系 统 科 学 与 数 学
J. Sys. Sci. & Math. Scis.
x(x) (20xx, x), 1–19

测量误差为Laplace分布的非线性统计推断*

史建红

(山西师范大学数学与计算机科学学院, 临汾 041004)

宋卫星

(Department of Statistics, Kansas State University, Manhattan, KS, 66503)

摘要 当p-维参数θ通过矩条件$Em(X,\theta)=0$定义，且X带有Laplace测量误差时，即我们只能观测到$Z=X+U$，文[7]构造了θ的 一个简单估计方法. 然而该估计方法仅适用于U的各分量服从Laplace分布且相互独立的情况. 本文我们将介绍一种一般的多元Laplace分布， 并将文[7] 的方法推广到具有这种多元Laplace分布的测量误差模型中. 另外，文[7]的方法是基于无条件期望关系 $Em(X,\theta)=EH(Z,\theta)$，其中$H$ 为某个形式已知的函数，而该方法对一些统计推断问题并不适用. 本文我们将构造一种基于条件期望 $E[m(X,\theta)|Z]$ 的估计方法. 当X为一维时，我们对这些估计的大样本性质进行了讨论.

关键词 非线性统计推断, 测量误差, Laplace分布, 偏差校正.

MR(2000)主题分类号 62J02, 62F10

Nonlinear Statistical Inferences With Laplace Measurement Error

SHI Jianhong

(School of Mathematics and Computer Science, Shanxi Normal University, Linfen, 041004)

SONG Weixing

(Department of Statistics, Kansas State University, Manhattan, KS, 66503)

Abstract When a p-dimensional parameter θ is defined through the moment condition $Em(X,\theta)=0$, a simple estimation procedure of θ was proposed in [7] when X,

*美国国家自然科学基金(NSF DMS 1205276)资助课题.
通讯作者: 宋卫星，weixing@ksu.edu
收稿日期: 200x-xx-xx, 收到修改稿日期: 200x-xx-xx.
编委:

a k-dimensional random vector, are contaminated with Laplace measurement error U, that is, we can only observe $Z = X + U$. However, the estimation procedure was designed particularly for the cases where the components of the measurement error vector U are independent. In this paper, we first introduce a general multivariate Laplace distribution, then extend the methodology in [7] to this general multivariate scenario. Moreover, the moment estimation procedure in [7] is based on the unconditional expectation $Em(X, \theta) = EH(X, \theta)$ for some function H. Example shows this techniques does not work in some cases. In this paper, we will propose an estimation procedure based on the condition expectation $E(m(X, \theta)|Z)$. Large sample properties of the proposed estimation procedure when X is one-dimensional are discussed.

Keywords Nonlinear Statistical Inference, Measurement Error, Laplace Distribution, Bias Correction.

1 引 言

非线性统计模型是统计学、经济计量学、生物统计学、农业统计学等众多学科中应用最为广泛的一类解释变量之间关系的统计模型. 较之于线性模型，非线性模型在刻划数据结构、拟合变量关系方面更加灵活，而相对于非参数模型来说，非线性模型的参数往往有着非常直观的实际解释，统计推断程序也往往由数据决定，鲜有人为介入. 在许多情况下，非线性模型中参数θ可由一系列总体矩估计方程决定：

$$Em(X;\theta) = 0 \tag{1.1}$$

其中X为一个k-维向量，$\theta \in \Theta \subset R^p$为待估参数向量，$m$为一个$d$维向量函数，通常有$d \geq p$.然而在实践中，可能由于测量工具的不精确，也可能由于测量成本的昂贵，抑或是由于人为因素，我们观测不到X的值，观测到的是一个与X有关的变量Z的值. 例如，在美国癌症研究所进行的蛋白与能量营养的研究(OPEN) 中，一个重要的观测指标为人们饮食中的蛋白摄入量. 然而这个指标的观测不但成本昂贵，而且程序繁琐. 最后，科学家选取了另外一个变量，即尿蛋白含量Z，来作为X 的一个替代变量. 关于OPEN 研究的具体情况，可参考文[13]的工作；文[14]研究了儿童喘鸣病与其接触二氧化氮量之间的关系，在该研究中，响应变量为Y, $Y = 1$表示儿童患有喘鸣病，$Y = 0$表示没有.如果解释变量接触二氧化氮量X可以观测的话，Logistic或者Probit 回归通常会用来对Y 与X的关系进行建模. 然而，在文[14]研究中， X的值观测不到，所能观测到的是儿童所在家庭中厨房与卧室里的二氧化氮含量Z. 专著[2]描述了许多来自营养学、生物检测学、医学等学科中涉及到测量误差的实例. 如何通过可观测数据(Y, Z)来研究Y与X 之间的关系即成为测量误差模型的重要的研究内容. 文献中关于Z与X之间的关系有很多讨论，有明确统计关系的包括可加结构$Z = X + U$，Berkson结构$X = Z + U$，回归结构$Z = \alpha + \gamma X + U$，与乘积结构$Z = XU$ 等，专著[2]的第一、二章对此也做了很多讨论. 没有明确统计关系的结构基本局限于一些辅助数据或者验证数据(Instrumental Data/Validation Data)存在的非参数推断中. 本文我们将采用常见的可加误差结构: $Z = X + U$，其中U 表示测量误差，X与U相互独立.

众所周知，简单的将Z取代X来进行统计推断会导致有偏的估计或低效的检验. 而要寻求

消除或降低测量误差对统计分析造成的影响，就需要加强对模型的假设或者收集更多的辅助数据或者验证数据. 在后者不存在的情况下，测量误差模型中通常会假设U的分布已知.在参数模型中，尤其是在线性或多项式模型中最常见的是U服从$N(0,\Sigma)$. 然而在其他稍微复杂的参数模型中，基于正态假设的统计推断程序并不简单，尤其是在非参数统计中，基于光滑技术的回归函数估计有着非常缓慢的收敛速度. 文[5]深入讨论了测量误差对非参数回归估计相合速度的影响，并且根据测量误差的特征函数对其进行了分类. 如果特征函数的尾部以指数速度衰减至0,则称该测量误差具有超光滑性(super smooth)；如果特征函数的尾部以多项式速度趋于0的话，则称该测量误差具有一般光滑性(ordinary smooth).正态分布是具有超光滑性的典型代表，而具有一般光滑性的典型例子是Laplace 分布.Laplace分布在语音识别与图像压缩方面有着广泛的应用. 在统计学中，由于最小绝对偏差估计与基于拉普拉斯分布的极大似然估计有着天然的联系，所以其在稳健性研究中也占有一席之地.

在假设U的分量相互独立，其各自服从Laplace分布的条件下，文[7]提出了估计模型(1.1)中参数θ的一种简单方法. 为介绍方便，记$U=(U_1,U_2,\ldots,U_k)$, U_j 的密度函数为$f(u)=\exp(-\sqrt{2}|u|/\sigma_j)/(\sqrt{2}\sigma_j)$, U_j, $j=1,2,\ldots,k$相互独立. 另记

$$\overline{m}(Z;\theta,\sigma)=m(Z;\theta)+\sum_{l=1}^{k}\left(-\frac{1}{2}\right)^{l}\sum_{j_1<\cdots<j_l}\sigma_{j_1}^2\cdots\sigma_{j_l}^2\left(\frac{\partial^{2l}m(Z;\theta)}{\partial Z_{j_l}^2\cdots\partial Z_{j_l}^2}\right).$$

在对函数$m(x;\theta)$ 施加了一些光滑性条件后，文[7]证明了如下的重要公式

$$Em(X;\theta)=E\overline{m}(Z;\theta,\sigma).$$

根据上述结论，矩条件$Em(X;\theta)=0$即可以由矩条件$E\overline{m}(Z;\theta,\sigma)=0$来代替. 值得注意的是，尽管这个新的矩条件只依赖于Z，但是却引入了另一个可能未知的参数σ. 如果$E\overline{m}(Z;\theta,\sigma)=0$可以唯一决定$\theta,\sigma$，文[7]给出了基于

$$(\widehat{\theta},\widehat{\sigma})=\operatorname{argmin}_{\theta,\sigma}\left(\sum_{i=1}^{n}\overline{m}(Z_i;\theta,\sigma)\right)'W_n\left(\sum_{i=1}^{n}\overline{m}(Z_i;\theta,\sigma)\right)$$

的修正矩估计. 在一些正则条件下，$\widehat{\theta},\widehat{\sigma}$的渐近正态性得到了证明. 如果$E\overline{m}(Z;\theta,\sigma)=0$不足以唯一决定$\theta,\sigma$，利用$\sigma_j^2$不超过$Z_j$方差的事实，文[7]构造了$\theta$的一个可能取值集合.

