

基于决策树利润优化模型的生产问题决策分析

摘要

为了优化产品生产流程中的质量控制体系，同时实现成本的有效削减和企业的长期可持续发展目标，本文通过构建合理的数学模型给出了不同情形下的最优决策分析。本文综合考虑了零配件和成品的相关参数对生产成本和企业信誉的影响，并建立起抽样检测下的决策分析模型，使该模型更符合生产实际。故该模型能很好地优化整个生产流程、提升质量控制效率与经济效益。

对于问题一，本文通过假设总体以及样本量足够大，将二项分布模型近似为**正态分布模型**。根据两种情形下的信度确定**显著性水平**，计算**置信区间**，最终推导得到了在一定的信度和误差限下的最小检测次数公式。最后通过带入数值得到了相应的最小检测次数，分别为**139**和**98**。并通过控制变量单独分析了检测次数与误差限、显著水平和次品率之间的关系，佐证了抽样方案的合理性。

对于问题二，本文建立了基于**决策树的利润最大化模型**，模拟了零配件检测、产品装配、成品检测及不合格品处理等环节，并考虑了退货损失、企业信誉等因素。我们通过**仿真模拟**，对16种不同的方案进行了检测，以最终利润来选择最优方案。其中我们为了体现企业信誉对企业营业的影响，构造了**企业信誉和企业销售量之间合理的函数关系**。经过优化后，我们得到了该情形下的最优决策方案，并制定了相应的图表。

对于问题三，题目要求在更多工序和零配件的情况下重复第二题的求解过程。由于此时仿真模拟量过大，本文在第二题模型的基础上又建立了**遗传算法模型**。…………

对于问题四，题目要求零配件和成品的次品率通过抽样检测给出。故本文在问题一的基础上，重新对总体进行抽样，结合第一问中得出的最小检测次数公式以及**贝叶斯更新模型**，求解不同情况下的零配件次品率和成品次品率，得到相应的**Beta函数**，再将对应的次品率作为参数传入到问题二和问题三构架的利润最大化模型中，重新得到基于抽样检测的最优决策方案。

关键词： 正态近似 决策树 利润最大化 仿真模拟 遗传算法 贝叶斯更新

1 问题重述

1.1 问题背景

某企业生产某种畅销的电子产品。在购买零件阶段，需要分别购买零配件1和零配件2，零配件的质量将直接决定生产成品的质量。因此两种零配件有必要在装配之前进行质量检测。在装配阶段，对于两种零配件装配的成品中，只要其中一个零配件不合格，则成品一定不合格；如果两个零配件均合格，装配出的成品也不一定合格。对于不合格成品，企业可以选择报废，或者对其进行拆解，拆解过程不会对零配件造成损坏，但需要花费拆解费用。

因此为了优化产品生产过程中的质量控制流程，减少成本投入以及实现企业的可持续发展，我们需要制定出一个良好的决策方案。一方面，我们需要在购买零配件时通过合理抽样来决定接受或拒收；另一方面，我们需要根据零配件以及成品的次品率采取合适的检测方案，以确保在进入市场的成品的次品率足够低的情况下，来尽量减少检测以及拆解等程序带来的成本消耗。在这种情形下，建立一个综合性的数学决策模型，对于优化产品生产流程十分重要。

1.2 题目重述

- 针对问题一：在供应商声称一批零配件（零配件1和零配件2）的次品率不会超过某个标称值的情况下，根据一个合适的抽样检测假设，分别在“在 95%的信度下认定零配件次品率超过标称值，则拒收这批零配件”和“在 90%的信度下认定零配件次品率不超过标称值，则接收这批零配件”两种情形下设计检测次数尽可能少的抽样检测方案。
- 针对问题二：在给定两种零配件和成品次品率、购买成本、检测成本等信息后，在生产过程中的4个具体阶段作出合理决策，并给出决策的依据及相应的指标结果。
- 针对问题三：在给出了2道工序、8个零配件的情况下重复问题2，给出具体的决策方案，以及决策的依据及相应指标。
- 针对问题四：假设问题 2 和问题 3 中零配件、半成品和成品的次品率均是通过抽样检测方法得到的，请重新完成问题2 和问题3。

2 问题分析

2.1 问题一的分析

问题一是在标称值给定的情况下，企业需要根据抽样检测结果决定是否接收零配件。我们可以利用二项分布和正态近似的方法，以及置信区间的构建进行分析。显然零配件的状态只有次品和合格品两类，因此零配件的分布满足二项分布，由于实际生产过程中的零配件数量很大，在这种情况下二项分布又可以用正态分布来近似。由于该题所给的两种情形都有相应的置信度要求，因此我们可以通过采用单侧检验的方法设置合理的零假设来确定接受域和拒绝域，并在此基础上进行相关的公式推导来计算检测次数较少的抽样检测方案。

2.2 问题二的分析

问题二要求我们为企业在生产过程中的关键决策点制定策略，包括零配件和成品的检测、不合格成品的处理，以及如何应对用户退换要求，以最大化利润并维护企业信誉。由于本题需要我们制定决策策略，我们可以采用决策树模型和利润最大化模型的方法，以及进行仿真模拟来加以分析。我们首先构建了四个决策变量： x_1, x_2, y, z ，分别代表零配件1、零配件2、产品是否检测，以及次品是否拆解。我们通过决策变量的不同取值来代表是否执行该决策，以此构建出决策树模型。通过产品的单价，检测费用，拆解费用，售价等信息，我们建立了利润最大化模型。我们还考虑到了产品合格率对应于产品销量的影响，建立了可能的影响曲线，以实现多目标优化。基于此我们通过大量的仿真模拟，获得利润最大的决策模型，进而获得最佳决策。

