

数学建模论文排版

摘要

这里写摘要，国赛论文摘要要求是一页最好，不要多也不要太少。

关键词： Fisher 精确检验 多元线性回归 系统聚类 灰色关联分析

一、问题重述

某电子产品的生产企业需要综合诸多考虑购置零部件、产品抽检、产品拆解、报废等问题，以确保产品质量的同时降低成本。

问题一：考虑到零配件供应商所述次品率不高于既定标称值，企业拟采用抽样检测方法以验收此批零配件。因为企业需承担检测费用，企业希望应用数学模型得到最少抽检次数的抽样方案。

已知标称值为 10%，结合以下两种不同情况，分别设计出具体的抽样检测方案：

1. 拒收条件：在 95% 的置信水平下，如果检测结果表明零配件的次品率超出了标称值，那么这批零配件将被拒收。

2. 接收条件：在 90% 的置信水平下，如果检测结果表明零配件的次品率未超过标称值，那么这批零配件将被接收。

问题二：在已知零配件及成品次品率情况下，在电子产品生产的零配件检测、装配、成品检测、不合格品拆解的各个阶段为企业作出最优决策。并且结合判断依据及相应的指标对表 1 中企业在生产中遇到的情况作出相应的最优决策方案。

问题三：在零配件、半成品和成品的次品率已知情况下，重复问题 2 的生产决策方案以适配有 m 道工序、 n 个零配件的问题。并且应用此方法针对表 2 中情况给出判断依据和指标得到最优的决策方案。

问题四：在零配件、半成品和成品的次品率均由抽样检测获得的情况下，重新考虑问题 2、3 的生产决策方案。

二、问题分析

2.1 问题一的分析

我们需要根据题目中给出的两种不同的情况，分别设计抽样检测方案。考虑单次检验为次品严格服从经典二项分布，拟采用异常检测的经典取样方法：序贯概率比检测 $SPRT$ 来动态地作出抽样决策。

2.2 问题二的分析

由于已知各零配件以及成品的次品率，依据排列组合我们能够得知在生产成品时零配件优劣的组合情况以及对应的概率分布。又由于可能出现两个好的零部件组成一个坏的成品，综合成本考量我们是否需要考虑对成品进行拆解。我们预先设置对零配件、

成品的抽样检测参数（即：抽检/不抽检），通过多次模拟迭代实验得到最优的检测方案，以此来减少检测次数。

2.3 问题三的分析

三、基本假设与符号说明

3.1 基本假设

- 假设理论物理跟泵不存在；
- 假设数据中未填写的数据项为 0
- 假设所提供的数据准确无误；
- 不考虑因检验手段等原因对数据值的影响。

3.2 符号说明

符号	含义
N	每批供应商提供的零配件总量
D_i	第 i 次抽检的零配件数量
μ	次品率
A	拒真的决策边界
B	纳伪的决策边界
LR	似然比

四、问题一模型的建立与求解

针对每批零配件，假定总量为 N ，我们考虑采用异常检测的经典取样方法：序贯概率比检测 $SPRT$ 作为抽检方案。在此之前我们考虑每次取样的样本量为 D_i ，令单个零件次品与否的布尔值为 x ，考虑其单次试验成功（为次品）概率的期望为 μ ，则其显然服从经典的二项分布表示：

$$Bern(x|\mu) = \mu^x(1 - \mu)^{1-x} \quad (1)$$

接下来考虑其在样本集上的对数似然函数，针对第 i 次取样 D_i ，对其中的每个样本取到观测 $x_1, x_2 \dots x_n$ ，根据题目要求样本集中零配件的次品产生事件可认定为相互独立的。

则其似然函数可写为：

$$\mathbf{P}(D_i|\mu) = \prod_{n=1}^N p(x_n|\mu) = \prod_{n=1}^N \mu^{x_n} (1-\mu)^{1-x_n} \quad (2)$$

为便于后续处理，我们取其对数似然：

$$\begin{aligned} \ln \mathbf{P}(D|\mu) &= \ln \prod_{n=1}^N \mu^{x_n} (1-\mu)^{1-x_n} = \ln \mu \sum_{n=1}^N x_n + \ln(1-\mu) \sum_{n=1}^N (1-x_n) \\ &= \ln \mu \sum_{n=1}^N x_n + \ln(1-\mu) (N - \sum_{n=1}^N x_n) = \sum_{n=1}^N x_n \ln \mu + (1-x_n) \ln(1-\mu) \end{aligned} \quad (3)$$

接下来我们依据题干给定零假设和备择假设：

$$\begin{cases} H_0 : \mu > 0.1 \\ H_1 : \mu \leq 0.1 \end{cases} \quad (4)$$

题干中的两种情况意味着拒真和纳伪的显著性水平 α 和 β 分别为 0.05 和 0.1。在 *SPRT* 语境下，考虑决策边界：

$$A = \ln \frac{\beta}{1-\alpha} \quad B = \ln \frac{1-\beta}{\alpha}$$

于是，针对每次采样 D_i ，我们需要求出在零假设和备择假设下的似然比 LR ：

$$LR = \frac{\sum_{n=1}^N x_n \ln \mu_0 + (1-x_n) \ln(1-\mu_0)}{\sum_{n=1}^N x_n \ln \mu_1 + (1-x_n) \ln(1-\mu_1)} \quad (5)$$

需要注意的是，在原生的 *SPRT* 场景中， H_0 和 H_1 一般被认定为较为复杂的参数估计 θ_0 和 θ_1 ，这取决于它们事先假定样本服从一个较为严谨且高度可表达的概率分布。然而基于问题一，在没有明确历史数据和概率分布的先验情况下，我们只能将其建模为一般二项分布，为了遵循 *SPRT* 的使用场景，我们将二项分布参数建模为 $\mu_0 = 0.1 + \Delta\mu_0$ ， $\mu_1 = 0.1 - \Delta\mu_1$ 。通过轻微扰动量来拟合样本的分布与所报称值的差异，扰动量的设置取决于样本量的大小，这点我们将在后续给出实验和说明。