尽管文[7]工作的重要性不言而喻，然而该方法也有着明显的缺陷. 首先该方法只适用于U的各分量相互独立的情形，而独立性对于涉及多元变量的统计推断来说是一个非常苛刻的假设；另外$Em(X;\theta)=E\overline{m}(Z;\theta,\sigma)$是有关无条件期望的等式. 基于无条件期望的统计推断在很多时候差强人意，尤其是在回归模型的讨论中，即使我们考虑的是X为一维的情形. 关于后者，我们会在后面的讨论中用Poisson回归模型作为例子来做进一步的阐述.

如果U服从正态分布，类似的无条件期望等式很难得到. 当$k=1$, $m(x;\theta)=x^p-\theta$时，文[7]给出了如下的等式，

$$Em(X;\theta)=EZ^p+\sum_{l=1}^{[p/2]}(-1)^l\left(\frac{\sigma^2}{2}\right)^l\frac{p!}{l!(p-2l)!}Z^{p-2l}-\theta.$$

如果U服从多元正态分布，类似的等式也很容易得到. 但是对于一般的非多项式线性函数$m(x,\theta)$，类似的结论还没有得到.

本文的结构安排如下，在第二节中，我们将介绍一种多元Laplace分布的定义及这种分布的简单性质，随后在第三节中，基于这种多元Laplace分布，我们会陈述一个类似

于$Em(X;\theta)=E\overline{m}(X;\theta,\sigma)$的无条件期望等式，并且建立平行于文[7]的估计方法与估计的大样本性质. 如此，我们就将在独立性假设之下建立的统计推断程序推广了一般多元的情况中. 在第四节，我们会在X 分布完全未知的情况下，给出$E(m(X;\theta)|Z)$的表达式，然后我们会利用这个条件期望的表达式，来改进基于无条件期望的估计方法. 所有理论结果的证明都放在了第五节.

2 多元Laplace分布

事实上，迄今为止，多元Laplace分布仍然没有统一的定义. 比较常见的是由文[4,8]提出的基于特征函数的定义. 事实上文[4]所定义的多元Laplace 分布是文[8]定义的多元Laplace分布的一个特例. 在本文的讨论中，我们采用文[4]的定义. 根据该定义，一个k 维向量X 的特征函数如果具有形式

$$\phi(t)=\frac{\exp(i\mu' t)}{1+t'\Sigma t/2},\quad t\in R^k,$$

则称X服从参数为(μ,Σ)的多元Laplace分布，记为$X\sim L_k(\mu,\Sigma)$，其中μ为一个k-维向量，Σ为一个$k\times k$正定矩阵. 很显然，如此定义的多元Laplace 分布是基于一维Laplace 分布特征函数的直接推广. 分布$L_k(\mu,\Sigma)$有如下的一些重要性质：

(1). 如果$U\sim L_k(\mu,\Sigma)$，那么从分布相等的意义上来说，$U=\mu+\sqrt{V}\Sigma^{1/2}W$，其中$V$服从均值为1的指数分布，$W$服从多元标准正态分布，$V$ 与W 相互独立;

(2). 如果$U\sim L_k(\mu,\Sigma)$，那么其密度函数具有如下形式

$$\psi(u)=\frac{2}{(2\pi)^{k/2}}\frac{B_{(d/2-1)}(\sqrt{2q(u;\mu,\Sigma)})}{(2q(u;\mu,\Sigma))^{d/2-1}},\quad u\in R^k, \tag{2.2}$$

其中$q(u;\mu,\Sigma)=(u-\mu)'\Sigma^{-1}(u-\mu)$，$B_r(x)$是阶为$r$的第二类修正Bessel函数；

(3). 如果$U\sim L_k(\mu,\Sigma)$，则$EU=\mu$, $\mathrm{Cov}(U)=\Sigma$.

性质(1)说明多元Laplace分布实际上是多元正态分布的一种刻度混合，该性质不仅提供了一种产生多元Laplace分布的方法，而且指出了利用该分布进行稳健统计推断时使用EM 算法的可能性. 这方面的工作可参看文[11,12]关于$k=1$时的讨论. 性质(3)表明参数Σ 确实刻画了U中各分量之间的关系.

由于U关于0对称是测量误差模型文献中常用的一个条件，所以在下面的讨论中， 我们假设$\mu=0$.

3 无条件矩估计

以f,g分别记X与Z的密度函数，以$\phi_z(t)$, $\phi_x(t)$与$\phi_u(t)$分别记Z,X与U的特征函数. 由可加结构$Z=X+U$，以及X与U的相互独立，易见$\phi_z(t)=\phi_x(t)\phi_u(t)$. 因为对所有的$t\in R^k$，$\phi_u(t)=(1+t'\Sigma t/2)^{-1}\neq 0$，所以很容易得到$\phi_x(t)=(1+t'\Sigma t/2)\phi_z(t)$.利用此事实，我们可以进一步得到下面的结论.

引理 3.1 假设$\phi_Z(t)$平方可积，则

$$f(x) = g(x) - \frac{1}{2}\sum_{j,l=1}^{k} \sigma_{jl} \frac{\partial^2 g(x)}{\partial x_j \partial x_l},$$

其中σ_{jl}为Σ的(j,l)元素，x_j为x的第j个分量.

基于上述引理，我们可以得到下面的无条件期望等式.

定理 3.1 假设参数函数$m(x,\theta)$与密度函数$g(z)$满足下面的条件

(C1). 对每个$\theta \in \Theta$，参数函数$m(x,\theta)$关于x二阶可导，并且当$\|x\| \to \infty$时，

$$m(x,\theta)g'(x) \to 0, \quad m'(x,\theta)g(x) \to 0.$$

(C2). 对每个$\theta \in \Theta$，

$$E\|m(Z,\theta)\| < \infty, \quad E\left\|\frac{\partial^2 m(Z,\theta)}{\partial Z_j \partial Z_l}\right\| < \infty,$$

则我们有

$$Em(X,\theta) = Em(Z,\theta) - \frac{1}{2}\sum_{j,l=1}^{k} \sigma_{jl} E\frac{\partial^2 m(Z,\theta)}{\partial Z_j \partial Z_l}. \tag{3.3}$$

借助引理3.1，定理3.1的证明非常简单. 事实上，利用定理条件并且进行分部积分，我们可以证明，对所有的$j,l=1,2,\dots,k$,

$$\int m(x,\theta)\frac{\partial^2 g(x)}{\partial x_j \partial x_l}dx = E\frac{\partial^2 m(Z,\theta)}{\partial Z_j \partial Z_l}.$$

和文[7]关于独立Laplace测量误差分量的结果相比较，定理3.1的结论从形式上看非常简单，无条件期望等式的右端只涉及到参数函数$m(z,\theta)$关于z的二阶偏导. 基于定理3.1，我们可以得到一系列平行于文[7]的结果. 为方便讨论，记

$$\widetilde{m}(Z;\theta,\Sigma) = m(Z,\theta) - \frac{1}{2}\sum_{j,l=1}^{k} \sigma_{jl} \frac{\partial^2 m(Z,\theta)}{\partial Z_j \partial Z_l}.$$

3.1 可识别情况下的点估计

假设矩条件$E\widetilde{m}(Z;\theta,\Sigma)=0$可以唯一确定$\theta$, Σ. 对任意的对称正定矩阵W_n, 定义

$$(\widehat{\theta},\widehat{\Sigma}) = \operatorname{argmin}_{\theta,\Sigma}\left(\sum_{i=1}^{n}\widetilde{m}(Z_i;\theta,\Sigma)\right)' W_n \left(\sum_{i=1}^{n}\widetilde{m}(Z_i;\theta,\Sigma)\right). \tag{3.4}$$

同时，我们假设

(C3). $E\widetilde{m}(Z;\theta,\Sigma)=0$ 当且仅当$\theta=\theta_0$, $\Sigma=\Sigma_0>0$, 其中θ_0,Σ_0为真实参数值；

(C4). 对某个正定对称矩阵，W_n依概率收敛于W;

(C5). 以σ记向量$(\sigma_{jl}, j\ge l)$, $\alpha=(\theta',\sigma')'.E\partial\widetilde{m}(Z;\theta,\Sigma)/\partial\alpha$ 存在，列满秩，并且在α_0的某个邻域内，$\partial\widetilde{m}(Z;\theta,\Sigma)/\partial\alpha$满足一阶Lipschitz条件.$E\|\widetilde{m}(Z;\theta_0,\Sigma_0)\|^2<\infty$.

在上述条件之下，我们有下面的定理.