2.3 问题三的分析

2.4 问题四的分析

问题四要求我们通过抽样检查的方法来确定零配件、成品和半成品的次品率。我们采用了第一问所使用的方法，在不知道抽样总体的次品率的情况下先进行小样本抽样（例如只抽20个），再根据这个先验概率构建预期的最小检测次数，由此决策第二次抽样的数目。再根据贝叶斯更新模型，每次抽样之后都会更新对应的次品率的Beta分布函数，当该函数的方差小于一个很小的值时认为找到了符合要求的Beta函数，由此得到每一步决策中相应的次品率。再将得到的次品率带入到第二问和第三问中建立起的数学模型中重新进行模拟，最终我们就能得到基于抽样检测的最佳决策方案。

3 模型假设

- 1 假设零配件的总数量足够大，使得抽样检测过程中的二项分布可以近似为正态分布。
- 2 假设出现次品的零配件均匀分布
- 3 假设零配件和成品的检测过程是独立的，即每个零配件或成品的检测结果不会受到其他零配件或成品检测结果的影响。
- 4 假设拆解之后的零件一定要经过检验才能决定是否进行回收，不合格的零件一定不能重新投入生产。

4 符号说明

符号	解释
\hat{p}	样本中抽到的次品的比例
n	抽样的样本数
p_0	标称的次品率
α	显著性水平
$Z_{\alpha/2}$	标准正态分布的分位数
E	误差限
x_i	是否检测零配件i
y_i, y	是否检测半成品i，成品
z_i, z	是否拆解半成品i，成品
c_i	第 i 种零配件的购买成本
p	成品的市场售价
k_i	第 i 种零配件的检测成本
a	每件产品的装配成本
d	拆解不合格产品的费用
l	每件退回的不合格产品产生的损失
q_i, q_p	配件i的次品率，成品的次品率。
$\sigma(x)$	客户满意度函数
N	产量
n	需求量
m	滞销量

5 模型建立

5.1 问题一模型的建立与求解

5.1.1 正态分布模型模拟抽取检测的样本

我们在这题中将供应商提供的零配件看做一个总体，其中某个样品是次品的概率是 $p_0 = 0.10$ 。我们从这个总体中随机抽取 n 个样本，并计算样本中抽到的次品的比例 \hat{p} 。这种抽样过程遵循二项分布，即：

$$\hat{p} \sim \text{Binomial}(n, p_0)$$

根据假设，我们认为零配件总体的总数很大。在大样本情况下，根据**中心极限定理**，二项分布可以近似为**正态分布**。因此，样本中次品的比例 \hat{p} 可以近似为正态分布，且其均值为 p_0 ，方差为 $\frac{p_0(1-p_0)}{n}$ 。也就是说：

$$\hat{p} \sim N\left(p_0, \frac{p_0(1-p_0)}{n}\right)$$

这将是后面进行最小样本数的公式推导提供基础。

5.1.2 构建置信区间

为了保证检测的准确性，我们需要给出样本比例 \hat{p} 的置信区间。置信区间的目的是根据样本比例 \hat{p} 来推测总体比例 p_0 的可能范围。对于比例问题，通常采用以下形式的置信区间：

$$\hat{p} \pm Z_{\alpha/2} \cdot \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

其中：

- $Z_{\alpha/2}$ 是标准正态分布的分位数，对应 $1 - \alpha$ 的置信水平；
- $\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$ 是比例的标准误。

在样本量足够大时， \hat{p} 可以用 p_0 来近似代替，因此标准误可以简化为 $\sqrt{\frac{p_0(1-p_0)}{n}}$ 。置信区间的上限和下限分别为：

$$\begin{aligned}\text{上限} &= \hat{p} + Z_{\alpha/2} \cdot \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \\ \text{下限} &= \hat{p} - Z_{\alpha/2} \cdot \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\end{aligned}$$

5.1.3 设定误差范围

在实际问题中，我们通常关心的是估计值 \hat{p} 与总体比例 p_0 之间的误差不超过某个指定范围 E （误差范围）。我们希望构造出一个置信区间，使得：

$$|\hat{p} - p_0| \leq E$$

这个式子可以转换为：

$$Z_{\alpha/2} \cdot \sqrt{\frac{p_0(1-p_0)}{n}} \leq E$$

为了使得估计结果较为精确，后文计算最小样本量时拟采用的误差范围为 $E = 0.01$ 到 $E = 0.10$

5.1.4 推导样本量公式

为了解决这个不等式，我们首先平方两边，得到：

$$Z_{\alpha/2}^2 \cdot \frac{p_0(1-p_0)}{n} \leq E^2$$

然后将 n 单独提出，得到样本量的公式：

$$n \geq \frac{Z_{\alpha/2}^2 \cdot p_0 \cdot (1-p_0)}{E^2}$$

根据上述公式，我们可以得到实际生产过程中（抽样检测的数目为整数）的最小样本数目表达式：

$$n = \left\lceil \frac{Z_{\alpha/2}^2 \cdot p_0 \cdot (1-p_0)}{E^2} \right\rceil \quad (1)$$

