

计数集团检查的抽查特征函数*

潘沈元 杨忠航 吕小军

(徐州师范大学生命科学学院, 徐州, 221116)

摘要

本文给出计数型集团检查一次、二次抽样方案的抽查特征函数及其算法, 并且通过与近似算法的比较和计算机抽样检查模拟, 验证了该方法的正确性, 并对该抽查特征函数性质和应用作了初步讨论.

关键词: 计数抽样检查, 集团检查, 接收概率, OC函数, 抽样方案.

学科分类号: O212.2.

§1. 引言

工业产品抽样检查的抽查特征函数(Operating-Characteristic Function, 简称OC函数)又称接收概率函数, 是抽样检查方案制定的主要依据. 根据OC函数绘制的抽查特征曲线(Operating-Characteristic Curve, 简称OC曲线)是评价某种抽样方案是否符合检验要求的唯一标准^{[1][2]}. 在工业产品的检查中, 一般OC函数的概率模型是超几何概率函数, 但在家蚕微粒子病检疫中采用的是集团母蛾抽样检查, 农业部部颁标准^{[3][4]}规定母蛾带毒率为0.5%的检查批, 其接收概率必须在1.5%以下, 为减少检查次数, 生产上规定以30只母蛾为1个集团, 混合碾磨后作为一个基本检查单位, 当带毒集团数超过某一指标后判该批蚕种不合格. 虽然这种集团抽样在蚕种生产上广泛采用, 但一直没有得到这种抽检方案的OC函数, 给抽样方案的制定带来了很大困难和误差^[5-8]. 作者在前文中提出了这种集团母蛾抽样检查的接收概率近似算法^[9]并应用于家蚕微粒子病检疫抽样方案的制定, 取得较好效果^{[10][11]}, 但对这种集团抽样检查未能有精确的OC函数, 总感到遗憾, 也不能对这种集团抽样检查进一步进行理论研究. 本文给出计数集团检查抽查特征函数的理论概率模型, 并通过计算机模拟验证.

§2. OC函数的导出

设检查批(总体)大小为 Nm , 其中 N 为批内集团数, m 为每个集团内的个体数, 批内带毒个体数为 D , 则批内带毒率为 $p = D/Nm$.

定义 2.1 1个集团内的 m 个个体全部合格为合格集团, 1个集团内的 m 个个体至少有1个带毒为带毒集团.

*国家蚕丝绸发展风险基金、国革协办(2002-36)资助.
本文2006年7月3日收到.

对于计数集团检查一次抽样方案 (N, n, m, A) 可表述为: 从检查批中随机抽取 nm 个个体, 组成 n 个集团, 其中带毒个体数为 d , 带毒集团数为 x , 当 x 小于等于合格判定标准 A , 则判该批产品为合格, 否则判为不合格.

对于计数集团检查二次抽样方案 $(N, n_1, n_2, m, A_1, A_2, R_1)$ 可表述为: 从检查批中随机抽取 n_1m 个个体, 组成 n_1 个集团, 其中带毒个体数为 d_1 , 带毒集团数为 x_1 , 当 x_1 小于等于合格判定标准 A_1 , 则判该批产品为合格, 当 x_1 大于等于不合格判定标准 R_1 , 则判该批产品为不合格, 否则继续随机抽取 n_2m 个个体, 组成 n_2 个集团, 其中带毒个体数为 d_2 , 带毒集团数为 x_2 , 当 $x_1 + x_2$ 小于等于合格判定标准 A_2 , 则判该批产品为合格, 否则判该批产品为不合格.

计数集团抽样检查的OC函数是指当检查批的批毒率为 p 时, 按照上述抽样方案对该批产品做检查, 接收该批产品的概率函数. 其推导过程如下.

规定当 $a < b$ 时, 组合数 $C_a^b = 0$.

从容量为 Nm 的总体中抽取容量为 nm 的样本, 随机组合成 n 个集团, 其中带毒个体数 d 服从超几何分布, 概率函数为:

$$h(d|Nm, D, nm) = \frac{C_D^d C_{Nm-D}^{nm-d}}{C_{Nm}^{nm}}, \quad 0 \leq d \leq D. \quad (2.1)$$

n 个集团中有 x 个带毒集团时, n 个集团中带毒个体数 d 可能是 $x \leq d \leq \min(D, xm)$ 个. 当 n 个集团中带毒个体数为 d 时, 带毒集团数为 x 的概率是:

$$s(x|n, m, d) = \frac{C_n^x \sum_{k=0}^x (-1)^k C_x^k C_{(x-k)m}^d}{C_{nm}^d}, \quad x \leq d \leq \min(D, xm). \quad (2.2)$$

