产数据的应用

在理论研究的结论中代入企业的实际生产数据，检验方案的实际改进效果，为企业提供决策支持。

## 第 4 章 优化潜力存在性

在本章中，我们将忽略相关性和提前期等一切影响因素，单纯从需求的概率分布变化来探究在制品库存前置策略对安全库存的优化作用。

### 4.1 安全库存的制定标准

(此处待补充一些文献，详细讨论什么是安全库存)

安全库存的计算方式是总库存减去需求的期望值。由于本文讨论的所有情况都不涉及需求期望值的变动，因此在比较策略实施前后安全库存的变化时，总库存的减少量与安全库存的减少量始终是等价的。本文后续讨论中不再刻意强调安全库存与总库存的区别，以减少繁琐的数学表达，使数学公式的推导更简洁。

在讨论安全库存时，我们常用的标准有惩罚成本和服务水平。

惩罚成本是企业缺货时需要付出的成本。合同罚款就是最常见的惩罚成本之一。在汽车配件的生产过程中，配件生产商和需求方签订合同，当生产商不能按时供货时，就需要承受一笔罚款。惩罚成本能够比较直观地反应缺货带来的损失，体现安全库存的作用。

然而，惩罚成本在使用过程中有一个难以解决的问题：如何准确确定惩罚成本。惩罚成本不仅包括实际的资金损失，还应该考虑其他隐性损失。以汽车配件生产商为例，合同规定的罚款只是表面上的惩罚成本，实际的损失还应该包括企业的声誉损失。汽车行业非常重视精益生产，缺货造成的声誉损失可能使企业丢失更多的潜在订单，甚至失去合作伙伴。这些损失比表面上的罚款更严重，但很难对其进行量化。

因此，如果仅仅使用合同的罚款数额作为惩罚成本，那么分析过程使用的惩罚成本是小于实际损失的。这种情况下，企业做出的决策，会更倾向于承受损失。这是对企业有害的。

本文中，我们将主要使用服务水平作为安全库存的制定标准。这样做的原因不仅包括以上讨论的惩罚成本的缺陷，还有给出更为科学的理由。在接下来的论述中，我们将会证明，如果使用惩罚成本作为安全库存的制定标准，与使

用服务水平相比，并不能带来更多的优势。

若只考虑惩罚成本，不考虑库存成本，则企业需要设定一个自己能接受的惩罚成本期望的最大值。我们首先分析惩罚成本的期望。设企业的库存为  $K$ ，惩罚成本为  $p$ ，需求服从分布  $P(X=x)=f(x)$ ，其累积分布为  $P(X<x)=F(x)$ 。则惩罚成本为

$$C(x, K) = \begin{cases} 0, & x \leq K \\ p(x - K), & x > K \end{cases} \quad (4-1)$$

如果期望存在的话，其对应的期望是

$$\begin{aligned} E[C(x, K)] &= \int_{-\infty}^{+\infty} C(x, K) f(x) dx \\ &= \int_K^{+\infty} p(x - K) f(x) dx \\ &= p(x - K) F(x) \Big|_K^{+\infty} - \int_K^{+\infty} p F(x) dx \\ &= \lim_{M \rightarrow +\infty} \left[ p(M - K) - \int_K^M p F(x) dx \right] \\ &= \lim_{M \rightarrow +\infty} \left[ \int_K^M p dx - \int_K^M p F(x) dx \right] \\ &= p \int_K^{+\infty} (1 - F(x)) dx \end{aligned} \quad (4-2)$$

另一方面，库存为  $K$  时，对应的服务水平是

$$\eta = \int_{-\infty}^K f(x) dx \quad (4-3)$$

由公式4-2和4-3可知，惩罚成本期望是随  $K$  单调递减的，服务水平是随  $K$  单调递增的。二者可以一一对应，即每个惩罚成本期望值都对应某个特定的服务水平。企业制定自己的惩罚成本期望值，和制定自己的服务水平，是等价的。

使用服务水平来计算安全库存时，不需要考虑库存成本。使用惩罚成本时，可以把库存成本也加入到模型中。接下来我们将证明，加入库存成本之后，惩罚成本与服务水平是一一对应的。以下推导过程中使用离散的库存量和需求量。

设企业的库存为  $K$ ，需求为  $D$ ，惩罚成本为  $p$ ，库存成本为  $h$ 。则总成本为

$$C(K) = p \max\{D - K, 0\} + h \left( \frac{1}{2} \min\{D, K\} + \max\{K - D, 0\} \right) \quad (4-4)$$

企业的目标是寻找最优的库存  $K$ ，使得库存成本和惩罚成本之和的期望最小。我们对公式4-4中的几个部分分别求期望。

$$\begin{aligned}
 E(\min\{D, K\}) &= \sum_{i=1}^{\infty} \min\{i, K\} P(D = i) \\
 &= \sum_{i=1}^{\infty} \sum_{j=1}^{\min\{i, K\}} P(D = i) \\
 &= \sum_{j=1}^K \sum_{i=j}^{\infty} P(D = i) \\
 &= \sum_{j=1}^K P(D \geq j) \tag{4-5}
 \end{aligned}$$

$$\begin{aligned}
 E(\max\{D - K, 0\}) &= -\min\{K - D, 0\} \\
 &= -(\min\{K, D\} - D) \\
 &= D - \sum_{j=1}^K P(D \geq j) \tag{4-6}
 \end{aligned}$$

$$\begin{aligned}
 E(\max\{K - D, 0\}) &= \sum_{i=1}^{\infty} \max\{K - i, 0\} P(D = i) \\
 &= \sum_{i=1}^{\infty} \sum_{j=1}^{\max\{K-i, 0\}} P(D = i) \\
 &= \sum_{j=1}^{K-1} \sum_{i=1}^j P(D = i) \\
 &= \sum_{j=1}^{K-1} P(D \leq j) \tag{4-7}
 \end{aligned}$$

将公式4-5、4-6和4-7代入公式4-4可得

$$E[C(K)] = pD + \left(\frac{h}{2} - p\right) \sum_{j=1}^K P(D \geq j) + h \sum_{j=1}^{K-1} P(D \leq j) \tag{4-8}$$

考虑库存增加时，总成本的增量

$$\begin{aligned}
 E[C(K+1)] - E[C(K)] &= \left(\frac{h}{2} - p\right) P(D \geq K+1) + h P(D \leq K) \\
 &= \frac{h}{2} - p + \left(\frac{h}{2} + p\right) P(D \leq K) \tag{4-9}
 \end{aligned}$$

公式4-9是随  $K$  单调递增的，因此，当  $E[C(K+1)] - E[C(K)]$  取得 0 值时， $E[C(K)]$

取得最小值。此时有

$$P(D \leq K) = \frac{2p - h}{2p + h} \quad (4-10)$$

其中  $P(D \leq K)$  即等于企业的服务水平  $\eta$ 。本文讨论的库存策略仅移动在制品库存在生产线上的位置，不影响库存费用和需求分布。因此，给出任意的惩罚成本，都有一个唯一的服务水平与之对应。

通过以上的讨论，我们已经证明了，以惩罚成本和库存成本来制定安全库存的问题，可以转化为以服务水平来制定安全库存的问题。为了使数学推导更加简洁，本文余下的部分将主要使用服务水平作为安全库存的制定标准。企业可以根据实际情况选取合适的制定方式，且不影响本文中结论的适用性。

## 4.2 两种成品需求独立正态分布

仍然以汽车保险杠生产为例，假设某种汽车保险杠只有两种颜色，且两种颜色的需求服从独立的正态分布。第一种颜色的保险杠，需求  $D_1$  服从正态分布  $N_1(\mu_1, \sigma_1^2)$ ；第二种颜色的保险杠，需求  $D_2$  服从正态分布  $N_2(\mu_2, \sigma_2^2)$ 。企业设定的服务水平为  $\eta$ 。设  $z_\eta$  是标准正态分布的  $\eta$  分位数，即  $z_\eta$  满足

$$\int_{-\infty}^{z_\eta} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx = \eta$$

则此服务水平下，两种成品的库存分别为

$$\xi_1 = \mu_1 + z_\eta \sigma_1 \quad (4-11)$$

$$\xi_2 = \mu_2 + z_\eta \sigma_2 \quad (4-12)$$

现在假设我们按照改进策略，取消两种颜色的成品库存，改为保留未喷涂的在制品库存。忽略供货提前期等变化的影响。企业的服务水平仍然为  $\eta$ 。为了满足该服务水平，所需的在制品库存为  $\xi$ 。此时的在制品库存需要同时应对两种颜色的成品需求，因此，对未喷涂的在制品需求为  $D = D_1 + D_2$ 。

设  $D_1$ 、 $D_2$  的联合分布为  $f(x_1, x_2)$ 。因为  $D_1$ 、 $D_2$  是独立的，所以  $f(x_1, x_2)$  的概率分布为

$$\begin{aligned} f(x_1, x_2) &= f(x_1) \cdot f(x_2) \\ &= \frac{1}{\sqrt{2\pi}\sigma_1} e^{-\frac{(x_1 - \mu_1)^2}{2\sigma_1^2}} \cdot \frac{1}{\sqrt{2\pi}\sigma_2} e^{-\frac{(x_2 - \mu_2)^2}{2\sigma_2^2}} \end{aligned}$$

$$= \frac{1}{2\pi\sigma_1\sigma_2} e^{-\frac{1}{2}\left[\frac{(x_1-\mu_1)^2}{\sigma_1^2} + \frac{(x_2-\mu_2)^2}{\sigma_2^2}\right]} \quad (4-13)$$

设  $D$ 、 $D_2$  的联合分布为  $g(y_1, y_2)$ 。由  $y_1 = x_1 + x_2, y_2 = x_2$  得  $x_1 = y_1 - y_2, x_2 = y_2$ 。因此，雅可比矩阵为

$$J = \begin{bmatrix} \frac{\partial x_1}{\partial y_1} & \frac{\partial x_1}{\partial y_2} \\ \frac{\partial x_2}{\partial y_1} & \frac{\partial x_2}{\partial y_2} \end{bmatrix} = \begin{bmatrix} 1 & -1 \\ 0 & 1 \end{bmatrix}$$

由此可得  $g(y_1, y_2)$  的概率分布为

$$\begin{aligned} g(y_1, y_2) &= f(x_1, x_2) \cdot |J| \\ &= f(y_1 - y_2, y_2) \cdot \begin{vmatrix} 1 & -1 \\ 0 & 1 \end{vmatrix} \\ &= \frac{1}{2\pi\sigma_1\sigma_2} e^{-\frac{1}{2}\left[\frac{(y_1-y_2-\mu_1)^2}{\sigma_1^2} + \frac{(y_2-\mu_2)^2}{\sigma_2^2}\right]} \end{aligned} \quad (4-14)$$