尽管在许多场景中单样本取样策略以及被证明取得了很好的效果，但考虑到题干背景，我们依然选择样本集作为采样标准。遵循 *SPRT* 方法，给定总零配件量 N ，初次取样 D_i 应为按照标称值所取的总样本配比，我们取 $D_1 = 0.01N$ ，而后计算出当前样本下的对数似然比 LR_1 。序贯检验比方法遵循以下停止法则：

$$\gamma = \inf \{n | n \geq 1, LR_n \in (A, B)\} \quad (6)$$

具体来说，若 $LR_1 \leq A$ ，接受 H_0 假设；若 $LR_1 \geq B$ ，接受 H_1 假设，否则继续采样。初次采样的样本量为 $D_1 = 0.01N$ ，假定每次采样的次品数为 n_i ，则此后每次采样量依据以下法则确定：

$$D_{i+1} = D_i - n_i \quad (7)$$

检验的完整流程可以作出如下表表示：

Algorithm 1: 序贯概率比检验 (SPRT) 流程

Input: 总零配件数量 N ，显著性水平 α ，第二类错误概率 β ，扰动量 $\Delta\mu_0, \Delta\mu_1$

Output: 接受的假设 (H_0 或 H_1)

1 计算决策边界 $A \leftarrow \ln \frac{\beta}{1-\alpha}$, $B \leftarrow \ln \frac{1-\beta}{\alpha}$;

2 设置 $\mu_0 \leftarrow 0.1 + \Delta\mu_0$, $\mu_1 \leftarrow 0.1 - \Delta\mu_1$;

3 初始样本量 $D_1 \leftarrow 0.01Nn$;

4 $i \leftarrow 1$;

5 **while** TRUE **do**

6 取样 D_i 个零配件，记录次品数量 n_i ;

7 计算 LR_i :

$$LR_i = \frac{\sum_{n=1}^{D_i} x_n \ln \mu_0 + (1 - x_n) \ln(1 - \mu_0)}{\sum_{n=1}^{D_i} x_n \ln \mu_1 + (1 - x_n) \ln(1 - \mu_1)}$$

if $LR_i \leq A$ **then**

8 | 接受零假设 H_0 ; **break**

9 **end**

10 **else if** $LR_i \geq B$ **then**

11 | 接受备择假设 H_1 ; **break**

12 **end**

13 **else**

14 | $i \leftarrow i + 1$; $D_{i+1} \leftarrow D_i - n_i$ 继续执行采样策略

15 **end**

16 **end**

17 **return** 接受的假设 (H_0 或 H_1)

接下来我们考虑之前我们搁置的 $\Delta\mu$ 的选取，这是该方案唯一的松弛参数；我们希望把它建模成一个函数而非定量。这是由于我们面临一个相当大的搜索空间，它主要取决于两个因素：（1）总零配件数量 N ，由于该变量我们是不可控且无先验的，所以我们暂且假定它的范围波动为 $[1000, 1000000]$ ；（2）真实标称值 (Ground Truth) 我们也预先假定它的范围波动为 $[5\%, 15\%]$ 。 $\Delta\mu$ 的确定直接意味着 SPRT 策略的固定，面对庞大的搜索空间这样做显然是欠鲁棒的，因此寻找一种对 $\Delta\mu$ 的拟合 $\Delta\mu = f(N, \theta)$ 是迫在眉睫的。其中 N 作为零配件总数是我们面临实际场景时的唯一自变量。 θ 是拟合函数的待优化参数。目标函数应该考虑到：（1）优化项：即尽可能减少取样量 D ；（2）惩罚项：即不出现误判的情况。于是它可以设计为：

$$\Delta\mu = f(N, \theta) \quad s.t. \quad \underset{\theta}{\operatorname{argmax}} \left(\frac{1}{D} \times \mathbf{1}_H \right) \quad (8)$$

这里的 $\mathbf{1}_H$ 是对假设 H 的指示函数，用以表示最终选取的假设是否为真。我们使用多层感知机 (MLP) 去拟合函数 $f(N, \theta)$ ，自定虚拟数据集的范围遵循前文中给出的零件数量 N 和真实标称值的波动范围。

在 1000 个 epoch 中，我们选取测试集中精度最高的 MLP 模型作为评估基线以衡量 SPRT 的表现情况。最终结果如图所示：

五、 问题二的模型建立与求解

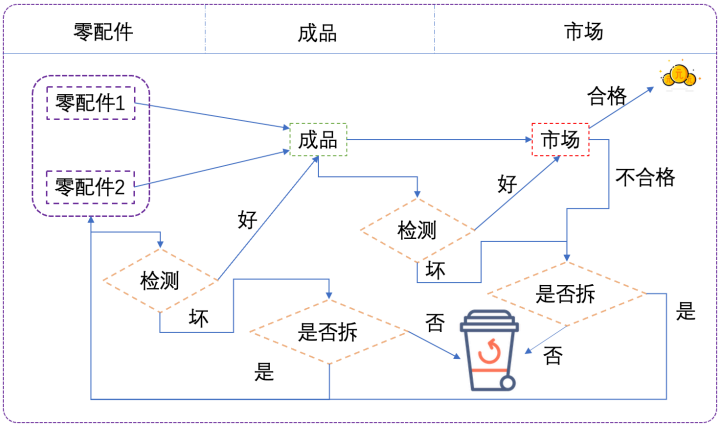


图 1 电子器件生产的流程图