# 线性模型下分布函数的估计

李 斌
(盐城工学院学生处,盐城 224003)

**摘 要** 利用辅助信息,针对线性模型 $y_i=\beta x_i+x_i^q\varepsilon_i$, $\varepsilon_i$ $i.i.d$, $E\varepsilon_i=0$,给出了总体分布函数的估计量:$\hat{F}(t)=\frac{1}{N}[\sum_{j\in s}\triangle(t-y_j)+\sum_{i\in\bar{s}}\frac{1}{n}\sum_{j\in s}\triangle(t-\hat{\beta}\cdot x_i-x_i^{q'}u_j)]$, $u_j=(y_i-\hat{\beta}\cdot x_j)/x_j^{q'}$, $j\in s$,改进了 Chambers—Dunstan 的结果。

**关键词** 抽样调查 分布函数 辅助信息 有限总体
**分类号** O51

随着社会的发展,抽样调查的应用日益广泛。但大多数关于抽样调查的文献及教科书都着重讨论总体均值的估计方法。然而在实际工作中,常要估计某一有限总体的分布函数。在简单随机样本的情况下,我们可以用经验分布函数作为总体分布函数的估计,Gliveko—Cantelli 定理[1]证明了其相合性。

估计某一总体 $Y$ 的分布函数,常有与之相关的总体 $X$(已知的辅助信息)可以利用,用 $Y_1,\cdots,Y_N$ 表示待估有限总体,$X_1,\cdots X_N$ 表示辅助信息,$(x_1,y_1),\cdots,(x_n,y_n)$为一样本,$F_Y(t)$表示 $Y$ 的分布函数。问题是如何给出 $F_Y(t)$的估计量? 近几年来有不少文献对此进行了讨论。

1986 年,Chambers、Dunstan 假设总体服从模型:

$$y_i=\beta x_i+x_i^q\varepsilon_i \qquad \varepsilon_i \quad i.i.d, E\varepsilon_i=0 \tag{1}$$

其中 $i=1,2,\cdots,N$, $0\leqslant q\leqslant 1$ 已知,$\beta$ 为待估参数。在一个具体问题中,取 $q=\frac{1}{2}$,给出了估计量

$$\hat{F}_m(t)=\frac{1}{N}[\sum_{j\in s}\triangle(t-y_j)+\sum_{i\in\bar{s}}\frac{1}{n}\sum_{j\in s}\triangle(t-\hat{\beta}x_i-\sqrt{x_i}u_j)] \tag{2}$$

其中,$\triangle(x)=\begin{cases}1 & x\geqslant 0\\ 0 & x<0\end{cases}$

$$\hat{\beta}=\frac{\sum_{i\in s}y_i}{\sum_{i\in s}x_i},\qquad u_j=\frac{y_j-\hat{\beta}x_j}{\sqrt{x_j}}$$

并证明了 $\hat{F}_m(t)$为模型无偏估计量[2]。

本文主要对 Chambers—Dunstan 方法进行改进,讨论线性模型(1)下,分布函数的估计。

对于(1),如果 $q$ 已知,利用最小二乘法估计 $\beta$,可得:

$$\hat{\beta}=\frac{\sum_{i=1}^{n}y_i x_i^{1-2q}}{\sum_{i=1}^{n}x_i^{2-2q}} \tag{3}$$

而 $q$ 未知时,可用 $t$ 替换

$$\hat{\beta}(t)=\frac{\sum_{i=1}^{n}y_i x_i^{1-2t}}{\sum_{i=1}^{n}x_i^{2-2t}} \tag{4}$$

如此,只要考虑 $t$ 对估计量的影响。

收稿日期:1999—05—31

定理1：设 $x_1<x_2\leqslant\cdots\leqslant x_{m-1}<x_{2m}$，$K=\max\{i\mid x_i<x_m\}$，$L=\min\{i\mid x_i>x_{m-1}\}$，如果 $\sum_{i=1}^{K}y_i^2>0$，且 $\sum_{i=L}^{2m}y_i^2>0$，则方程

$$\sum_{i=1}^{m}\left(\frac{y_i}{x_i^t}\right)^2=\sum_{i=m+1}^{2m}\left(\frac{y_i}{x_i^t}\right)^2 \tag{5}$$

存在唯一解 $q$。

证明：不妨 令

$$S(t)=\frac{\sum_{i=1}^{m}\left(\frac{y_i}{x_i^t}\right)^2}{\sum_{i=m+1}^{2m}\left(\frac{y_i}{x_i^t}\right)^2} \tag{6}$$