$g(y_1, y_2)$  的边缘分布  $h(y_1)$  即是在制品需求  $D$  的概率分布。对  $y_2$  积分可得

$$h(y_1) = \int_{-\infty}^{+\infty} \frac{1}{2\pi\sigma_1\sigma_2} e^{-\frac{1}{2}\left[\frac{(y_1-y_2-\mu_1)^2}{\sigma_1^2} + \frac{(y_2-\mu_2)^2}{\sigma_2^2}\right]} dy_2 \quad (4-15)$$

令

$$A = \frac{1}{\sqrt{2\pi(\sigma_1^2 + \sigma_2^2)}} e^{-\frac{[y_1 - (\mu_1 + \mu_2)]^2}{2(\sigma_1^2 + \sigma_2^2)}}$$

则公式4-15可变形为

$$\begin{aligned} h(y_1) &= A \int_{-\infty}^{+\infty} \frac{\sqrt{2\pi(\sigma_1^2 + \sigma_2^2)}}{2\pi\sigma_1\sigma_2} e^{-\frac{1}{2}\left[\frac{(y_1-y_2-\mu_1)^2}{\sigma_1^2} + \frac{(y_2-\mu_2)^2}{\sigma_2^2}\right] + \frac{[y_1 - (\mu_1 + \mu_2)]^2}{2(\sigma_1^2 + \sigma_2^2)}} dy_2 \\ &= A \int_{-\infty}^{+\infty} \frac{\sqrt{\sigma_1^2 + \sigma_2^2}}{\sqrt{2\pi}\sigma_1\sigma_2} e^{-\frac{1}{2}\left[\frac{\sqrt{\sigma_1^2 + \sigma_2^2}}{\sigma_1\sigma_2} y_2 - \left(\frac{\sigma_2}{\sigma_1\sqrt{\sigma_1^2 + \sigma_2^2}}(y_1 - \mu_1) + \frac{\sigma_1}{\sigma_2\sqrt{\sigma_1^2 + \sigma_2^2}}\mu_2\right)\right]^2} dy_2 \end{aligned} \quad (4-16)$$

再令

$$t = \frac{\sqrt{\sigma_1^2 + \sigma_2^2}}{\sigma_1\sigma_2} y_2 - \left( \frac{\sigma_2}{\sigma_1\sqrt{\sigma_1^2 + \sigma_2^2}}(y_1 - \mu_1) + \frac{\sigma_1}{\sigma_2\sqrt{\sigma_1^2 + \sigma_2^2}}\mu_2 \right)$$

则公式4-16可变形为

$$h(y_1) = A \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt \quad (4-17)$$

公式4-17的右半部分是标准正态分布的累积概率函数，总累积概率应该为 1，即

$$\int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt = 1$$

因此得到  $h(y_1) = A$ ，即

$$h(y_1) = \frac{1}{\sqrt{2\pi(\sigma_1^2 + \sigma_2^2)}} e^{-\frac{[y_1 - (\mu_1 + \mu_2)]^2}{2(\sigma_1^2 + \sigma_2^2)}} \quad (4-18)$$

由公式4-18可知，在制品的需求  $D$  服从正态分布  $N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$ 。为了满足服务水平  $\eta$ ，所需的在制品库存为

$$\xi = \mu_1 + \mu_2 + z_\eta \sqrt{\sigma_1^2 + \sigma_2^2} \quad (4-19)$$

至此，我们可以利用公式4-11、4-12和4-19将改进前后的库存总量进行比较

$$\begin{aligned} \xi_1 + \xi_2 - \xi &= \mu_1 + z_\eta \sigma_1 + \mu_2 + z_\eta \sigma_2 - \left( \mu_1 + \mu_2 + z_\eta \sqrt{\sigma_1^2 + \sigma_2^2} \right) \\ &= z_\eta \left( \sigma_1 + \sigma_2 - \sqrt{\sigma_1^2 + \sigma_2^2} \right) \\ &> z_\eta \left( \sigma_1 + \sigma_2 - \sqrt{\sigma_1^2 + \sigma_2^2 + 2\sigma_1\sigma_2} \right) \\ &= z_\eta \left( \sigma_1 + \sigma_2 - \sqrt{(\sigma_1 + \sigma_2)^2} \right) \\ &= 0 \end{aligned} \quad (4-20)$$

由公式4-20可知  $\xi_1 + \xi_2 > \xi$ ，即改进后的在制品库存严格小于改进前的成品库存之和。这个结果说明，当两种不同颜色的成品需求服从独立正态分布时，通过保留在制品库存能够降低企业的库存总量。

### 4.3 两种以上的成品需求独立正态分布

现在我们将上述结果进行推广。假设某种汽车保险杠有  $N$  种颜色 ( $N \geq 2$ )，并且服从两两相互独立的正态分布。第  $i$  种颜色的保险杠，需求  $D_i$  服从正态分

布  $N_i(\mu_i, \sigma_i^2)$ ,  $i = 1, 2, \dots, N$ 。企业的服务水平仍为  $\eta$ ，其对应的标准正态分布  $\eta$  分位数仍为  $z_\eta$ 。

为了满足此服务水平，各成品需要保留的库存为

$$\xi_i = \mu_i + z_\eta \sigma_i, \quad i = 1, 2, \dots, N \quad (4-21)$$

现在假设我们按照改进策略，取消所有颜色的成品库存，改为保留未喷涂的在制品库存。为了满足服务水平  $\eta$ ，所需的在制品库存为  $\xi$ 。此时在制品库存的需求为  $D = \sum_{i=1}^N D_i$ 。前面已证任意两个独立的正态分布之和仍然服从正态分布，由于所有的  $D_i$  都是相互独立的，因此它们可以连续加和，其结果仍然服从正态分布。所以我们有  $D \sim N(\sum_{i=1}^N \mu_i, \sum_{i=1}^N \sigma_i^2)$ 。进而可以得到在制品库存为

$$\xi = \sum_{i=1}^N \mu_i + z_\eta \sqrt{\sum_{i=1}^N \sigma_i^2} \quad (4-22)$$

通过公式4-21和4-22比较改进前后的库存变化

$$\begin{aligned} \sum_{i=1}^N \xi_i - \xi &= \sum_{i=1}^N (\mu_i + z_\eta \sigma_i) - \left( \sum_{i=1}^N \mu_i + z_\eta \sqrt{\sum_{i=1}^N \sigma_i^2} \right) \\ &= z_\eta \left( \sum_{i=1}^N \sigma_i - \sqrt{\sum_{i=1}^N \sigma_i^2} \right) \\ &> z_\eta \left( \sum_{i=1}^N \sigma_i - \sqrt{\sum_{i=1}^N \sigma_i^2 + 2 \sum_{i=1}^{N-1} \sum_{j=i+1}^N \sigma_i \sigma_j} \right) \\ &= z_\eta \left( \sum_{i=1}^N \sigma_i - \sqrt{\left( \sum_{i=1}^N \sigma_i \right)^2} \right) \\ &= 0 \end{aligned} \quad (4-23)$$

由公式4-23可知  $\sum_{i=1}^N \xi_i > \xi$ ，即改进后的在制品库存严格小于改进前的成品库存。至此，我们已证明，当成品库存服从两两相互独立的正态分布时，通过保留在制品库存能够降低企业的库存总量。

#### 4.4 一些参数对改进效果的影响



## 第 5 章 改进失效现象

在第4章的证明过程中，我们通过添加项的方式对不等式进行放缩，从而证明了  $\sum_{i=1}^N \xi_i > \xi$  是严格成立的。但是这一推导过程中，我们默认了  $z_\eta > 0$ 。由  $z_\eta$  的定义可知

$$z_\eta > 0 \iff \eta > 0.5$$

也就是说，只有当企业设定的服务水平高于 50% 时，才能够通过保留在制品库存降低企业库存总量。

当  $\eta = 0.5$  时，有  $z_\eta = 0$ ，参考公式4-23的推导过程，可知此时  $\sum_{i=1}^N \xi_i - \xi = 0$ ，即改进前后企业的总库存没有发生变化。

当  $\eta < 0.5$  时，有  $z_\eta < 0$ ，参考公式4-23的推导过程，可以发现此时的不等式放缩方向是相反的，应该得到  $\sum_{i=1}^N \xi_i - \xi < 0$ ，即改进后反而使得企业的总库存增加。

由此可见，当需求服从正态分布时，在某些服务水平下，将成品库存合并前移到在制品库存不仅没有起到改进效果，反而会使总库存增大。我们将这种情况称为改进失效。

本章将结合常用的需求分布函数，详细讨论改进失效现象与需求分布、服务水平等因素的关系。

### 5.1 对称稳定分布的参数与改进失效的关系

前面的部分主要考虑了需求为正态分布时，改进策略对库存总量的影响。之所以研究正态分布，是因为正态分布确实广泛存在于现实生活中。实际生产中的需求分布可能是各种各样的，甚至可能写不出解析式，但只要是方差有限的独立同分布，根据中心极限定理，在进行叠加的时候总会趋近于正态分布。而正态分布自身叠加仍是正态分布。因此，正态分布就像一个吸引域，从各种分布出发最终都落入正态分布。正态分布下的库存改进效果，也能在很大程度上反应实际的改进效果。

第4章中已经提到，安全库存等于总库存与需求期望值之差。回顾公式4-21和4-22可知，改进前和改进后的安全库存分别为  $z_\eta \sum_{i=1}^n \sigma_i$  和  $z_\eta \sqrt{\sum_{i=1}^n \sigma_i^2}$ 。为了使考虑的问题更简单，我们假设所有的需求是独立同分布的  $N(\mu, \sigma^2)$ ，则改进前和改进后的安全库存分别为  $nz_\eta \sigma$  和  $\sqrt{n}z_\eta \sigma$ 。

当  $\eta < 0.5$  的时候，安全库存是负值，这是实际生产中一般不会出现的情况。当  $\eta > 0.5$  时，可以看到， $n$  种成品合并之后的安全库存是单个成品安全库存的  $\sqrt{n}$  倍，而不合并的情况下则是单个成品的  $n$  倍。此时改进是一定有效果的。

(Feller 1971) 提到，对称稳定分布有  $f_{n,\alpha}(x) = f_\alpha(n^{1/\alpha}x)$  的性质。也就是说， $n$  个独立同分布的对称稳定分布的随机变量  $X$  相加，和的概率分布与随机变量  $n^{1/\alpha}X$  是相同的。而正态分布正是对称稳定分布取  $\alpha = 2$  的一个特例。事实上，合并后安全库存是单个成品安全库存的  $n^{1/\alpha}$  倍的性质，同样出现在其他的对称稳定分布中。

