

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

2.4 问题四的分析

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
C	决策向量

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所示：

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：

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

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1所示。具体来说，我们在

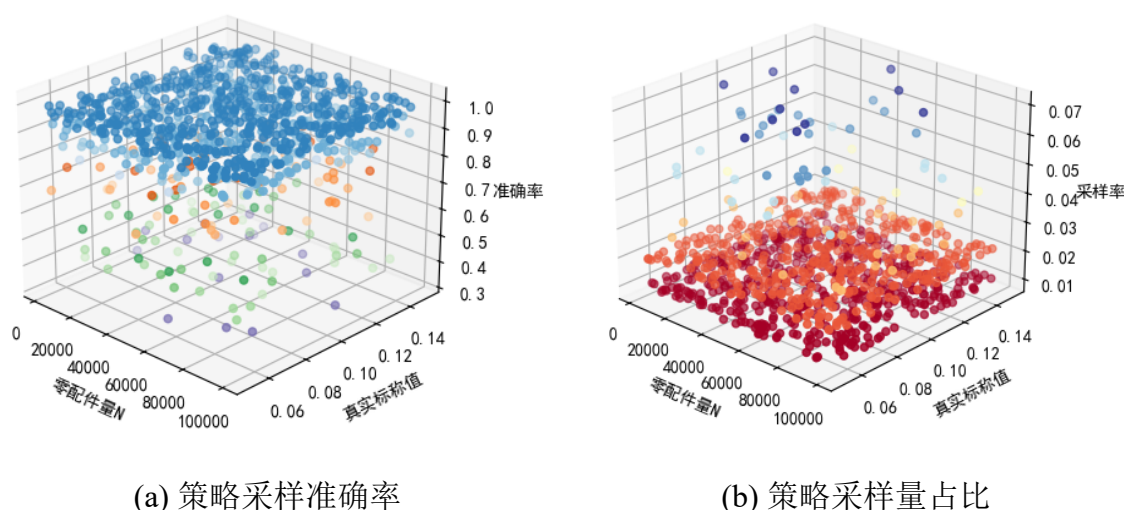


图 1 策略采样结果

N 和真实标称值的预测范围内均匀采样；对于每个样本点，我们模拟 100 种样本集情况（这体现在零配件次品的分布次序上）；在 100 种样本集中由于 N 固定，我们拟合的 $\Delta\mu$ 也固定，因此 $SPRT$ 策略也是固定的，我们充分评估此基准上的假设判断成功率，发现在绝大多数情况下我们的策略都具备 80% 以上的成功率；由于总样本量的跨度较大，我们用策略共采集的样本量在总样本中的占比作为衡量指标，我们的采样占比绝大多数情况小于 3%，这充分体现了我们采样策略的高效。此外根据我国工业检测国标：对于一般工业零件需要达到 2%-5% 的抽检率，对于高风险零件需要达到 10% 以上的抽检率；我们的采样策略是合乎标准且稳健的。

五、问题二的模型建立与求解

首先，我们可以将电子产品加工过程分为三个阶段，即：零配件、成品、市场。每当生产过程从一个阶段进入到另一个阶段均需要决策是否需要进行抽检，将此过程抽象为图2。

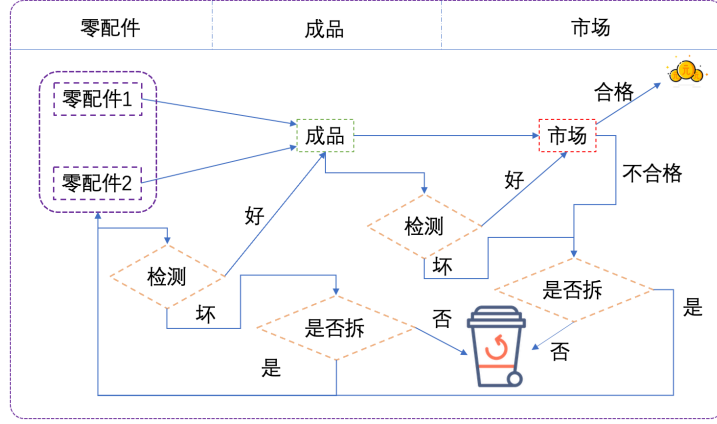


图 2 电子器件生产的流程图

5.1 模型的建立

在生产过程中我们设置一个列向量 $\vec{C} = (c_1, c_2, c_3, c_4)^T$ ，其中的元素 c_i 分别表示零配件 1、零配件 2、成品是否抽检及检测到不合格的成品是否拆解 (c_i 的值为 0 时表示为不抽检/不拆解 (丢弃)，值为 1 时表示检测/拆解)。我们的目标是通过合理的决策方案使得纯利润的期望值最最大。