说明: C_{nm}^d 为从 n 个集团的 nm 个座位中选出 d 个座位放置带毒个体的方法数; C_n^x 为从 n 个集团中选出 x 个集团的方法数; 从 x 个集团的 xm 个座位中选出 d 个座位放置带毒个体的方法数为 C_{xm}^d , 其中至少有一个集团不含带毒个体的方法数为: $\sum_{k=1}^x (-1)^{k-1} C_x^k C_{(x-k)m}^d$, 所以, x 个集团全部含有带毒个体的方法数为:

$$C_{xm}^d - \sum_{k=1}^x (-1)^{k-1} C_x^k C_{(x-k)m}^d = \sum_{k=0}^x (-1)^k C_x^k C_{(x-k)m}^d,$$

n 个集团中有 x 个集团含有带毒个体的方法数为: $C_n^x \sum_{k=0}^x (-1)^k C_x^k C_{(x-k)m}^d$.

于是, 从总体的 Nm 个个体中, 随机抽取 nm 个样本, 组成 n 个集团, 其中带毒集团数为 x 的概率可由全概率公式得出:

$$\begin{aligned} u(x|N, D, n, m) &= \sum_{d=x}^{\min(D, xm)} h(d|Nm, D, nm) s(x|n, m, d) \\ &= \sum_{d=x}^{\min(D, xm)} \frac{C_D^d C_{Nm-D}^{nm-d} C_n^x \sum_{k=0}^x (-1)^k C_x^k C_{(x-k)m}^d}{C_{Nm}^{nm} C_{nm}^d}. \end{aligned} \quad (2.3)$$

又由Bayes定理和(2.1)、(2.2)得, 在带毒集团数为 x 的条件下, x 个集团中带毒个体数为 d 的概率:

$$t(d|n, m, x) = \frac{h(d|Nm, D, nm)s(x|n, m, d)}{\sum_{k=x}^{\min(D, xm)} h(k|Nm, D, nm)s(x|n, m, k)}, \quad x \leq d \leq \min(D, xm). \quad (2.4)$$

根据OC函数的定义和上面的结论, 一次抽样方案 (N, n, m, A) 的OC函数为:

$$L_1(A|N, D, n, m) = \sum_{x=0}^A u(x|N, D, n, m). \quad (2.5)$$

二次抽样方案 $(N, n_1, n_2, m, A_1, A_2, R_1)$ 的OC函数为:

$$\begin{aligned} & L_2(A_1, A_2, R_1|N, D, n_1, n_2, m) \\ &= \sum_{x_1=0}^{A_1} u(x_1|N, D, n_1, m) + \sum_{x_1=A_1+1}^{R_1-1} \left\{ u(x_1|N, D, n_1, m) \right. \\ & \quad \left. \cdot \sum_{x_2=0}^{A_2-x_1} \left[\sum_{d_1=x_1}^{\min(D, x_1 m)} t(d_1|n, m, x_1) u(x_2|N - n_1, D - d_1, n_2, m) \right] \right\}. \end{aligned} \quad (2.6)$$

二次抽样方案的平均检查次数为:

$$ASN(A_1, A_2, R_1|N, D, n_1, n_2, m) = n_1 + n_2 \sum_{x_1=A_1+1}^{R_1-1} u(x_1|N, D, n_1, m). \quad (2.7)$$

(2.6)式虽然是二次抽样方案OC函数的精确表达式, 但不易用计算机实现其算法, 更不易实现多次抽样方案的计算机算法, 这里给出二次抽样方案OC函数的近似公式:

$$\begin{aligned} & L_2(A_1, A_2, R_1|N, D, n_1, n_2, m) \\ &= \sum_{x_1=0}^{A_1} u(x_1|N, D, n_1, m) \\ & \quad + \sum_{x_1=A_1+1}^{R_1-1} \left[u(x_1|N, D, n_1, m) \sum_{x_2=0}^{A_2-x_1} u(x_2|N - n_1, D - \bar{d}_1, n_2, m) \right]. \end{aligned} \quad (2.8)$$

(2.8)式中 \bar{d}_1 为第1次抽验的平均带毒个体数, 在第1检带毒集团数为 x_1 的情况下, 带毒个体数 d_1 不确定, 但有 $x_1 \leq d_1 \leq \min(x_1 m, D)$, 根据超几何分布的数学期望公式 $\mu = nmD/(Nm)$, d_1 的平均数可估计为:

$$\bar{d}_1 = x_1 + (x_1 m - x_1) \cdot \frac{D - x_1}{Nm} = x_1 \left[1 + \frac{(m-1)(D-x_1)}{Nm} \right]. \quad (2.9)$$

§3. 接收概率的递推算法

事实上, 直接应用以上导出的(2.3) (2.5) (2.8)式, 并不适合于接收概率的计算, 因为当总体很大、样本数很多时, 计算组合数时将产生计算机溢出. 以下给出接收概率的递推公式及计算机算法.

