

多档次试验结果下的一种罐子模型 *

陈桂景¹ 胡舒合¹ 洪圣岩^{2,1}

(¹ 安徽大学数学与计算科学学院, 合肥, 230039)

(² Lilly Research Laboratories, Eli Lilly and Company, Indianapolis, IN 46285, USA)

摘要

针对多档次试验结果的情况, 本文构造了一种罐子模型, 并在此模型中建立了极限理论。理论和例子表明, 这种临床试验设计, 既符合人道精神, 又保持较好的统计效率。因此该文有应用参考价值。

关键词: 自适应设计, 罐子模型, 极限定理, 渐近正态性。

学科分类号: O212.7.

§1. 引言

二战以来, 随着各种新的药品和治疗方法的不断涌现, 临床试验设计越来越受到重视, 现在一些国家设立了以统计人才为主而构成的专门机构, 用以来对临床试验设计工作进行培训, 指导, 鉴定和研究。传统的临床试验设计是随机化 50-50 设计, 这种设计的优点是操作简便, 且当各治疗方案成功率比较接近时, 也具有较好的统计效率。但是, 当各治疗方案治愈率相差较大, 存在有的处理对病员有严重负面效果时, 则将有近半数病员受到这种处理的损害, 此时这样的设计是不人道的(见 Connor (1994) 的例子)。因此, Robbins (1952) 提出了自适应设计的思想。所谓自适应设计是, 根据前面的试验结果, 逐步修正后面的设计, 以期使较好的治疗能以较大的机会分配给病员, 而使较差的治疗减少分配病员的机会, 为此, Zelen (1969), Wei and Durhan (1978), Athreya and Karlin (1968) 先后提出了“胜者优先”(Play-the-Winner Rule) 设计, “随机化胜者优先”设计以及广义 Friedman 罐子模型设计。关于罐子模型, 文献中已很多工作, 比如可参看文献 [1], [2], [3], [4], [7], [9] 等。

但是文献中考虑的以上这些模型, 大都考虑的是两档次结果(成功与失败)的情况。显然, 在临床试验中, 多档次试验结果甚至连续量的试验结果也是常见的。对于这种情况的设计, 文献中也有些研究, 比如 Rosenberger (1993) 利用非参数秩统计方法提出了一些设计。但可惜的是, 在那种设计中渐近理论尚未能完全建立, 至今未能解决(见 [8])。本文要利用罐子模型, 为多档次结果的临床试验提出了一种设计, 并建立渐近定理。理论和数值例子表明, 这种设计更符合人道精神, 仍保持了较高的统计效率, 因此文中方法可作为应用参考。

* 国家自然科学基金资助项目(项目编号 10571001)。

本文 2006 年 3 月 22 日收到。

§ 2. 罐子模型

为简明计,本文只考虑两个处理(分别记为 A, B)试验问题,所得结果不难推广到多处理试验中去,假定临床试验结果可划分为 $2k$ 个档次: $F_k, F_{k-1}, \dots, F_1, S_1, \dots, S_k$. 其中 F_1, F_2, \dots, F_k 表示负面结果, 负面程度由轻到重, 而 S_1, S_2, \dots, S_k 表示正面结果, 正面程度也是由轻到重. 显然对于具有连续量试验结果的情况, 通过离散化, 也可转变为这种多档次情况. 对病员有人文关怀的设计应是这样, 当处理 A 的试验结果为 F_i ($i = 1, \dots, k$) 时, 下一步则应减少处理 A 的试验机会, 同时增加处理 B 的机会; 当处理 A 的试验结果为 S_i ($i = 1, \dots, k$) 时, 下一步则应增加 A 的机会, 而减少 B 的机会. 而且应随着负面(正面)结果程度的增加, 相应的减少(增加)机会的力度也应增大. 对于处理 B 的结果也应对称地设计.