因为零部件及成品的次品率已知，在每批零配件数目 N 足够大的情况下，我们可以通过对正次品进行排列组合，算出成品中的正次品零部件各种组合的概率。在这里假设每批投入零部件 1 和零部件 2 的数目分别为 N_1 和 N_2 (这里的 N 足够大)，其中 μ_1 和 μ_2 分别为零部件 1、2 的次品率，则近似可以认为供应商提供的零部件 1、2 的次品数目为 $N_1 \cdot \mu_1$ 和 $N_2 \cdot \mu_2$ 。先考虑对零部件不进行检测直接拼装为成品并且成品不拆除的情况 $\vec{C} = (0, 0, 0, 0)^T$ ，可以得出存在零部件次品的成品数目 k 的值域为 $[\max\{N_1 \cdot \mu_1, N_2 \cdot \mu_2\}, N_1 \cdot \mu_1 + N_2 \cdot \mu_2]$ 。令 k 个成品中存在 i 和 j 个零配件 1、2 的次品个数。可以得到成品中有 i 个零配件 1 和 j 个零配件 2 的概率为 $P_{i,j}$ ，计算公式如下：

$$P_{i,j} = C_{N_1 \cdot \mu_1}^i C_{N_2 \cdot \mu_2}^j \mu_1^i \mu_2^j (1 - \mu_1)^{k-i} (1 - \mu_2)^{k-j} \quad (9)$$

则 i 和 j 应当满足 $i + j \geq k, i \leq k, j \leq k$ ，经过计算得到 i 和 j 的取值范围为 $[k - j, k]$ 和 $[\frac{k}{2}, k]$ 。我们假设 P_k 为存在次品零部件的成品数目为 k 的概率，计算公式如下：

$$P_k = \sum_{i=k-j}^k \sum_{j=\frac{k}{2}}^k C_{N_1 \cdot \mu_1}^i C_{N_2 \cdot \mu_2}^j \mu_1^i \mu_2^j (1 - \mu_1)^{k-i} (1 - \mu_2)^{k-j} \quad (10)$$

由此可以得到成品中的不合格产品的期望值 M 为 (μ_3 为成品次品率)：

$$M = \sum_{k=\max\{N_1 \cdot \mu_1, N_2 \cdot \mu_2\}}^{N_1 \cdot \mu_1 + N_2 \cdot \mu_2} k \cdot P_k + (\min\{N_1, N_2\} - k) \cdot \mu_3 \quad (11)$$

令: 产品售价为 p_1 检测零配件 1、2 及成品的费用分别为 f_1 、 f_2 和 f_3 , f_4 和 f_5 为调换损失及拆解费用, 则可以构造产生费用的行向量 $\vec{F} = [N_1 \cdot f_1, N_2 \cdot f_2, \min\{N_1, N_2\} \cdot f_3, M \cdot f_5]$ 由此能够得到此轮次获利 $Profit$:

$$Profit = (\min\{N_1, N_2\} - M) \cdot p_1 - \vec{F} \cdot \vec{C} - M \cdot f_4 \quad (12)$$

使函数适配 16 中不同的决策方案, 我们需要定义每轮零配件池中所含数量 N'_1 、 N'_2 即包含每轮次固定补充数量 N_1 、 N_2 也包含上一轮次零产品拆解后补进的配件 (即: $N' = N + M$)。 M' 为新轮次中成品次品数目。则零配件池中为次品的平配件数目 n_1 、 n_2 的计算公式如下:

$$n_1 = N_1 \cdot \mu_1 + \sum_{j=\frac{k}{2}}^k \cdot P_{i,j} \quad n_2 = N_2 \cdot \mu_2 + \sum_{i=\frac{k}{2}}^k \cdot P_{i,j} \quad (13)$$

在此基础上我们可以构造每轮次的收益 $Profit$:

$$\begin{cases} Profit = (\min\{N'_1 - n_1 \cdot c_1, N'_2 - n_2 \cdot c_2\} - M') \cdot p_1 - \vec{F} \cdot \vec{C} - M' \cdot f_4 \\ M' = [\sum_{k=\max\{n_1, n_2\}}^{n_1+n_2} k \cdot P_k + (\min\{N'_1 - n_1 \cdot c_1, N'_2 - n_2 \cdot c_2\} - k) \cdot \mu_3] \cdot c_3 \end{cases} \quad (14)$$