稳定分布的概率密度函数为  $f(x; \alpha, \beta, c, \mu)$ ，其中参数  $\alpha \in (0, 2]$ ，影响曲线的陡峭程度； $\beta \in [-1, 1]$ ，影响曲线的偏斜程度； $c \in (0, +\infty)$ ，影响曲线的水平尺度； $\mu \in (-\infty, +\infty)$ ，影响曲线的水平中心。当  $\beta = 0$  时，即为对称稳定分布。图 5.1 展示了  $\beta = 0, c = 1, \mu = 0$  的对称稳定分布在不同的  $\alpha$  取值下的图形，其中  $\alpha = 2$  时即为正态分布。

广义中心极限定理指出， $n$  个独立同分布的随机变量（方差可以无限）相加，最终会收敛到一个稳定分布。因此，和正态分布一样，稳定分布也广泛存在于实际生活中。(Newman 2005) 的一系列实证分析证实了这一点。

假设单个成品需求服从对称稳定分布  $f(x; \alpha, 0, c, \mu)$ ，则它的安全库存服从对称稳定分布  $f(x; \alpha, 0, c, 0)$ 。设此需求分布下，满足服务水平  $\eta$  需要的安全库存为  $s$ ， $n$  个成品需要的总安全库存是  $ns$ 。根据 (Feller) 所述的性质可以推知，改进后需要的安全库存为  $n^{1/\alpha}s$ 。 $\eta > 0.5$  时  $s > 0$ ，故改进是否有效就取决于  $\alpha$  的取值。将改进前后的库存作比较：

$$\frac{ns}{n^{1/\alpha}s} = n^{1-1/\alpha}$$

其中  $\alpha$  的取值范围是  $(0, 2]$ 。

当  $\alpha \in (1, 2]$  时， $n^{1-1/\alpha} > 1$ ，改进前的库存大于改进后的库存，改进有效。

当  $\alpha \in (0, 1]$  时， $n^{1-1/\alpha} \leq 1$ ，改进前的库存不大于改进后的库存，改进无

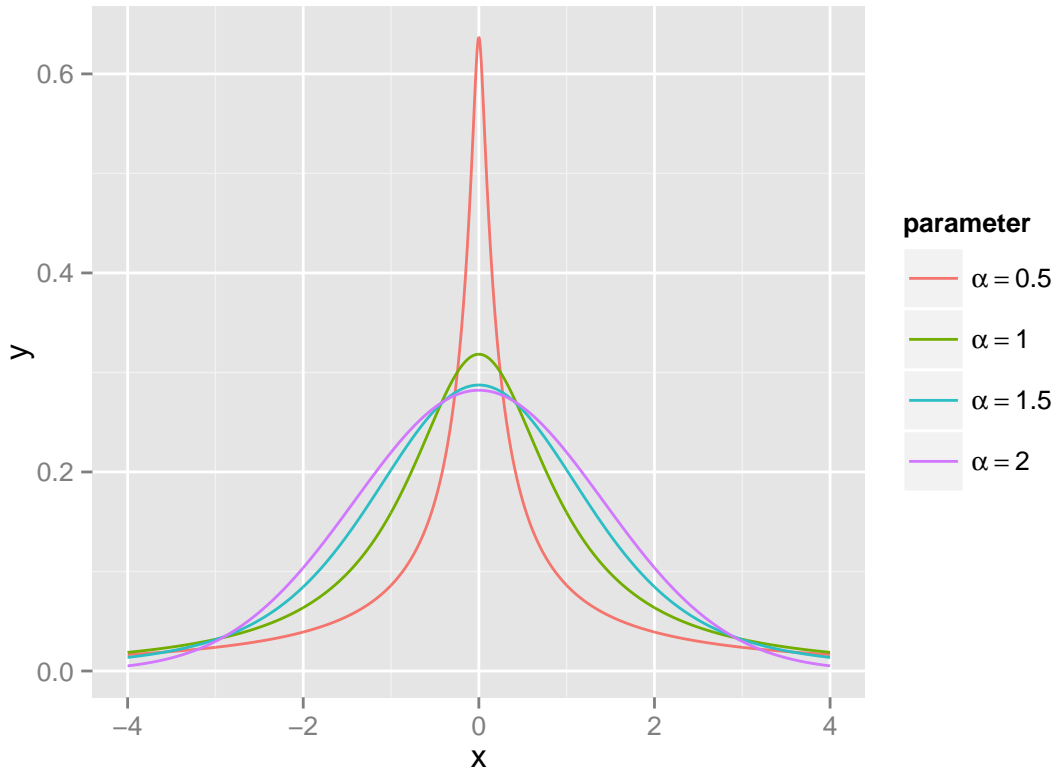


图 5.1 不同参数  $\alpha$  下的对称稳定分布

效。

## 5.2 右偏斜的需求分布一定存在改进失效的可能

对称分布存在一个普遍的缺点——需求可能出现负值，因此我们在实际生产中也常常使用一些其他的分布函数来描述需求的分布。现实的需求不可能为负值，其概率密度函数一般不会有左侧的长尾。因此，需求分布一般都是右偏斜的。(Agrawal 1996) 使用一些实际的需求数据，拟合一些常用的需求分布，包括泊松分布、指数分布、负二项分布等。它们都是右偏斜的分布。

(此处补充一些文献提到右偏斜的需求分布对 substitution 的影响)

假设有  $n$  种颜色的成品，其需求  $D_i (i = 1, 2, \dots, n)$  独立同分布且该分布是右偏斜的，累积分布函数都为  $F$ ，均值为  $\mu$ ，方差为  $\sigma^2$ 。将成品库存合并前移到在制品库存后，在制品的总需求为  $D_n = \sum_{i=1}^n D_i$ ，其累积分布函数为  $F_n$ ，均值为  $n\mu$ ，方差为  $n\sigma^2$ 。

设每种颜色的成品库存量为  $s$ ，则对应的服务水平为  $\eta = F(s)$ 。由  $F_n$  的定

义知

$$F_n(ns) = P(D_n < ns) = P\left(\frac{\sum_{i=1}^n D_i}{n} < s\right) \quad (5-1)$$

根据中心极限定理，当  $n \rightarrow \infty$  时， $\frac{1}{n} \sum_{i=1}^n D_i$  的分布趋近于正态分布  $N(\mu, \sigma^2/n)$ 。设正态分布  $N(\mu, \sigma^2/n)$  的累积分布函数为  $G$ 。由中心极限定理得

$$\lim_{n \rightarrow \infty} P\left(\frac{\sum_{i=1}^n D_i}{n} < s\right) = \lim_{n \rightarrow \infty} \Phi\left(\frac{s - \mu}{\sigma/\sqrt{n}}\right) = \lim_{n \rightarrow \infty} G(s) \quad (5-2)$$

其中  $\Phi$  为标准正态分布的累积分布函数。由公式5-1和5-2可知

$$\lim_{n \rightarrow \infty} [F_n(ns) - G(s)] = 0 \quad (5-3)$$

接下来我们将证明，对任何的右偏斜需求分布，一定存在一个区间，当服务水平在此区间内时，就存在改进失效的可能。已知需求  $D_i$  的分布是右偏的，因此有

$$F(\mu) > 0.5 = \Phi(0) = G(\mu) \quad (5-4)$$

由  $F$  和  $G$  的连续性，至少存在一个  $\mu$  的领域  $[\mu, s_0)$ ，使得这个区间内的所有库存值  $s$  都满足

$$F(s) > G(s), \quad \forall s \in [\mu, s_0)$$

令  $\delta = F(s) - G(s) > 0$ 。在公式5-3中，根据极限的定义，存在一个正数  $N$ ，使得对任意的  $n > N$ ，都有

$$|F_n(ns) - G(s)| < \delta = F(s) - G(s) \quad (5-5)$$

公式5-5显示，对于任意的  $s \in [\mu, s_0)$ ，都存在一个  $N$ ，使得  $n > N$  时恒有  $F_n(ns) < F(s)$ 。

我们知道  $F(s)$  代表需求量小于  $s$  的概率，即库存量  $s$  对应的服务水平。反过来，某个服务水平  $\eta$  也对应着一个库存量，我们将这个对应关系定义为函数  $F^{-1}$ 。函数  $F^{-1}$  的表达式为

$$F^{-1}(\eta) = \inf \{s | F(s) \geq \eta\}$$

其中  $\inf$  表示集合的下确界。同理可定义  $F_n^{-1}$ 。

若企业需要满足的服务水平为  $\eta$ ，则改进前和改进后的库存分别为  $F^{-1}(\eta)$  和  $F_n^{-1}(\eta)$ 。若改进后的在制品库存大于改进前的各颜色成品库存之和，就说明改进失效了。也就是说，改进失效的具体表现是  $F_n^{-1}(\eta) > nF^{-1}(\eta)$ 。下面我们将证明，当  $F^{-1}(\eta) \in [\mu, s_0)$  时，存在改进失效的可能。

现在假设企业的设定的服务水平为  $\eta$ ，满足  $\eta \in [F(\mu), F(s_0))$ ，则改进前各颜色成品的库存为  $s = F^{-1}(\eta) \in [\mu, s_0)$ 。前面已经证明，对于任意的  $s \in [\mu, s_0)$ ，都存在一个  $N$ ，使得  $n > N$  时恒有

$$F_n(ns) < F(s) = \eta \quad (5-6)$$

根据反函数的性质，由于累积分布函数  $F$  和  $F_n$  是单调递增的，所以  $F^{-1}$  和  $F_n^{-1}$  也是单调递增的。公式5-6显示  $F_n(ns) < \eta$ ，结合  $F_n^{-1}$  的单调性可知

$$F_n^{-1}(\eta) > ns \quad (5-7)$$

把  $s = F^{-1}(\eta)$  代入公式5-7有

$$F_n^{-1}(\eta) > nF^{-1}(\eta)$$

前面已经提到，上式即为改进失效的具体表现。

通过本节分析，我们证明了，当各颜色成品需求服从独立同分布的任意一种右偏斜分布时，存在一个大于  $F^{-1}(\mu)$  的服务水平区间，企业的服务水平位于这个区间时，存在改进失效的可能。当服务水平小于  $F^{-1}(\mu)$  时，安全库存是负值，无需过多讨论。