为此, 取定 $2k$ 个正数 $0 < \beta_k < \beta_{k-1} < \dots < \beta_1 < 1/2 < \alpha_1 < \dots < \alpha_k$, 并约定 $\alpha_i + \beta_i = 1$, $i = 1, \dots, k$. 预先在罐子中放入 y_{01} 个黑球和 y_{02} 个白球, 其中黑球对应处理 A, 白球对应 B. 为简单且不失本质, 可假定 $y_{01} = y_{02} = 1/2$. 试验开始时, 随机地从罐子中抽取一个球, 若抽得黑球, 则对进入试验的病员施加处理 A, 这试得结果如果是 S_i , 则在罐中同时添加 α_i 个黑球和 β_i 个白球; 若试得结果为 F_i 时, 则在罐中添加 β_i 个黑球和 α_i 个白球; 反之若抽得白球, 则施加处理 B, 若试得结果是 S_i , 则加入 α_i 个白球和 β_i 个黑球, 若试得结果是 F_i , 加入 β_i 个白球和 α_i 个黑球, $i = 1, \dots, k$.

这样的试验设计可以重复地, 递推地进行, 到第 n 步时, 记罐子中成份为 $Y_n = (y_{n1}, y_{n2})'$, 其中 y_{n1}, y_{n2} 分别表示黑球与白球的个数. 现用 $X^{(1)}, X^{(2)}$ 分别表示在 A, B 处理下治疗结果. 记

$$p_{ij} = P\{X^{(i)} = S_j\}, \quad q_{ij} = P\{X^{(i)} = F_j\},$$

$i = 1, 2, j = 1, \dots, k$. 那么 $p_1 = \sum_{j=1}^k p_{1j}$, $p_2 = \sum_{j=1}^k p_{2j}$ 分别表示 A, B 的成功率. 于是, 上述罐子模型可以表示成如下递推形式:

$$Y_i = \left(I + \frac{1}{i} H\right) Y_{i-1} + Q_i, \quad i = 1, 2, \dots \quad (1)$$

其中

$$H = \begin{bmatrix} \sum_{i=1}^k (p_{1i}\alpha_i + q_{1i}\beta_i) & \sum_{i=1}^k (p_{2i}\beta_i + q_{2i}\alpha_i) \\ \sum_{i=1}^k (p_{1i}\beta_i + q_{1i}\alpha_i) & \sum_{i=1}^k (p_{2i}\alpha_i + q_{2i}\beta_i) \end{bmatrix},$$

它是模型的生成矩阵.

$$Q_i = (Y_i - Y_{i-1}) - E\{(Y_i - Y_{i-1}) | \mathcal{F}_{i-1}\}, \quad i = 1, 2, \dots$$

是一个二维的鞅差序列, 相应的 σ -域序列为 $\mathcal{F}_0, \mathcal{F}_1, \mathcal{F}_2, \dots, \mathcal{F}_n, \dots$. \mathcal{F}_0 是平凡 σ -域, $\mathcal{F}_i = \sigma\{Y_0, Y_1, \dots, Y_i\}$, $i = 1, 2, \dots$. 矩阵 H 有下面谱分解:

$$THT^{-1} = \begin{pmatrix} 1 & 0 \\ 0 & \lambda_2 \end{pmatrix} \cong J,$$

其中

$$T = \begin{pmatrix} 1 & 1 \\ -a_2 & a_1 \end{pmatrix}, \quad T^{-1} = \begin{pmatrix} a_1 & -1 \\ a_2 & 1 \end{pmatrix}, \quad (2)$$

$$\lambda_2 = \sum_{i=1}^k [\alpha_i(p_{1i} - q_{2i}) + \beta_i(q_{1i} - p_{2i})], \quad (3)$$

$$a_1 = \frac{\sum_{i=1}^k (p_{2i}\beta_i + q_{2i}\alpha_i)}{\sum_{i=1}^k [\beta_i(p_{1i} + p_{2i}) + \alpha_i(q_{1i} + q_{2i})]}, \quad a_2 = \frac{\sum_{i=1}^k (p_{1i}\beta_i + q_{1i}\alpha_i)}{\sum_{i=1}^k [\beta_i(p_{1i} + p_{2i}) + \alpha_i(q_{1i} + q_{2i})]}. \quad (4)$$

§ 3. 极限定理

记 $a_n = (a_{n1}, a_{n2})' \equiv Y_n/n$, a_{n1}, a_{n2} 分别表示第 n 步时罐内两种球所占的比例

引理 1 当 $n \rightarrow \infty$ 时, 有

$$a_n \rightarrow a, \quad \text{a.s.}, \quad (5)$$

其中 $a' = (a_1, a_2)$ 由 (4) 式定义.

