从“许氏模型”想到的……*

张尧庭

（武汉大学）

许宝騄先生（1910—1970）逝世已有十年了．有些同志要我写文章，介绍许先生教学、科研工作的特点和经验，我感到很困难．主要的原因是自己学得不好，自己看到的、理解得可能不对，说错了反而造成对许先生的误解．1980年9月，北京大学举行了“许宝騄教授诞生七十周年纪念会”，会上不少先生作了报告，全面地介绍、评述了许先生的一生，从他的政治思想，到他的学术活动，从他的教学工作，到他的科研成果都已说得很清楚了．那么，我为什么又想起来要写这样一篇文章呢？近几年来，国内外的杂志上刊登了好几篇讨论“许氏模型”的文章，这个课题已成了数理统计中比较受人注意的专题之一，我在学习这些论文的时候，脑子里不断地想起许先生生前言谈的某些片断，我想介绍一下还是很有意义的，这就是我写这篇文章的目的．

“一篇论文的价值，不是在它刚刚发表的时候，而是在它不断地被人引证的时候，才能得到完全的确定．”这是许先生生前讲过的关于如何评价学术论文的一个意见．从“许氏模型”的发展过程可以看出许先生的这一意见是非常正确的．

早在1938年，许先生写出了他的第三篇论文[1]，当时并没有引起多大的重视，他所提出的问题是新的，解决问题的想法也是很巧的，但是，并没有人在这一方面继续往下深入地做更多的研究．直到1952年，才有*C. R. Rao*[2]，讨论了相同的问题，得到了另一个结果．这一段时期内，大多数的统计学家都关心的是线性模型中期望值参数的估计和检验的问题，而不关心线性模型中方差的估计问题．许先生是第一个证明了用残差平方和除以自由度去估计方差在什么条件下是“最优的”．到了五十年代，由于线性模型的发展，关于方差分量的估计出现了一些新的问题，例如用通常的方法获得的某个方差的估计量可能会取负值，这是无法接受的．因此，在六十年代就引起了一些人的注意和研究．到了七十年代，欧美的、苏联东欧的一些统计学家都从许先生的1938年的文章[1]中，得了启发，把〔1〕的想法一般化，就导出了一系列新的结果，形成了“许氏模型”和“离差——均值模型”，直到现在，还有很多问题还没有解决．Lehmann在〔3〕中介绍许先生的工作时说：“可以认为许的文章（即指〔1〕）是近来关于方差分量和方差的最佳二次估计的大量文献的起点”．

许先生在文〔1〕中主要是提出了什么问题，用了什么方法呢？这里用现在大家熟悉的符号来介绍一下．

* 1980年12月11日收到．

假定随机向量 y($n\times1$ 的向量）的一阶矩、二阶矩满足条件：

$$E(y)=C\theta,\qquad V(y)=\sigma^2 I_n, \tag{1}$$

其中 C 是已知的常数矩阵，θ 及 σ^2 是未知的参数。

我们都知道，当 C 阵满秩时（设 C 是 $n\times k$ 的阵），则最小二乘估计

$$\hat\theta\ (C'C)^{-1}C'y \tag{2}$$

是 θ 的最佳线性无偏估计，而

$$\hat\sigma^2=\frac{1}{n-k}y'(I-C(C'C)^{-1}C')y \tag{3}$$

是 σ^2 的一个无偏估计．许先生在〔1〕中讨论(3)式给出的 $\hat\sigma^2$ 是否具有某种优良性．值得注意的是，他是在怎样的范围内讨论 $\hat\sigma^2$ 的优良性的．他提出了两个条件：估计量要满足