5.1.5 模型求解

- 对于情形一：我们的要求是在95%的信度下认定零配件次品率超过标称值，则拒收这批零配件；据此我们做出零假设：抽样检测的零配件次品率不超过标称值，即

$$H_0 : p \leq p_0 \quad (2)$$

可画出对应的接受域与拒绝域的图像：

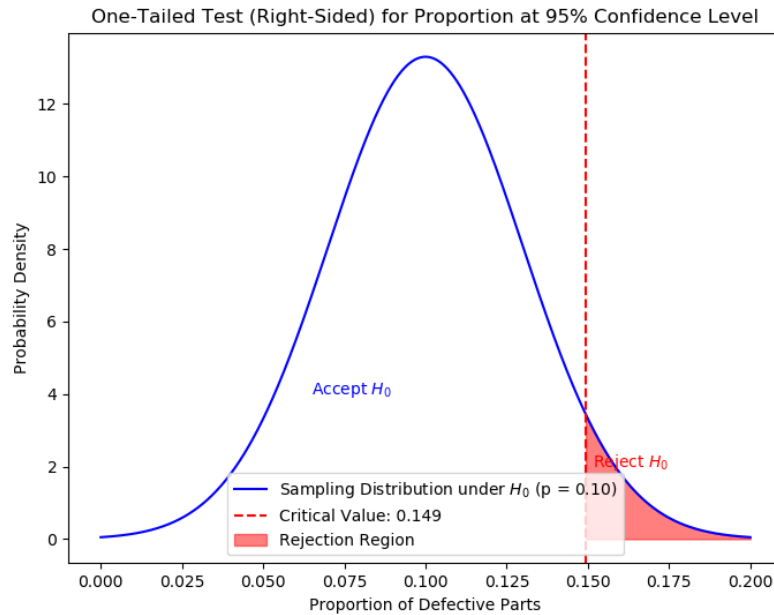


Figure 1: 95%的信度下的接受域与拒绝域

当信度为95%时，取得 $Z_{\alpha/2} = 1.645$ ，同时取 $E = 0.05$ ，代入公式 (1)可得到最终的求解到的最小检测次数为 **139**。

- 对于情形二：我们的要求是在 90%的信度下认定零配件次品率不超过标称值，则接收这批零配件；据此我们也做出零假设：抽样检测的零配件次品率不超过标称值，同 (1)，可画出对应的接受域与拒绝域的图像：

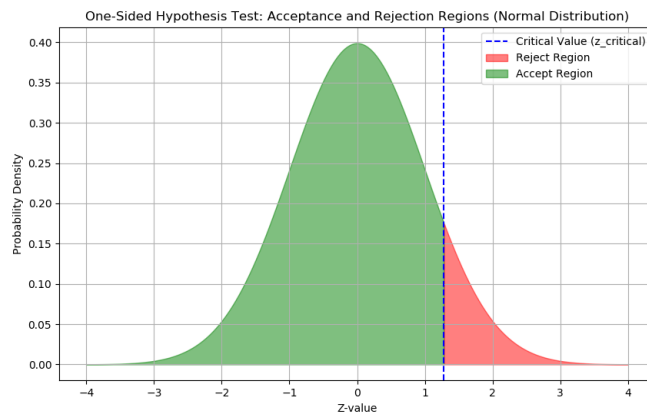


Figure 2: 90%的信度下的接受域与拒绝域

当信度为90%时，取得 $Z_{\alpha/2} = 1.28$ ，同时取 $E = 0.05$ ，代入公式 (1)可得到最终的求解到的最小检测次数为 **98**。

5.1.6 抽样检测方案合理性检验

根据公式 (1)，我们可以作出抽样检测次数 n 与显著性水平 α 、标称的次品率 p_0 ，误差限 E 三者之间的关系。

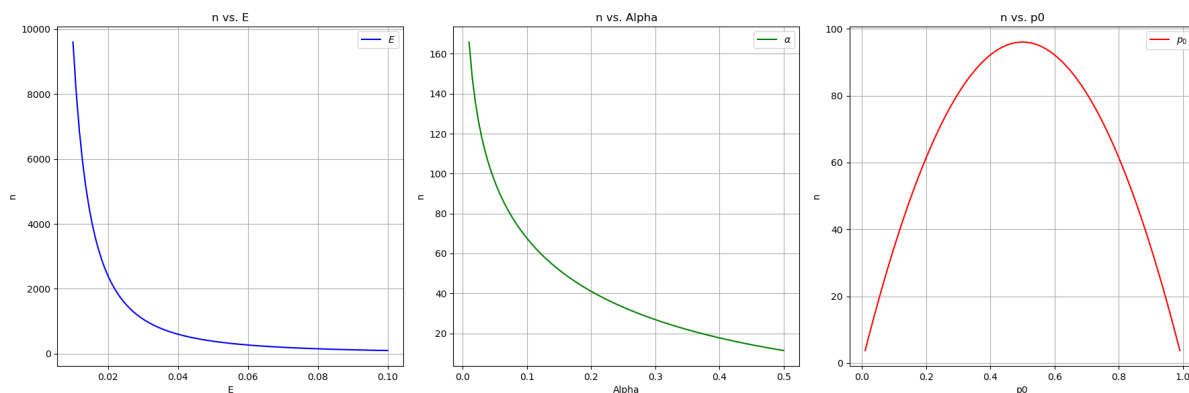


Figure 3: 样本量 n 与误差限 E 、显著性水平 α 、假设的次品率 p_0 的关系

- **与误差限 E 的关系：**误差限 E 表示我们对产品次品率估计的精度要求。由于 E^2 在分母中，误差限 E 的减少将导致样本量 n 的增大。这说明如果我们希望减少测量次品率的误差，则需要增加样本量。这对于确保生产过程符合严格的质量标准是非常重要的。
- **与显著性水平 α 的关系：**显著性水平 α 定义了置信区间的范围。显著性水平 α 越小， $Z_{\alpha/2}$ 值越大，样本量 n 也会增加。这说明提高置信水平会要求更大的样本量来确认产品质量，更高的置信度要求更大的样本量以减少错误判断的风险。
- **与标称的次品率 p_0 的关系：**当 p_0 接近 0 或 1 时，因此样本量 n 较小；当 p_0 接近 0.5 时，样本量 n 最大。这说明如果初期预期次品率约为50%，则需要较大的样本量来确保对次品率的准确估计；而若次品率预期接近0或1，则样本量可以相对较小。