为了更直观的得到不同决策向量对电子产品生产的收益的影响, 我们将图2中的单次生产流程抽象为一个生产轮次。在每个轮次中我们假设每轮接受供应商提供的配件数目为固定的 N_1 、 N_2 , 经过上一轮被检测为次品的零部件全部被舍弃, 被拆解的成品中的零配件投入下一轮次的零配件池进入生产。在每轮次生产中我们需要确定不同决策向量以保证电子产品生产的收益, 于是我们拟定 t 个生产轮次为一个生产周期, 每个生产周期中决策向量 $\vec{C}^* = (c_1, c_2, \dots, c_{4t})^T$ 为 t 个生产轮次中决策列向量 C 的叠加, 如图3。

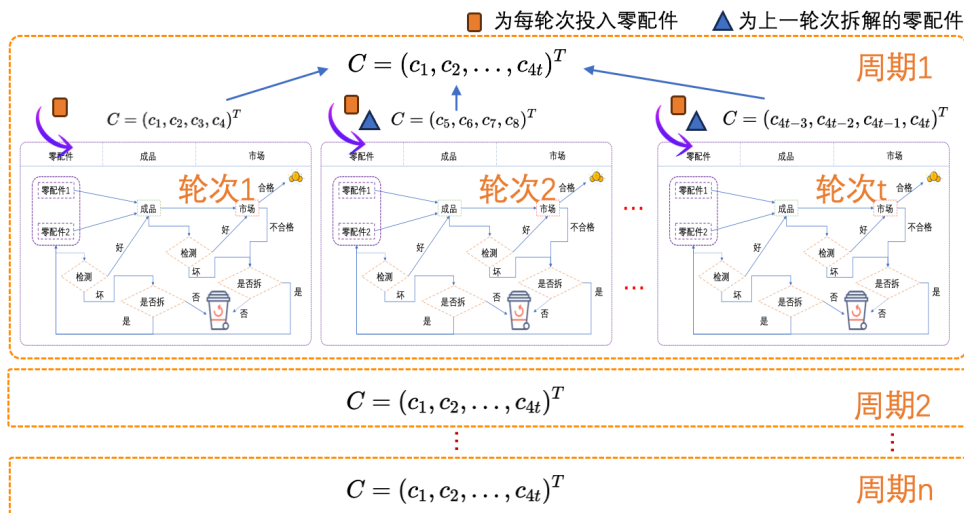


图3 产品决策周期图

综上所述我们构建了一个基于生产周期的决策模型, 通过对每个生产周期中的决策向量

C^* 进行优化，就能够得到最优的决策方案。

$$\begin{aligned} \max \quad & Profit = \sum_{n=1}^t [(\min\{N'_1 - n_1 \cdot c_n, N'_2 - n_2 \cdot c_n\} - M') \cdot p_1 - M' \cdot f_4] - \vec{F} \cdot \vec{C}^* \\ s.t. \quad & \begin{cases} M' = [\sum_{k=\max\{n_1, n_2\}}^{n_1+n_2} k \cdot P_k + (\min\{N'_1 - n_1 \cdot c_1, N'_2 - n_2 \cdot c_2\} - k) \cdot \mu_3] \cdot c_3 \\ n_1 = N_1 \cdot \mu_1 + \sum_{j=\frac{k}{2}}^k P_{i,j} \quad n_2 = N_2 \cdot \mu_2 + \sum_{i=\frac{k}{2}}^k P_{i,j} \\ N'_1 = N_1 + M \quad N'_2 = N_2 + M \\ \vec{C}^* = (c_1, c_2, \dots, c_{4t}) \\ \vec{F}^* = [N_1 \cdot f_1, N_2 \cdot f_2, \min\{N_1, N_2\} \cdot f_3, M \cdot f_5, N'_1 \cdot f_1, N'_2 \cdot f_2, \min\{N'_1, N'_2\} \cdot f_3, M' \cdot f_5, \dots] \\ P_k = \sum_{i=k-j}^k \sum_{j=\frac{k}{2}}^k C_{N_1 \cdot \mu_1}^i C_{N_2 \cdot \mu_2}^j \mu_1^i \mu_2^j (1 - \mu_1)^{k-i} (1 - \mu_2)^{k-j} \\ P_{i,j} = C_{N_1 \cdot \mu_1}^i C_{N_2 \cdot \mu_2}^j \mu_1^i \mu_2^j (1 - \mu_1)^{k-i} (1 - \mu_2)^{k-j} \end{cases} \end{aligned}$$