### 5.3 任何服务水平都存在改进失效的可能

在上一节的讨论中，我们已经证明，如果需求服从独立同分布的右偏斜分布，则存在一个大于  $F^{-1}(\mu)$  的服务水平区间，使得这个区间内的服务水平下可能发生改进失效的风险。那么，这个区间是否存在上限？当服务水平远高于  $F^{-1}(\mu)$  的时候，是否存在某些服务水平，使得在这些服务水平下一定不会发生改进失效？我们以最简单的伯努利分布为例，来研究这个问题。

如果单个产品的需求服从伯努利分布：需求量为  $d_1$  的概率为  $\alpha$ ，需求量为  $d_2$  的概率为  $1 - \alpha$ 。为了保证需求是右偏的，我们设定  $d_1 < d_2$ ，且  $\alpha > 0.5$ 。

假设有两种颜色的成品，需求相互独立，且都服从上述的伯努利分布。企业的服务水平为  $\eta$ ，则  $\eta$  和  $\alpha$  的相对大小会对库存产生影响：当  $\eta > \alpha$  时，每种成品需要保留的库存为  $d_2$ ，总库存为  $2d_2$ ；当  $\eta \leq \alpha$  时，每种成品需要保留的库存为  $d_1$ ，总库存为  $2d_1$ 。因此，改进前的服务水平与库存的关系如表5.1所示。

表 5.1 改进前服务水平与库存的关系

服务水平	单个成品库存	总库存
$\eta > \alpha$	$d_2$	$2d_2$
$\eta \leq \alpha$	$d_1$	$2d_1$

现将这两种成品的库存合并保留到在制品库存。首先计算这两种成品需求的联合分布

表 5.2 两种成品需求的联合分布

第一种成品需求	第二种成品需求	概率
$d_1$	$d_1$	$\alpha^2$
$d_1$	$d_2$	$\alpha(1 - \alpha)$
$d_2$	$d_1$	$\alpha(1 - \alpha)$
$d_2$	$d_2$	$(1 - \alpha)^2$

由表5.2可以得到合并后在制品的需求分布，如表5.3所示。

表 5.3 在制品需求分布

需求量	概率	累积概率
$2d_1$	$\alpha^2$	$\alpha^2$
$d_1 + d_2$	$2\alpha(1 - \alpha)$	$2\alpha - \alpha^2$
$2d_2$	$(1 - \alpha)^2$	1

企业的服务水平仍为  $\eta$ ，根据表5.3在制品的需求累积概率分布与  $\eta$  的相对大小，可以得出需要保留的在制品库存：当  $\eta > 2\alpha - \alpha^2$  时，在制品需要保留的库存为  $2d_2$ ；当  $\alpha^2 < \eta \leq 2\alpha - \alpha^2$  时，在制品需要保留的库存为  $d_1 + d_2$ ；当  $\eta \leq \alpha^2$  时，在制品需要保留的库存为  $2d_1$ 。因此，改进后的服务水平与库存的关系如表5.4所示。

表 5.4 改进后服务水平与库存的关系

服务水平	在制品库存
$\eta > 2\alpha - \alpha^2$	$2d_2$
$\alpha^2 < \eta \leq 2\alpha - \alpha^2$	$d_1 + d_2$
$\eta \leq \alpha^2$	$2d_1$

对比表5.1和表5.4可知：当  $\eta \leq \alpha^2$  时，改进前后的库存都为  $2d_1$ ；当  $\alpha^2 < \eta \leq \alpha$  时，改进后的库存  $2d_2$  大于改进前的库存  $2d_1$ ； $\alpha < \eta \leq 2\alpha - \alpha^2$  时，改进后的库存  $d_1 + d_2$  小于改进前的库存  $2d_2$ ；当  $\eta > 2\alpha - \alpha^2$  时，改进前后的库存都为  $2d_2$ 。汇总如表5.5所示。

表 5.5 不同服务水平下改进前后库存对比

服务水平	改进前库存量	改进后库存量
$\eta \leq \alpha^2$	$2d_1$	$2d_1$
$\alpha^2 < \eta \leq \alpha$	$2d_1$	$2d_2$
$\alpha < \eta \leq 2\alpha - \alpha^2$	$2d_2$	$d_1 + d_2$
$\eta > 2\alpha - \alpha^2$	$2d_2$	$2d_2$

从表5.5中可以看出，当  $\alpha^2 < \eta \leq \alpha$  时，一定会发生改进失效。显然，对任意的服务水平  $\eta$ ，一定存在某些  $\alpha$ ，使得  $\alpha^2 < \eta \leq \alpha$ 。因此，任意的服务水平下都能够找到某些右偏斜的伯努利分布，使得改进失效。

上述结论很容易推广到三种以上成品合并库存的情况。假设有  $n$  种颜色的成品，需求服从独立同分布的伯努利分布，需求量为  $d_1$  的概率为  $\alpha$ ，需求量为  $d_2$  的概率为  $1 - \alpha$ ， $d_1 < d_2$ ， $\alpha > 0.5$ 。易知，当服务水平  $\eta \leq \alpha$  时，需要保留的单个成品库存量为  $d_1$ ，总库存量为  $nd_1$ 。将它们的库存合并到在制品库存后，在制品的需求服从二项分布，即

$$P(kd_1 + (n - k)d_2) = C_n^k \alpha^k (1 - \alpha)^{n-k}$$

取  $k = n$  可知  $P(nd_1) = \alpha^n$ ，因此当服务水平  $\eta > \alpha^n$  时，需要保留的在制品库存量是大于  $nd_1$  的。当服务水平满足  $\alpha^n < \eta \leq \alpha^n$  时，一定发生改进失效。

通过以上讨论我们证明了，将  $n$  种颜色的成品库存合并为在制品库存时，对于任意的服务水平  $\eta$ ，都可能在某些右偏的需求分布下发生改进失效。因此，

在做出改进的决策之前，企业必须根据自身的服务水平和实际的需求分布，谨慎地进行模拟和推算，防止发生改进失效。

#### 5.4 泊松分布下发生改进失效的区间图示

在前两节的讨论中，我们从理论上证明了改进失效现象的存在性。但是，对改进失效现象发生的具体范围和如何避免，仍然没有清晰直观的认识。本节中，我们将以实际生产中常用的泊松分布为例，展示改进失效的实际影响和避免改进失效的方法。

为了精确描述库存总量在改进前后的变化，我们首先需要知道改进前后需求分布的变化。设需求  $D$  服从泊松分布  $Po(\lambda)$ ，则其概率为

$$P(D=x) = \frac{e^{-\lambda} \lambda^x}{x!}, \quad x \in \mathbb{N}$$

设企业的服务水平为  $\eta$ ，满足此服务水平所需的库存为  $\xi$ 。则库存  $\xi$  的表达式为

$$\xi = \inf \left\{ \xi \in \mathbb{N} \left| \sum_{x=0}^{\xi} \frac{e^{-\lambda} \lambda^x}{x!} \geq \eta \right. \right\} \quad (5-8)$$

其中符号  $\inf$  表示集合的下确界。

设两种颜色的成品需求服从相互独立的泊松分布  $Po(\lambda_1)$ 、 $Po(\lambda_2)$ 。企业的服务水平为  $\eta$ 。则两种成品需要保持的库存  $\xi_1$ 、 $\xi_2$  分别为

$$\xi_1 = \inf \left\{ \xi_1 \in \mathbb{N} \left| \sum_{x=0}^{\xi_1} \frac{e^{-\lambda_1} \lambda_1^x}{x!} \geq \eta \right. \right\} \quad (5-9)$$

$$\xi_2 = \inf \left\{ \xi_2 \in \mathbb{N} \left| \sum_{x=0}^{\xi_2} \frac{e^{-\lambda_2} \lambda_2^x}{x!} \geq \eta \right. \right\} \quad (5-10)$$

现在假设取消两种颜色的成品库存，改为保留共同的在制品库存。企业的服务水平仍然为  $\eta$ 。为了满足该服务水平，所需的在制品库存为  $\xi$ 。此时的在制品库存需要同时应对两种颜色的成品需求，因此，对未喷涂的在制品需求为  $D = D_1 + D_2$ 。

设  $D_1$ 、 $D_2$  的联合分布为  $f(x_1, x_2)$ 。因为  $D_1$ 、 $D_2$  是独立的，所以  $f(x_1, x_2)$  的概率分布为

$$f(x_1, x_2) = f(x_1)f(x_2)$$



$$\begin{aligned}
&= \frac{e^{-\lambda_1} \lambda_1^{x_1}}{x_1!} \cdot \frac{e^{-\lambda_2} \lambda_2^{x_2}}{x_2!} \\
&= \frac{e^{-(\lambda_1+\lambda_2)} \lambda_1^{x_1} \lambda_2^{x_2}}{x_1! x_2!}
\end{aligned} \tag{5-11}$$

设  $D, D_2$  的联合分布为  $g(y_1, y_2)$ 。由  $y_1 = x_1 + x_2, y_2 = x_2$  得  $x_1 = y_1 - y_2, x_2 = y_2$ 。与正态分布下的证明过程类似，我们求出雅可比矩阵

$$J = \begin{bmatrix} \frac{\partial x_1}{\partial y_1} & \frac{\partial x_1}{\partial y_2} \\ \frac{\partial x_2}{\partial y_1} & \frac{\partial x_2}{\partial y_2} \end{bmatrix} = \begin{bmatrix} 1 & -1 \\ 0 & 1 \end{bmatrix}$$

由此可得  $g(y_1, y_2)$  的概率分布为

$$\begin{aligned}
g(y_1, y_2) &= f(y_1 - y_2, y_2) \cdot |J| \\
&= \frac{e^{-(\lambda_1+\lambda_2)} \lambda_1^{y_1-y_2} \lambda_2^{y_2}}{(y_1 - y_2)! y_2!}
\end{aligned} \tag{5-12}$$

然后求边缘分布  $h(y_1)$ 。与前面不同的是，泊松分布的自变量  $y_1 - y_2 \geq 0$ ，因此有  $y_2 \leq y_1$ 。

$$\begin{aligned}
h(y_1) &= \sum_{y_2=0}^{y_1} g(y_1, y_2) = \sum_{y_2=0}^{y_1} \frac{e^{-(\lambda_1+\lambda_2)} \lambda_1^{y_1-y_2} \lambda_2^{y_2}}{(y_1 - y_2)! y_2!} \\
&= \frac{e^{-(\lambda_1+\lambda_2)} (\lambda_1 + \lambda_2)^{y_1}}{y_1!} \sum_{y_2=0}^{y_1} \frac{y_1!}{y_2! (y_1 - y_2)!} \left( \frac{\lambda_1}{\lambda_1 + \lambda_2} \right)^{y_1-y_2} \left( \frac{\lambda_2}{\lambda_1 + \lambda_2} \right)^{y_2}
\end{aligned} \tag{5-13}$$