定理 3.2 假设条件(C1)-(C5)成立，则$(\widehat{\theta},\widehat{\sigma})$依概率趋于$(\theta_0,\sigma_0)$，并且

$$\sqrt{n}\begin{pmatrix}\widehat{\theta}-\theta_0\\ \widehat{\sigma}-\sigma_0\end{pmatrix}\Longrightarrow N\left(0,(A'WA)^{-1}(A'W\Omega WA)(A'WA)^{-1}\right),$$

其中

$$A=\left[\frac{\partial m(Z;\theta)}{\partial\theta}-\frac{1}{2}\sum_{j,l=1}^{k}\sigma_{jl}\frac{\partial m^3(Z;\theta)}{\partial Z_j\partial Z_l\partial\theta},\quad\left(-\frac{1}{2^{\delta(j,l)}}\frac{\partial^2 m(Z;\theta)}{\partial Z_j\partial Z_l},j\geq l\right)\right],$$

$\Omega=E\widetilde{m}(Z;\theta,\Sigma)\widetilde{m}'(Z;\theta,\Sigma)$; 如果$j=l$, $\delta(j,l)=1$, 否则$\delta(j,l)=0$.

由$\widehat{\theta},\widehat{\Sigma}$的相合性，我们不难得到下面的结论.

定理 3.3 假设条件(C1)-(C5)成立，取$\widehat{W}_n=n(\sum_{i=1}^n\widetilde{m}(Z_i;\widehat{\theta},\widehat{\Sigma})\widetilde{m}'(Z_i;\widehat{\theta},\widehat{\Sigma}))^{-1}$, 然后取

$$(\widetilde{\theta},\widetilde{\Sigma})=\mathrm{argmin}_{\theta,\sigma}\left(\sum_{i=1}^n\widetilde{m}(Z_i;\theta,\sigma)\right)'\widehat{W}_n\left(\sum_{i=1}^n\widetilde{m}(Z_i;\theta,\sigma)\right)$$

作为(θ,Σ)的估计，则$\widehat{W}_n$依概率趋于Ω^{-1}，$(\widetilde{\theta},\widetilde{\Sigma})$仍为$(\theta_0,\Sigma_0)$的相合估计，并且

$$\sqrt{n}\begin{pmatrix}\widetilde{\theta}-\theta_0\\ \widetilde{\sigma}-\sigma_0\end{pmatrix}\Longrightarrow N\left(0,(A'\Omega A)^{-1}\right).$$

其中$\widetilde{\sigma}=(\widetilde{\sigma}_{jl},j\geq l)$, $\widetilde{\sigma}_{jl}$为$\widetilde{\Sigma}$的(j,l)分量.

循标准的关于矩估计相合性与标准正态的论证，我们可以得到定理3.2与3.3的证明. 为简洁起见，本文我们省略详细的证明过程.

3.2 不可识别情况下的可能值估计

如果矩方程$E\widetilde{m}(Z;\theta,\Sigma)=0$不足以唯一决定$\theta,\Sigma$，也就是说，条件(C3)不成立，那么按照文[7]提供的思路，我们可以构造一个关于θ的一个可能的取值集合. 定义

$$\eta_0(\theta)=Em(Z,\theta)-\frac{1}{2}\sum_{j,l=1}^{k}\sqrt{\tau_j^2\tau_l^2}\left[E\frac{\partial^2 m(Z,\theta)}{\partial Z_j\partial Z_l}\right]^-.$$

$$\eta_1(\theta)=Em(Z,\theta)+\frac{1}{2}\sum_{j,l=1}^{k}\sqrt{\tau_j^2\tau_l^2}\left[E\frac{\partial^2 m(Z,\theta)}{\partial Z_j\partial Z_l}\right]^+.$$

其中τ_j^2为Z的第j个分量的方差，$j=1,2,\ldots,k$. 上式右端出现在指数位置上的$+$, $-$表示相应函数的正部与负部. 由于对任意的$j,l=1,2,\ldots,k$，都有$|\sigma_{jl}|\leq\sqrt{\sigma_j^2\sigma_l^2}\leq\sqrt{\tau_j^2\tau_l^2}$，所以我们有

$$\eta_0(\theta_0)\leq E\widetilde{m}(Z;\theta_0,\Sigma_0)=0\leq\eta_1(\theta_0),$$

而此不等式意味着参数θ的一个可能取值集合可选为

$$\Theta_0=\{\theta|\theta\in\Theta,\eta_0(\theta)\leq 0\leq\theta_1(\theta)\}.$$

为统计推断计，取任一正定矩阵W，及趋于W的任一可能依赖数据的正定矩阵序列W_n，定义

$$Q(\theta)=[\eta_0(\theta)I(\eta_0(\theta)>0)]'W[\eta_0(\theta)I(\eta_0(\theta)>0)]+[\eta_1(\theta)I(\eta_1(\theta)<0)]'W[\eta_1(\theta)I(\eta_1(\theta)<0)],$$

与

$$\begin{aligned} Q_n(\theta) &= [\eta_{n0}(\theta)I(\eta_{n0}(\theta)>0)]'W_n[\eta_{n0}(\theta)I(\eta_{n0}(\theta)>0)] \\ &\quad +[\eta_{n1}(\theta)I(\eta_{n1}(\theta)<0)]'W_n[\eta_{n1}(\theta)I(\eta_{n1}(\theta)<0)], \end{aligned}$$

其中

$$\eta_{n0}(\theta)=\frac{1}{n}\sum_{i=1}^{n}m(Z_i,\theta)-\frac{1}{2}\sum_{j,l=1}^{k}\sqrt{\widehat{\tau}_j^2\widehat{\tau}_l^2}\left[\frac{1}{n}\sum_{i=1}^{n}\frac{\partial^2 m(Z_i,\theta)}{\partial Z_{ij}\partial Z_{il}}\right]^{-}.$$

$$\eta_{n1}(\theta)=\frac{1}{n}\sum_{i=1}^{n}m(Z_i,\theta)+\frac{1}{2}\sum_{j,l=1}^{k}\sqrt{\widehat{\tau}_j^2\widehat{\tau}_l^2}\left[\frac{1}{n}\sum_{i=1}^{n}\frac{\partial^2 m(Z_i,\theta)}{\partial Z_{ij}\partial Z_{il}}\right]^{+},$$

这里$\widehat{\tau}_j^2$为$Z_{ij}, i=1,2,\ldots,n$的样本方差. 在下面的集合距离定义之下

$$\rho(A,B)=\sup_{a\in A}\inf_{b\in B}|a-b|,$$

而$|a-b|$为a,b两点之间的欧式距离，我们有下面的结论.

定理 3.4 假设条件(C1)-(C3)，(C5)成立. 对于趋于0的一个正数序列γ_n，定义集合

$$\Theta_n=\{\theta|\theta\in\Theta, Q_n(\theta)\le \mathrm{argmin}_{b\in\Theta}Q_n(b)+\gamma_n\},$$

则当$n\to\infty$, $\rho(\Theta_n,\Theta_0)\to 0$, a.s..进一步，如果我们有

$$\left\{\sup_{\theta\in\Theta}Q_n(\theta)-Q(\theta)\right\}\Big/\gamma_n\to 0,\quad \text{a.s.},$$

那么$\rho(\Theta_0,\Theta_n)\to 0$, a.s..

定理3.4的证明类似于文[9]中命题3与5的证明，故此省略.

4 条件矩估计

我们以带有测量误差的Poisson回归作为一个例子来开始本节的讨论. 在给定随机变量X的条件下，假设Y服从均值为$\exp(X\theta)$的Poisson分布.即

$$P(Y=y|X)=\frac{\exp(yX\theta)\exp(-\exp(X\theta))}{y!},\quad y=0,1,\ldots.$$

如果可以得到(X,Y)的一组观测$(X_i,Y_i), i=1,2,\ldots,n$，那么$\theta$最好的估计应该是极大似然估计，也就是使

$$L_n(\theta)=\frac{1}{n}\sum_{i=1}^{n}[-\exp(X_i\theta)+Y_iX_i\theta-\log Y_i!]$$

达到最大的θ值. 在X的值不可观测且服从测量误差模型$Z=X+U$时，我们可能想到是否可以用使得

$$\widetilde{L}_n(\theta)=\frac{1}{n}\sum_{i=1}^{n}[-\exp(Z_i\theta)+Y_iZ_i\theta-\log Y_i!]$$

达到最大的θ值来作为θ的估计. 这就是所谓的Naive估计. 注意到当$n\to\infty$ 时，

$$\widetilde{L}_n(\theta)\to E[-\exp(X\theta)+YX\theta-\log Y!]+E[-\exp(Z\theta)+\exp(X\theta)],$$

也就是说，如果我们定义

$$\mathcal{L}_{0n}(\theta)=\frac{1}{n}\sum_{i=1}^{n}[Y_iZ_i\theta-\log Y_i!]-E\exp(X\theta)$$

则$\mathcal{L}_{0n}(\theta)$与$L_n(\theta)$皆依概率收敛于$E[-\exp(X\theta)+YX\theta-\log Y!]$. 基于这一修正的似然函数$\mathcal{L}(\theta)$，文[6]指出，如果$E\exp(X\theta)$已知，则$\mathcal{L}_{0n}(\theta)$的最大值点可以作为$\theta$的估计. 文[6]称这一估计为校正得分估计.在一定的正则条件之下，校正得分估计的相合性与渐近正态性得到了证明.

