精确分布推导的一个基本的方法*

张尧庭

（武汉大学）

寻求统计量的精确分布，在数理统计的理论中是一个重要的问题。然而，在多元统计分析中，往往很不容易得到统计量的精确分布，有些统计量的分布，例如非中心的Wishart分布，持续了几十年，还没有获得完满的结果。我国的许宝騄先生早在1939年，就得到了正态总体样本协差阵的特征根的联合分布（参看[1]）。据徐钟济先生的回忆，当时R. A. Fisher和S. N. Roy也正在考虑这一问题，许先生用了四天的时间就完全解决了，这一结果通常的书上都称为"Fisher—Hsu—Roy"的分布。

实际上，许先生在三十年代末期，整理发展了当时关于精确分布推导的一系列成果。在他之前，Fisher, Wishart, Mahalanobis, Bose以及Roy, Hotelling等著名的学者，推导精确分布的基本方法都是几何方法。强有力的分析方法是许先生首先引进的，他熟练地把矩阵与分析相结合，对一些重要分布的推导，如Wishart分布，Bartlett分解，T^2分布，…等等都给出了严格而清晰的证明，发展了计算积分变数替换的雅可比行列式的技巧。四十年代，他在北卡罗来纳大学讲课时，系统地讲授了这些内容。1951年Deemer和Olkin根据他们自己听课的笔记详细整理发表了，这就是[2]。多元分析方面一本重要的著作[3]，作者T. W. Anderson说明他写的第十三章就是根据许先生的方法、参看[2]写成的。六十年代初期，许先生领导了一个关于次序统计量的极限分布的讨论班，我第一次接触到许先生推导精确分布的方法（在讨论班中主题是求极限分布的律型），当时，并没有什么体会。七十年代中期，实际工作中提出了大量的统计问题大多数与多元分析有关，自己再重新学习有关的内容时，深深感到这种方法的好处，它只用一些分析和线性代数的知识，用比较初等的方法就可导出一些相当复杂的分布。这些年来，根据许先生在一本书的扉页上留下的提纲，我和一些同志整理了一些材料，这就是[4]。在这个基础上，还进一步做了一点工作（见[5]）。弄清这个方法，对于学习数理统计和进行这方面的科研工作，我想都是有益的。

假定随机向量 $x=\begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$ 的联合分布密度是已知函数 $P(x)$，$\underset{n\times 1}{y}=f(x)=(f_1(x), f_2(x), \cdots, f_n(x))'$，这些 f_i 对 x 有直到二阶的各种连续的偏微商，且作为普通变量，

* 1981年4月7日收到。

y 对 x 的雅可比行列式 $J(y|x)$ 不等于 0，于是从积分变换的公式就知道，随机向量 y 的联合密度

$$q(y)=p(x(y))\,|J(x|y)|\,, \tag{1}$$

其中 $x(y)$ 是 $y=f(x)$ 的反函数。这一结果是一般的概率统计的书上都能见到的。从(1)式明显地可以看到如何求得雅可比行列式 $J(x|y)$ 是关键的一步。更麻烦的是，在通常情况下，y 的维数并不和 x 一样，这时(1)式就无法使用了；或者 $y=f(x)$ 并不是 x 的可导函数，这时(1)式更不能用了。许先生的方法不是从(1)式出发，而是从更基本的一个事实出发。

要求 y 的分布密度，就是求

$$P(f_1(x)<t_1,\ f_2(x)<t_2\cdots,\ f_n(x)<t_n)$$

的值。如果用一个简单的符号表示，就是 $P(y<t)$。引入集合的特征函数 $I_{\{y<t\}}(x)$，于是

$$I_{\{y<t\}}(x)=\begin{cases}1\ ,&\text{当} y(X)<t\ ,\\ 0\ ,&\text{其他}\ .\end{cases}$$