公式5-13的右半部分是二项分布  $B(y_1, \frac{\lambda_2}{\lambda_1+\lambda_2})$  的全部累积概率，其和应该为1，即

$$\sum_{y_2=0}^{y_1} \frac{y_1!}{y_2! (y_1 - y_2)!} \left( \frac{\lambda_1}{\lambda_1 + \lambda_2} \right)^{y_1-y_2} \left( \frac{\lambda_2}{\lambda_1 + \lambda_2} \right)^{y_2} = 1$$

因此得到边缘分布为

$$h(y_1) = \frac{e^{-(\lambda_1+\lambda_2)} (\lambda_1 + \lambda_2)^{y_1}}{y_1!} \tag{5-14}$$

由公式5-14可知，在制品的总需求  $D$  服从泊松分布  $Po(\lambda_1 + \lambda_2)$ 。所以应保留的在制品库存为

$$\xi = \inf \left\{ \xi \in \mathbb{N} \left| \sum_{x=0}^{\xi} \frac{e^{-(\lambda_1+\lambda_2)} (\lambda_1 + \lambda_2)^x}{x!} \geq \eta \right. \right\} \tag{5-15}$$

根据公式5-9、5-10和5-15，如果改进前的需求分布服从  $Po(\lambda_1)$  和  $Po(\lambda_2)$ ，则改进后的需求服从  $Po(\lambda_1 + \lambda_2)$ 。据此，我们就可以算出各种情况下改进前后的库存具体数值，从而进行比较。

为了方便作图，我们假设两种成品的需求是独立同分布的，都服从泊松分布  $Po(\lambda)$ ，则改进后的在制品需求服从泊松分布  $Po(2\lambda)$ 。企业的服务水平为  $\eta$ 。  $Po_{\lambda}^{-1}(\eta)$  表示改进前的每种成品的库存，  $Po_{2\lambda}^{-1}(\eta)$  表示改进后的在制品总库存。每一个参数  $\lambda$  和每一个服务水平  $\eta$ ，都对应着一个改进前后的库存差值  $\Delta = 2Po_{\lambda}^{-1}(\eta) - Po_{2\lambda}^{-1}(\eta)$ 。利用计算机程序求出  $\lambda \in (0, 100)$ 、 $\eta \in (0.5, 1)$  的所有对应  $\Delta$  值，并作图如图5.2所示。

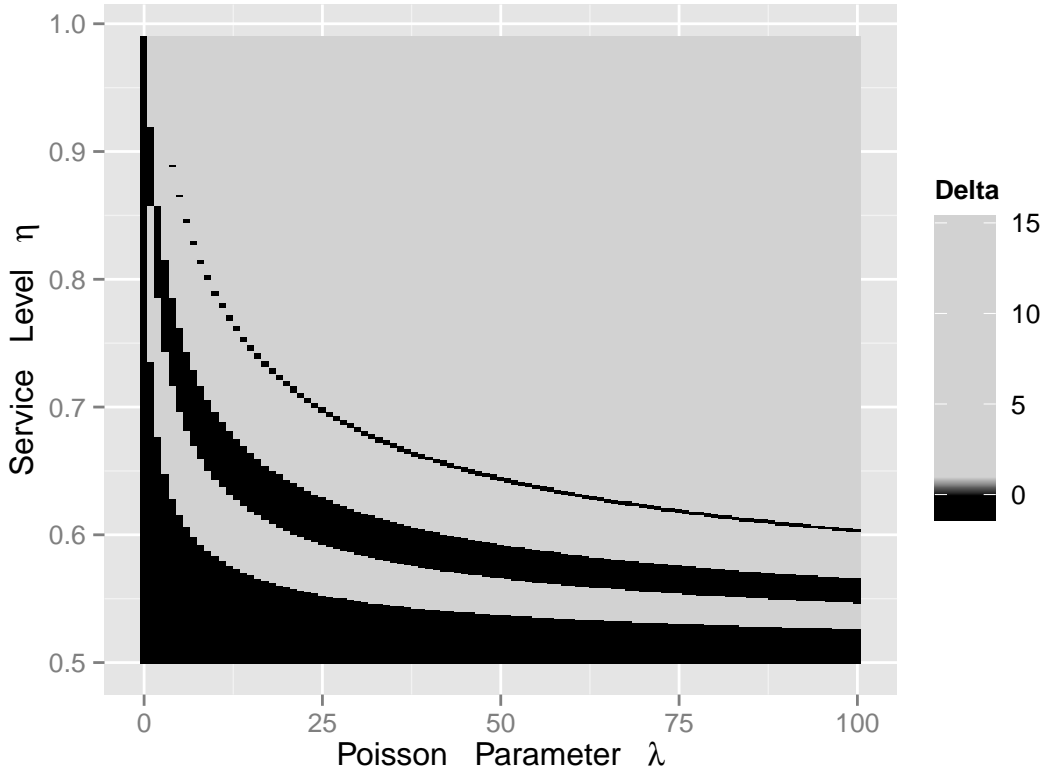


图 5.2 泊松分布下的改进失效区间图示

图5.2中每个点的位置代表一种  $(\lambda, \eta)$  组合，每个点的颜色代表该处的  $\Delta$  值。如果  $\Delta \leq 0$ ，则该点为黑色；如果  $\Delta > 0$ ，则该点为浅灰色。因此，图中的黑色区域代表发生改进失效的区域，灰色区域代表改进有效果的区域。

由于泊松分布的离散性，图5.2中的黑色区域边缘存在锯齿，甚至是断断续

续的。但是从这幅图形中，我们能够得到很多有效的信息。

首先，它验证了前两节证明的一些论断。作为右偏斜分布的泊松分布，在服务水平大于 0.5 的某个区间内，确实出现了改进失效的一片区域；另外，当  $\lambda \rightarrow 0$  时，黑色区域的延伸趋势也验证了上一节的论断，即不存在“不会发生改进失效的服务水平”。

其次，黑色区域在横轴方向的变化趋势告诉我们，需求量越大，发生改进失效的服务水平区间就越小，也就是说越不容易发生改进失效。当  $\lambda$  增大到一定程度时，泊松分布会趋近于正态分布。我们已经了解到，正态分布在  $\eta > 0.5$  的时候是不会出现改进失效的。因此，随着  $\lambda$  的增长，黑色区域会越来越小，最终小到可以忽略的程度。结合上一节中伯努利分布的讨论可以看出，对于右偏斜的需求分布，偏斜程度越大，越可能发生改进失效。

最后，黑色区域在纵轴方向的变化趋势显示，服务水平越高，发生改进失效的区域越少。也就是说，这样的改进方法对于服务水平高的企业是有利的。虽然理论上不存在“不会发生改进失效的服务水平”，但是如果已经知道了需求的参数  $\lambda$  的一个确定的下限，就能够指出一个对应的服务水平的安全线，使得服务水平高于这个安全线时不可能发生改进失效。

#### 5.4.1 第二类服务水平与改进失效

到目前为止，我们在提到服务水平时，都是指第一类服务水平，即得到满足的订单数占总订单数的比例。在实际生产中，也有一些企业采用第二类服务水平作为库存的制定标准。第二类服务水平是指得到满足的需求量占总需求量的比例。如果库存不足以满足需求，就从需求中减去现有的全部库存，剩余的量才作为未满足的需求量。下面我们将证明，如果改进前后都以相同的第二类服务水平作为制定库存的标准，则不可能发生改进失效。

我们知道，发生改进失效的判定标准是“服务水平不变的情况下，改进后的总库存大于改进前的总库存”。事实上，因为服务水平是随库存增加而单调增加的，因此上述判定标准等价于“总库存不变的情况下，改进后的服务水平低于改进前的服务水平”。所以，如果保持改进前后库存总量不变，只需证明改进后的第二类服务水平不可能低于改进前的第二类服务水平即可。

设两种成品的需求  $D_1$ 、 $D_2$  相互独立，概率密度函数分别为  $f_1(x_1)$  和  $f_2(x_2)$ 。企业的服务水平（第二类服务水平）为  $\eta$ 。该服务水平下，两种成品的库存分

别为  $\xi_1$  和  $\xi_2$ 。根据第二类服务水平的定义，得到满足的需求量与总需求量之比的期望值为第二类服务水平。两种成品得到满足的需求量分别为  $\min(x_1, \xi_1)$  和  $\min(x_2, \xi_2)$ ，因此有

$$\eta = \int_{-\infty}^{+\infty} \frac{\min(x_1, \xi_1)}{x_1} f(x_1) dx_1 = \int_{-\infty}^{+\infty} \frac{\min(x_2, \xi_2)}{x_2} f(x_2) dx_2$$

考虑两种成品需求的联合分布  $f(x_1, x_2)$ 。由于需求是相互独立的，因此有  $f(x_1, x_2) = f(x_1)f(x_2)$ 。将两种成品看做一个整体，则得到满足的需求量为  $\min(x_1, \xi_1) + \min(x_2, \xi_2)$ ，总需求量为  $x_1 + x_2$ 。因为它们各自的服务水平都是  $\eta$ ，所以综合服务水平也是  $\eta$ 。因此有

$$\eta = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \frac{\min(x_1, \xi_1) + \min(x_2, \xi_2)}{x_1 + x_2} f(x_1)f(x_2) dx_1 dx_2 \quad (5-16)$$

现将两种成品库存合并到在制品库存，保持总量不变，即合并后的库存为  $\xi = \xi_1 + \xi_2$ 。合并后的服务水平为  $\eta'$ 。合并后的总需求量为  $x_1 + x_2$ ，其中得到满足的需求量为  $\min(x_1 + x_2, \xi_1 + \xi_2)$ ，因此合并后的服务水平为

$$\eta' = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \frac{\min(x_1 + x_2, \xi_1 + \xi_2)}{x_1 + x_2} f(x_1)f(x_2) dx_1 dx_2 \quad (5-17)$$

为了比较  $\eta$  和  $\eta'$  的大小，我们先证明  $\min(x_1, \xi_1) + \min(x_2, \xi_2) \leq \min(x_1 + x_2, \xi_1 + \xi_2)$ 。设  $m_1 = \min(x_1, \xi_1)$ ， $m_2 = \min(x_2, \xi_2)$ ，则有  $m_1 \leq x_1$ ， $m_1 \leq \xi_1$ ， $m_2 \leq x_2$ ， $m_2 \leq \xi_2$ 。因此  $x_1 + x_2 \geq m_1 + m_2$ ， $\xi_1 + \xi_2 \geq m_1 + m_2$ 。由此可证

$$\begin{aligned} \min(x_1 + x_2, \xi_1 + \xi_2) &\geq \min(m_1 + m_2, m_1 + m_2) \\ &= m_1 + m_2 \\ &= \min(x_1, \xi_1) + \min(x_2, \xi_2) \end{aligned} \quad (5-18)$$