$\forall\ \triangle>0$，$S(t)>0$

$$S(t+\triangle)=\sum_{i=1}^{m}\left(\frac{y_i}{x_i^{t+\triangle}}\right)^2\Big/\sum_{i=m+1}^{2m}\left(\frac{y_i}{x_i^{t+\triangle}}\right)^2=$$

$$\sum_{i=1}^{m}\left(\frac{y_i}{x_i^t}\right)^2\cdot\left(\frac{1}{x_i^{\triangle}}\right)^2\Big/\sum_{i=m+1}^{2m}\left(\frac{y_i}{x_i^t}\right)^2\cdot\left(\frac{1}{x_i^{\triangle}}\right)^2=$$

$$\sum_{i=1}^{m}\left(\frac{y_i}{x_i^t}\right)^2\cdot\left(\frac{x_m}{x_i}\right)^{2\triangle}\Big/\sum_{i=m+1}^{2m}\left(\frac{y_i}{x_i^t}\right)^2\cdot\left(\frac{x_m}{x_i}\right)^{2\triangle} \tag{7}$$

$$\sum_{i=1}^{K}\left(\frac{y_i}{x_i^t}\right)^2\cdot\left(\frac{x_m}{x_i}\right)^{2\triangle}>\sum_{i=1}^{K}\left(\frac{y_i}{x_i^t}\right)^2>0$$

$$\sum_{i=K+1}^{m}\left(\frac{y_i}{x_i^t}\right)^2\cdot\left(\frac{x_m}{x_i}\right)^{2\triangle}=\sum_{i=K+1}^{m}\left(\frac{y_i}{x_i^t}\right)^2$$

因此 $$\sum_{i=1}^{m}\left(\frac{y_i}{x_i^t}\right)^2\cdot\left(\frac{x_m}{x_i}\right)^{2\triangle}>\sum_{i=1}^{m}\left(\frac{y_i}{x_i^t}\right)^2>0 \tag{8}$$

同理 $$0<\sum_{i=m+1}^{2m}\left(\frac{y_i}{x_i^t}\right)^2\cdot\left(\frac{x_m}{x_i}\right)^{2\triangle}<\sum_{i=m+1}^{2m}\left(\frac{y_i}{x_i}\right)^2 \tag{9}$$

从(7)(8)(9)知，$S(t+\triangle)>S(t)$，即 $S(t)$ 严格单调增加。

又 $S(+\infty)=\lim_{t\to+\infty}\left[\sum_{i=1}^{m}\left(\frac{y_i}{x_i^t}\right)^2\Big/\sum_{i=m+1}^{2m}\left(\frac{y_i}{x_i^t}\right)^2\right]=\lim_{t\to+\infty}\left\{\left[\sum_{i=1}^{m}y_i^2\left(\frac{x_k}{x^t}\right)^{2t}\right]\Big/\left[\sum_{i=1}^{2m}y_i^2\left(\frac{x_K}{x^t}\right)^{2t}\right]\right\}=+\infty$　　$\left(\sum_{i=1}^{m}y_i^2\left(\frac{x_K}{x^t}\right)^{2t}>0,\ \lim_{t\to+\infty}\sum_{i=m+1}^{2m}y_i^2\left(\frac{x_k}{x^t}\right)^{2t}=0\right)$

同理 $S(-\infty)=0$，由介值定理知 $S(t)=1$ 有唯一解，即(5)有唯一解 $q$。

定理2：设 $2<X<M$，$x_1\leqslant x_2\leqslant\cdots\leqslant x_n$ 且等号不全成立，则下述结论成立：①任意实数 $t$，$\hat\beta(t)$ 是 $\beta$ 无偏估计量。②任意实数 $t$，$\hat\beta(t)$ 是 $\beta$ 的弱相合估计。③$n$ 固定时，任意 $t\neq q$，有 $Var(\hat\beta(q))<Var(\hat\beta(t))$

证明：① $$\hat\beta(t)=\sum_{i=1}^{n}\left[(\beta x_i+x_i^q\epsilon_i)x_i^{1-2t}\right]\Big/\sum_{i=1}^{n}x_i^{2-2t}=\sum_{i=1}^{n}\beta x_i^{2-2t}\Big/\sum_{i=1}^{n}x_i^{2-2t}+\sum_{i=1}^{n}x_i^{1+q-2t}\epsilon_i\Big/\sum_{i=1}^{n}x_i^{2-2t}=\beta+\sum_{i=1}^{n}x_i^{1+q-2t}\epsilon_i\Big/\sum_{i=1}^{n}x_i^{2-2t}$$

$$E(\hat\beta(t))=\beta+\sum_{i=1}^{n}x_i^{1+q-2t}E\epsilon_i\Big/\sum_{i=1}^{n}x_i^{2-2t}=\beta$$