就有

$$P(y<t)=EI_{\{y<t\}}(x)\ .$$

因此，求 $P(f_1(x)<t_1,\cdots,f_n(x)<t_n)$ 的值，就是求期望值 $EI_{\{y<t\}}(x)$。注意到 $I_{\{y<t\}}(x)$ 是一个非负 Borel 可测函数，而且它一定是 $g(y(x))$ 形式的函数，于是只要能对任一非负 Borel 可测函数 $g(y)$，求出 $Eg(y(x))$ 的值，我们自然就求出了 $y(x)$ 的分布密度。下面我们用两个典型的例子来说明这一点。

例 1．Wishart 分布．

从[3]的第十三章或[4]，已知有公式：对任何一非负的 Borel 可测函数 $g(Y)$，Y 是一 $p\times p$ 的对称阵，有

$$\int_{R_{mp}} g(\underset{m\times p}{X'X})\,dX=\frac{\pi^{\frac{1}{2}pm-\frac{1}{4}p(p-1)}}{\prod_{i=1}^{p}\Gamma\left(\frac{m+1-i}{2}\right)}\int_{Y>0}|Y|^{\frac{m-p-1}{2}}\,g(Y)\,d\underset{p\times p}{Y}\,, \tag{2}$$

其中 $m\geqslant p\geqslant 1$，R_{mp} 表示 mp 维欧氏空间，$\int_G(\quad)dX$ 表示在区域 G 上的重积分，当 $\underset{n\times n}{X}$ 是对称阵时，dX 只表示 X 中上三角部分的变量 $\frac{n}{2}(n+1)$ 个，因而 $\int$ 也只表示 $\frac{n}{2}(n+1)$ 重积分，$B>0$ 表示 B 是正定阵，(2) 式右端表示 Y 中的 $\frac{p}{2}(p+1)$ 个变量在区域 $Y>0$ 上积分。

在(2)式中取

$$g(X'X)=\left(\frac{1}{\sqrt{2\pi}}\right)^{mp}\ e^{-\frac{1}{2}trX'X}f(X'X)\ ,$$

其中f是任一非负的Borel可测函数，于是有：当$\underset{m\times p}{X}$中的mp个随机变量相互独立，都遵从标准正态分布$N(0,1)$时，$Y=X'X$的任一非负Borel可测函数$f(Y)$的期望值

$$Ef(Y)=\int_{R^{mp}} g(X'X)\,dX$$

$$=\frac{\pi^{\frac{1}{2}pm-\frac{1}{4}p(p-1)}}{\prod_{i=1}^{p}\Gamma\left(\frac{m+1-i}{2}\right)}\cdot\left(\frac{1}{2\pi}\right)^{\frac{mp}{2}}\int_{Y>0} f(Y)\,|Y|^{\frac{m-p-1}{2}}e^{-\frac{1}{2}trY}dY\ .$$

$$=\frac{\pi^{\frac{1}{2}pm-\frac{1}{4}p(p-1)-\frac{1}{2}mp}}{2^{\frac{mp}{2}}\prod_{i=1}^{p}\Gamma\left(\frac{m+1-i}{2}\right)}\int_{Y>0} f(Y)\,|Y|^{\frac{m-p-1}{2}}e^{-\frac{1}{2}trY}dY\ .$$

从概率论的知识我们已知：若对任一y的非负可测的(Borel可测)函数$f(y)$，有

$$Ef(y)=\int f(y)q(y)\,dy\ ,$$

则y的分布密度就是$q(y)$。因为对任一Borel集E，考虑特征函数$I_{\{y\in E\}}(y)$，于是

$$P(y\in E)=EI_{\{y\in E\}}(y)=\int_{y\in E} q(y)\,dy\ .$$

于是就可得到$Y=X'X$的联合分布密度是

$$\frac{|Y|^{\frac{m-p-1}{2}}e^{\frac{1}{2}trY}}{2^{\frac{mp}{2}}\pi^{\frac{1}{4}p(p-1)}\prod_{i=1}^{p}\Gamma\left(\frac{m+1-i}{2}\right)},\ Y>0\ .$$

这就是著名的Wishart分布。

从这个推导过程可以看出，这时不涉及通常求积分(分布函数相应的积分)要遇到的不等式，因为$P(f_1(x)<t_1,\cdots,f_n(x)<t_n)$要能积分，必需将区域$\{X:f_1(x)<t_1,\cdots,f_n(x)<t_n\}$表示成$x$的明显的不等式，然后才能求出积分，而这里却避免了这一麻烦的步骤。