综上所述，最终的抽样检测方案为在情形一下抽取139个样本，在情形二下抽取98个样本。同时抽样过程中不必要一次性抽满139个或98个样本，秉持着检测次数最少的原则，可以在抽样过程中不断利用公式 (1)更新预计的抽样个数，直到抽样所产生的次品率概率分布非常集中时就有足够把握认为零假设的真假（可能此时并没有达到预期的最小抽样检测数，因此能减少检测次数）。

5.2 问题二模型的建立与求解

5.2.1 决策树的建立

在解决B题第二问的过程中，我们所构建的决策树从零配件1的检测决策 (x_1) 开始，此决策点分为“检测”和“不检测”两个分支。对于每个零配件1的检测决策后，我们进一步决定是否检测零配件2 (x_2)。接着，每种零配件组合后，决定是否对成品进行检测 (y)。最后，对于检测出的不合格成品或者是顾客的退还产品，我们需要作出是否拆解的决策 (z)，决策结果为直接报废或进行拆解。这些决策将直接影响到终端节点所显示的经济后果，包括成本和收益的计算。在建立的决策树中，通过考虑是否检测零配件1 (x_1)、是否检测零配件2 (x_2)、是否检测成品 (y)，以及对不合格成品是否进行拆解 (z)，我们得到了四个二元决策变量。每个决策变量有两个可能的选项（检测或不检测/拆解或不拆解），因此我们一共有16种决策方案。

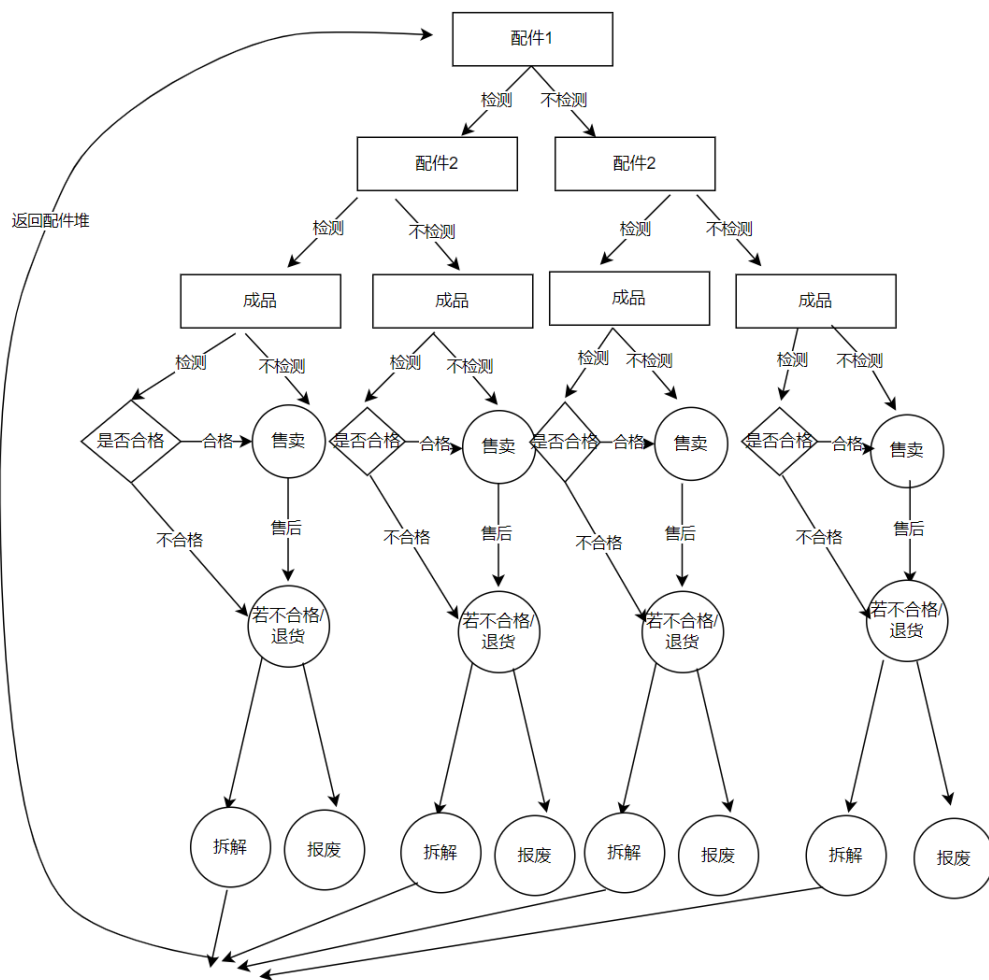


Figure 4: 决策树

Table 2: 对应方案的各个决策变量取值

Plan	x_1	x_2	y	z
1	0	0	0	0
2	0	0	0	1
3	0	0	1	0
4	0	0	1	1
5	0	1	0	0
6	0	1	0	1
7	0	1	1	0
8	0	1	1	1
9	1	0	0	0
10	1	0	0	1
11	1	0	1	0
12	1	0	1	1
13	1	1	0	0
14	1	1	0	1
15	1	1	1	0
16	1	1	1	1