(i) 是 y 的齐二次函数，即形如 $y'Ay$ 的估计量，$A'=A$，并且是无偏的，即有

$$Ey'Ay=\sigma^2;$$

(ii) 估计量的方差与未知期望值参数 θ 无关．条件(i)是很自然的要求，条件 (ii) 的限制有些不一般，Rao 在 1952 年〔2〕中将(ii)改为 $A\geqslant0$，得到了另一个结果．事实上后来的研究证明了许在〔1〕所加的条件(ii)就是不变性，这是 Seely（1971 年）在〔4〕中阐明的．与这个条件(ii)相似的条件在许先生后来的工作中还一再地出现过，只是形式不同，基本思想始终是一致的．他在讨论方差分析似然比检验势的优良性时，也提出了一个类似的限制条件，后来的研究证明，他提出的条件也是与不变性概念等价的．如果我们注意到不变性概念是在五十年代才明确起来的这一事实时，就会深深感到许先生在1938年就能认识到这一条件的作用，这不能不使人感到他确有深刻的见解．他在〔1〕中找到了，(3) 式 $\hat\sigma^2$ 是上述 (i)、(ii) 条件都满足的估计量中达到最小方差的充要条件是什么．他的方法是把二次估计的问题变成一次的估计问题．为了便于说明，我们引入一些记号，设 A 是一 $n\times m$ 的矩阵，可以将 A 写成

$$A=\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1m}\\ a_{21} & a_{22} & \cdots & a_{2m}\\ \vdots & \vdots & & \vdots\\ a_{n1} & a_{n2} & \cdots & a_{nm}\end{bmatrix}.$$

用 $\vec{A}$ 表示把 A 中的元素依行的顺序逐个地排下来写成一列，是一个 $nm\times1$ 的向量，即

$$\vec{A}=\begin{bmatrix}a_{11}\\ a_{12}\\ \vdots\\ a_{1m}\\ a_{21}\\ a_{22}\\ \vdots\\ a_{2m}\\ \vdots\\ a_{nm}\end{bmatrix}.$$

对(1)式中的y进行变换，可得一更为简明的模式。令$z=(I-C(C'C)^{-1}C')y$，于是记$P=I-C(C'C)^{-1}C'$后，就有$z=Py$，此时

$$E(z)=PE(y)=PC\theta=0,$$

$$V(z)=Ezz'=\sigma^2PV(y)P'=\sigma^2P,$$

也即有

$$E\,\overrightarrow{zz'}=\vec{P}\sigma^2, \tag{4}$$

将(4)与(1)相比较，$\overrightarrow{zz'}$相当于y，$\vec{P}$相当于C，σ^2相当于θ。Gauss—Markov定理关于(1)中θ的最佳线性无偏估计（即方差最小的线性无偏估计）已有明确的结论，因此，可以讨论（4）中σ^2的关于$\overrightarrow{zz'}$的线性无偏估计中方差最小的估计。由于对矩阵来说，有

$$\vec{A}'\vec{B}=t_rA'B,$$

因此，$\overrightarrow{zz'}$的线性函数

$$\vec{A}'\overrightarrow{zz}=t_rA'zz'=t_rz'A'z=z'A'z.$$

可见$\overrightarrow{zz'}$的线性函数就是z的二次函数，并且无妨假定$A'=A$。反之，z的二次函数$z'Az=t_rAzz'=\vec{A}'\overrightarrow{zz'}$也就是$\overrightarrow{zz'}$的线性函数。这样，就把二次估计的问题化成了线性估计的问题。容易看出，要使$y'Ay$的方差与θ无关（即不涉及y期望值的参数），$y'Ay$必须是$z'Bz$的形式，这样，所要考虑的全部的估计就是$z'Bz$而且$Ez'Bz=\sigma^2$的全体估计。(3)式的

$$\hat{\sigma}^2=\frac{1}{n-k}y'(I-C(C'C)^{-1}C')y$$

$$=\frac{1}{n-k}y'Py=\frac{1}{n-k}t_rPyy'=\frac{1}{n-k}t_rPzz',$$

而$\vec{P}'\vec{P}=t_rP'P=t_rP=n-k$，于是

$$\hat{\sigma}^2=(\vec{P}'\vec{P})^{-1}\vec{P}'\overrightarrow{yy'}=(\vec{P}'\vec{P})^{-1}\vec{P}'\overrightarrow{zz'}. \tag{5}$$

将(2)与(5)对比一下，就可看出$\hat{\sigma}^2$就相当于(1)中θ的最小二乘估计。可见只要弄清最小二乘估计在什么条件下是最佳线性无偏估计，就可以导出(5)式给出的最小二乘估计在什么条件下是σ^2的最佳二次无偏的估计（满足条件(ii)的）。这样，就把方差中的参数的估计化成了相当于线性模型中期望值参数的估计。这个想法既是这样的巧妙，而又是这样的自然，似乎很不费力就可以了解到，这正体现了许先生解决许多问题的特点。