② $$Var(\hat\beta(t))=\sum_{i=1}^{n}x_i^{2+2q-4t}Var(\epsilon_i)\Big/\left(\sum_{i=1}^{n}x_i^{2-2t}\right)^2=\sigma_0^2\cdot\frac{\sum_{i=1}^{n}x_i^{2+2q-4t}}{\left(\sum_{i=1}^{n}x_i^{2-2t}\right)^2}$$

由于 $2<X<M$，任取 $n$、$i$、$j$ 有：$0<\frac{x_j^{1+q-2t}}{x_i^{2-2t}}<M<+\infty$

即：$0<x_j^{1+q-2t}<Mx_i^{2-2t}$ 对任意 $n$、$i$、$j$

令 $x_K^{2-2t}=\min\{x_i^{2-2t}\mid i=1,2,\cdots,n\}$

$x_L^{1+q-2t}=\max\{x_i^{1+q-2t}\mid i=1,2,\cdots,n\}$

则 $$Var(\hat\beta(t))\leqslant\frac{\sum_{i=1}^{n}(x_L^{1+q-2t})^2}{\left(\sum_{i=1}^{n}x_K^{2-2t}\right)^2}\cdot\sigma_0^2=\frac{n(x_L^{1+q-2t})^2}{n^2(x_K^{2-2t})^2}\cdot\sigma_0^2\leqslant\frac{(M')^2}{n}\cdot\sigma_0^2\to0\ (n\to+\infty\text{时})$$

又 $\hat{\beta}(t)$ 无偏，故 $\hat{\beta}(t)$ 为 $\beta$ 的弱相合估计。

③ 令 $V(t)\triangleq Var(\hat{\beta}(t))/\sigma_0^2$

则 $V(t)=\sum_{i=1}^{n}(x_i^{1+q-2t})^2/(\sum_{i=1}^{n}x_i^{2-2t})^2=(\sum_{i=1}^{n}x_i^{2+2q-4t})(\sum_{i=1}^{n}x_i^{2-2t})^{-2}$

$$\frac{dV(t)}{dt}=(\sum_{i=1}^{n}x_i^{2+2q-4t}\cdot lnx_i\cdot(-4))(\sum_{i=1}^{n}x_i^{2-2t})^{-2}-(\sum_{i=1}^{n}x_i^{2+2q-4t})(-2)(\sum_{i=1}^{n}x_i^{2-2t})^{-3}(\sum_{i=1}^{n}x_i^{2-2t}lnx_i(-2))=$$

$$4[\sum_{i=1}^{n}x_i^{2-2t}]^{-3}\{(\sum_{i=1}^{n}x_i^{2+2q-4t})(\sum_{i=1}^{n}x_i^{2-2t}lnx_i)-(\sum_{i=1}^{n}x_i^{2+2q-4t}lnx_i)(\sum_{i=1}^{n}x_i^{2-2t})\}\triangleq$$

$$4[\sum_{i=1}^{n}x_i^{2-2t}]^{-3}W(t)$$

$4[\sum_{i=1}^{n}x_i^{2-2t}]^{-3}>0$，$W(t)$、$\frac{dV(t)}{dt}$、$\frac{dVar(\hat{\beta}(t))}{dt}$ 同号

$$W(t)=\sum_{i,j=1}^{n}[x_i^{2+2q-4t}x_j^{2-2t}lnx_j]-\sum_{i,j=1}^{n}[x_j^{2+2q-4t}x_i^{2-2t}lnx_j]=\sum_{i,j=1}^{n}x_i^{2-2t}x_j^{2-2t}lnx_j[x_i^{2q-2t}-x_j^{2q-2t}]=$$

$$\sum_{i=j}^{n}\{x_i^{2-2t}x_j^{2-2t}[x_j^{q-t}+x_i^{q-t}]\}lnx_j[x_i^{q-t}-x_j^{q-t}]=\sum_{i\neq j}^{n}D(i,j)lnx_j[x_i^{q-t}-x_j^{q-t}]=$$