由(2.2)式得

$$s(x|n, m, d) = \sum_{k=0}^x (-1)^k \cdot \frac{C_n^x C_x^k C_{(x-k)m}^d}{C_{nm}^d}. \quad (3.1)$$

令

$$g(k|n, m, d, x) = \frac{C_n^x C_x^k C_{(x-k)m}^d}{C_{nm}^d}, \quad (3.2)$$

则

$$s(x|n, m, d) = \sum_{k=0}^x (-1)^k g(k|n, m, d, x). \quad (3.3)$$

代入(2.4)式得

$$u(x|N, D, n, m) = \sum_{d=x}^{\min(D, xm)} \left[h(d|Nm, D, nm) \sum_{k=0}^x (-1)^k g(k|n, m, d, x) \right]. \quad (3.4)$$

对于式中的超几何分布的概率函数 $h(d|N, D, n)$, 已有如下递推公式^[10]

$$h(0|N, D, n) = \prod_{i=0}^{n-1} \frac{N-D-i}{N-i} = \prod_{i=0}^{n-1} \left(1 - \frac{D}{N-i} \right), \quad (3.5)$$

$$h(i+1|N, D, n) = \frac{(D-i)(n-i)}{(i+1)(N-D-n+i+1)} h(i|N, D, n), \quad i = 0, 1, 2, \dots, n-1, \quad (3.6)$$

$$h(0|N, D, n+1) = \left(1 - \frac{D}{N-n} \right) h(0|N, D, n), \quad n = 1, 2, 3, \dots, N-D. \quad (3.7)$$

而对于(3.2)式,

$$g(k|n, m, d, x) = \frac{n(n-1) \cdots (n-x+1) \cdot [(x-k)m][(x-k)m-1] \cdots [(x-k)m-d+1]}{k!(x-k)!nm(nm-1) \cdots (nm-d+1)}. \quad (3.8)$$

这里规定当 $a > b$ 时, $\prod_a^b x = 1$, 于是有下列递推公式:

$$\begin{aligned} g(k|n, m, x, x) &= \frac{n(n-1) \cdots (n-x+1) \cdot [(x-k)m][(x-k)m-1] \cdots [(x-k)m-x+1]}{k!(x-k)!nm(nm-1) \cdots (nm-x+1)} \\ &= \prod_{i=0}^{k-1} \frac{(n-i)[(x-k)m-i]}{(k-i)(nm-i)} \prod_{i=k}^{x-1} \frac{(n-i)[(x-k)m-i]}{(x-i)(nm-i)}, \end{aligned} \quad (3.9)$$

$$g(0|n, m, 0, 0) = 1, \quad (3.10)$$

$$\begin{aligned} g(0|n, m, x, x) &= \frac{n(n-1) \cdots (n-x+1) \cdot xm(xm-1) \cdots (xm-x+1)}{x!nm(nm-1) \cdots (nm-x+1)} \\ &= \prod_{i=0}^{x-1} \frac{(n-i)(xm-i)}{(x-i)(nm-i)}, \end{aligned} \quad (3.11)$$

$$\begin{aligned} g(k|n, m, d+1, x) &= \frac{n(n-1) \cdots (n-x+1) \cdot [(x-k)m][(x-k)m-1] \cdots [(x-k)m-d]}{k!(x-k)!nm(nm-1) \cdots (nm-d)} \\ &= \frac{(x-k)m-d}{nm-d} g(k|n, m, d, x). \end{aligned} \quad (3.12)$$

《应用概率统计》版权所有

有下面给出(3.4)式的计算机算法:

(a) 由(3.5) (3.6) (3.7)计算 $h(d|Nm, D, nm)$, $d = 0, 1, 2, \dots, \min(D, xm)$, 结果存于数组中;

(b) 由(3.9)计算 $d = x$ 时的 $g(k|n, m, d, x)$, $k = 0, 1, 2, \dots, x$, 结果存于数组中;

(c) 计算 $h(d|Nm, D, nm) \sum_{k=0}^x (-1)^k g(k|n, m, d, x)$, 结果存于累加器中;

(d) 若 $d > \min(D, xm)$, 结束; 否则 $d = d + 1$, 由(3.12)计算 $g(k|n, m, d, x)$, $k = 0, 1, 2, \dots, x$, 转(c);

将(3.4)式分别代入(2.5)和(2.8), 可得一次抽样、二次抽样的接收概率.