证明: 记 $M_i = I + (1/i)H$, 由模型 (1), 用递推法, 可得

$$Y_i = M_i M_{i-1} \cdots M_1 Y_0 + \sum_{j=1}^i B_{i,j} Q_j. \quad (6)$$

其中记 $B_{i,j} = M_i \cdots M_{j+1}$, $B_{i,i} = I$, 从而有

$$\mathbb{E}(Y_i) = M_i M_{i-1} \cdots M_1 Y_0, \quad Y_i = \mathbb{E}(Y_i) + \sum_{j=1}^i B_{i,j} Q_j.$$

由此, 为证引理仅需证明

$$\frac{1}{n} \mathbb{E}(Y_n) \rightarrow a \quad (n \rightarrow \infty), \quad (7)$$

$$\frac{1}{n} \sum_{j=1}^n B_{n,j} Q_j \rightarrow 0, \quad \text{a.s.} \quad (n \rightarrow \infty). \quad (8)$$

因为

$$\lambda_2 < 1, \quad \frac{1}{n} \prod_{j=1}^n \left(1 + \frac{1}{j}\right) \rightarrow 1, \quad \frac{1}{n} \prod_{j=1}^n \left(1 + \frac{\lambda_2}{j}\right) \rightarrow 0.$$

故有: 当 $n \rightarrow \infty$ 时,

$$\begin{aligned} \frac{1}{n} T \mathbb{E}(Y_n) &= \frac{1}{n} T M_n T^{-1} T M_{n-1} T^{-1} \cdots T M_1 T^{-1} T Y_0 \\ &= \frac{1}{n} \prod_{j=1}^n \begin{bmatrix} 1 + \frac{1}{j} & 0 \\ 0 & 1 + \frac{\lambda_2}{j} \end{bmatrix} \begin{pmatrix} 1 \\ * \end{pmatrix} \rightarrow \begin{pmatrix} 1 \\ 0 \end{pmatrix}. \end{aligned}$$

于是有 $\frac{1}{n} \mathbb{E}(Y_n) \rightarrow T^{-1} \begin{pmatrix} 1 \\ 0 \end{pmatrix} = a$, 即 (7) 式成立.

注意 $TQ_j Q'_j T' = \begin{pmatrix} 0 & 0 \\ 0 & * \end{pmatrix}$, 那么有

$$\begin{aligned} & \left(\frac{1}{n} T \sum_{j=1}^n B_{n,j} Q_j \right) \left(\frac{1}{n} T \sum_{j=1}^n B_{n,j} Q_j \right)' \\ &= \frac{1}{n^2} \sum_{j=1}^n \left\{ \begin{pmatrix} 1 + \frac{1}{n} & 0 \\ 0 & 1 + \frac{\lambda_2}{n} \end{pmatrix} \cdots \begin{pmatrix} 1 + \frac{1}{j+1} & 0 \\ 0 & 1 + \frac{\lambda_2}{j+1} \end{pmatrix} \right. \\ &\quad \cdot \left. \begin{pmatrix} 0 & 0 \\ 0 & * \end{pmatrix} \begin{pmatrix} 1 + \frac{1}{j+1} & 0 \\ 0 & 1 + \frac{\lambda_2}{j+1} \end{pmatrix} \cdots \begin{pmatrix} 1 + \frac{1}{n} & 0 \\ 0 & 1 + \frac{\lambda_2}{n} \end{pmatrix} \right\} \\ &= \begin{pmatrix} 0 & 0 \\ 0 & \frac{1}{n^2} \sum_{j=1}^n \left(\frac{n}{j}\right)^{2\lambda_2} \end{pmatrix}. \end{aligned}$$

记 $\sigma_n^2 = (1/n^2) \sum_{j=1}^n (n/j)^{2\lambda_2}$, 那么由有界鞅差序列的 Hoeffding 不等式 (见 [5], P135, 定理 9.1) 对任 $\varepsilon > 0$

$$\sum_{n=1}^{\infty} \mathbb{P} \left\{ \left| \frac{1}{n} T \sum_{j=1}^n B_{n,j} Q_j \right| > \varepsilon \right\} \leq \sum_{n=1}^{\infty} 2e^{-2\varepsilon^2/\sigma_n^2} < \infty,$$

