DOI:10.13747/j.cnki.bdxyxb.2011.03.001

【数学研究】

多元分布拟合优度检验研究进展

苏 岩

摘 要:探讨多元概率分布的构造理论与多元概率分布拟合优度的各种检验方法,指出拟合优度检验在多元数据分析中的重要作用.阐述基于单位球的多元概率分布拟合优度检验的独特的理论意义,探讨回归模型有效性检验的研究路径,强调小样本理论与大样本理论在统计模型检验中的综合应用.

关键词:多元分布;均匀分布;单位球;拟合优度检验
中图分类号:O212 文献标识码:A 文章编号:1674-2494(2011)03-0001-07

作者简介:苏 岩(1963-),男,北京市人,副教授,理学博士,主要研究方向为应用统计;华北电力大学数理学院,河北保定 071003.

总体分布是构成统计模型的基本要素,统计推断离不开对总体分布的假设,拟合优度检验实质上就是模型检验. Karl Pearson在1900年发表的卡方(chi-square)拟合优度检验的论文,被认为是现代统计学理论的开端. 2000年在巴黎召开“拟合优度检验与统计模型有效性”国际会议以纪念这篇里程碑式的论文发表100周年[1].

一元分布的拟合优度检验方法相对成熟,主要包括作图法、Pearson χ^2型检验、EDF型检验、条件积分变换方法等. 多元分布的拟合优度检验较为复杂,检验方法尚处于研究探讨阶段. 在这里将介绍一些学者的相关工作以及笔者在多元分布拟合优度检验方面所做的探讨. 首先讨论多元概率分布的构造理论[2].

1 多元概率分布的构造

ⅰ)球对称分布:d维随机变量X称为服从球对称分布,当且仅当$X\overset{d}{=}RU^{(d)}$,其中R是非负随机变量,$U^{(d)}$是单位球面上的均匀分布随机变量,R与$U^{(d)}$相互独立. 球对称分布等价表达形式为$X\overset{d}{=}R^*V^{(d)}$,其中$R^*$是非负随机变量,$V^{(d)}$是单位球上的均匀分布随机变量,$R^*$与$V^{(d)}$相互独立.

ⅱ)椭球对称分布:d维随机变量Y称为服从椭球对称分布,当且仅当$Y\overset{d}{=}\mu+AX$,其中X服从d维球对称分布,A为$d\times d$阶非奇异矩阵. 标准多元正态分布属于球对称分布,当$X\sim N_d(0,I_d)$时,$Y\sim N_d(\mu,AA')$. 多元正态分布是椭球对称分布的特殊情形.

ⅲ)中心相似分布:Yang等[3](2003)基于垂直密度表示(VDR),提出了中心相似分布. 定义d维基本集D_0(包含0)及其界函数$b(x)$,$x\in\mathbf{R}^d$. X称为服从中心相似分布,当且仅当$X\overset{d}{=}RZ^{(d)}$,其中$Z^{(d)}$在基本集D_0上服从均匀分布,R是非负随机变量且与$Z^{(d)}$相互独立. 也称X服从标准中心相似分布,记为$X\sim C(0_d,I_d,D_0)$. 此时X具有垂直表示的独立可分解结构. 中心相似分布等价表达形式为

$$X\overset{d}{=}R^*W^{(d)},R^*=R\cdot\|Z^{(d)}\|,W^{(d)}=Z^{(d)}/\|Z^{(d)}\|.$$

其中R^*是非负随机变量且R^*与$W^{(d)}$相互独立. 此时,单位球面上的随机变量$W^{(d)}$不再服从球面上的均匀分布.

收稿日期:2011-02-11
基金项目:华北电力大学博士学位教师基金“概率分布拟合优度检验”(932090905)

令$Y=\alpha+MX$，其中$\alpha\in\mathbf{R}^d$，M为$d\times d$阶非奇异矩阵，$X\sim C(0_d,I_d,D_0)$．称Y服从中心为α，变换矩阵为M的中心相似分布（带有参数）[4]，记做$Y\sim C(\alpha,M_d,D_0)$．设$X\overset{d}{=}RZ^{(d)}$，$R^2\sim\chi^2_{d+2}$（自由度为$d+2$的卡方分布），$Z^{(d)}\sim U(D_0)$（D_0上的均匀分布），当D_0为d维单位球时，$Y\sim N_d(\alpha,MM')$．因此，多元正态分布是含参数的中心相似分布的特例．改变中心相似分布中随机变量R，$Z^{(d)}$的分布，就可以得到不同的多元分布．