§4. 计算机模拟验证

为了验证本文提出的集团检查OC函数的正确性, 这里以家蚕微粒子病母蛾集团检查为例, 分别采用本文提出的OC函数(2.5)式和(2.8)式—方法1和前文作者提出的近似算法^[9]—方法2, 计算集团检查一次抽样方案和二次抽样方案的接收概率, 与计算机模拟方法—方法3得出的结果相比较. 模拟算法如下:

首先设计随机函数产生随机数, 按检查批毒率 p 产生带毒个体数为 D 、批量为 Nm 的检查批, 再利用随机函数在检查批中抽取容量为 n_1m 和 n_2m 的样本, 根据一次、二次抽样方案和模拟产生的带毒集团数, 判断检查批是否合格, 记录判为合格的次数, 如此重复模拟1000次, 统计计算接收频率. 因为这里计算的接收频率是随机变量, 所以进一步以这1000次模拟为单位, 重复计算统计接收频率61次, 计算接收频率的平均数、标准差和置信度为99%的置信区间, 作为对接收概率的估计.

数据分析计算在Visual C++ 6.0 for Windows环境下编程实现^[12]. 一次抽检方案的分析结果如表1, 二次抽检方案的分析结果如表2所示, 根据一次抽样方案(5000, 280, 20, 16)和二次抽检方案(6000, 110, 110, 30, 5, 19, 14)采用三种方法绘制的OC曲线见图1和图2 (计算机模拟次数10000次).

表1 集团检查一次抽样方案接收概率算法的比较

一次抽样方案 (N, n, m, A)	批毒率 ($p = D/Nm$)	接收概率(%)		
		方法1	方法2	方法3
(5000, 280, 20, 16)	0.002	95.2985	95.1676	95.35 [95.10, 95.60]
	0.005	1.23345	1.22402	1.244 [1.122, 1.366]
(5000, 200, 30, 17)	0.002	95.7655	95.7135	95.92 [95.70, 96.12]
	0.005	1.14963	1.12042	1.200 [1.074, 1.326]
(5000, 150, 40, 17)	0.002	96.1816	96.1761	96.24 [96.04, 96.45]
	0.005	1.44729	1.42480	1.454 [1.323, 1.586]

注: 方法3中方括号[]内是置信度为99%的置信区间, 括号前是平均数, 下同.

表2 集团检查二次抽样方案接收概率算法的比较

二次抽样方案 ($N, n_1, n_2, m, A_1, A_2, R_1$)	批毒率 ($p = D/Nm$)	接收概率(%)		
		方法1	方法2	方法3
(6000, 150, 150, 20, 5, 17, 13)	0.002	95.3060	95.2480	95.31 [95.08, 95.55]
	0.005	1.10114	1.07295	1.118 [0.988, 1.248]
(6000, 110, 110, 30, 5, 19, 14)	0.002	96.8055	96.7790	96.68 [96.47, 96.90]
	0.005	1.05068	1.03393	1.043 [0.931, 1.154]
(6000, 80, 80, 40, 5, 18, 12)	0.002	95.4209	95.3954	95.27 [95.03, 95.50]
	0.005	1.12666	1.10174	1.152 [1.046, 1.258]