从而由 Borel-Cantelli 引理知, $(1/n)T \sum_{j=1}^n B_{n,j} Q_j \rightarrow 0$, a.s., 于是 (8) 式成立, 引理 1 得证. #

令 $\zeta_{1i} = 1$, 如果第 i 步时抽得黑球; $= 0$, 第 i 步时抽得白球, $\zeta_{2i} = 1 - \zeta_{1i}$, $i = 1, 2, \dots$. 并记

$$M_{n1} = \sum_{i=1}^n \zeta_{1i}, \quad M_{n2} = n - M_{n1},$$

则它们分别表示到 n 步为止, A, B 两种处理试验的次数, 令 $M_n = (M_{n1}, M_{n2})'$.

定理 1 当 $n \rightarrow \infty$ 时, 有

$$\frac{M_n}{n} \rightarrow a, \quad \text{a.s..} \tag{9}$$

证明: 记 $W_i = (M_i - M_{i-1}) - \mathbb{E}(M_i - M_{i-1} | \mathcal{F}_{i-1})$, $i = 1, 2, \dots$, 则 $\{W_i, i = 1, 2, \dots\}$ 是一二维鞅差序列, 并且 $\|W_i\| \leq 1$. 注意有

$$\mathbb{E}[M_i - M_{i-1} | \mathcal{F}_{i-1}] = \frac{Y_{i-1}}{i}, \quad i = 1, 2, \dots$$

我们有递推公式:

$$M_i = M_{i-1} + \frac{Y_{i-1}}{i} + W_i, \quad i = 1, 2, \dots$$

两边从 1 到 n 求和, 有

$$M_n = M_0 + \sum_{i=1}^n \frac{Y_{i-1}}{i} + \sum_{i=1}^n W_i. \tag{10}$$

因为

$$\sum_{i=1}^{\infty} \mathbb{E} \left(\frac{W_i^2}{i^2} \right) \leq \sum_{i=1}^{\infty} \frac{1}{i^2} < \infty,$$

由鞅的收敛定理(见[6], 定理2.15)则有

$$\frac{1}{n} \sum_{i=1}^n W_i \rightarrow 0, \quad \text{a.s. } (n \rightarrow \infty).$$

再根据引理1, 有

$$\frac{1}{n} \sum_{i=1}^n \frac{Y_{i-1}}{i} \rightarrow a, \quad \text{a.s. } (n \rightarrow \infty).$$

由(10)式, 定理1成立. #

记 $\eta_i(S_j) = 1$, 如果第 i 次试验结果为 S_j ; $= 0$, 其它. 又记 $\eta_i(F_j) = 1$, 如果第 i 次试验结果为 F_j ; $= 0$, 其它. $j = 1, \dots, k$, $i = 1, 2, \dots$. 令

$$\begin{aligned} \hat{p}_{1j} &= \frac{\sum_{i=1}^n \zeta_{1i} \eta_i(S_j)}{M_{n1}}, & q_{1j} &= \frac{\sum_{i=1}^n \zeta_{1i} \eta_i(F_j)}{M_{n1}}, \\ \hat{p}_{2j} &= \frac{\sum_{i=1}^n \zeta_{2i} \eta_i(S_j)}{M_{n2}}, & \hat{q}_{2j} &= \frac{\sum_{i=1}^n \zeta_{2i} \eta_i(F_j)}{M_{n2}}, \\ \hat{p}_1 &= \sum_{j=1}^k \hat{p}_{1j}, & \hat{p}_2 &= \sum_{j=1}^k \hat{p}_{2j}, \end{aligned}$$

则它们分别是对 p_{1j} , q_{1j} , p_{2j} , q_{2j} , p_1 和 p_2 的最大似然估计. 记

$$\begin{aligned} P &= (p_{11}, \dots, p_{1k}, q_{11}, \dots, q_{1k}, p_{21}, \dots, p_{2k}, q_{21}, \dots, q_{2k})', \\ \hat{P}_n &= (\hat{p}_{11}, \dots, \hat{p}_{1k}, \hat{q}_{11}, \dots, \hat{q}_{1k}, \hat{p}_{21}, \dots, \hat{p}_{2k}, \hat{q}_{21}, \dots, \hat{q}_{2k})'. \end{aligned}$$

估计量 \hat{P}_n 有渐近正态分布.