标准中心相似分布可视为球对称分布的推广，因此椭球对称分布是中心相似分布的特例．中心相似分布可用来描述非对称性多元数据，其在多元数据分析中具有潜在的应用价值，相应的理论分析有待深入研究．由多元概率分布的构造理论可以看到，基于单位球（或单位球面）上的均匀分布，可以构造椭球对称分布，基于d维实心区域D_0上的均匀分布（或单位球面上非均匀分布），可以构造中心相似分布．

2 多元分布拟合优度检验主线

球对称分布、椭球对称分布的拟合优度检验可以转换为单位球上均匀分布的拟合优度检验，中心相似分布的拟合优度检验可以转换为单位球上非均匀分布的拟合优度检验．由此，得出一条多元分布拟合优度检验主线[5]：单位球面（或球体）上均匀分布的拟合优度检验⇒球对称分布的拟合优度检验⇒椭球对称分布（包含多元正态分布）的拟合优度检验．单位球面（或球体）上非均匀分布（或基本集D_0上均匀分布）的拟合优度检验⇒中心相似分布的拟合优度检验．因此，单位球面上多元分布的拟合优度检验具有重要意义．

下面简述(0，1)区间上均匀分布$U(0,1)$的基于间隔的检验方法及光滑检验，这2种检验的基本思想可应用到多元分布的拟合优度检验．设$X_1,\cdots,X_n$是抽自一维随机变量概率分布$F(\cdot)$的随机样本，$F_0(x)$是已知的连续分布函数．现欲检验假设

$$H_0:F=F_0. \tag{1}$$

令$U_i=F(X)$，$i=1,\cdots,n$，则可视$U_1,\cdots,U_n$为抽自(0，1)上均匀分布的随机样本．以$U_1,\cdots,U_n$为样本化检验(1)为

$$H_0:f(u)\equiv1,0<u<1, \tag{2}$$

其中$f(u)$是U_1的概率密度函数．

设$U_1,\cdots,U_n$是$i.i.d.$ $U(0,1)$随机变量，$U_{(i)}$，$i=1,\cdots,n$是其顺序统计量．令$U_{(0)}=0$，$U_{(n+1)}=1$，定义间隔$D_i=U_{(i)}-U_{(i-1)}$，$i=1,\cdots,n+1$．

Greenwood统计量G_n定义为

$$G_n=\sum_{i=1}^{n}D_i^2, \tag{3}$$

当G_n偏大时，拒绝样本$U_1,\cdots,U_n$为来自均匀分布$U(0,1)$．Jammalamadaka等[6]（2004）提出了基于Gini间隔指标的一维概率分布的拟合优度检验统计量

$$G_n^*=\sum_{i=1}^{n}\sum_{j=1}^{n}|D_i-D_j|. \tag{4}$$

Gini指标是观察到的一步间隔对（pair）所有绝对偏差和，其统计意义较Greenwood检验统计量G_n更为合理．Gini检验统计量G_n^*具有简洁性且能够快速收敛到其渐近分布．模拟功效表明Gini检验优于已有的基于间隔指标（包括Greenwood统计量）的主要拟合优度检验方法．

将Pearson χ^2检验应用于连续型总体，需对总体分布进行有限离散化．为了克服离散化过程的不确定性，Neyman[7]（1937）提出了光滑检验，即将原假设中的概率密度f_0嵌入到对立假设中的概率分布族中，化总体分布的拟合优度检验为参数检验．由此，参数检验的一些优良的统计性质，可转化为拟合优度检验的优良性．设$\{\pi_i(y)\}$为关于(0，1)区间上均匀分布的正交函数系，对立分布概率密度族定义为

$$g_k(y,\theta)=C(\theta)\exp\{\sum_{i=1}^{k}\theta_i\pi_i(y)\},0<y<1,\tag{5}$$

则(0,1)区间上均匀分布的拟合优度检验转化为

$$H_0:\theta=0,H_1:\theta\neq 0.\tag{6}$$

其中参数$\theta=(\theta_1,\cdots,\theta_i)^{\mathrm{T}}$. 设$Y_1,\cdots,Y_n$为来自总体的样本,Neyman构造检验统计量为

$$\Psi_k^2=\sum_{i=1}^{k}U_i^2,U_i=\sum_{j=1}^{n}\pi_i(Y_j)/\sqrt{n},\tag{7}$$

当Ψ_k^2取值偏大时,拒绝原假设$H_0:\theta=0$. 这里Ψ_k^2的渐近分布是自由度为k的卡方分布.