将不等式5-18代入公式5-16和5-17中，就能比较  $\eta$  和  $\eta'$  的大小，具体过程如下：

$$\begin{aligned} \eta' &= \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \frac{\min(x_1 + x_2, \xi_1 + \xi_2)}{x_1 + x_2} f(x_1)f(x_2) dx_1 dx_2 \\ &\geq \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \frac{\min(x_1, \xi_1) + \min(x_2, \xi_2)}{x_1 + x_2} f(x_1)f(x_2) dx_1 dx_2 \\ &= \eta \end{aligned} \quad (5-19)$$

由公式5-19可知，总库存保持不变的情况下，改进后的第二类服务水平不低于改进前的第二类服务水平。因此，保持第二类服务水平不变的情况下，改进后的库存不会大于改进前的库存。

至此，我们证明了本小节开头提出的观点，即“如果改进前后都以相同的第二类服务水平作为制定库存的标准，则不可能发生改进失效”。因此，对于使用第二类服务水平作为库存制定标准的企业来说，这样的改进是值得考虑的。

## 5.5 避免改进失效

通过本章的探讨，我们初步了解了改进失效现象的影响因素和具体表现。作为对本章的一个总结，我们在此归纳各种参数对改进失效的影响，作为企业改进前需要考虑的因素：

1. 如果需求服从正态分布，只需保证企业的服务水平大于 0.5，即可避免改进失效；
2. 如果需求服从对称稳定分布，需要保证决定分布陡峭程度的参数  $\alpha$  的值在 (1,2] 区间内，才能避免改进失效；
3. 如果需求服从右偏斜的分布，需要关注需求分布的偏斜程度和企业的服务水平。分布的偏斜程度越大，越可能发生改进失效；企业的服务水平越低，越可能发生改进失效。

总而言之，企业做出改进决策之前，首先需要考虑的就是需求分布的特点，其次需要检查自身的服务水平在这个需求分布下有无改进失效的可能。

## 第6章 需求相关性对改进效果的影响

本章将讨论成品需求之间的相关性对改进方案的效果的影响。本章中仍然不考虑提前期等其他因素的影响。

### 6.1 需求相关性的现实意义

此处待补充实际例子，证明同一型号不同颜色的配件需求之间可能存在相关性。

车型的销售数据容易获得，但是各颜色的销售数据找不到。考虑用“高配/低配”或者其他存在后期分化的产品的相关性来代替。

### 6.2 需求服从二维正态分布的情况

为了构造相关系数为  $\rho$  的两个正态分布的需求变量，我们假设两种颜色的成品需求  $(D_1, D_2)$  服从二维正态分布  $N(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \rho)$ ，则它们的联合分布为

$$f(x_1, x_2) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} e^{-\frac{1}{2(1-\rho^2)} \left[ \frac{(x_1-\mu_1)^2}{\sigma_1^2} + \frac{(x_2-\mu_2)^2}{\sigma_2^2} - \frac{2\rho(x_1-\mu_1)(x_2-\mu_2)}{\sigma_1\sigma_2} \right]} \quad (6-1)$$

对  $x_2$  积分，可以得到  $D_1$  的边缘分布

$$f_1(x_1) = \int_{-\infty}^{+\infty} \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} e^{-\frac{1}{2(1-\rho^2)} \left[ \frac{(x_1-\mu_1)^2}{\sigma_1^2} + \frac{(x_2-\mu_2)^2}{\sigma_2^2} - \frac{2\rho(x_1-\mu_1)(x_2-\mu_2)}{\sigma_1\sigma_2} \right]} dx_2 \quad (6-2)$$

令

$$A = \frac{1}{\sqrt{2\pi}\sigma_1} e^{-\frac{(x_1-\mu_1)^2}{2\sigma_1^2}}$$

则公式6-2可变形为

$$f_1(x_1) = A \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi(1-\rho^2)}\sigma_2} e^{-\frac{1}{2} \left[ \frac{\sigma_1(x_2-\mu_2) - \rho\sigma_2(x_1-\mu_1)}{\sqrt{1-\rho^2}\sigma_1\sigma_2} \right]^2} dx_2 \quad (6-3)$$

再令

$$t = \frac{\sigma_1(x_2 - \mu_2) - \rho\sigma_2(x_1 - \mu_1)}{\sqrt{1-\rho^2}\sigma_1\sigma_2}$$

则公式6-3可变形为

$$f_1(x_1) = A \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt \quad (6-4)$$

公式6-4的有半部分是标准正态分布的累积概率，总累积概率应为 1，即

$$\int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt = 1$$

因此得到  $f_1(x_1) = A$ ，即

$$f_1(x_1) = \frac{1}{\sqrt{2\pi}\sigma_1} e^{-\frac{(x_1-\mu_1)^2}{2\sigma_1^2}} \quad (6-5)$$

同理有

$$f_2(x_2) = \frac{1}{\sqrt{2\pi}\sigma_2} e^{-\frac{(x_2-\mu_2)^2}{2\sigma_2^2}} \quad (6-6)$$

公式6-5和6-6表明，当需求  $D_1$ 、 $D_2$  服从二维正态分布  $N(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \rho)$  时，它们分别服从一维正态分布  $N(\mu_1, \sigma_1^2)$  和  $N(\mu_2, \sigma_2^2)$ 。下面证明，正态分布的两个变量  $D_1$ 、 $D_2$  之间的相关系数为  $\rho$ 。

由相关系数的定义可知

$$\begin{aligned} \text{Corr}(D_1, D_2) &= \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \frac{(x_1 - \mu_1)(x_2 - \mu_2)}{\sigma_1 \sigma_2} f(x_1, x_2) dx_1 dx_2 \\ &= \int_{-\infty}^{+\infty} \frac{x_2 - \mu_2}{\sigma_2} \int_{-\infty}^{+\infty} \frac{x_1 - \mu_1}{\sigma_1} f(x_1, x_2) dx_1 dx_2 \end{aligned} \quad (6-7)$$

与公式6-3类似，令

$$t = \frac{\sigma_2(x_1 - \mu_1) - \rho\sigma_1(x_2 - \mu_2)}{\sqrt{1 - \rho^2}\sigma_1\sigma_2}$$

则公式6-7变形为

$$\text{Corr}(D_1, D_2) = \int_{-\infty}^{+\infty} \frac{x_2 - \mu_2}{\sigma_2} \int_{-\infty}^{+\infty} \frac{e^{-\frac{(x_2 - \mu_2)^2}{2\sigma_2^2}}}{2\pi\sigma_2} \left( t \sqrt{1 - \rho^2} + \frac{\rho(x_2 - \mu_2)}{\sigma_2} \right) e^{-\frac{t^2}{2}} dt dx_2 \quad (6-8)$$

易知

$$\int_{-\infty}^{+\infty} t e^{-\frac{t^2}{2}} dt = 0$$

且

$$\int_{-\infty}^{+\infty} e^{-\frac{t^2}{2}} = \sqrt{2\pi}$$

因此公式6-8可得

$$\begin{aligned} \text{Corr}(D_1, D_2) &= \int_{-\infty}^{+\infty} \frac{x_2 - \mu_2}{\sigma_2} \left[ \frac{\rho(x_2 - \mu_2)}{\sqrt{2\pi}\sigma_2^2} \right] dx_2 \\ &= \frac{\rho}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} \left( \frac{x_2 - \mu_2}{\sigma_2} \right)^2 e^{-\frac{1}{2} \left( \frac{x_2 - \mu_2}{\sigma_2} \right)^2} d\left( \frac{x_2 - \mu_2}{\sigma_2} \right) \end{aligned} \quad (6-9)$$

令

$$w = \frac{x_2 - \mu_2}{\sigma_2}$$

则公式6-9可变形为

$$\begin{aligned} \text{Corr}(D_1, D_2) &= \frac{\rho}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} w^2 e^{-\frac{w^2}{2}} dw \\ &= -\frac{\rho}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} w d\left( e^{-\frac{w^2}{2}} \right) \\ &= -\frac{\rho}{\sqrt{2\pi}} \left[ w e^{-\frac{w^2}{2}} \Big|_{-\infty}^{+\infty} - \int_{-\infty}^{+\infty} e^{-\frac{w^2}{2}} dw \right] \\ &= -\frac{\rho}{\sqrt{2\pi}} (0 - \sqrt{2\pi}) \\ &= \rho \end{aligned} \quad (6-10)$$

(证两个边缘分布都是正态，且相关系数为 rho)

设企业的服务水平为  $\eta = 1 - \alpha$ ，对应的标准正态分布分位数为  $z_\alpha$ 。则两种颜色的成品需要保留的库存  $\xi_1$ 、 $\xi_2$  分别为

$$\xi_1 = \mu_1 + z_\alpha \sigma_1 \quad (6-11)$$

$$\xi_2 = \mu_2 + z_\alpha \sigma_2 \quad (6-12)$$

现在假设我们按照改进策略，取消两种颜色的成品库存，改为保留未喷涂的在制品库存。忽略供货提前期等变化的影响。企业的服务水平仍然为  $\eta = 1 - \alpha$ 。为了满足该服务水平，所需的在制品库存为  $\xi$ 。此时的在制品库存需要同时应对两种颜色的成品需求，因此，对未喷涂的在制品需求为  $D = D_1 + D_2$ 。



(证二维正态分布下，边缘分布之和仍然服从正态分布。)

由以上证明可知，在制品需求  $D$  服从正态分布  $N(\mu, \sigma^2)$ ，且  $\mu = \mu_1 + \mu_2$ ， $\sigma^2 = \sigma_1^2 + \sigma_2^2 + 2\rho\sigma_1\sigma_2$ 。因此，需要保留的在制品库存  $\xi$  为

$$\xi = \mu + z_\alpha \sigma = \mu_1 + \mu_2 + z_\alpha \sqrt{\sigma_1^2 + \sigma_2^2 + 2\rho\sigma_1\sigma_2} \quad (6-13)$$

根据公式6-11、6-12和6-13，可以比较改进前后的库存变化。

$$\begin{aligned} \xi_1 + \xi_2 - \xi &= \mu_1 + z_\alpha \sigma_1 + \mu_2 + z_\alpha \sigma_2 - \left( \mu_1 + \mu_2 + z_\alpha \sqrt{\sigma_1^2 + \sigma_2^2 + 2\rho\sigma_1\sigma_2} \right) \\ &= z_\alpha \left( \sigma_1 + \sigma_2 - \sqrt{\sigma_1^2 + \sigma_2^2 + 2\rho\sigma_1\sigma_2} \right) \end{aligned} \quad (6-14)$$