例2．次序统计量的联合分布．

设$x_1,x_2,\cdots,x_n$是来自分布$F(x)$的一个独立样本。令$\xi_1=\min\limits_{1\le i\le n}x_i$，$\xi_2=\{x_1,\cdots,x_n\}$中第二小的，…，$\xi_n=\max\limits_{1\le i\le n}x_i$，于是有$\xi_1\le\xi_2\cdots\le\xi_n$．很明显，这些$\xi_i$都是$x_1,\cdots,x_n$的连续函数，但不是可导的，因此，用求雅可比的方法是无法进行的。但是，用例1类似的方法，仍然是可以直接求得的。

假定$g(y_1,\cdots y_n)$是非负的Borel可测函数，于是有

$$Eg(\xi_1,\cdots,\xi_n)=\int g(\xi_1(x),\cdots,\xi_n(x))\prod_{i=1}^{n}dF(x_i)\ .$$

当 $F(x)$ 是连续函数时，$x_i=X_j(i\neq j)$ 的情况发生的概率是0，因而只需考虑 $x_1,\cdots,x_n$ 全不相等的情形，我们用 $(i_1,\cdots,i_n)$ 表示 $(1,2,\cdots,n)$ 的一个置换，于是有

$$
\begin{aligned}
Eg(\xi_1,\cdots,\xi_n) &= \int g(\xi_1(x),\cdots,\xi_n(x))\prod_{i=1}^{n}dF(x_i)\\
&= \sum_{(i_1,\cdots,i_n)}\int_{x_{i_1}<x_{i_2}<\cdots<x_{i_n}} g(\xi_1(x),\cdots,\xi_n(x))\prod_{i=1}^{n}dF(x_i)\\
&= \sum_{(i_1,\cdots,i_n)}\int_{x_{i_1}<x_{i_2}<\cdots<x_{i_n}} g(x_{i_1},x_{i_2},\cdots,x_{i_n})\prod_{i=1}^{n}dF(x_i)\\
&= \sum_{(i_1,\cdots,i_n)}\int_{x_1<\cdots<x_n} g(x_1,\cdots,x_n)\prod_{i=1}^{n}dF(x_i)\\
&= \int_{x_1<x_2<\cdots<x_n} g(x_1,\cdots,x_n)\left(\frac{1}{n!}\right)\prod_{i=1}^{n}dF(x_i)\,.
\end{aligned}
$$

这样就导出了次序统计量 $\xi_1,\cdots,\xi_n$ 相应的分布。

上面这两个例子是非常典型的。例1中原有的随机矩阵 $\underset{m\times p}{X}$ 中共有 mp 个随机变量，而 $Y=X'X$ 只有 $\frac{p}{2}(p+1)$ 个随机变量(由于 $m\geqslant p$，Y 中的随机变量的个数少于 X，这用(1)式的解法是不好办的，而例2正好是无法求雅可比的例子，实际上例2中的方法，可以求任意指定的次序统计量 $\xi,\cdots,\xi_n$ 中的某一部份 $\xi_{i_1}<\xi_{i_2}<\cdots<\xi_{i_k}$ 的联合分布。即使(1)式是完全可用的场合，如何来求 $J(x|y)$ 也是很有意思的，下面我们也举一个例子来说明它。

在讨论矩阵随机变量的分布时，如果 $\underset{n\times n}{X}$ 是 n^2 个独立标准正态变量组成的矩阵，显然 $|X|=0$ 的概率是0，因此可以求 $Y=X^{-1}$ 的分布，这时只需求出 $J(Y|X)$，用(1)就行了，这一雅可比行列式在[3]的附录中就已给出了，在[5]中，已推导出 Y 的密度。然而当 $\underset{p\times p}{X}$ 是一个正定阵时，如果 $X\sim W(n,I_p)$(即 X 是 n 个自由度，协差阵为 I_p 的Wishart分布)，要求出 $Y=X^{-1}$ 的雅可比行列式，却相当麻烦，因为 Y 和 X 都只有 $\frac{p}{2}(p+1)$ 个随机变量，而不是 p^2 个随机变量。然而，极其有意义的是，我们可以利用(2)式来导出当 $X>0$ 时，$Y=X^{-1}$ 的雅可比行列式，从而我们就可以很容易地得到 X^{-1} 的分布(这就是通常所说的反Wishart分布)。下面我们用一个求Wishart分布的“平方根”矩阵的分布来说明这一点。