Boulerice等[8](1997)研究了d维球面Ω_d上的概率分布的光滑检验. 此时,(7)式中的$\{\pi_i(y)\}$被替换为关于球面上概率密度f_0的正交函数系$\{h_i(y),y\in\Omega_d\}$. 多元概率分布的光滑检验是值得关注的一个研究方向.

3　单位球均匀分布的拟合优度检验

杨振海等[9](2007)研究了单位球均匀分布的拟合优度检验. 基于条件积分变换及单位球均匀分布的充要条件表示定理,构造球均匀性检验统计量,得到其渐近分布为卡方分布,随机模拟表明,检验统计量能够快速收敛到其渐近分布. 证明了拟合优度检验的相合性,并做拟合优度检验相合性的随机模拟. 我们提出的球均匀性检验方法,具有同Gini检验方法的检验功效. 单位球均匀分布的拟合优度检验的主要结论,可以推广L_α到模单位球均匀分布的拟合优度检验. 设$x=(x_1,\cdots,x_d)^{\mathrm{T}}$,定义$L_\alpha$模为

$$\|\boldsymbol{x}\|_\alpha=\left(\sum_{i=1}^{d}|x_i|_\alpha\right)^{\frac{1}{\alpha}},\alpha>0.\tag{8}$$

半径为r的L_α模球为$S_d^\alpha(r)=\{\boldsymbol{x}:\|\boldsymbol{x}\|_\alpha\leq r\}$. L_α模球$S_d^\alpha(r)$的体积[10]为

$$r^d\left(\frac{2}{\alpha}\right)^d\frac{\Gamma^d(1/\alpha)}{\Gamma(d/\alpha+1)}.\tag{9}$$

我们得到了L_α模单位球均匀分布检验统计量的渐近分布,并证明了该检验具有相合性.

Su等[11](2010)推广L_α模单位球均匀分布的拟合优度检验到任意有界实心区域D_0上均匀分布的拟合优度检验. 在边界已知条件下,较好地解决了中心相似分布中基本集D_0的均匀性拟合优度检验问题. 举例说明L_1模$^+$单位球均匀分布拟合优度检验统计量的构造,L_1模$^+$单位球均匀分布即为参数为(1,1,…,1)的Dirichlet分布,其在Bayes统计分析及成分数据的统计分析中有重要应用[2].

4　多元正态分布的拟合优度检验

多元数据分析是SAS软件的核心模块,多元正态分布是经典多元分析中的基本假定. 无论在理论研究还是在实际应用中,多元正态分布的拟合优度检验始终受到高度重视. Mardia提出了多元正态性偏度、峰度检验统计量. 基于球对称分布下t统计量具有统计不变性的特点,Liang J.等[12](2004)给出了多元正态性的$Q-Q$图检验. Székely等[13](2005)提出了多元正态性检验统计量$\hat{\varepsilon}_{n,d}$. 设$X_1,\cdots,X_n$为来自分布F的样本,F_0为$N_d(\mu,\Sigma)$. 做变换$Y_j=\hat{\Sigma}^{-1/2}(X_j-\hat{\mu})$.

$$\hat{\varepsilon}_{n,d}=n\left(\frac{2}{n}\sum_{j=1}^{n}\|y_j-Z\|-2\frac{\Gamma\left(\frac{d+1}{2}\right)}{\Gamma\left(\frac{d}{2}\right)}-\frac{1}{n^2}\sum_{j,k=1}^{n}\|y_j-y_k\|\right),$$

$$E\|a-Z\|=\frac{\sqrt{2}\,\Gamma\left(\frac{d+1}{2}\right)}{\Gamma\left(\frac{d}{2}\right)}+\sqrt{\frac{2}{\pi}}\sum_{k=0}^{\infty}\frac{(-1)^k}{k!2^k}\frac{\|a\|^{2k+2}}{(2k+1)(2k+2)}\frac{\Gamma\left(\frac{d+1}{2}\right)\Gamma\left(k+\frac{3}{2}\right)}{\Gamma\left(k+\frac{d}{2}+1\right)}.$$

其中$Z\sim N_d(O_d,I_d)$.