$$\sum_{i>j}^{n}D(i,j)lnx_j[x_i^{q-t}-x_j^{q-t}]+\sum_{i<j}^{n}D(i,j)lnx_j[x_i^{q-t}-x_j^{q-t}]=$$

$$\sum_{i>j}^{n}D(j,i)lnx_i[x_i^{q-t}-x_j^{q-t}]+\sum_{j>i}^{n}D(i,j)lnx_j[x_i^{q-t}-x_j^{q-t}]=$$

$$\sum_{i>j}^{n}D(i,j)[lnx_j-lnx_i][x_i^{q-t}-x_j^{q-t}]$$

其中 $D(i,j)=x_i^{2-2t}x_j^{2-2t}[x_j^{q-t}+x_i^{q-t}]=D(j,i)>0$

由此可得，当 $t=q$ 时，$W(t)=0$；$t>q$ 时，$q-t<0$，$x_i^{q-t}-x_j^{q-t}\geqslant 0(i>j)$，等号成立的充要条件是 $x_i=x_j$，于是当 $x_1,\cdots,x_n$ 不全相等时，$W(t)>0$。同理，$t<q$ 时，$W(t)<0$，又 $W(t)$、$d(Var(\hat{\beta}(t)))/dt$ 同号，故 $Var(\hat{\beta}(t))$ 在 $t=q$ 处取最小值。

综上，如果有 $\hat{q}$，可用 $\hat{\beta}(\hat{q})$ 作为 $\beta$ 的估计量。而对给定的 $\hat{\beta}$，对 $y_i-\hat{\beta}x_i,x_i$ 用定理1解方程 $\sum_{i=1}^{m}(\frac{y_i-\hat{\beta}x_i}{x_i^t})^2=\sum_{i=m+1}^{2m}(\frac{y_i-\hat{\beta}x_i}{x_i^t})^2$ (10)可得解 $\hat{q}(\hat{\beta})$。

任取 $t_0\in[0,1]$ 作为迭代初值，由定理2知 $\hat{\beta}(t_0)\triangleq\hat{\beta}_1$ 是 $\beta$ 的无偏、相合估计，只是估计量的方差较 $Var(\hat{\beta}(q))$ 大些，再利用方程(10)得到 $\hat{q}_1=\hat{q}(\hat{\beta}(t_0))$，依此类推反复迭代，对 $\hat{\beta}_K,\hat{q}_K$，令 $\hat{\beta}_{K+1}=\hat{\beta}(\hat{q}_K)$，$\hat{q}_{K+1}=\hat{q}(\hat{\beta}_{K+1})$，……设 $\hat{\beta}_K\to\hat{\beta}^*$，$\hat{q}_K\to\hat{q}^*$，$\hat{\beta}^*$，$\hat{q}^*$ 就是最终相应参数的估计量，据此给出有限总体分布函数的估计量：

$$\hat{F}(t)=\frac{1}{N}[\sum_{j\in s}\triangle(t-y_j)]+\sum_{i\in\bar{s}}\frac{1}{n}\sum_{j\in s}\triangle(t-\hat{\beta}^*x_i-x_i^{\hat{q}^*}u_j)] \qquad (11)$$

其中 $u_j=(y_j-\hat{\beta}^*x_j)/x_j^{\hat{q}^*}$，$j\in s$。

## 参 考 文 献

1 陈家鼎，李东风. 数理统计学讲义. 北京：高等教育出版社，1993

2 Chambers, R. L & Dunstan, R. (1986) Estimating distribution functions from survey data. Biometrika, Vol. 73

# The Estimation of Scattering Functions Under the Lined Model

*Li Bin*

(The Students Department of Yancheng Institute of Technology, Yancheng 224003, PRC)

**Abstract** Using assisted message, the writing gives the estimating quality of the total scattering functions: $\hat{F}(t)=\frac{1}{N}[\sum_{j\in s}\triangle(t-y_j)+\sum_{i\in\bar{s}}\frac{1}{n}\sum_{j\in s}\triangle(t-\hat{\beta}^*x_i-x_i^{\hat{q}^*}u_j)]$, $u_j=(y_j-\hat{\beta}^*x_j)/x_j^{\hat{q}^*}$, $j\in s$, under the lined model: $y_i=\beta x_i+x_i\varepsilon_i$, $\varepsilon_i$ i.i.d, $E\varepsilon_i=0$ and improves the result of Chambers—Dunrtan.

**Keywords** selective examination; scattering functions; assisted message; limited emtity