但是，$E\exp(X\theta)$已知的条件往往意味着我们需要对X的分布有所了解，而在测量误差的文献中，研究人员很少对X的分布做出假设，所以如果要应用文[6]的方法， 唯一可行的做法是得到$E\exp(X\theta)$ 的估计. 这样一来，在测量误差服从Laplace分布的情况下，文[7]的工作，或者我们在第三节得到的结果，便提供了$E\exp(X\theta)$的一种可能的估计. 事实上，我们有

$$E\exp(X\theta)=\left[1-\frac{1}{2}\sigma^2\theta^2\right]E\exp(Z\theta).$$

如此，如果进一步假设测量误差U的方差σ^2已知，我们即可以通过极大化

$$\mathcal{L}_{1n}(\theta)=\frac{1}{n}\sum_{i=1}^{n}\left[Y_iZ_i\theta-\log Y_i!-\left(1-\frac{1}{2}\sigma^2\theta^2\right)\exp(Z_i\theta)\right]$$

来求得θ的估计. 然而仔细分析可以看到，对于一组给定的$(Z_i,Y_i),i=1,2,\ldots,n$，作为θ函数，当$\theta\to\pm\infty$时，$\mathcal{L}_{1n}(\theta)\to+\infty$. 也就是说，$\mathcal{L}_{1n}(\theta)$的最大值点并不存在. 这也意味着在上述情况中，文[7]的工作，或者我们在第三节得到的结果并不适用.

4.1. 一个条件期望公式

在测量误差模型的统计推断中，回归校正(Regression Calibration)也是一种能有效减低估计偏差的方法. 使用该方法首先要得到$E(X|Z)$的一个估计，然后在相应的统计推断程序中，以$E(X|Z)$的估计取代X. 关于该方法的详细情况，可以参看专著[2]的第4章，或者文[3]中的讨论. 事实上，回归校正并不能完全的消除测量误差带来的影响，欲使其达到令人满意的效果，除了要有较为精确$E(X|Z)$ 的估计，还需要$m(x,\theta)$为x颇为光滑的函数, 或者有着较小的测量误差. 在下面的讨论中会看到，当测量误差模型具有Laplace分布时，我们不仅能得到$E(X|Z)$的精确表达式，而且对较为光滑的函数$m(x,\theta)$，我们也能得到$E[m(X,\theta)|Z]$的精确表达式. 而使用后者往往可以得到更为有效的估计.

我们首先陈述一个适用于多维X的结果. 对于参数函数$m(x;\theta)$，如果$E(m(X;\theta))=0$，那么也意味着$E(E[m(X;\theta)|Z])=0$. 所以如果我们能得到$E[m(X;\theta)|Z]$ 的表达式(显然这样的表达式也依赖于Σ)，则在3.1 节中取$\widetilde{m}(Z;\theta,\Sigma)$，循类似的思路也可以构造$\theta,\Sigma$的估计，并且直观上来说，由条件期望$E[m(X;\theta|Z)]$出发构造的估计要比根据无条件期望关系中$\widetilde{m}(Z;\theta,\Sigma)$ 构造的估计更好一些，毕竟在所有Z 的可测函数中，条件期望$E[m(X;\theta|Z)]$在均方误差意义下更接近于$m(X;\theta)$.

假设$m(x)$为一可测函数. 下面的定理给出了条件期望$E(m(X)|Z)$ 的一个表达式.

定理 4.1 假设测量误差U服从由(2.2)定义的均值为0的多元Laplace分布，Z的特征函

数$\phi_Z(t)$平方可积，则

$$E[m(X)|Z]=\frac{1}{g(Z)}\left[\int m(x)\psi(Z-x)g(x)dx-\frac{1}{2}\sum_{j,l=1}^{n}\sigma_{jl}\int m(x)\psi(Z-x)\frac{\partial^2 g(x)}{\partial x_j\partial x_l}dx\right]. \quad (4.5)$$

其中ψ为由(2.2)定义的U的密度函数.

如果$k=1$，则我们有下面的推论

推论 4.1 假设函数m，g二阶可导，并且对$b(x)=m(x)g(x)$, $m'(x)g(x)$或$m(x)g'(x)$来说

$$\lim_{x\to\pm\infty} b(x)\exp(-|x|)\to 0.$$

则我们有

$$E[m(X)|Z]=m(Z)+l(Z), \quad (4.6)$$

其中

$$l(z)=\frac{e^{z/b}}{g(z)}\int_z^{\infty}\left[m'(x)-\frac{bm''(x)}{2}\right]g(x)e^{-x/b}dx-\frac{e^{-z/b}}{g(z)}\int_{-\infty}^{z}\left[m'(x)+\frac{bm''(x)}{2}\right]g(x)e^{x/b}dx,$$

$b=\sigma/\sqrt{2}$.

注1. 当U服从正态分布时，无论X分布为何，我们总有$E(X|Z)=Z+\sigma^2 g'(Z)/g(Z)$. 文[10]提到此公式是Maurice Tweedie在与该文作者的私人通信中提出的，所以该公式被称为Tweedie公式. 由于其与经验Bayes理论的密切关系，Tweedie公式得到了非常广泛的应用，详细情况请参考文[1]的工作. 定理4.1 的一个重要意义在于当测量误差服从Laplace 分布时，我们得到了类似的Tweedie 公式(4.5)，其中$m(x)$ 不仅仅限于x 的线性函数.

注2. 不难证明，对(4.5)中的$l(Z)$取期望后即可得到

$$El(Z)=-\frac{\sigma^2}{2}E\left[\frac{\partial^2 m(Z)}{\partial Z^2}\right],$$

由此即得(3.3)当$k=1$的情形.

为陈述方便与简洁起见，我们接下来只考虑一维的情形. 这样做的原因有两点，一是在多维情况下，由(4.7)可知，g的二阶导数出现在了条件期望的表达式中，直接用g的核密度估计的二阶导数取代g很可能会对估计的理论探讨带来麻烦. 二是在一定的条件之下，我们虽然可以将(4.7) 化为类似于(4.8) 的形式的，也就是说在$E(m(X)|Z)$的表达式中出现的是g本身，而不是g的二阶导数，但是在多维情况的处理中，除了记号的繁杂以外，大多数的理论论证过程与一维情况大同小异.

4.2 条件矩估计

由于X的密度函数未知，从而Z的密度函数也无从知晓.如果以$Z_1,Z_2,\ldots,Z_n$表示Z 的一组样本，那么我们就可以使用g的核密度估计.以K记一选定的核函数，h为窗宽，则g的核密度估计定义为

$$\widehat{g}_h(z)=\frac{1}{nh}\sum_{i=1}^{n}K\left(\frac{Z_i-z}{h}\right).$$

为方便计，记

$$\mu_l(x;\theta,\sigma)=\left[m'(x;\theta)+(-1)^l\frac{\sigma}{2\sqrt{2}}m''(x;\theta)\right]\exp\left((-1)^l\frac{x\sqrt{2}}{\sigma}\right),$$

$l=1,2$.根据推论4.1, 对固定的θ,σ，$E(m(X,\theta)|Z)$可以用下式估计

$$\widehat{E}(m(X,\theta)|Z)=m(Z;\theta)+\frac{e^{Z\sqrt{2}/\sigma}}{\widehat{g}_h(Z)}\int_Z^{\infty}\mu_1(x;\theta,\sigma)\widehat{g}_h(x)dx-\frac{e^{-Z\sqrt{2}/\sigma}}{\widehat{g}_h(Z)}\int_{-\infty}^{Z}\mu_2(x;\theta,\sigma)\widehat{g}_h(x)dx.$$

需要注意的是，当分母出现核估计时，在理论论证时我们往往要保证核估计有非零下界. 但是这一要求对支撑为整个实数集合的分布并不合理. 文献中已经有一些方法来处理这个问题. 在本文中，我们采用的作法是将参与运算的样本$Z_i, i=1,2,\ldots,n$限制在某个闭区间内，这样只需要假设相关函数的连续性，即可回避掉分母接近零带来的理论困难. 为此，取一个足够大的正数a, 令$H(Z;\theta,\sigma)=I_{[-a,a]}(Z)E(m(X,\theta)|Z)$，并且用

$$\widehat{H}(Z;\theta,\sigma)=I_{[-a,a]}(Z)\widehat{E}(m(X,\theta)|Z)$$

作为$H(Z;\theta,\sigma)$的估计，这里$I_{[-a,a]}(\cdot)$为闭区间$[-a,a]$的示性函数. 为简便计，在下面的讨论中我们用$I_a(\cdot)$来表示$[-a,a]$的示性函数.