在原假设下，通过随机模拟得到$\hat{\varepsilon}_{n,d}$的有限样本经验分位数. 当$\hat{\varepsilon}_{n,d}$取值偏大时，拒绝数据来自多元正态分布. 文献[13]的功效模拟表明，多元正态性的$\hat{\varepsilon}_{n,d}$检验，优于Mardia提出的偏度、峰度检验，基于经验特征函数的BHEP检验等多元正态性检验.

基于单位球均匀分布拟合优度检验，苏岩等(2010)研究了多元正态分布的$N_d(\mu,\Sigma)$的VDR条件拟合优度检验[14]. 基于垂直密度表示(VDR)，转换多元正态性检验为球均匀性检验. 多元正态分布转换样本$Y\overset{d}{=}RV^{(d)}$服从Pearson Ⅱ型分布，证明了R^2服从贝塔分布. 基于贝塔分布和单位球均匀分布，提出了多元正态性检验统计量χ^2，证明了该检验统计量的渐近分布为卡方分布. 在检验统计量χ^2的构造中，不用估计未知参数μ,Σ，故在理论上具有较高的检验功效. 功效模拟显示，χ^2检验统计量优于Székely提出的$\hat{\varepsilon}_{n,d}$检验等已有主要多元正态性检验统计量.

5 球面均匀分布的拟合优度检验

Justel等[15](1997)提出了多元分布的Kolmogorov-Smirnov型拟合优度检验. 利用条件积分变换及一维概率分布的Kolmogorov-Smirnov型检验，构造多元分布的拟合优度检验统计量. 当维数$d>2$时，存在检验统计量数值计算量过大的问题. Huffer等[16](2002)研究了多元数据的结构检验. 基于Pearson χ^2检验，构造检验统计量. 方开泰等[17](1990)提出了基于惯量矩的三维球面上均匀分布的拟合优度检验，指出三维球面上样本点惯量矩对应矩阵的特征根近似相等时，样本点来自均匀分布.

设$\boldsymbol{X}_i^{(d)}=(X_{1i},\cdots,X_{di})^{\mathrm{T}},i=1,\cdots,n$为$n$个单位质量的单位球面上的质点，$\boldsymbol{H}=(\omega_1,\cdots,\omega_d)^{\mathrm{T}}$是空间任一固定方向. n个样本点关于$\boldsymbol{H}$的惯量矩$\boldsymbol{M}$定义为这些点到方向$\boldsymbol{H}$的垂直距离平方之和

$$\boldsymbol{M}=\boldsymbol{H}^{\mathrm{T}}\boldsymbol{B}\boldsymbol{H},\boldsymbol{B}=n\boldsymbol{I}_d-\boldsymbol{A},\boldsymbol{A}=\sum_{i=1}^{n}\boldsymbol{X}_i^{(d)}(\boldsymbol{X}_i^{(d)})^{\mathrm{T}}. \tag{10}$$

引理1[5] 设$X=(X_1,\cdots,X_d)^{\mathrm{T}}$是$d$维随机变量，$\|X\|=1$，$Cov(X)$存在，则$X$对任意方向$\boldsymbol{H}=(\omega_1,\cdots,\omega_d)^{\mathrm{T}}$的期望惯量矩与$H$无关的充要条件是

$$E(X_i^2)=\frac{1}{d},i=1,\cdots,d;E(X_iX_j)=0,i\neq j,i,j=1,\cdots,d. \tag{11}$$

引理2[5] 设$X_i^{(d)}=(X_{1i},\cdots,X_{di})^{\mathrm{T}},i=1,\cdots,n$为$n$个单位球面上的样本点，$H$是空间任一固定方向，则当矩阵$\boldsymbol{B}$的特征根$\beta_1,\cdots,\beta_d$相等时，球面上的$n$个样本点对任意方向$H$的惯量矩恒相等.

引理1，引理2表明球面上的n样本点对任意方向H的惯量矩恒相等，是样本点服从球面均匀分布的重要特征. 基于质心和惯量矩，苏岩等提出了球面均匀分布拟合优度检验统计量[5]. 证明了d维单位球面上样本服从均匀分布的基本特征. 得到球面上随机变量的期望惯量矩对任意方向恒相等的充要条件. 证明了球面均匀分布协差阵特征根估计的强相合性及渐近多元正态性. 证明了检验统计量的极限分布为卡方分布及拟合优度检验的相合性. 样本质心，惯量矩对应着球面均匀分布的一阶矩，二阶矩. 拟合优度检验相合性的随机模拟表明，基于质心检验与惯量矩检验的球面均匀性的联合检验是必要的.