假定 $\underset{p\times p}{A}\sim W(n,I_p)$，由于 A 的正定性，可知 $A=\underset{p\times p}{C'}\ \underset{p\times p}{C}$，$|C|\neq 0$，现在要求 C 的联合密度的表达式，通常把 C 称为 A 的平方根矩阵，或记为 $A^{\frac{1}{2}}$。因此 A 可以写为 $A=(A^{\frac{1}{2}})'A^{\frac{1}{2}}$。

无妨设C的联合密度为$p(C)$，于是对任给的非负Borel可测函数$g(Y)$，就有

$$\begin{aligned}&\int g(C'C)P(C)dC\\&=Eg(C'C)\\&=Eg(A)\\&=\int_{A>0}K(n,I_p)\,|A|^{\frac{n-p-1}{2}}e^{-\frac{1}{2}trA}g(A)\,dA\,,\end{aligned}$$

其中
$$K(n,I_p)=\left(2^{\frac{np}{2}}\pi^{\frac{p}{4}(p-1)}\prod_{i=1}^{p}\Gamma\left(\frac{n+1-i}{2}\right)\right)^{-1}.$$

但是由$A>0$，$(n\geqslant p)$，因此可以利用(2)式将上面最后一式写成dC的积分式，即有

$$\begin{aligned}&\int_{A>0}K(n,I_p)\,|A|^{\frac{n-p-1}{2}}e^{-\frac{1}{2}trA}g(A)dA\\&=\int_{R_{p^2}}ag(C'C)\,|C'C|^{\frac{n-p}{2}}e^{-\frac{1}{2}trC'C}dC\end{aligned}$$

其中a是某一个确定的常数．注意到

$$|C'C|=|C|^2\,,$$

再有g的任意性，我们就得到c的联合分布密度是

$$a|C|^{n-p}e^{-\frac{1}{2}trC'C},$$

其中a的值可以直接利用公式(2)得到，或利用

$$a\int|C|^{n-p}e^{-\frac{1}{2}trC'C}dC=1$$

求得．

很明显，当$A\sim W(n,\underset{p\times p}{\Sigma})$，$n\geqslant p$时，完全类似的方法可以求得相应的$C$的分布．

从(1)式还可以看出，随机变量之间的变换与积分变量之间的变换，有着本质的联系，是否可以不经过积分变量的变换，直接处理随机变量之间的变换而导出分布呢？一般地说，这是不行的．然而，在一些常见的特殊情况下，这是完全行得通的．现以非中心T^2的分布为例来说明这一点．假定$z\sim N(\theta,I_p)$，z与$u_1,\cdots u_n$独立，且u_i是来自$N(0,I_p)$的一个样本，现在要求$t=z'A^{-1}z$的分布，其中$A=\sum_{i=1}^{n}u_iu_i'$．

考虑$y=Qz$，Q是一正交阵，且有形状$Q=\begin{pmatrix}z'/\|z\|^2\\ *\end{pmatrix}$，可见$Q$是一随机矩阵．记$U=(u_1\,u_2\cdots u_n)$，$V=QU$．任给二个非负的Borel可测函数$f_1$，$f_2$，

$$Ef_1(y)f_2(V)=Ef_1(Qz)f_2(QU)\triangleq Ef(Qz,QU)$$

$$=K\int f(Qz,QU)\,e^{-\frac{1}{2}trC[(z-\theta)(z-\theta)'+UU']}dzdU$$

作积分变换 $y=Qz$，$V=QU$，于是

$$z=Q'y,\quad U=Q'V,$$

注意到变换相应的雅可比并非简单的计算就可获得的，但是把上式积分分开后分别作变换却是方便的。因为利用上式就有

$$Ef_1(y)f_2(V)=K\int f_1(Qz)\,e^{-\frac{1}{2}tr(z-\theta)(z-\theta)'}dz\int f_2(QU)\,e^{-\frac{1}{2}trUU'}\,dU,$$