如3.1节中的讨论，如果矩条件$EH(Z;\theta,\sigma)=0$可以唯一确定θ,σ，对任意只依赖于数据的正定矩阵W_n，我们可以用

$$(\widehat{\theta},\widehat{\sigma})=\mathrm{argmin}_{\theta,\sigma}\left(\sum_{i=1}^n\widehat{H}(Z_i;\theta,\sigma)\right)'W_n\left(\sum_{i=1}^n\widehat{H}(Z_i;\theta,\sigma)\right).\tag{4.7}$$

同时，我们假设

(S1). θ, σ的参数空间分别为R^p, R^+中的紧子集；

(S2). $EH(Z;\theta,\sigma)=0$ 当且仅当$\theta=\theta_0$, $\sigma=\sigma_0>0$, 其中θ_0,σ_0为真实参数值；

(S3). 对某个正定对称矩阵，W_n依概率收敛于W；

(S4). 以记$\alpha=(\theta',\sigma)'$. $EH(Z;\theta,\sigma)/\partial\alpha$ 存在，列满秩，并且在α_0的某个邻域内，$\partial H(Z;\theta,\sigma)/\partial\alpha$满足一阶Lipschitz条件. $E\|H(Z;\theta_0,\sigma_0)\|^2<\infty$.

(S5). 核函数K满足

$$K(x)\geq 0,\quad \int K(x)dx=1,\quad \int xK(x)dx=0,\quad 0<\int x^2K(x)dx<\infty.$$

(S6). 窗宽h满足$h\to 0$, $nh^2\to\infty$, $nh^4\to 0$.

在上述条件之下，我们有下面的定理.

定理 4.2 假设条件(S1)-(S6)成立，则$(\widehat{\theta},\widehat{\sigma})$依概率趋于$(\theta_0,\sigma_0)$，并且

$$\sqrt{n}\begin{pmatrix}\widehat{\theta}-\theta_0\\ \widehat{\sigma}-\sigma_0\end{pmatrix}\Longrightarrow N\left(0,(A'WA)^{-1}(A'W\Omega WA)(A'WA)^{-1}\right),$$

其中$A=E\partial H(Z;\theta,\sigma)/\partial\alpha$, $\Omega=\mathrm{Cov}(S_{11}(Z)-S_{12}(Z)+S_{21}(Z)-S_{22}(Z))$,其中

$$S_{11}(Z)\quad=\quad\mu_1(Z;\theta_0,\sigma_0)\int_{-a}^{a}I(u<Z)\exp(u\sqrt{2}/\sigma_0)du;$$

$$\begin{aligned} S_{12}(Z) &= \frac{\exp(Z\sqrt{2}/\sigma)}{g(Z)} I_a(Z) \int_Z^\infty \mu_1(x;\theta_0,\sigma_0) g(x) dx; \\ S_{21}(Z) &= \mu_2(Z;\theta_0,\sigma_0) \int_{-a}^{a} I(u>Z) \exp(-u\sqrt{2}/\sigma_0) du; \\ S_{22}(Z) &= \frac{\exp(-Z\sqrt{2}/\sigma)}{g(Z)} I_a(Z) \int_{-\infty}^{Z} \mu_2(x;\theta_0,\sigma_0) g(x) dx. \end{aligned}$$

由$\widehat{\theta},\widehat{\sigma}$的相合性，我们不难得到下面的结论.

定理 4.3 假设条件(S1)-(S6)成立，取$\widehat{W}_n = n(\sum_{i=1}^n \widehat{H}(Z_i;\widehat{\theta},\widehat{\sigma})\widehat{H}'(Z_i;\widehat{\theta},\widehat{\sigma}))^{-1}$, 然后取

$$(\widetilde{\theta},\widetilde{\sigma}) = \operatorname{argmin}_{\theta,\sigma} \left(\sum_{i=1}^n \widehat{H}(Z_i;\theta,\sigma) \right)' \widehat{W}_n \left(\sum_{i=1}^n \widehat{h}(Z_i;\theta,\sigma) \right)$$

作为(β,σ)的估计，则$\widehat{W}_n$依概率趋于Ω^{-1}，$(\widetilde{\theta},\widetilde{\sigma})$仍为$(\theta_0,\sigma_0)$的相合估计，并且

$$\sqrt{n} \begin{pmatrix} \widehat{\theta}-\theta_0 \\ \widehat{\sigma}-\sigma_0 \end{pmatrix} \Longrightarrow N\left(0,(A'\Omega A)^{-1}\right).$$

5 主要结果的证明

我们首先来证明引理3.1.

证明： 以$\phi_Z(t)$, $\phi_u(t)$分别表示Z与U的特征函数. 由特征函数的反演公式，X的密度函数为

$$\begin{aligned} f(x) &= \frac{1}{(2\pi)^k} \int \exp(-it'x)\phi_Z(t)\phi_u^{-1}(t) dt = \frac{1}{(2\pi)^k} \int \exp(-it'x)\phi_Z(t) \left[1+\frac{1}{2}t'\Sigma t\right] dt \\ &= \frac{1}{(2\pi)^k} \int \exp(-it'x)\phi_Z(t) dt + \frac{1}{2(2\pi)^k} \sum_{j,l=1}^k \sigma_{jl} \int \exp(-it'x) t_j t_l \phi_Z(t)\phi_u^{-1}(t) dt \\ &= g(x) + \frac{1}{2(-i)^2} \sum_{j,l=1}^k \sigma_{jl} \frac{\partial^2}{\partial x_j \partial x_l} \left[\frac{1}{(2\pi)^k} \int \exp(-it'x)\phi_Z(t) dt \right] \\ &= g(x) - \frac{1}{2} \sum_{j,l=1}^k \sigma_{jl} \frac{\partial^2 g(x)}{\partial x_j \partial x_l}. \end{aligned}$$

接下来我们证明推论4.1.

证明： 为方便起见，我们将$m(x;\theta)$写为$m(x)$, $b=\sigma/\sqrt{2}$. 利用条件期望的定义，我们有

$$E[m(X)|Z] = \int m(x) f(x|z) dx = \frac{\int m(x)\frac{1}{2b}e^{\frac{-|x-z|}{b}}[g(x)-b^2g''(x)]dx}{g(z)}.$$

上式中的分子可以进一步写为下面两式的和：

$$\begin{aligned} S_{1n}(z) &= \frac{1}{2b} e^{z/b} \int_z^\infty m(x) e^{-x/b} [g(x)-b^2g''(x)] dx, \\ S_{2n}(z) &= \frac{1}{2b} e^{-z/b} \int_{-\infty}^{z} m(x) e^{x/b} [g(x)-b^2g''(x)] dx. \end{aligned}$$

利用定理条件并进行分部积分，可以得到

$$\begin{aligned}
&\int_z^{\infty} m(x)e^{-x/b}g''(x)dx = \int_z^{\infty} m(x)e^{-x/b}dg'(x)\\
= &-m(z)e^{-z/b}g'(z) - \int_z^{\infty} g'(x)\left[m'(x)e^{-x} - \frac{1}{b}m(x)e^{-x/b}\right]dx\\
= &-m(z)e^{-z/b}g'(z) - \int_z^{\infty} g'(x)m'(x)e^{-x/b}dx + \frac{1}{b}\int_z^{\infty} g'(x)m(x)e^{-x/b}dx\\
= &-m(z)e^{-z/b}g'(z) + g(z)m'(z)e^{-z/b} + \int_z^{\infty} g(x)m''(x)e^{-x/b}dx - \frac{1}{b}g(z)m(z)e^{-z/b}\\
&-\frac{2}{b}\int_z^{\infty} g(x)m'(x)e^{-x/b}dx + \frac{1}{b^2}\int_z^{\infty} g(x)m(x)e^{-xb}dx.
\end{aligned}$$

类似可以得到

$$\begin{aligned}
&\int_{-\infty}^{z} m(x)e^{x/b}g''(x)dx = \int_{-\infty}^{z} m(x)e^{x/b}dg'(x)\\
= \ &m(z)e^{z/b}g'(z) - g(z)m'(z)e^{z/b} + \int_{-\infty}^{z} g(x)m''(x)e^{x/b}dx\\
&-\frac{1}{b}g(z)m(z)e^{z/b} + \frac{2}{b}\int_{-\infty}^{z} g(x)m'(x)e^{x/b}dx + \frac{1}{b^2}\int_{-\infty}^{z} g(x)m(x)e^{x/b}dx.
\end{aligned}$$

将上述两个式子代入$S_{n1}(z)$, $S_{n2}(z)$的表达式中，我们可以得到

$$\begin{aligned}
S_{n1}(z) &= -\frac{b}{2}e^{z/b}\int_z^{\infty} g(x)m''(x)e^{-x/b}dx + e^{z/b}\int_z^{\infty} g(x)m'(x)e^{-x/b}dx\\
&\quad +\frac{b}{2}m(z)g'(z) - \frac{b}{2}g(z)m'(z) + \frac{1}{2}g(z)m(z),\\
S_{n2}(z) &= -\frac{b}{2}e^{-z/b}\int_{-\infty}^{z} g(x)m''(x)e^{x/b}dx - e^{-z/b}\int_{-\infty}^{z} g(x)m'(x)e^{x/b}dx\\
&\quad -\frac{b}{2}m(z)g'(z) + \frac{b}{2}g(z)m'(z) + \frac{1}{2}g(z)m(z).
\end{aligned}$$

将上述两个式子代入$E[m(X)|Z]$的表达式中, 稍加整理即可得到推论4.1的结论.