为了更好地反映球面均匀分布的概率特性，笔者做了进一步的研究工作，将广义逆引入到球面均匀分布的拟合优度检验.

6 椭球对称分布的拟合优度检验

椭球对称分布是广义多元分析的基本假设，椭球对称分布具有许多类似于多元正态分布的优良性质. 当数据来自椭球对称分布时，经典数据分析中的一些方法可用来做服从椭球对称分布的数据分析. Zhu等[18](2003)提出了椭球对称分布的条件拟合优度检验，利用球面上均匀分布随机数，生成新的样本. 基于条件

经验过程及新生成的样本构造检验统计量，得到造检验统计量依分布收敛到高斯过程. 设X_i，$i=1$，…，n为来自椭球对称分布的样本

$$\bar{X}=n^{-1}\sum_{i=1}^{n}X_i, S=n^{-1}\sum_{i=1}^{n}(X_i-\bar{X})(X_i-\bar{X})^{\mathrm{T}},$$

$$Z_i=G(X_i-\bar{X}), i=1,\cdots,n, \tag{12}$$

其中，矩阵G，满足$GSG=I$.

当样本X_1，…，X_n来自椭球对称分布时，其球化转换样本$Y_j=S^{-1/2}(X_j-\bar{X})$应渐近服从球面上的均匀分布. 据此，Manzotti等[19]（2002）提出了基于球调和函数的椭球对称分布拟合优度检验. Fred W. Huffer等[20]（2007）研究了椭球对称分布的拟合优度检验，基于Pearson χ^2检验及球化样本Z_i，$i=1$，…，n，构造检验统计量.

7　回归模型的有效性检验

多元概率分布的拟合优度检验可应用于回归模型的有效性检验. M. D. Jimenez Gamero等[21]（2005）研究了多元线性模型中误差分布的拟合优度检验，利用经验特征函数建立检验统计量，得到原假设下检验统计量的渐近分布. 设Y_1，…，Y_n为取值于$\mathbf{R}^d$中的独立随机变量，满足模型

$$Y_j^{\mathrm{T}}=X_j^{\mathrm{T}}\beta+\varepsilon_j^{\mathrm{T}}\Sigma^{1/2}, j=1,2,\cdots,n. \tag{13}$$

随机变量$X_j\in\mathbf{R}^p$不具随机性，β为未知的$p\times d$矩阵，Σ为未知的$d\times d$矩阵，ε_j为未知的d维误差随机变量. $E(\varepsilon_j)=0$，$Var(\varepsilon_j)=I_d$，$j=1,2,\cdots,n$. 欲检验ε_j是否来自给定的分布，即原假设为

$$H_0:F=F_0.$$

设$\hat{\beta}$，$\hat{\Sigma}$为β，Σ的估计，$e_j^{\mathrm{T}}=(Y_j^{\mathrm{T}}-X_j^{\mathrm{T}}\hat{\beta})\hat{\Sigma}^{-1/2}$，基于$e_1$，…，$e_n$检验$H_0$，即基于残差集进行$d$维误差分布的拟合优度检验.

Zhu等[22]（2008）提出基于非参数Monte Carlo检验（NMCT），进行多元回归模型诊断. 在误差分布为椭球对称分布条件下，给出了检验回归函数$m(\cdot)$为某给定形式的检验步骤及算法，原假设为

$$H_0:m(\cdot)=G(\beta,\cdot), \tag{14}$$

NMCT借助生成球面上均匀分布随机数，避开了对检验统计量渐近协差阵的估计. Baraud等[1]（2002）研究了《A New Test of Linear Hypothesis in Regression》，给出了误差分布为正态分布时的模型检验. Anderson等[23]研究了误差分布为球对称分布时线性模型回归参数的假设检验，得到似然比检验统计量服从F分布. Crainiceanu等[24]（2004）研究了非线性回归模型似然比拟合优度检验，将非线性回归模型嵌入到一类半参数模型中，通过回归模型参数MLE的一阶Taylor逼近，转化非线性回归模型的拟合优度检验为线性回归模型的拟合优度检验. Chai等[25]（1992）基于概率密度核估计，给出了线性回归模型中误差分布的非参数估计，可据此进行误差分布的拟合优度检验. 苏岩等[26]（2007）研究了趋势时间序列模型，进行了白噪声残差检验的实证分析. Paula等[27]（2009）研究了具有一阶自回归椭球误差分布的线性模型诊断. Sun等[28]研究了随机删失下偏线性模型的假设检验，基于经验过程构造检验统计量.