定理2 当 $n \rightarrow \infty$ 时, 有

$$\sqrt{n}(\hat{P}_n - P) \xrightarrow{D} N(0, \Sigma),$$

其中

$$\Sigma = \begin{bmatrix} \frac{1}{a_1} \Sigma_{11} & 0 \\ 0 & \frac{1}{a_2} \Sigma_{22} \end{bmatrix},$$

$$\Sigma_{ii} = \begin{pmatrix} p_{i1}(1-p_{i1}) & -p_{i1}p_{i2} & \cdots & -p_{i1}p_{ik} & -p_{i1}q_{i1} & \cdots & -p_{i1}q_{ik} \\ -p_{i2}p_{i1} & p_{i2}(1-p_{i2}) & \cdots & -p_{i2}p_{ik} & -p_{i2}q_{i1} & \cdots & -p_{i2}q_{ik} \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ -p_{ik}p_{i1} & -p_{ik}p_{i2} & \cdots & p_{ik}(1-p_{ik}) & -p_{ik}q_{i1} & \cdots & -p_{ik}q_{ik} \\ -q_{i1}p_{i1} & -q_{i1}p_{i2} & \cdots & -q_{i1}p_{ik} & q_{i1}(1-q_{i1}) & \cdots & -q_{ik}q_{ik} \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ -q_{ik}p_{i1} & -q_{ik}p_{i2} & \cdots & -q_{ik}p_{ik} & -q_{ik}q_{i1} & \cdots & q_{ik}(1-q_{ik}) \end{pmatrix},$$

$i = 1, 2$.

证明：由定理 1，有

$$\begin{aligned}
 & \sqrt{n}(\widehat{P}_n - P) \\
 = & \left(\frac{\sum_{i=1}^n \zeta_{1i}(\eta_i(S_1) - p_{11})}{M_{n1}}, \dots, \frac{\sum_{i=1}^n \zeta_{1i}(\eta_i(S_k) - p_{1k})}{M_{n1}}, \frac{\sum_{i=1}^n \zeta_{1i}(\eta_i(F_1) - q_{11})}{M_{n1}}, \right. \\
 & \dots, \frac{\sum_{i=1}^n \zeta_{1i}(\eta_i(F_k) - q_{1k})}{M_{n1}}, \frac{\sum_{i=1}^n \zeta_{2i}(\eta_i(S_1) - p_{21})}{M_{n2}}, \dots, \frac{\sum_{i=1}^n \zeta_{2i}(\eta_i(S_k) - p_{2k})}{M_{n2}}, \\
 & \left. \frac{\sum_{i=1}^n \zeta_{2i}(\eta_i(F_1) - q_{21})}{M_{n2}}, \dots, \frac{\sum_{i=1}^n \zeta_{2i}(\eta_i(F_k) - q_{2k})}{M_{n2}} \right)' \\
 \cong & \left(\sum_{i=1}^n \zeta_{n1i}^{(1)}, \dots, \sum_{i=1}^n \zeta_{nki}^{(1)}, \sum_{i=1}^n \zeta_{n1i}^{(2)}, \dots, \sum_{i=1}^n \zeta_{nki}^{(2)}, \sum_{i=1}^n \zeta_{n1i}^{(3)}, \dots, \sum_{i=1}^n \zeta_{nki}^{(3)}, \sum_{i=1}^n \zeta_{n1i}^{(4)}, \dots, \sum_{i=1}^n \zeta_{nki}^{(4)} \right)' \\
 & \cdot (1 + o_p(1)).
 \end{aligned}$$

由鞅的中心极限定理（见 [6]，定理 3.2），可以证明定理 2 成立。#

推论 当 $n \rightarrow \infty$ 时，有

$$\sqrt{n} \begin{pmatrix} \widehat{p}_1 - p_1 \\ \widehat{p}_2 - p_2 \end{pmatrix} \xrightarrow{D} N \left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \frac{p_1(1-p_1)}{a_1} & 0 \\ 0 & \frac{p_2(1-p_2)}{a_2} \end{pmatrix} \right).$$