5.2.2 最大化利润模型的建立

我们通过模拟的方式构建了最大化利润的模型，该模型基于以下几个步骤进行具体推导。

1. 零配件的购买 假设企业需要购买 N 套零配件来生产 N 个成品。对于每个零配件 i （包括零配件1和零配件2），其次品率为 q_i 。购买零配件的总成本可以表示为：

$$\text{零配件总成本} = N \cdot (c_1 + c_2)$$

其中， c_1 和 c_2 分别为零配件1和零配件2的单位购买成本。

2. 零配件的检测 若企业决定检测零配件，则会产生检测成本，并根据检测结果决定是否丢弃不合格的零配件。假设企业选择检测零配件1的概率为 x_1 ，检测零配件2的概率

为 x_2 ，那么检测成本为：

$$\text{检测成本} = N \cdot (x_1 \cdot k_1 + x_2 \cdot k_2)$$

其中， k_1 和 k_2 分别为零配件1和零配件2的单位检测成本。如果检测出零配件不合格，我们假设其会被丢弃并替换，因此有效的次品率在检测后为：

$$q'_1 = q_1 \cdot (1 - x_1)$$

$$q'_2 = q_2 \cdot (1 - x_2)$$

经过检测后，次品率被降低，但仍可能存在次品零配件进入装配环节。

3. 产品的装配 在装配过程中，我们将零配件1和零配件2进行组合，生成成品。每次装配的成本为 a 。成品的质量由两个零配件的质量和成品自身的次品率 q_p 决定。成品的次品率综合考虑了零配件的次品率以及成品自身的装配问题，可以表示为：

$$q'_p = 1 - (1 - q'_1)(1 - q'_2)(1 - q_p)$$

即在装配过程中，成品的最终次品率 q'_p 是由零配件1和零配件2的次品率以及装配过程中出现次品的概率共同决定的。

4. 成品的检测与处理 企业可以选择对成品进行检测，以避免不合格产品流入市场。检测成品的成本为：

$$\text{成品检测成本} = N \cdot y \cdot p_t$$

其中， y 是是否检测成品的决策变量， p_t 是成品的检测成本。如果检测出的成品不合格，企业可以选择直接报废或支付拆解费用 d 以回收可用的零配件：

$$\text{处理成本} = N \cdot (q'_p \cdot (y \cdot l + z \cdot d))$$

同时，对于拆解后的零配件，如果选择再次进行检测，则需要支付额外的检测成本，表示为：

$$\text{拆解后零件的检测成本} = z \cdot N \cdot (x_1 \cdot k_1 + x_2 \cdot k_2)$$

其中， l 是每件退回不合格产品产生的损失， z 是是否拆解不合格成品的决策变量， d 是拆解费用。

5. 利润的计算 最终的利润 Π 可以表示为销售收入减去所有相关成本，包括零配件购买成本、检测成本、装配成本、退货损失和拆解费用：

$$\Pi = \text{销售收入} - \text{总成本}$$

销售收入是由合格产品的市场售价决定的，而总成本包括上述所有成本,化简后，利润函数为：

$$\Pi = n \cdot (1 - q'_p) \cdot p - N [c_1 + c_2 + a + y \cdot p_t + q'_p \cdot (y \cdot l + z \cdot d) + (1 + z) \cdot (x_1 \cdot k_1 + x_2 \cdot k_2)]$$

6. 目标 我们的目标是通过选择合适的检测和拆解策略 (x_1, x_2, y, z) 来最大化利润函数 Π ，从而优化企业的生产决策，实现利润最大化。

5.2.3 通过仿真模拟来选择最优策略

在仿真模拟中，为了确保模型的可靠性和结果的统计显著性，我们通过次品率 q_i 模拟出足够大的样本量的配件空间,结合之前所建立的决策树模型和最大化利润模型来寻找最佳策略。

1. 样本量的确定 在模拟中，我们选择一个足够大的样本量 $N = 100000$ ，以确保样本能够涵盖各种可能的情况。我们所选区的样本量 N 足够大，以减少由于随机性带来的误差，同时能反映出生产过程中的真实情况。

2. 基于次品率生成零配件 对于每个零配件 i （如零配件1和零配件2），次品率 q_i 表示该零配件出现次品的概率，我们主要假定次品零配件在总体中成均匀分布。我们通过随机数生成器生成 N 个零配件，每个零配件的质量状态根据次品率 q_i 进行判断：

具体而言，对于每个零配件，我们生成一个0到1之间的随机数。如果该随机数小于次品率 q_i ，则将该零配件标记为次品（即不合格）；否则标记为正品（即合格）。

3.其它关联条件的加入 由于产品的销售量同时会受到产品总次品率的影响，产品总次品率越高，顾客的退换率越高，企业信誉随之下降，产品销售量也会大打折扣。为此我们寻找了相应的满意度-购买意愿函数，满意度越高，购买意愿越强，滞销率越低。