对任意d元非零随机向量X_d，均有

$$X_d=\|X_d\|\cdot\frac{X_d}{\|X_d\|}, \tag{15}$$

即X_d等于其模长与方向向量的乘积.因此，球面数据均匀分布（或非均匀分布）的拟合优度检验，对椭球对称分布（或中心相似分布）假设下的多元分析及时间序列分析，具有重要意义.（15）式表明，将对多元分析的研究可以转换为球面上的统计分析. 球极投影变换将$\mathbf{R}^d$中的点一一映射到$d+1$维单位球面上. 赵颖等[29]（2005）利用一维核函数构造多维概率密度的球极投影变换核估计，将多维问题转化为一维问题.模拟表明，球极投影变换核估计优于普通的多维核密度估计. 球极投影变换核估计，可以改进高维数据相对稀疏

情形下的非参数统计推断. 基于球极投影变换核估计,Su[30] (2010) 给出了一种新的非正态总体假设下的非参数判别准则,得到其条件错判概率的几乎处处收敛性.故基于球极投影变换及球调和分析,可进行非对称分布假设下的回归模型的有效性分析.

8 讨论

综上所述可以看到,统计学者从不同角度,使用不同方法探讨多元分布的拟合优度检验. 包括Pearson χ^2型检验,EDF型检验,非参数U-统计量、V-统计量检验,多元分布特征检验,Monte Carlo检验,样本变换检验,经验特征函数检验等. 统计研究不但关注优良的理论结果,更关注统计理论的可实际应用性. 从一元到多元,统计研究会变得非常复杂,关键是"维数"问题. 在应用中,这直接反映为对样本容量的需求状况. 在多元分布的拟合优度检验中,可能从不同侧面拒绝原假设. 提出转换多元分布的拟合优度检验为球面 (或单位球) 上概率分布的拟合优度检验. 关于球面 (或单位球) 上均匀分布的拟合优度检验,得到了一些研究成果.

椭球对称分布,中心相似分布为半参数结构模型. 故在广义多元数据研究中,需将分布模型与非参数统计方法有机地结合起来,即需使用小样本理论与大样本理论. 推广球面上概率分布的拟合优度检验至各种回归模型 (包括时间序列模型) 的有效性检验,这方面的工作还有待做深入研究.

参考文献:

[1] HUBER-CAROL C,BALAKRISHNAN N,NIKULIN M S,et al. Goodness-of-fit tests and model validity [M]. Boston·Basel·Berlin:Birkhäuser,2002.

[2] FANG K T,KOTZ S,NG K W. Symmetric multivariate and related distributions [M]. London·New York:Chapman and Hall,1990.

[3] YANG Z H,KOTZ S. Center-similar distributions with applications in multivariate analysis [J]. Statistics & Probability Letters, 2003 (64):335-345.

[4] 戴家佳,苏 岩,杨爱军,等. 中心相似分布的参数估计 [J]. 应用数学学报,2008,31 (3):480-491.

[5] 苏 岩,杨振海. 球面均匀分布的拟合优度检验 [J]. 应用数学学报,2009,32 (1):93-105.

[6] JAMMALAMADAKA S R,GORIA M N. A test of goodness-of-fit based on Gini's index of spacings [J]. Statistics & Probability Letters,2004 (68):177-187.

[7] RAYNER J C W,BEST D J. Smooth tests of goodness of fit [M]. New York,Oxford:Oxford University Press,1989.

[8] BOULERICE B,DUCHARME G R. Smooth tests of goodness-of-fit for directional and axial data [J]. Journal of Multivariate Analysis,1997 (60):154-175.

[9] 杨振海,苏 岩. 单位球均匀分布的拟合优度检验 [J]. 北京工业大学学报,2007,33 (7):771-777.

[10] SZABLOWSKI P J. Uniform distributions on spheres in finite dimensional and their generalizations [J]. Journal of multivariate analysis,1998 (64):103-117.