§ 4. 例子和结论

考虑 $k = 2$ 的情况。在表 1 中给出若干罐子模型设计结果的例子。比较第 1, 4, 9, 10 列，我们可以看到在这些自适应设计下，治疗效果较好的处理自动向病员倾斜，使病员以很大的机会受到较好处理的试验，因此，这样的设计体现了对受试病员的人文关怀。又从定理 2 及其推论可以看到，所有的估计量的渐近方差具有通常的 $O(1/n)$ 的数量级，从而这样设计仍保持了较好的统计效率，从“人道”与“效率”两个方面综合考虑，本设计具有实际应用参考价值。

表 1 罐子模型设计的几个例子 ($k = 2$)

p_1	p_{11} p_{12}	q_{11} q_{12}	p_2	p_{21} p_{22}	q_{21} q_{22}	α_1 α_2	β_1 β_2	a_1	a_2
0.9	0.4	0.06	0.6	0.4	0.3	0.6	0.4	0.58	0.42
	0.5	0.04		0.2	0.1	0.8	0.2		
0.9	0.4	0.06	0.6	0.4	0.3	0.7	0.3	0.64	0.36
	0.5	0.04		0.2	0.1	0.9	0.1		
0.9	0.4	0.06	0.6	0.4	0.3	0.8	0.2	0.69	0.31
	0.5	0.04		0.2	0.1	0.95	0.05		
0.8	0.5	0.12	0.4	0.3	0.4	0.7	0.3	0.63	0.37
	0.3	0.08		0.1	0.2	0.9	0.1		
0.8	0.5	0.12	0.4	0.3	0.4	0.65	0.35	0.65	0.35
	0.3	0.08		0.1	0.2	0.85	0.15		
0.8	0.5	0.12	0.4	0.3	0.4	0.8	0.2	0.67	0.33
	0.3	0.08		0.1	0.2	0.95	0.05		

后记 陈希孺老师对我们的教诲历历在目，难以忘怀。仅以此文表示对于一代宗师的敬仰与怀念。

参 考 文 献

- [1] Althreya, K.B. & Karlin, S., Embedding of urn scheme into continuous time branching processes and related limit theorems, *Ann. Math. Statist.*, **39**(1968), 1801–1817.
- [2] Bai, Z.D. & Hu, F.F., Asymptotic theorems for urn models with monhomogeneous generating matrices, *Stochastic Processes and Their Applications*, **80**(1)(1999), 87–101.
- [3] Bai, Z.D., Chen G.J. & Hu, F.F., Some theorems, under urn models with time trebds, *Chinese Annals of Mathematics*, **22**(A)(2001), 89–96.
- [4] Chen, G.J., Zhu, C.H. & Wang, Y.H., Limit theorems and optimal design with adaptive urn models, *Journal of Systems Science and Complexity*, **18**(2005), 347–360.
- [5] Devroye, L., Lasjlo, G., Lugosi, G., *A Probabilistic Theory of Pattern Recognition*, Springer, 1996.
- [6] Hall, P. & Hegde, C.C., *Martingale Limit Theory and Its Application*, Academic Press, London, 1980.
- [7] Rosenberger, W.F., Flournoy, N. & Durham, S.D., Asymptotic normality of maximum likelihood estimators from multiparameter response-driven designs, *J. Statist. Plann. Inf.*, **60**(1997), 69–76.
- [8] Rosenberger, W.F., Asymptotic inference with response adaptive treatment allocation designs, *Ann. Statist.*, **21**(1993), 2098–2107.
- [9] Wei, L.J. & Durham, S., The randomized play-the-winner rule in medical trials, *J. Amer. Statist. Assoc.*, **73**(1978), 840–843.

A Kind of Urn Models in Clinical Trials With Multi-Outcome

CHEN GUIJING¹ HU SHUHE¹ HONG SHENGYAN^{2,1}

(¹School of Mathematics and Computational Science, Anhui University, Hefei, 230039)

(²Lilly Research Laboratories, Eli Lilly and Company, Indianapolis, IN 46285, USA)

A kind of adaptive designs in clinical trials with multi-out come have been constructed using urn models some limit theorems have been shown in such models, some examples are given, which indicate that the better treatment will be allocated to patients with large chance, and the statistical efficiency in these designs is still keeping better. So these adaptive designs are feasible.