$$\sigma(p) = \frac{1}{1 + \exp(-5 \cdot (p - 0.5))} (0 < p < 1)$$

而满意度反应了需求量的变化情况：

$$\sigma(p) = 1 - \frac{n}{N}$$

通过该函数，我们可以更好的拟合出市场的真实状况，并通过滞销量 m 这一指标反应销售状况。图示如下：

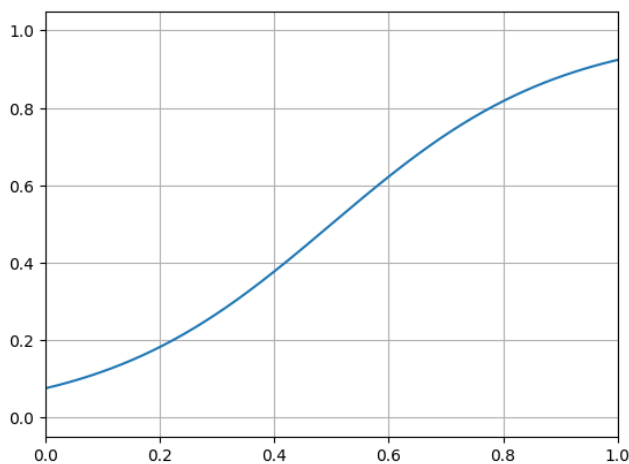


Figure 5: 次品率-满意度函数

4.结合决策树和利润最大化模型求解 综合上述分析，我们可以对本题的六种情况进行仿真模拟，对应的模拟结果如下：



(a) Situation 1



(b) Situation 2

Figure 6: Situations 1 and 2



(a) Situation 3



(b) Situation 4

Figure 7: Situations 3 and 4



(a) Situation 5



(b) Situation 6

Figure 8: Situations 5 and 6

通过数据整体可以看出，采取检测产品的策略具有明显的优越性，既保证了足够高的满意度，又可以获得高额利润；而如果不加以检测，在产品次品率过高的情况下会严重影响满意度，甚至会亏本，是极为不推荐的销售策略。但是当产品的拆解率过高的时候，选择拆解的决策方案的优越性显著下降，这从情况六中体现的非常明显。对于每种情况单独分析的最优选择结果如下表所示：

Table 3: 不同情况下可采取的方案

情况	x1	x2	y	z
情况一	0	0	1	1
情况二	1	0	1	1
情况三	0	0	1	1
情况四（策略一）	0	1	1	1
情况四（策略二）	1	0	1	1
情况五	0	1	1	1
情况六	0	0	1	0

5.3 问题三模型的建立与求解

5.3.1 决策树的建立

该题我们同样先构建了相对应的决策树，决策变量一共有16种，主要对应的是配件，半成品，成品是否检测，半成品，成品是否拆解，具体情况如下表所示：

Table 4: 决策变量的取值情况

x1	x2	x3	...	x7	x8	y1	y2	y3	z1	z2	z3	y	z
0	0	0	...	0	0	0	0	0	0	0	0	0	0
0	0	0	...	0	1	0	0	0	0	0	0	0	1
0	0	0	...	1	0	0	0	0	0	0	1	0	0
0	0	0	...	1	1	0	0	0	0	1	0	0	0
⋮	⋮	⋮	...	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
1	1	1	...	1	0	1	1	1	1	1	0	1	1
1	1	1	...	1	1	1	1	1	1	1	1	1	1

可以从表格看出，总的可能决策种类一共有 2^{16} ，过于多的决策种类不利于我们直接进行仿真模拟，故而我们采取遗传算法来找寻最优的决策。

5.3.2 遗传模型的建立

遗传算法是一种基于自然选择和遗传机制的优化算法。我们通过以下几个方面建立对应的数学模型，包括种群、个体、基因、染色体、交叉操作、变异操作、适应度计算和选择。

种群 在该遗传模型中，我们建立出一个个体数为 M 的种群，通过观察每次迭代后种群的优势基因型的变化来找寻最佳决策方案。

个体 个体是遗传模型中的基本单位，代表一个可能的决策方案。每个个体由一组基因组成，基因的排列顺序和取值构成了染色体。在该问题中，每个个体对应于一个可行解，其优劣通过适应度函数进行评估。

基因与染色体 基因是个体的最小组成部分，在该模型中我们将每个决策变量视为一个基因。染色体是基因的排列序列，用于表示一个完整的解。在本文中，染色体由十六个基因组成，每个变量的取值为 0 或 1。

$$\mathbf{c}_i = (x_1, x_2, \dots, x_8, y_1, y_2, y_3, z_1, z_2, z_3, y, z)$$

交叉操作 交叉操作是遗传算法中的主要操作之一，通过交换两个个体的部分染色体来生成新的个体。常见的交叉方式包括单点交叉和多点交叉。交叉操作旨在通过组合父代个体的基因，生成具有不同特征的后代个体。我们设置了合适的交叉率，来实现遗传效果。设两个个体 \mathbf{x}_1 和 \mathbf{x}_2 进行单点交叉，交叉点为 p ，则新的子代个体 \mathbf{x}'_1 和 \mathbf{x}'_2 可以表示为：

$$\mathbf{x}'_1 = (x_{1,1}, x_{1,2}, \dots, x_{1,p}, x_{2,p+1}, \dots, x_{2,N})$$

$$\mathbf{x}'_2 = (x_{2,1}, x_{2,2}, \dots, x_{2,p}, x_{1,p+1}, \dots, x_{1,N})$$

其中 p 为交叉点的位置。

变异操作 变异操作用于保持种群的多样性，通过随机改变个体的一个或多个基因的值来产生新的个体。变异通常以较低的概率发生。变异操作能够避免算法陷入局部最优解，并促进全局搜索，帮助我们找寻到最优策略。

适应度计算 适应度函数用于评估个体的优劣。适应度值越高，表示该个体的解越优。我们采用了第二问中的最大利润模型来评估个体对应的适应度，当某方案的最大利润高，请对应的适应度相对提高。