[11] SU Y,SHI H F. Goodness-of-fit analysis for uniformity in a bounded solid region,in:Proceeding of the Ninth International Conference on Machine Learning and Cybernetics,Qingdao,2010 (1):231-235.

[12] LIANG J J,PAN W S Y,YANG Z H. Characterization-based Q-Q plots for testing multinormality [J]. Statistics & Probability Letters,2004 (70):183-190.

[13] SZÉKELY G J,RIZZO M L. A new test for multivariate normality [J]. Journal of Multivariate Analysis,2005 (93):58-80.

[14] 苏 岩,杨振海. 多元正态分布的VDR条件拟合优度检验 [J]. 应用概率统计,2010 (26):234-244.

[15] JUSTEL A,PEÑA D,ZAMAR R. A multivariate kolmogorov-Smirnov test of goodness of fit [J]. Statistics & Probability Letters, 1997 (35):251-259.

[16] HUFFER F W,PARK C. The limiting distribution of a test for multivariate structure [J]. Journal of Statistical Planning and Inference,2002 (105):417-431.

[17] 方开泰,范剑青,金 辉,等. 方向数据的统计分析 [J]. 数理统计与管理,1990 (2):59-65.

[18] ZHU L X,NEUHAUS G. Conditional tests for elliptical symmetry [J]. Journal of Multivariate Analysis,2003 (84):284-298.

[19] MANZOTTI A,PÉREZ F J. A Statistic for testing the null hypothesis of elliptical symmetry [J]. Journal of Multivariate Analysis, 2002 (81):274-285.

[20] HUFFER F W,PARK C. A test for elliptical symmetry [J]. Journal of multivariate analysis,2007 (98):256-281.

[21] JIMÉNEZ GAMERO M D,MUÑOZ GARCÍA J,PINO MEJÍAS R. Testing goodness of fit for the distribution of errors in multivariate linear models [J]. Journal of multivariate analysis,2005 (95):301-322.

[22] ZHU L X,ZHU R Q,SONG S. Diagnostic checking for multivariate regression models [J]. Journal of Multivariate Analysis,2008 (99):1841-1859.

[23] 方开泰,张尧庭. 广义多元分析 [M]. 北京:科学出版社,1993.

[24] CRAINICEANU C M,RUPPERT D. Likelihood ratio tests for goodness-of-fit of a nonlinear regression model [J]. Journal of Multivariate Analysis,2004 (91):35-52.

[25] CHAI G X,SU Y,SHI S J. A nonparametric approach to the estimation of error distributions in linear model,in:Probability and statistics [M]. Edited by JIANG Z P,YAN S J,CHENG P,WU R. Singapore·New Jersey·London·Hong Kong:World Scientific,1992:1-11.

[26] 苏 岩,杨振海,李双杰. 增长型经济变量的趋势时间序列预测模型 [J]. 数学的实践与认识,2007,37 (3):4-8.

[27] PAULA G A,MEDEIROS M,VILCA-LABRA F E. Influence diagnostics for linear models with first-order autoregressive elliptical errors [J]. Statistics and Probability Letters,2009 (79):339-346.

[28] SUN Z H,Wang Q H,Dai P J. Model checking for partially linear models with missing responses at random [J]. Journal of Multivariate Analysis,2009 (100):636-651.

[29] 赵 颖,杨振海. 球极投影变换核估计及其逐点收敛速度 [J]. 数学年刊,2005 (26A,1):19-30.

[30] SU Y,YANG Z H. The strong consistency of the conditional probability of error in discrimination based on kernel stereographic projection density estimator,in:Data Processing and Quantitative Economy Modeling [M]. Edited by ZHU K L, ZHANG H. Sydney:Aussino Academic Publishing House,2010:508-511.

The Study of Goodness-of-Fit Tests for Multivariate Distributions

Su Yan

(School of Mathematics and Physics,North China Electric Power University, Baoding 071003,China)

Abstract: The construction of multivariate distributions and the various methods of goodness-of-fit tests for multivariate distributions are introduced,the important action of goodness-of-fit tests in multivariate analysis is pointed out. The distinctive theoretical meaning of goodness-of-fit tests based on unit sphere for multivariate distributions is expounded,the exploratory road map of model validity tests is stated,the synthetic use of small sample theory and large sample theory in statistical model tests is emphasized.

Key words: multivariate distribution; uniform distribution; unit sphere; goodness-of-fit test