为证明定理4.2，我们首先给出两个引理. 假设$X_1, X_2, \ldots, X_n$为服从密度函数$f(x)$的独立同分布随机变量，$\widehat{f}_n(x)$为$f(x)$的核密度估计，其中核函数为K，窗宽为h.

引理 5.1 假设密度函数$f(x)$的支撑为R，满足$f(x)$二阶可导，且其二阶导数有界.则当K 满足条件(S5)，则对任意正数a,

$$\sup_{x\in[-a,a]}|\widehat{f}_n(x) - f(x)| = O\left(\sqrt{\frac{\log n}{nh}}\right) + O(h^2), \quad \text{a.s.}.$$

引理5.1陈述的是核密度估计中一个广为人知的结果，所以我们述而不证.

引理 5.2 假设$\eta(x)$为一连续函数，密度函数$f(x)$二阶可导且二阶导数有界. 我们有

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\eta(X_i)I_a(X_i)[\widehat{f}_n(X_i) - f(X_i)] = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}[\eta(X_i)f(X_i)I_a(X_i) - E\eta(X)f(X)I_a(X)] + o_p(1).$$

证明： 稍加整理，我们有

$$\begin{aligned}\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\eta(X_i)I_a(X_i)\widehat{f}_n(X_i) &= \frac{1}{\sqrt{n}(n-1)h}\sum_{i\neq j}\eta(X_i)I_a(X_i)K\left(\frac{X_i-X_j}{h}\right)+\frac{K(0)}{\sqrt{n}nh}\sum_{i=1}^{n}\eta(X_i)I_a(X_i)\\ &\quad -\frac{1}{\sqrt{n}n(n-1)h}\sum_{i\neq j}\eta(X_i)I_a(X_i)K\left(\frac{X_i-X_j}{h}\right),\end{aligned}$$

由于$E|\eta(X)I_a(X)|<\infty$，$E|\eta(X)f(X)I_a(X)|<\infty$下，后两项均为$o_p(1)$.记

$$h(X_i,X_j)=\frac{1}{2}[\eta(X_i)I_a(X_i)+\eta(X_j)I_a(X_j)]\frac{1}{h}K\left(\frac{X_i-X_j}{h}\right),$$

则我们有

$$\frac{1}{n(n-1)}\sum_{i\neq j}\eta(X_i)I_a(X_i)\frac{1}{h}K\left(\frac{X_i-X_j}{h}\right)=\frac{2}{n(n-1)}\sum_{i<j}h(X_i,X_j).$$

记上式为U_n，则不难看出U_n为一U-统计量，且

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\eta(X_i)I_a(X_i)\widehat{f}_n(X_i)=\sqrt{n}U_n+o_p(1).$$

定义$\theta=Eh(X_1,X_2)$. 我们来考虑U_n在每个样本点上的投影，$E(U_n|X_i)$, $i=1,2,\ldots,n$. 简单计算可以发现，对每个$i=1,2,\ldots,n$,

$$E(U_n|X_i)=\frac{2}{n}E[h(X,X_i)|X_i]+\frac{n-2}{n}\theta,$$

上式右端第一项中的X与X_i独立同分布.令 $\widehat{U}_n=\sum_{i=1}^{n}E(U_n|X_i)-(n-1)\theta$, 易见

$$\widehat{U}_n-\theta=\frac{2}{n}\sum_{i=1}^{n}\left(E[h(X,X_i)|X_i]-\theta\right).$$

记$\widetilde{h}(X_i)=E[h(X,X_i)|X_i]$，$H(X_i,X_j)=h(X_i,X_j)-\widetilde{h}(X_i)-\widetilde{h}(X_j)-\theta$.则

$$U_n-\widehat{U}_n=\frac{2}{n(n-1)}\sum_{i<j}[h(X_i,X_j)-\widetilde{h}(X_i)-\widetilde{h}(X_j)-\theta]=\frac{2}{n(n-1)}\sum_{i<j}H(X_i,X_j).$$

注意到当$i,j,k(=1,2,\ldots,n)$两两不等时，我们有$EH(X_i,X_j)=0$, $E[H(X_i,X_j)|X_i]=0$，而后者也意味着$EH(X_i,X_j)H(X_j,X_k)=0$，从而可以推出

$$E(U_n-\widehat{U}_n)^2=\frac{2}{n(n-1)}EH^2(X_1,X_2).$$

简单计算后可以发现$EH^2(X_1,X_2)=O(1/h)$，如此我们即得到

$$E(U_n-\widehat{U}_n)^2=O\left(\frac{1}{n^2h}\right),\quad \text{or}\quad U_n=\widehat{U}_n+O_p\left(\frac{1}{n\sqrt{h}}\right).$$

由于$nh\to\infty$，所以这个事实意味着$\sqrt{n}(U_n-\theta)=\sqrt{n}(\widehat{U}_n-\theta)+o_p(1)$. 另外，我们有

$$\begin{aligned}\theta &= E\left(\eta(X_1)I_a(X_1)\frac{1}{h}K\left(\frac{X_1-X_2}{h}\right)\right)=\iint_{-a}^{a}\eta(u)f(v)f(u)\frac{1}{h}K\left(\frac{u-v}{h}\right)dudv\\ &= \int_{-a}^{a}\eta(u)f(u)\left[\int K(v)f(u+vh)dv\right]du=Ef(X)\eta(X)I_a(X)+O(h^2).\end{aligned}$$

注意到$E[h(X_i,X_j)|X_i]$可以表示为

$$\frac{1}{2}E\left[\eta(X_i)I_a(X_i)\frac{1}{h}K\left(\frac{X_i-X_j}{h}\right)\Big|X_i\right]+\frac{1}{2}E\left[\eta(X_j)I_a(X_j)\frac{1}{h}K\left(\frac{X_i-X_j}{h}\right)\Big|X_i\right],$$

简单计算可以求得

$$E\left[\eta(X_i)I_a(X_i)\frac{1}{h}K\left(\frac{X_i-X_j}{h}\right)\Big|X_i\right]=\eta(X_i)f(X_i)I_a(X_i)+O_p(h^2)$$

下面来证明下式也成立

$$E\left[\eta(X_j)I_a(X_j)\frac{1}{h}K\left(\frac{X_i-X_j}{h}\right)\Big|X_i\right]=\eta(X_i)f(X_i)I_a(X_i)+O_p(h^2).$$

为简洁起见，记上式左端为$Q_h(X_i)$.事实上，我们有

$$\begin{aligned}&n^{-1/2}\sum_{i=1}^{n}[Q_h(X_i)-\eta(X_i)f(X_i)I_a(X_i)]\\=&\frac{1}{\sqrt{n}}\sum_{i=1}^{n}[Q_h(X_i)-\theta]-\frac{1}{\sqrt{n}}\sum_{i=1}^{n}[\eta(X_i)f(X_i)I_a(X_i)-E\eta(X)f(X)I_a(X)]+o_p(1).\end{aligned}$$

而

$$\begin{aligned}&\mathrm{Var}\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}[Q_h(X_i)-\theta]-\frac{1}{\sqrt{n}}\sum_{i=1}^{n}[\eta(X_i)f(X_i)I_a(X_i)-E\eta(X)f(X)I_a(X)]\right)\\=&E\left([Q_h(X_1)-\theta]-[\eta(X_1)f(X_1)I_a(X_1)-E\eta(X)f(X)I_a(X)]\right)^2\\=&E[Q_h(X_1)-\theta]^2+E[\eta(X_1)f(X_1)I_a(X_1)-E\eta(X)f(X)I_a(X)]^2\\&-2E[Q_h(X_1)-\theta][\eta(X_1)f(X_1)I_a(X_1)-E\eta(X)f(X)I_a(X)].\end{aligned}$$

可以证明当$n\to\infty$时，上式右端的三个期望都收敛于$\mathrm{Var}(\eta(X)f(X)I_a(X))$. 这意味着

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}E\left[\eta(X_j)I_a(X_j)\frac{1}{h}K\left(\frac{X_i-X_j}{h}\right)\Big|X_i\right]=\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\eta(X_i)f(X_i)I_a(X_i)]+o_p(1).$$

综上所述，我们有

$$\sqrt{n}(\widehat{U}_n-\theta)=\frac{2}{\sqrt{n}}\sum_{i=1}^{n}[\eta(X_i)f(X_i)I_a(X_i)-E\eta(X)f(X)I_a(X)]+O_p(\sqrt{n}h^2).$$

如果条件$nh^4\to 0$成立，则我们最终得到

$$\begin{aligned}\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\eta(X_i)[\widehat{f}_n(X_i)-f(X_i)]&=\sqrt{n}U_n-\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\eta(X_i)f(X_i)I_a(X_i)+o_p(1)\\&=\sqrt{n}(U_n-\theta)-\frac{1}{\sqrt{n}}\sum_{i=1}^{n}[\eta(X_i)f(X_i)I_a(X_i)-\theta]+o_p(1).\end{aligned}$$

利用前述所得，对上式右端稍加整理即得引理结论.