注意到 $V=QU$ 相应的雅可比行列式的绝对值恒为1，因此作$V=QU$变换后，就有$U=Q'V$

$$\int f_2(QU)\,e^{-\frac{1}{2}trUU'}\,dU$$

$$=\int f_2(V)\,e^{-\frac{1}{2}trVV'}dV,$$

这是因为 $trQ'VV'Q=trVV'QQ'=trVV'$。因此

$$Ef_1(y)\,f_2(V)=K\int f_1(Qz)\,e^{-\frac{1}{2}tr(z-\theta)z(-)'\theta}\,dz\int f_2(V)\,e^{-\frac{1}{2}trVV'}\,dV$$

$$=Ef_1(Qz)Ef_2(V)$$

$$=Ef_1(y)Ef_2(V),$$

也即经过形如Q的随机正交矩阵的变换，这两个“随机向量”Qz与QU仍然是独立的。于是

$$t=z'A^{-1}z=z'Q'QA^{-1}Q'Qz$$

$$=y'(\sum_{i=1}^{n}Qu_iu_i'Q')^{-1}y$$

$$\triangleq y'B^{-1}y,$$

而

$$\begin{pmatrix}y_1\\ \vdots\\ y_p\end{pmatrix}=y=Qz=\begin{pmatrix}z'/\|z\|\\ *\end{pmatrix}z=\begin{pmatrix}\|z\|\\ 0\end{pmatrix},$$

记 B^{-1} 的元素为 b^{ij}，于是

$$t=y_1b^{11}y_1=y_1^2b^{11}$$

y_1^2 是非中心 χ^2 分布，$(b^{11})^{-1}$ 是中心 χ^2 分布，且 $y_1^{2\,11}$ 与 b 独立，因此 t 是非中心的 F 分布。

把上面论证 $Ef_1(Qz)f_2(QU)=Ef_1(Qz)Ef_2(QU)$ 的方法一般化，就得到[6]中的一条引理，这就是1958年以后，由Wijsman提出的随机变量变换的方法，从上面的介绍也可以明显地看到这一方法确实是由计算雅可比的方法演变出来的。

当然，到了今天，八十年代会有什么新的更好的方法，还不好说。但是在六十年代到七十年代，采用外微分形式，张量以及带形多项式等工具，解决了一些求分布的问题，这可参看[6]，[7]，这些已超出本文的范围了。

参　考　文　献

[1] Hsu, P. L., On the distribution of roots of certain determinantal equations, Ann. Eugen., 9(1939), 250-258.

[2] Deemer, Walter L. and Olkin, Ingram, The Jacobians of certain matrix transformations useful in multivariate analysis, Biometrika, 38(1951), 345-367.

[3] Anderson, T. W., "An Introduction to Multivariate Statistical Analysis, John Wiley and Sons, New York (1958).

[4] 张尧庭，邹新堤，肖佑恩，关于某些统计量分布的推导，武汉大学学报(自然科学版)(1978)第1期.

[5] 黄秦、张尧庭，"矩阵变量的 t，F 分布"(即将发表).

[6] Farrell, Roger H., "Techniques of Multivariate Calculation, Springer-Verlag (1976).

[7] Springer M. D., The Algebra of Random Variables, John Wiley and Sons, New York (1979).

An Element Method for the Derivation of Exact Sampling Distributions

By Zhang Yaoding (张尧庭)

Abstract

In this paper, we introduce a method for derivating the exact sampling distributions. The method, which was introduced to the multivariate analysis first by P. L. Hsu in 1938-1939, is a straightforward derivation, and was extended to derivate the exact distribution of order statistics by P.L.Hsu in 1960-1961.

By this method, we get some new distributions, which will be printed in [5].