我们知道  $\sigma_1 > 0$ ， $\sigma_2 > 0$  以及  $|\rho| \leq 1$ ，因此有

$$\sqrt{\sigma_1^2 + \sigma_2^2 + 2\rho\sigma_1\sigma_2} \leq \sqrt{\sigma_1^2 + \sigma_2^2 + 2\sigma_1\sigma_2} = \sigma_1 + \sigma_2 \quad (6-15)$$

将公式6-15代入公式6-14，得

$$\xi_1 + \xi_2 - \xi \geq z_\alpha [\sigma_1 + \sigma_2 - (\sigma_1 + \sigma_2)] = 0 \quad (6-16)$$

公式6-16的结果表明，对于需求有相关性、服从二维正态分布的两种成品，如果取消成品库存，改为保留共同的在制品库存，在相同的服务水平下，改进后的在制品库存不会比改进前的两种成品库存之和更大。

### 6.3 需求服从多维正态分布的情况

现在我们将上述结果进行推广。假设某成品有  $N$  种颜色 ( $N \geq 2$ )，它们的需求服从多维正态分布，且两两之间的相关系数为  $\rho_{ij}$ 。则每种颜色的库存为

$$\xi_i = \mu_i + z_\alpha \sigma_i, \quad i = 1, 2, \dots, N \quad (6-17)$$

假设取消全部成品库存，改为保留共同的在制品库存，则在制品库存  $\xi$  应为

$$\xi = \sum_{i=1}^N \mu_i + z_\alpha \sqrt{\sum_{i=1}^N \sigma_i^2 + 2 \sum_{i=1}^{N-1} \sum_{j=i+1}^N \rho_{ij} \sigma_i \sigma_j} \quad (6-18)$$

根据公式6-17和6-18作比较

$$\begin{aligned}
\sum_{i=1}^N \xi_i - \xi &= \sum_{i=1}^N (\mu_i + z_\alpha \sigma_i) - \left( \sum_{i=1}^N \mu_i + z_\alpha \sqrt{\sum_{i=0}^N \sigma_i^2 + 2 \sum_{i=1}^{N-1} \sum_{j=i+1}^N \rho_{ij} \sigma_i \sigma_j} \right) \\
&= z_\alpha \left( \sum_{i=1}^N \sigma_i - \sqrt{\sum_{i=0}^N \sigma_i^2 + 2 \sum_{i=1}^{N-1} \sum_{j=i+1}^N \rho_{ij} \sigma_i \sigma_j} \right) \\
&\geq z_\alpha \left( \sum_{i=1}^N \sigma_i - \sqrt{\sum_{i=0}^N \sigma_i^2 + 2 \sum_{i=1}^{N-1} \sum_{j=i+1}^N \sigma_i \sigma_j} \right) \\
&= 0
\end{aligned} \tag{6-19}$$

公式6-19表明，对于两种以上颜色、需求具有相关性的成品，改进后的在制品库存不会比改进前的各颜色成品库存之和更大。

## 6.4 极端情况的讨论

公式6-18中， $\rho_{ij}$  的取值范围是  $[-1, 1]$ 。容易发现，在其他参数不变的情况下，当所有的  $\rho_{ij}$  都同时满足  $\rho_{ij} = -1$  时，在制品库存  $\xi$  取得最小值；当所有的  $\rho_{ij}$  都同时满足  $\rho_{ij} = 1$  时，在制品库存  $\xi$  取得最大值。

$\rho_{ij} = -1$  时，各种颜色的成品需求都是完全负相关的。一种颜色的成品需求增加，就会抑制其他颜色的需求。这种情况主要发生在需求总量稳定、需求偏好不确定的条件下。在制品库存  $\xi$  取得最小值，说明此时采取保留在制品库存而不是成品库存，优化的收益最大。这与我们生活经验是相符的。比如全校毕业生征订毕业衫，总需求量几乎是确定的，每个人一般只订一件，且每个人偏好的颜色不同。此时，提前给每种颜色的毕业衫印制足量的库存显然是不妥的，一般会采用先收集订单再印制的方式。

$\rho_{ij} = 1$  时，各种颜色成品需求都是完全正相关的。一种颜色的成品需求增加，会导致其他颜色的需求同步增加。在制品库存  $\xi$  取得最大值，说明此事采取保留在制品库存的策略，收益最小。事实上， $\rho_{ij} = 1$  时，公式6-19中的等号是成立的，即  $\sum_{i=1}^N \xi_i - \xi = 0$ 。换句话说，此时采取保留在制品库存的改进措施毫无效果。

其他情况下，不论各成品的需求是正相关还是负相关，只要不是完全正相关（ $\rho_{ij} = 1$ ），都能从保留在制品库存的策略中获得收益。并且，与需求独立的

情况相比，负相关时的收益更大，正相关时的收益更小。这是企业在做出改进决策之前需要考虑的重要因素。

## 第 7 章 Backorder 对改进效果的影响

作为供应商，如果不能按时交货，有时需要将订单记下，尽快补货，并且承受一定的惩罚。这种订单称为 backorder。本章中，我们将讨论 backorder 对我们的方案改进效果的影响。

### 7.1 考虑 Backorder 成本的任意更新过程

以汽车保险杠的生产线为例，为了使讨论简化，我们假设注塑环节需要的时间远大于喷涂环节，从而忽略喷涂环节的生产时间。也就是说，改进前后相比，单件产品的生产时间是不变的。

假设有  $N$  种颜色的成品，每种产品的需求都是独立到来的。每当需求到来时，如果该产品有库存，则立即满足该需求，且注塑机产生一个生产订单；如果没有库存，则产生一个 backorder，并且在注塑机处产生一个生产订单。所有的生产订单采用先到先服务的策略。

定义：

$s_i$ ：每种颜色初始库存

$h$ ：库存成本率

$b$ ：backorder 成本率

$D_i(t)$ ：直到  $t$  时刻， $i$  产品的总需求

$C_i(t)$ ：直到  $t$  时刻， $i$  产品的总产量

$I_i(t)$ ： $t$  时刻的库存， $I_i(t) = s_i + C_i(t) - \min\{s_i + C_i(t), D_i(t)\}$

$B_i(t)$ ： $t$  时刻的 backorder， $B_i(t) = D_i(t) - \min\{s_i + C_i(t), D_i(t)\}$

$\phi_{i,N}(t)$ ： $t$  时刻  $i$  产品的成本率， $\phi_{i,N} = hI_i(t) + bB_i(t)$

$\phi_N(t)$ ： $t$  时刻所有产品的成本率， $\phi_N(t) = \sum_{i=1}^N \phi_{i,N}(t)$

$\Phi_N(t)$ ：从初始时刻到  $t$  时刻的总成本， $\Phi_N(t) = \int_0^t \phi_N(u) du$

以上所有变量去掉角标，即对应改进后的变量。

很容易得出以下结论：

$$\phi(t) \leq \phi_N(t), \quad \forall t \geq 0 \quad (7-1)$$

$$\Phi(t) \leq \Phi_N(t), \quad \forall t \geq 0 \quad (7-2)$$

设  $Q_i(t) = D_i(t) - C_i(t)$ ，且所有的到达都是独立同分布的更新过程，到达速率严格小于生产速率。根据 Wolff(1989)， $Q_i(t)$  是存在极限分布的。设此极限分布为  $Q_i$ 。那么极限情况下的总成本率为

$$\psi_N((s_1, s_2, \dots, s_N)) = \sum_{i=1}^N bE(Q_i) + hs_i - (h+b)E(\min\{s_i, Q_i\}) \quad (7-3)$$

公式7-3是初始库存  $(s_1, s_2, \dots, s_N)$  的函数，因此存在一组  $(s_1^*, s_2^*, \dots, s_N^*)$  使  $\psi_N((s_1, s_2, \dots, s_N))$  取得最优值。

同理，对于改进后的系统，极限状态下的成本率为  $\psi(s)$ ，使其取得最优值的  $s$  记为  $s^*$ 。从而有

$$\psi(s^*) \leq \psi\left(\sum_{i=1}^N s_i^*\right) \leq \psi_N((s_1^*, s_2^*, \dots, s_N^*)) \quad (7-4)$$

因此，即使在分别采取最优初始库存的情况下，改进后的成本率仍然比改进前要低。

## 7.2 考虑 Backorder 成本的泊松过程

假设需求和生产都是泊松过程。

定义：

$\lambda_i$ :  $i$  产品的到达速率，且总速率为  $\lambda = \sum_{i=1}^N \lambda_i$

$p_i$ :  $i$  产品的需求占比， $p_i = \frac{\lambda_i}{\lambda}$

$\mu$ : 生产速率

$\rho$ : 需求产能比， $\rho = \frac{\lambda}{\mu}$ ，且  $\rho < 1$

则  $Q_i$  极限分布是几何分布

$$P(Q_i = n) = (1 - r_i)r_i^n$$

其中

$$r_i = \frac{p_i \rho}{1 - \rho + p_i \rho}$$

由此算出库存和 backorder 的极限分布期望

$$E(I_i(s_i)) = s_i - \frac{r_i(1 - r_i^{s_i})}{1 - r_i}$$

$$E(B_i(s_i)) = \frac{r_i^{s_i+1}}{1-r_i}$$

然后得到极限成本率

$$\psi_N((s_1, s_2, \dots, s_N)) = E \left( \sum_{i=1}^N [hI_i(s_i) + bB_i(s_i)] \right) \quad (7-5)$$

公式7-5中, 对于每一个  $hI_i(s_i) + bB_i(s_i)$ , 可以采用  $[hI_i(s_i+1) + bB_i(s_i+1)] - [hI_i(s_i) + bB_i(s_i)]$  的方式找到最优的  $s_i^*$ 。根据 Veatch 和 Wein(1996), 这个最优值为

$$s_i^* = \left\lfloor \frac{\ln \gamma}{\ln r_i} \right\rfloor$$

其中  $\gamma = \frac{h}{h+b}$ 。

作为一种无伤大雅的近似, 我们不对  $s_i^*$  向下取整, 而是直接代入  $\psi_N(\cdot)$ , 得到的是

$$\psi_N((s_1^*, s_2^*, \dots, s_N^*)) = h \sum_{i=1}^N \frac{\ln \gamma}{\ln \frac{p_i \rho}{1-\rho+p_i \rho}} \quad (7-6)$$