引理 5.3 假设$\mu(x),\eta(x)$为连续函数，且

$$\int_{-a}^{a}\eta^2(u)\left[\int_{u}^{\infty}f(x)\mu^2(x)dx\right]f(u)du<\infty,$$

则我们有

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\eta(X_i)I_a(X_i)\int_{X_i}^{\infty}[\widehat{f}_n(x)-f(x)]\mu(x)dx=\frac{1}{\sqrt{n}}\sum_{i=1}^{n}[P(X_i)-EP(X)]+o_p(1).$$

其中

$$P(X_i)=\mu(X_i)\int_{-a}^{a}\eta(u)I(u<X_i)f(u)du.$$

证明： 由于本引理的证明类似于引理5.2的证明，所以我们仅叙述不同之处. 首先，

$$\begin{aligned}\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\eta(X_i)I_a(X_i)\int_{X_i}^{\infty}\widehat{f}_n(x)\mu(x)dx &= \frac{1}{nh\sqrt{n}}\sum_{i=1}^{n}\eta(X_i)I_a(X_i)\int_{X_i}^{\infty}K\left(\frac{X_i-x}{h}\right)\mu(x)dx\\ &-\frac{1}{n(n-1)h\sqrt{n}}\sum_{i\neq j}\eta(X_i)I_a(X_i)\int_{X_i}^{\infty}K\left(\frac{X_j-x}{h}\right)\mu(x)dx\\ &+\frac{1}{(n-1)h\sqrt{n}}\sum_{i\neq j}\eta(X_i)I_a(X_i)\int_{X_i}^{\infty}K\left(\frac{X_j-x}{h}\right)\mu(x)dx.\end{aligned}$$

由于

$$E\left|\frac{1}{nh}\sum_{i=1}^{n}\eta(X_i)I_a(X_i)\int_{X_i}^{\infty}K\left(\frac{X_i-x}{h}\right)\mu(x)dx\right| \le \int_{-a}^{a}|\eta(u)|f(u)\left[\int_{0}^{\infty}K(v)\mu|(v+uh)|dv\right]du,$$

而上式右端的阶为$O(1)$, 所以

$$\frac{1}{nh\sqrt{n}}\sum_{i=1}^{n}\eta(X_i)I_a(X_i)\int_{X_i}^{\infty}K\left(\frac{X_i-x}{h}\right)\mu(x)dx = o_p(1).$$

我们也可推出

$$\begin{aligned}&E\left|\frac{1}{n(n-1)h}\sum_{i\neq j}\eta(X_i)I_a(X_i)\int_{X_i}^{\infty}K\left(\frac{X_j-x}{h}\right)\mu(x)dx\right|\\ \le\ &\int_{-a}^{a}|\eta(u)|f(u)\left[\int_{u}^{\infty}f(x)|\mu(x)|dx\right]du + O(h^2) = O(1),\end{aligned}$$

所以

$$\frac{1}{n(n-1)h\sqrt{n}}\sum_{i\neq j}\eta(X_i)I_a(X_i)\int_{X_i}^{\infty}K\left(\frac{X_j-x}{h}\right)\mu(x)dx = o_p(1).$$

如此，我们就可以得到

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\eta(X_i)\int_{X_i}^{\infty}\widehat{f}_n(x)\mu(x)dx = \sqrt{n}U_n + o_p(1),$$

其中$U_n = 2n^{-1}(n-1)^{-1}\sum_{i=1}^{n}h(X_i,X_j)$，而

$$h(X_i,X_j) = \frac{1}{2}\left[\eta(X_i)\int_{X_i}^{\infty}\frac{1}{h}K\left(\frac{X_j-x}{h}\right)\mu(x)dx + \eta(X_j)\int_{X_j}^{\infty}\frac{1}{h}K\left(\frac{X_i-x}{h}\right)\mu(x)dx\right].$$

定义$\theta = Eh(X_1,X_2)$，并且考虑U_n在每个样本点上的投影，我们可以得到熟悉的表达式

$$\widehat{U}_n - \theta = \sum_{i=1}^{n}E[U_n|X_i] - (n-1)\theta = \frac{2}{n}\sum_{i=1}^{n}\left(E[h(X,X_i)|X_i] - \theta\right),$$

与

$$U_n - \widehat{U}_n = \frac{2}{n(n-1)}\sum_{i<j}[h(X_i,X_j) - \widetilde{h}(X_i) - \widetilde{h}(X_j) - \theta] = \frac{2}{n(n-1)}\sum_{i<j}H(X_i,X_j).$$

并且我们有

$$E[U_n - \widehat{U}_n]^2 = \frac{2}{n(n-1)}EH^2(X_1,X_2).$$

在引理条件之下，可以证明$EH^2(X_1,X_2) = O(1)$. 由此即可得

$$\sqrt{n}(U_n - \theta) = \sqrt{n}(\widehat{U}_n - \theta) + o_p(1).$$

若记

$$R_h(X_i) = E\left[\eta(X_i)I_a(X_i)\int_{X_i}^{\infty}\frac{1}{h}K\left(\frac{X-x}{h}\right)\mu(x)dx\Big|X_i\right]$$

$$P_h(X_i) = E\left[\eta(X)I_a(X)\int_{X}^{\infty}\frac{1}{h}K\left(\frac{X_i-x}{h}\right)\mu(x)dx\Big|X_i\right]$$

易证

$$R_h(X_i) = \eta(X_i)I_a(X_i)\int_{X_i}^{\infty}f(u)\mu(u)du + O_p(h^2).$$

我们也可以得到

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}P_h(X_i) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}P(X_i) + o_p(1),$$

事实上，我们只需验证$E[P_h(X_i) - P(X_i)]^2 = o(1)$即可.

另外，我们也很容易看到$E[h(X, X_i)|X_i] = [P_h(X_i)+R_h(X_i)]/2$，$EP_h(X) = ER_h(X)$，$\theta = [EP_h(X) + ER_h(X)]/2$. 所以有

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\eta(X_i)I_a(X_i)\int_{X_i}^{\infty}[\widehat{f}_n(x)-f(x)]\mu(x)dx = \sqrt{n}U_n - \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\eta(X_i)I_a(X_i)\int_{X_i}^{\infty}f(x)\mu(x)dx + o_p(1),$$

将前面所得事实带入上式的右端，即可得引理的结论.

推论 5.2 假设$\mu(x), \eta(x)$为连续函数，且

$$\int_{-a}^{a}\eta^2(u)\left[\int_{-\infty}^{u}f(x)\mu^2(x)dx\right]f(u)du < \infty,$$

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\eta(X_i)I_a(X_i)\int_{-\infty}^{X_i}[\widehat{f}_n(x) - f(x)]\mu(x)dx = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}[Q(X_i) - EQ(X)] + o_p(1).$$

其中

$$Q(X_i) = \mu(X_i)\int_{-a}^{a}\eta(u)I(u > X_i)f(u)du.$$

推论5.2的证明过程类似于引理5.3的证明，故述而不证.

下面我们来完成定理4.2的证明.

定理4.2的证明： 由于参数空间的紧性以及相关函数关于参数连续性，相合性的证明比较简单, 故此省略. 我们仅需证明估计的渐近正态性.