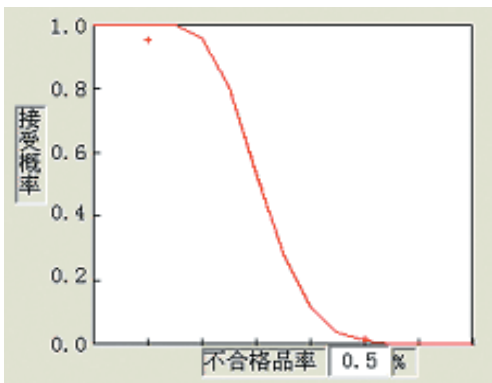


图1 三种方法绘制的一次方案OC曲线

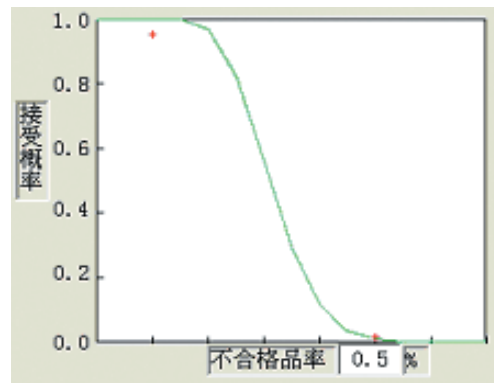


图2 三种方法绘制的二次方案OC曲线

从表1和表2可以看到, 方法1和方法2计算的接收概率与计算机模拟的结果已相当一致, 且都在一个较小的置信区间内, OC曲线基本重叠, 说明了本文提出的OC函数和前文^[9]所提出的接收概率近似算法的正确性和一致性. 应该指出, 尽管式(2.1)–(2.6)给出了计数集团抽样检查OC函数的精确公式, 但计算机的计算结果不免要出现舍入误差.

§5. 应用讨论

关于计数集团抽样检查OC函数在抽样方案设计中的意义, 已经在前文^[9]中作了介绍, 在此不再赘述. 本文给出的OC函数是计数集团抽样检查的OC函数, 是集团抽样检查的理论概率模型, 其导出过程也解决了某些复杂概率问题的求解, 这些概率模型的性质、特征还有待于进一步研究. 相对于前文^[9]提出的接收概率近似算法, 本文提出的OC函数的计算方法较为复杂, 计算时间较长, 当总体很大, 集团内个体数很多时, 会产生较大的舍入误差, 此时的计算精度可能反而不如近似算法. 从本文和前文的计算结果比较来看, 前文提出的近似算法完全可以达到集团抽样方案设计的要求, 较之工业产品抽检方案设计中, 接受概率计算用Poisson分布或二项分布代替超几何分布所产生的误差^{[1][2]}将远大于作者提出的两

《应用概率统计》版权所有

种概率模型. 虽然接收概率的近似算法已经能达到计数集团抽检方案设计的要求, 但是本文提出的概率模型无疑将为进一步探讨集团抽样检查的特性, 探讨具有较高计算精度的近似算法提供理论依据.

参 考 文 献

- [1] Grant, E.L. and Leavenworth, R.S., *Statistical Quality Control*, McGraw-Hill Book, Company, 1980.
- [2] Hald, A., *Statistical Theory of Sampling Inspection by Attribute*, London: Academic Press, 1981.
- [3] 中华人民共和国农业部, 桑蚕一代杂交种, NY/T326-1997.
- [4] 中华人民共和国农业部, 桑蚕一代杂交种检验规程, NY/T327-1997.
- [5] 藤原公, 家蚕微粒子病の集团蛾検査法に関する研究, 蚕丝试验场汇报, **120**(1984), 146-159.
- [6] 栗栖弋彦ほか, 普通蚕種の母蛾検査に関する研究(1)モンテカルロ法による電算機シミュレーション, 京都工艺纤维大学纤维学报, **10**(2)(1983), 151-162.
- [7] 杨明观等, 集团母蛾检查抽样方案的改进, 蚕桑通报, **(3)**(1986), 4-9.
- [8] 杨义群等, 稀有事件的简捷集团抽检法及其应用, 浙江农业大学学报, **16**(4)(1990), 436-440.
- [9] 潘沈元等, 记数集团抽样检查接收概率的近似算法及其应用, 应用概率统计, **22**(4)(2006), 438-442.
- [10] 潘沈元等, 增加家蚕微粒子病母蛾抽样检查集团蛾数的探讨, 蚕业科学, **30**(1)(2004), 64-68.
- [11] 潘沈元等, 家蚕微粒子病母蛾检查五次抽样方案研究, 蚕业科学, **29**(4)(2004), 407-411.
- [12] 潘沈元等, 记数抽样检查方案设计与分析软件研制, 数理统计与管理, **22**(5)(2003), 43-47.

The Operating Characteristic Functions for Sampling Inspection by Attributes Based on the Grouping Method

PAN SHENYUAN YANG ZHONGHANG LU XIAOJUN

(School of Life Science, Xuzhou Normal University, Xuzhou, 221116)

Several operating characteristic (OC) functions and the according algorithm for single and double sampling inspection by attributes based on the grouping method are provided in this paper. The validity of the OC functions is proved by comparing with the foregoing approximate algorithm and computer simulation. The property and application of the OC functions are simply discussed at last.

Keywords: Sampling inspection by attributes, grouping method, probability of acceptance, OC function, sampling plan.

AMS Subject Classification: 62D05, 62P10.