选择 选择操作用于从当前种群中选择个体参与交叉和变异。选择的概率通常与个体的适应度成正比。我们采用了轮盘赌选择的方法，使得每个个体被选择的概率与其适应度值成正比。适应度值越高的个体，被选中的概率就越大。我们采用轮盘赌选择法，个体 \mathbf{x}_i 被选中的概率 p_i 与其适应度值成正比，定义为：

$$p_i = \frac{f(\mathbf{x}_i)}{\sum_{j=1}^M f(\mathbf{x}_j)}$$

其中 $\sum_{j=1}^M f(\mathbf{x}_j)$ 表示种群中所有个体的适应度总和。

5.3.3 具体的求解与计算

5.4 问题四模型的建立与求解

5.4.1 基于贝叶斯更新思想建立次品率的概率模型

本题的次品率需要通过抽样检测获得。由于无论是零配件、半成品还是成品，获得次品率的抽样方法都相同，因此这里只讨论对零配件的抽样情形。假设初始时我们没有任何关于零配件的先验信息，因此初始的次品率期望 p_0 要靠第一次抽样检测获取。这里我们可以使用一个较小的样本数目 $n = 20$ 进行抽样，抽样所得结果作为 p_0 的初始值。即假设第一次抽样中抽到了 k 个次品，那么 p_0 的值应该为

$$p_0 = \frac{k}{n}$$

此时次品率满足的 *Beta* 分布为

$$\text{Beta}(k, n - k)$$

结合公式 (1)，在得到次品率期望 p_0 的情况下，我们可以计算出在保证信度为 95%、误差 $E = 0.01$ 的情况下，下一次抽样的最小样本量 n_1 可以计算得出。

为了尽量减少检测次数，下一次抽样不应该直接抽取 $n_1 - n$ 个样本。我们可以折中选取 $\frac{n_1 - n}{2}$ 个样本，在这次抽样检测之后再更新次品率期望 p_0 ，以此再次计算得出下一次抽样的最小样本量。也就是说我们并不是一一次性将样品取完，而是采取多次抽样，不断在抽样过程中更新下一次的预期抽样，以此来达到减少抽样次数的目的。

同时，为了有一个标准衡量样本的次品率能多大程度上代表总体的次品率。这里我们基于贝叶斯更新的思想，在每次抽样结束之后更新次品率满足的 *Beta* 分布，当次品率满足的 *Beta* 分布的方差小于我们的误差值（这里采用较小值 0.01）时，我们有足够的理由认为当前次品率的 *Beta* 分布能较好地反映出总体抽样时得到对应次品率的概率分布。所以对于次品率的概率分布的迭代终止条件应该为

$$\frac{\alpha\beta}{(\alpha + \beta)^2(\alpha + \beta + 1)} < 0.01$$

对于上述更新次品率的概率模型的建立，可以做出以下的流程图加以说明：

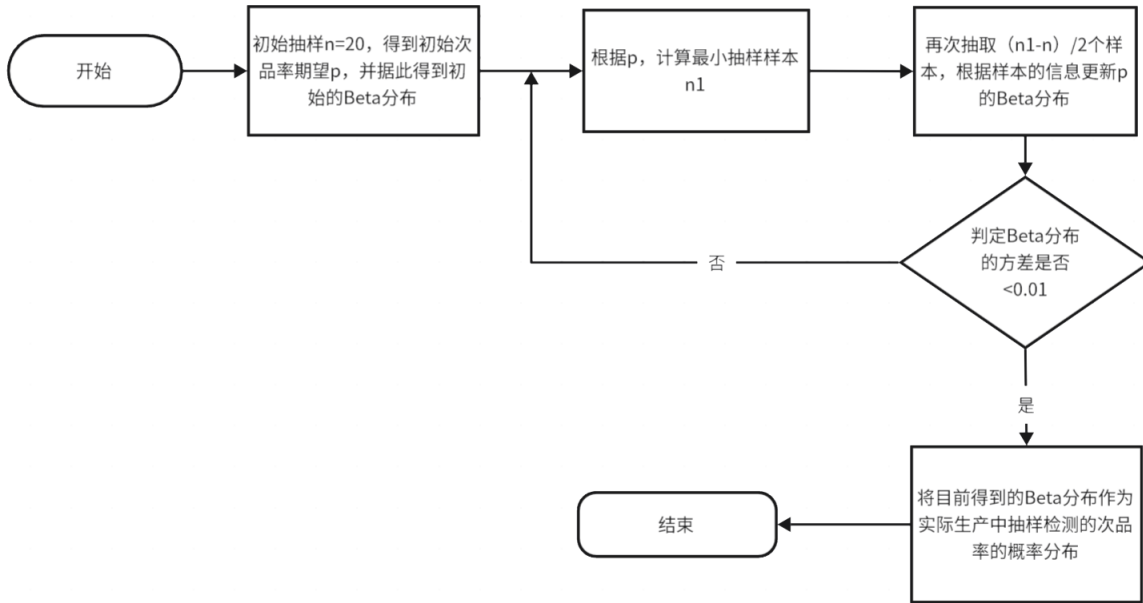
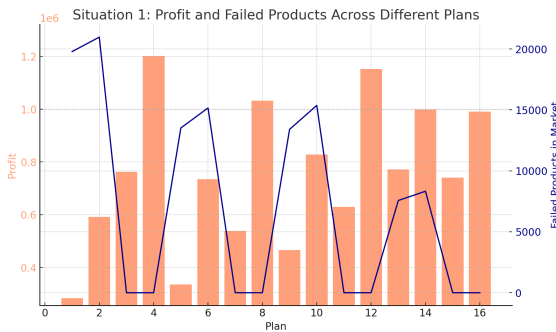