5.2 模型的求解

我们拟应用微生物遗传算法进行不断的迭代优化来寻找最优解，算法流程：

Algorithm 2: 微生物遗传算法 (MGA) 流程

Input: 每个段产生成本、检测成本、不合格品调换拆解费用和次品率 μ

Output: 单周期的成品收益 $Profit$ 最优的决策向量 \vec{C}^*

```

1 初始化一个周期为  $t$  轮次的决策列向量  $C^*$  的种群，种群中的个体是随机生成的；
2 while TRUE do
3   从种群中随机选择两个个体  $\vec{C}_A^*$  和  $\vec{C}_B^*$ ；
4   计算两个个体的  $Profit$ ；
5   if  $Profit_A \leq Profit_B$  then
6     winner  $\leftarrow \vec{C}_B^*$ 
7     loser  $\leftarrow \vec{C}_A^*$ 
8   end
9   else
10    winner  $\leftarrow \vec{C}_A^*$ 
11    loser  $\leftarrow \vec{C}_B^*$ 
12  end
13  for 每个在 loser 的向量  $g$  do
14    if 随机数  $r < crossoverRate$  then
15      随机选择 winner 中的一个向量  $h$ ；
16      对  $g$  和  $h$  进行交叉操作；
17    end
18  end
19 end
20 return 最优的决策向量  $\vec{C}^*$ 

```

六、问题三的模型建立与求解

6.1 模型的建立

6.2 模型的求解

七、问题四的模型建立与求解

7.1 模型的建立

本问题中我们需要考虑到抽样检测的准确性和应对风险的鲁棒性，而不能直接获取真实次品率：相比于对每个阶段的抽样进行风险建模，一个更为直观的想法是对每一步的预测次品率 μ 进行概率建模，这样做的好处是便于拟合我们在工序每阶段的概率分布（这可以被看作是相互独立事件间的概率分布的传播）——我们选用高斯分布拟合之。事实上，如果我们假定阶段 p 的期望为问题二及三中给定的次品率，那么：

$$z_p = \mu_p + \sigma_p \odot \epsilon, \epsilon \sim \mathcal{N}(0, \mathbf{I}) \quad (15)$$

对于每次评估的 μ_p ，我们只需计算出相应的风险参数 σ_p ，就可以重新从标准高斯分布中采样出噪声施加在其上：

$$z \sim \mathcal{N}(z; \mu_p, \sigma_p^2 \mathbf{I}) \quad (16)$$

这样，实际的工序流程中次品率就被建模成了高斯分布，在优化过程中从中采样即可。接下来，我们论证该分布在工序流程中的可传播性。事实上，如果我们把零配件的组装和成品看作一个马尔可夫过程，那么，概率分布的传播就可以视为一个前向动力学模型，针对任意的相邻两个工序（我们暂且搁置对于一般的偏置项的考虑）：

$$\begin{aligned} P(z_{p+1}|z_p) &= \mathcal{N}(z_{p+1}; z_p, \sigma_p^2 \mathbf{I}) \\ &= \mu_{p+1} z_p + \sigma_{p+1} \odot \epsilon \\ &= \mu_{p+1}(\mu_p + \sigma_p \odot \epsilon) + \sigma_{p+1} \odot \epsilon \\ &= \mu_{p+1} \mu_p + (\mu_{p+1} \sigma_p + \sigma_{p+1}) \odot \epsilon \end{aligned} \quad (17)$$

而又根据独立高斯分布的可加性： $\mathcal{N}(0, \sigma_1^2 \mathbf{I}) + \mathcal{N}(0, \sigma_2^2 \mathbf{I}) \sim \mathcal{N}(0, (\sigma_1^2 + \sigma_2^2) \mathbf{I})$ 我们便可以得到 z_{p+1} 的分布：

$$z_{p+1} \sim \mathcal{N}(z_{p+1}; \mu_{p+1} \mu_p, \sqrt{(\mu_{p+1} \sigma_p)^2 + \sigma_{p+1}^2} \mathbf{I}) \quad (18)$$

于是，我们的风险就可以通过前向动力学模型来传播了， μ 的期望依然遵循于预先设定的标准，而风险可以通过标准差逐层传递。

7.2 模型的求解