同时可以推出改进后的极限成本率为

$$\psi(s^*) = \frac{h \ln \gamma}{\ln \rho} \quad (7-7)$$

由公式7-6和7-7可以比较改进前后的库存差, 即  $\Delta\psi = \psi_N((s_1^*, s_2^*, \dots, s_N^*)) - \psi(s^*)$ 。数值实验如图7.1所示。图中  $h = 1$ ,  $b = 5$ ,  $p_i = \frac{1}{N}$ 。

图7.1显示,  $\rho$  越接近 1, 改进前后的库存之差越大。我们求出  $\rho$  接近 1 时,  $\Delta\psi$  的极限, 可以作为改进效果的上限。设所有需求的到达速率相同, 即  $p_i = \frac{1}{N}$ 。

$$\lim_{\rho \rightarrow 1} \Delta\psi = h \ln \gamma \lim_{\rho \rightarrow 1} \sum_{i=1}^N \left[ \frac{1}{\ln \frac{p_i \rho}{1-\rho+p_i \rho}} - \frac{p_i}{\ln \rho} \right] \quad (7-8)$$

对公式7-8使用洛必达法则, 可以得到

$$\lim_{\rho \rightarrow 1} \Delta\psi = \frac{1}{2} h (N-1) \ln \frac{h+b}{h} \quad (7-9)$$

公式7-9可以作为估算改进效果上限的一种手段。

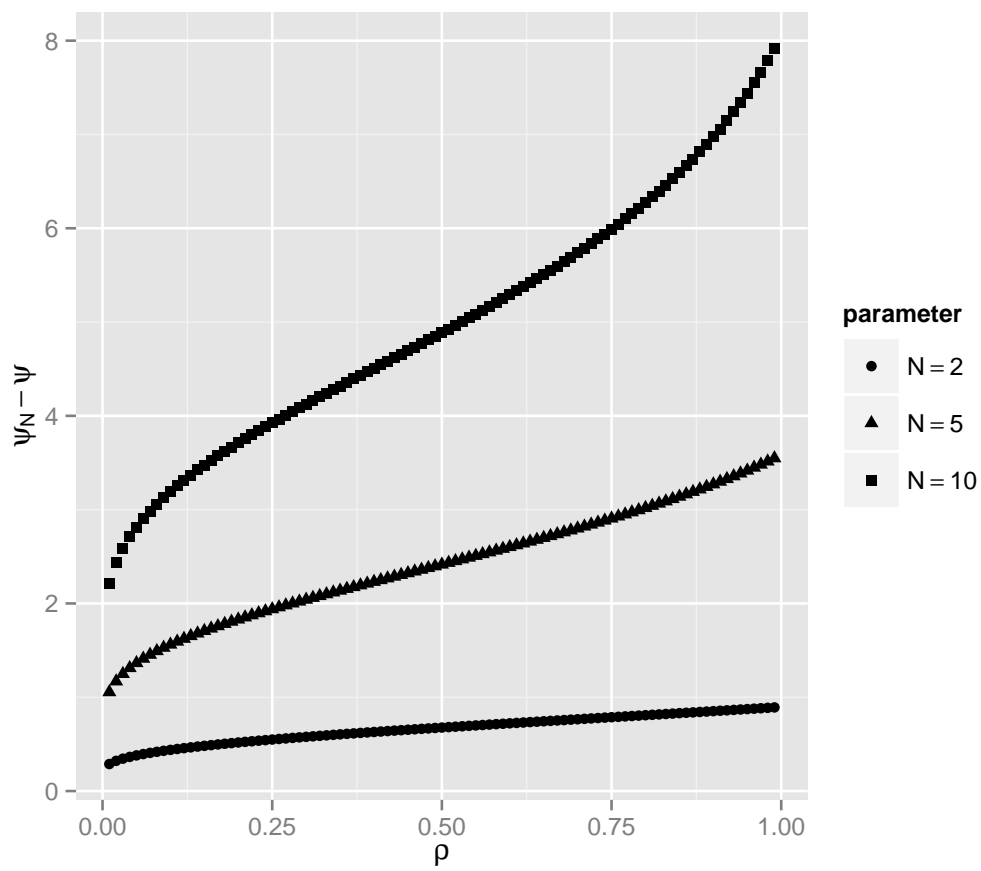


图 7.1 改进前后极限成本率之差随  $\rho$  的变化

### 7.3 服务水平约束的影响

见文献 «On the benefits of ...»

## 第 8 章 提前期对改进效果的影响

### 8.1 需求正态分布的情况

仍然以汽车保险杠生产为例。假设生产一个注塑件耗时  $\tau_1$ ，喷涂一个注塑件耗时  $\tau_2$ 。客户允许的备货时间是  $\tau$ 。

$\tau_1$ 、 $\tau_2$ 、 $\tau$  三者的大小关系发生变化的时候，方案的改进效果也会发生变化。  
(此处需要分类讨论很多条)

### 8.2 需求泊松到达的情况

设需求到达的速率为  $\lambda_1$ 、 $\lambda_2$ ，注塑速率  $\mu_1$ ，喷涂速率  $\mu_2$ 。建模成马尔科夫(生灭问题)，可以求稳态概率，State 0 的概率即为缺货概率。

需要证明两个独立的泊松流合并后仍然是泊松流，且到达速率为  $\lambda_1 + \lambda_2$ 。

还需证明两个输入输出均为泊松过程的服务台，可以合并视为一个泊松过程的服务台(参考 Jackson 网络相关章节)。

### 8.3 (s,S) 库存策略下的情况

由于换模等因素的影响，企业的实际生产中不可能随时对某种产品补货。往往采用的是 (s,S) 策略，即库存降到 s 后开始生产该产品，一直补货到库存为 S 为止。这种策略下也能建模成马尔科夫模型。



## 插图索引

图 2.1	单产品多阶段生产 .....	5
图 2.2	拉动式生产系统流程图 .....	6
图 2.3	推动式生产系统流程图 .....	7
图 5.1	不同参数 $\alpha$ 下的对称稳定分布 .....	21
图 5.2	泊松分布下的改进失效区间图示 .....	28
图 7.1	改进前后极限成本率之差随 $\rho$ 的变化 .....	41

## 表格索引

表 5.1	改进前服务水平与库存的关系 .....	24
表 5.2	两种成品需求的联合分布 .....	24
表 5.3	在制品需求分布 .....	24
表 5.4	改进后服务水平与库存的关系 .....	25
表 5.5	不同服务水平下改进前后库存对比 .....	25

## 公式索引

公式 2-1 .....	3
公式 2-2 .....	3
公式 2-3 .....	3
公式 4-1 .....	13
公式 4-2 .....	13
公式 4-3 .....	13
公式 4-4 .....	13
公式 4-5 .....	14
公式 4-6 .....	14
公式 4-7 .....	14
公式 4-8 .....	14
公式 4-9 .....	14
公式 4-10 .....	15
公式 4-11 .....	15
公式 4-12 .....	15
公式 4-13 .....	16
公式 4-14 .....	16
公式 4-15 .....	16
公式 4-16 .....	16
公式 4-17 .....	17
公式 4-18 .....	17
公式 4-19 .....	17
公式 4-20 .....	17
公式 4-21 .....	18
公式 4-22 .....	18
公式 4-23 .....	18
公式 5-1 .....	22
公式 5-2 .....	22

公式 5-3 .....	22
公式 5-4 .....	22
公式 5-5 .....	22
公式 5-6 .....	23
公式 5-7 .....	23
公式 5-8 .....	26
公式 5-9 .....	26
公式 5-10.....	26
公式 5-11.....	27
公式 5-12.....	27
公式 5-13.....	27
公式 5-14.....	27
公式 5-15.....	27
公式 5-16.....	30
公式 5-17.....	30
公式 5-18.....	30
公式 5-19.....	30
公式 6-1 .....	32
公式 6-2 .....	32
公式 6-3 .....	32
公式 6-4 .....	33
公式 6-5 .....	33
公式 6-6 .....	33
公式 6-7 .....	33
公式 6-8 .....	33
公式 6-9 .....	34
公式 6-10.....	34
公式 6-11.....	34
公式 6-12.....	34
公式 6-13.....	35
公式 6-14.....	35

公式 6-15.....	35
公式 6-16.....	35
公式 6-17.....	35
公式 6-18.....	35
公式 6-19.....	36
公式 7-1 .....	38
公式 7-2 .....	38
公式 7-3 .....	39
公式 7-4 .....	39
公式 7-5 .....	40
公式 7-6 .....	40
公式 7-7 .....	40
公式 7-8 .....	40
公式 7-9 .....	40

## 参考文献

- [1] Fetter R B, Dalleck W C. Decision models for inventory management. RD Irwin, 1961
- [2] Lambrecht M R, Muckstadt J A, Luyten R. Protective stocks in multi-stage production systems. THE INTERNATIONAL JOURNAL OF PRODUCTION RESEARCH, 1984, 22(6):1001–1025
- [3] Clark A J, Scarf H. Approximate solutions to a simple multi-echelon inventory problem. Studies in Applied Probability and Management Science, 1962. 88–110
- [4] Baker K R, Magazine M J, Nuttle H L. The effect of commonality on safety stock in a simple inventory model. Management Science, 1986, 32(8):982–988
- [5] Eppen G D, Martin R K. Determining safety stock in the presence of stochastic lead time and demand. Management Science, 1988, 34(11):1380–1390
- [6] Molinder A. Joint optimization of lot-sizes, safety stocks and safety lead times in an MRP system. International Journal of Production Research, 1997, 35(4):983–994
- [7] Chopra S, Reinhardt G, Dada M. The Effect of Lead Time Uncertainty on Safety Stocks. Decision Sciences, 2004, 35(1):1–24
- [8] Fang X, Zhang C, Robb D J, et al. Decision support for lead time and demand variability reduction. Omega, 2013, 41(2):390–396
- [9] Gershwin S B. Manufacturing systems engineering. PTR Prentice Hall (Englewood Cliffs, NJ), 1994
- [10] Little J D. A Proof for the Queuing Formula:  $L = \lambda W$ . Operations research, 1961, 9(3):383–387
- [11] Goyal S K. Economic batch quantity in a multi-stage production system. THE INTERNATIONAL JOURNAL OF PRODUCTION RESEARCH, 1978, 16(4):267–273
- [12] Kimura O, Terada H. Design and analysis of Pull System, a method of multi-stage production control. The International Journal Of Production Research, 1981, 19(3):241–253
- [13] Tabe T, Muramatsu R, Tanaka Y. Analysis of production ordering quantities and inventory variations in a multi-stage production ordering system. International Journal of Production Research, 1980, 18(2):245
- [14] Conway R, Maxwell W, McClain J O, et al. The role of work-in-process inventory in serial production lines. Operations Research, 1988, 36(2):229–241