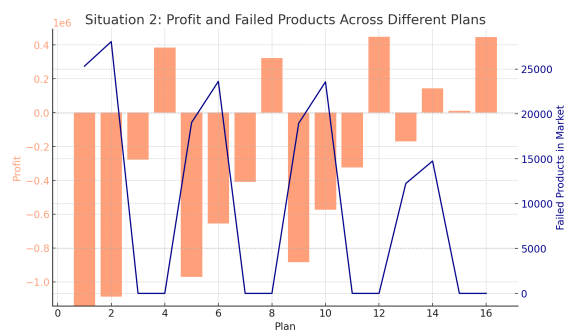
Figure 9: 次品率的概率模型

5.4.2 问题二的重新求解

通过前述方法所建立的模型，我们来解决问题二。我们可以获得对应的策略-利润-滞销量图示，具体如下：



(a) Situation 1

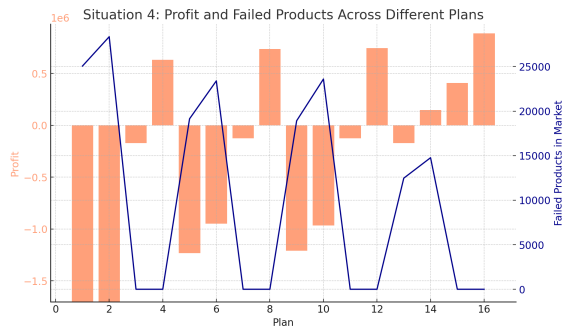


(b) Situation 2

Figure 10: Situations 1 and 2

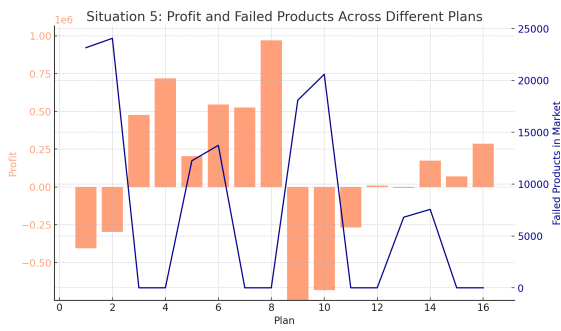


(a) Situation 3

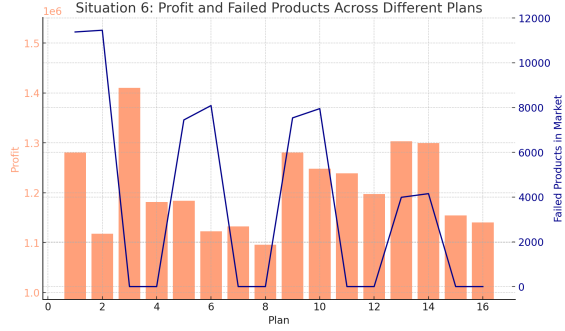


(b) Situation 4

Figure 11: Situations 3 and 4



(a) Situation 5



(b) Situation 6

Figure 12: Situations 5 and 6

从图示中可以看出，虽然我们采用抽样检测方法得到的次品率有一定的波动，但是整体的趋势并没有大幅度的改变，仍然是采取检验产品的策略更为优越，不检测的策略可能造成大量积压，乃至亏本，具体每种情况的策略选择如下表所示：

Table 5: 不同情况下可采取的方案

情况	x1	x2	y	z
情况一	0	0	1	1
情况二（策略一）	1	0	1	1
情况二（策略二）	1	1	1	1
情况三	0	0	1	1
情况四	1	1	1	1
情况五	0	1	1	1
情况六	0	0	1	0

5.4.3 问题三的重新求解

基于次品率概率模型和第三问中建立的基于遗传算法的利润最优化模型，我们可以对问题三重新求解，最后求解得到的结论如下图所示：

6 模型的评价

6.1 模型的优点

- (1) 模型结合了二项分布的正态近似、决策树、利润最大化模型以及遗传算法等多种方法，形成了一个综合性的决策支持系统，能够全面考虑生产过程中的多种因素。
- (2) 模型考虑了企业信誉对产品销量的影响，通过构建信誉与销量的函数关系，优化决策方案以维护企业信誉。
- (3) 模型可以根据不同的生产工序和零配件数量进行调整，适用于多种生产场景。同时通过仿真模拟，能够应对多种可能的生产情况，选择最优方案。
- (4) 模型通过优化抽样检测次数，减少不必要的检测成本，同时确保产品质量的可靠性。在生产过程中，通过决策树模型选择最优的检测和拆解策略，有效降低了因次品带来的额外成本。
- (5) 模型结合了实际的生产情况进行了剪枝，效率较高且更符合实际生产应用。

6.2 模型的缺点

- (1) 模型的准确性高度依赖于输入参数的准确性，如次品率、检测成本、拆解费用等，这些参数的微小变化都可能对最终决策产生较大影响。

- (2) 模型主要关注生产过程中的质量控制和成本优化，未充分考虑市场变化对生产和销售的影响。

6.3 模型的推广

- (1) 在实际生产中，市场需求、原材料供应等因素可能发生变化。该模型可以引入动态调整机制，实时优化生产和检测策略。通过定期更新次品率、成本等数据，并重新运行模型进行仿真模拟，企业可以快速响应市场变化，保持竞争优势。
- (2) 该模型可以纳入风险因素分析，评估不同策略下的风险水平。通过构建风险应对策略库，企业可以在面临突发情况时迅速制定应对措施，降低损失并保障生产连续性。

参考文献

- [1] 杨学兵,张俊.决策树算法及其核心技术[J].计算机技术与发展,2007,(01):43-45.
- [2] 许珂.基于贝叶斯和动态规划的成本差异控制模型及应用研究[D].天津大学,2008.
- [3] 葛继科,邱玉辉,吴春明,等.遗传算法研究综述[J].计算机应用研究,2008,(10):2911-2916.