许先生经常要我们注意，那些表面上看来似乎是不同的东西，实质上是一样的，也即从某一个观点看来，就是一回事，他认为这就是数学所需要的一种抽象能力。二次型$y'Ay$与线性型$a'y$表面上看来似乎是很不一样的对象，但是当我们把矩阵看成向量时，那么$y'Ay$就是A与yy'的内积，它是$\overrightarrow{yy'}$的线性函数。

事实上，不难将（1）的形式改变为更复杂的情形：

$$E(y)=C\theta, \qquad V(y)=\sum_{i=1}^{t}\xi_i V_i, \tag{6}$$

其中C及V_1，…，V_t均为已知的，θ及ξ_1，…，ξ_t是未知参数。

仿照上面的方法，作变换$z=Py$，令$W_i=PV_iP$，则有

$$\overrightarrow{Ezz'}=\sum_{i=1}^{t}\xi_i\overrightarrow{PV_iP}=\sum_{i=1}^{t}\xi_i\overrightarrow{W}_i$$

$$=\left(\overrightarrow{W}_1\ \overrightarrow{W}_2\cdots\overrightarrow{W}_t\right)\begin{pmatrix}\xi_1\\ \vdots\\ \xi_t\end{pmatrix},$$

即

$$\overrightarrow{Ezz'}\triangleq W\xi. \tag{7}$$

它完全是线性模型的形式，很快就可以得到ξ的最小二乘估计

$$\hat{\xi}=(W'W)^{-}W'\overrightarrow{zz'}, \tag{8}$$

并且可以讨论$\hat{\xi}$是最佳二次不变估计的充要条件是什么。目前看到的一些关于这方面的文献，都是从这样的想法开始的，然后去寻求各种优良的估计。

许先生的教学和科研一向是为人称道的。他经常把极其繁复的结果用一个很简洁的、初等的方法给以证明。就如同一个熟练的老师傅，他能把一个极其粗糙的铸件加工成非常光洁的成品，在这一方面，也可以说，他更像一个艺术家，可以把一块璞玉，制作成一件精美的艺术品。我感到他所以能这样，是他阅读了有关这一问题的大量的文献，经过了仔细的分析比较，弄清了这个问题应该怎么提，怎样来解决它。他的研究工作的特点是往往在一个问题的开始部份，他的看法和别人就不一样，因此，他处理问题的方法也随着就会不同。即使是同一个问题，同一个定理，由他来讲，往往会给出一个新的证明。在他感到身体很不好之后，他想系统地讲授几门基础课，能够体现出他的处理方法的特点，在1964年主持了三门课：点集拓扑、复变函数、平稳过程。可惜的是由于政治运动，这些课程都

只进行了一部份，中途停顿了，但是，就是从这一部份不全的讲课笔记中，我们仍然能够学到不少有用的方法。

以上这些，就是我在学习"许氏模型"时所想到的，可能有不对的地方，欢迎批评指正。

参　考　文　献

[1] Hsu, P. L. (许宝騄), On the best quadratic estimate of variance, Statist. Res. Mem., 2, (1938), 91—104.

[2] Rao, C. R., Some theorems on minimum variance estimation, Sankhya, 12 (1952), 27—44.

[3] Lehmann, E. L., 许宝騄在统计推断方面的工作，原文刊登在 Ann. statist., 3 : 7 (1979) 上，译文刊登在数学实践与认识，1980年第 3 期，6—7.

[4] Seely, J., Quadratic subspaces and completeness, Ann. Math. Statist., 42(1971), 710—721.

What Could Be Realized From P. L. Hsu's Model

By Zhang Yaoting (张尧庭)

Abstract

This article gives a brief exposition of P. L. Hsu's model of the best quadratic estimate of variance. As is well known, the idea of methodology suggested by Hsu's model has been proved quite successful and fruitful in many subsequent investigations and is therefore worth noticing and learning.