由多元函数的泰勒展开，$\widehat{\theta}, \widehat{\sigma}$满足下述方程

$$\begin{aligned} 0 \quad = \quad & \left(\frac{1}{n}\sum_{i=1}^{n}\dot{\widehat{H}}(Z_i;\theta_0,\sigma_0)\right)' W_n\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\widehat{H}(Z_i;\theta_0,\sigma_0)\right) \\ & + \left(\frac{1}{n}\sum_{i=1}^{n}\ddot{\widehat{H}}(Z_i;\widetilde{\theta},\widetilde{\sigma})\right)'\left\{W_n\left(\frac{1}{n}\sum_{i=1}^{n}\widehat{H}(Z_i;\widetilde{\theta},\widetilde{\sigma})\right)\otimes I_{(p+1)\times(p+1)}\right\}(\widehat{\alpha} - \alpha_0) \\ & + \left(\frac{1}{n}\sum_{i=1}^{n}\dot{\widehat{H}}(Z_i;\widetilde{\theta},\widetilde{\sigma})\right)' W_n\left(\frac{1}{n}\sum_{i=1}^{n}\dot{\widehat{H}}(Z_i;\widetilde{\theta},\widetilde{\sigma})\right)\sqrt{n}(\widehat{\alpha} - \alpha_0), \end{aligned}$$

其中$\dot{\widehat{H}}(Z;\theta,\sigma)$, $\ddot{\widehat{H}}(Z;\theta,\sigma)$分别表示$\widehat{H}(Z;\theta,\sigma)$对$(\theta,\sigma)$的一阶与二阶偏导，$\widetilde{\theta}$介于$\widehat{\theta}$与$\theta_0$之间，$\sigma$介于$\widehat{\sigma}$与$\sigma_0$之间. 由$\widehat{\theta}$, $\widehat{\sigma}$的相合性，及条件(S2)可知，我们只需证明$n^{-1/2}\left(\sum_{i=1}^n \widehat{H}(Z_i;\theta_0,\sigma_0)\right)$具有所需的渐近正态性.

根据$\widehat{H}(Z_i;\theta_0,\sigma_0)$的定义，我们可以将 $n^{-1/2}\left(\sum_{i=1}^n \widehat{H}(Z_i;\theta_0,\sigma_0)\right)$写为

$$\frac{1}{\sqrt{n}}\sum_{i=1}^n \widehat{H}(Z_i;\theta_0,\sigma_0)=\frac{1}{\sqrt{n}}\sum_{i=1}^n H(Z_i;\theta_0,\sigma_0)+\sum_{j=1}^6 T_{jn},$$

其中，

$$\begin{aligned}
T_{1n} &= \frac{1}{\sqrt{n}}\sum_{i=1}^n\left[\frac{1}{\widehat{g}_n(Z_i)}-\frac{1}{g(Z_i)}\right]\exp(Z_i\sqrt{2}/\sigma)I_a(Z_i)\int_{Z_i}^{\infty}\mu_1(x;\theta_0,\sigma_0)[\widehat{g}_n(x)-g(x)]dx\\
T_{2n} &= \frac{1}{\sqrt{n}}\sum_{i=1}^n g^{-1}(Z_i)\exp(Z_i\sqrt{2}/\sigma)I_a(Z_i)\int_{Z_i}^{\infty}\mu_1(x;\theta_0,\sigma_0)[\widehat{g}_n(x)-g(x)]dx\\
T_{3n} &= \frac{1}{\sqrt{n}}\sum_{i=1}^n\left[\frac{1}{\widehat{g}_n(Z_i)}-\frac{1}{g(Z_i)}\right]\exp(Z_i\sqrt{2}/\sigma)I_a(Z_i)\int_{Z_i}^{\infty}\mu_1(x;\theta_0,\sigma_0)g(x)dx\\
T_{4n} &= \frac{1}{\sqrt{n}}\sum_{i=1}^n\left[\frac{1}{\widehat{g}_n(Z_i)}-\frac{1}{g(Z_i)}\right]\exp(-Z_i\sqrt{2}/\sigma)I_a(Z_i)\int_{-\infty}^{Z_i}\mu_2(x;\theta_0,\sigma_0)[\widehat{g}_n(x)-g(x)]dx\\
T_{5n} &= \frac{1}{\sqrt{n}}\sum_{i=1}^n g^{-1}(Z_i)\exp(-Z_i\sqrt{2}/\sigma)I_a(Z_i)\int_{-\infty}^{Z_i}\mu_2(x;\theta_0,\sigma_0)[\widehat{g}_n(x)-g(x)]dx\\
T_{6n} &= \frac{1}{\sqrt{n}}\sum_{i=1}^n\left[\frac{1}{\widehat{g}_n(Z_i)}-\frac{1}{g(Z_i)}\right]\exp(-Z_i\sqrt{2}/\sigma)I_a(Z_i)\int_{-\infty}^{Z_i}\mu_2(x;\theta_0,\sigma_0)g(x)dx.
\end{aligned}$$

由引理5.1可以推得$T_{1n}=o_p(1)$, $T_{4n}=o_p(1)$. 在引理5.3中，取$\eta(Z)=g^{-1}(Z)\exp(Z\sqrt{2}/\sigma_0)$, $\mu(x)=\mu_1(x;\theta_0,\sigma_0)$, 利用定理中的记号$S_{11}(Z)$, 则有

$$T_{2n}=\frac{1}{\sqrt{n}}\sum_{i=1}^n[S_{11}(Z_i)-ES_{11}(Z)]+o_p(1),$$

由引理5.1可得

$$T_{3n}=\frac{1}{\sqrt{n}}\sum_{i=1}^n\left[\frac{[g(Z_i)-\widehat{g}_n(Z_i)]}{g^2(Z_i)}\right]\exp(Z_i\sqrt{2}/\sigma)I_a(Z_i)\int_{Z_i}^{\infty}\mu_1(x;\theta_0,\sigma_0)g(x)dx+o_p(1).$$

在引理5.2中取

$$\eta(Z)=g^{-2}(Z_i)\exp(Z_i\sqrt{2}/\sigma)I_a(Z_i)\int_{Z_i}^{\infty}\mu_1(x;\theta_0,\sigma_0)g(x)dx,$$

则有

$$T_{2n}=-\frac{1}{\sqrt{n}}\sum_{i=1}^n[S_{12}(Z_i)-ES_{12}(Z)]+o_p(1),$$

利用引理5.2，以及推论5.2，类似可得

$$T_{5n}=\frac{1}{\sqrt{n}}\sum_{i=1}^n[S_{21}(Z_i)-ES_{21}(Z)]+o_p(1),$$

$$T_{6n}=-\frac{1}{\sqrt{n}}\sum_{i=1}^n[S_{22}(Z_i)-ES_{22}(Z)]+o_p(1).$$

由此即得定理结论.

参 考 文 献

[1] Efron, G. Tweedie's formula and selection bias. *JASA*, 2011, **106**(496): 1602-1614.

[2] Carroll, R., Ruppert, D., Stefanski, L.A., Crainiceanu,C. *Measurement Error in Nonlinear Models: A Modern Perspective*, Second Edition, 2006, Chapman and Hall/CRC.

[3] Carroll, R., Stefanski, L.A. Approximate quasi-likelihood estimation in models with surrogate predictors. *JASA*, 1990, **85**(411): 652-662.

[4] Eltoft, T.On the Multivariate Laplace Distribution. *IEEE Signal processing letters*, 2006, **13**(5): 300-303.

[5] Fan, J.and Truong, Y. Nonparametric Regression with Errors in Variables. *Ann.Statist.*, 1993, **21**(4): 1900-1925.

[6] Guo, J., Li, T. Poisson regression models with errors-in-variables: implication and treatment. *Journal of Statistical Planning and Inferences*, 2002, **104**(2): 391-401.

[7] Hong, H.and Tamer, E.A. Simple Estimator for Nonlinear Error in Variable Models. *Journal of Econometrics*, 2003, **117**(1): 1-19.

[8] Kotz, S., Kozubowski, T., Podgórski, K. *The Laplace Distribution and Generalizations*, 2001, Birkhäuser Boston, 239-272.

[9] Manski, C., Tamer, E. Inference on regressions with interval data on a regressor or outcome. *Econometrica*, **70**: 519 – 546.

[10] Robbins, H. An empirical Bayes approach to statistics. In Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability, 1954 – 1955, vol.I. University of California Press, Berkeley and Los Angeles, 1956， 157 – 163.

[11] Shi, J., Chen, K., Song, W. Robust errors-in-variables linear regression via Laplace distribution. *Statistics & Probability Letters*, 2014, **84**: 113-120.

[12] Song, W., Yao, W., Xin, Y. Robust mixture regression model fitting by Laplace distribution. *Computational Statistics & Data Analysis*, 2014, **71**, 128 – 137.

[13] Subar A.F., Thompson F.E., Kipnis V., Midthune D., Hurwitz P., McNutt S., McIntosh A., Rosenfeld S. Comparative validation of the Block, Willett, and National Cancer Institute food frequency questionnaires: the Eating at America's Table Study. *Am J Epidemiol.*, 2001, **154**(12): 1089-1099.

[14] Tosteson T.D., Stefanski L.A., Schafer D.W. A measurement-error model for binary and ordinal regression. *Stat Med.*, 1989 **8**(9):1139-1147.

谨以此文深切怀念我的导师成平先生. 感念先生纳我为学生，授我以为学之道，为人之本！ —— 学生：宋卫星