第10卷 无锡轻工业学院学报 Vol.10
第1期 JOURNAL OF THE WUXI INSTITUTE OF LIGHT INDUSTRY 1991 No.1

代表点与总体之间的统计关系

费荣昌

（基础课部）

摘要 本文研究了代表点与总体之间的统计关系以及损失函数的某些性质。

关键词 总体；代表点；损失函数；统计关系

0 引 言

文[1—3]研究了在连续型总体中如何选取代表点的问题，即从总体 X 选取 n 个代表点 x_{n1}，x_{n2}，…，x_{nn}，使损失函数

$$L(x_1,x_2,\cdots,x_n)=\frac{1}{DX}\int_{-\infty}^{+\infty}\min_{1\leqslant i\leqslant n}(x-x_i)^2\varphi(x)\mathrm{d}x$$

在 x_{n1}，x_{n2}，…，x_{nn} 取得最小值，其中 $\varphi(x)$ 是 X 的密度函数，DX 是 X 的方差。

记

$$f_n(x)=\begin{cases}x_{n1}, & x\leqslant(x_{n1}+x_{n2})/2;\\ x_{ni}, & (x_{n,i-1}+x_{ni})/2<x\leqslant(x_{ni}+x_{n,i+1})/2\ (i=2,3,\cdots,n-1);\\ x_{nn}, & (x_{n,n-1}+x_{nn})/2<x,\end{cases}$$

则 $X_n=f_n(X)$ 的概率分布为

$$P\{X_n=x_{ni}\}=p_{ni}\ (i=1,2,\cdots,n),$$

其中

$$p_{n1}=\int_{-\infty}^{(x_{n1}+x_{n2})/2}\varphi(x)\mathrm{d}x,$$

$$p_{ni}=\int_{(x_{n,i-1}+x_{ni})/2}^{(x_{ni}+x_{n,i+1})/2}\varphi(x)\mathrm{d}x\ (i=2,3,\cdots,n-1),$$

$$p_{nn}=\int_{(x_{n,n-1}+x_{nn})/2}^{+\infty}\varphi(x)\mathrm{d}x.$$

本文研究代表点 X_n 与总体 X 之间的统计关系以及损失函数 $L(x_{n1},x_{n2},\cdots,x_{nn})$ 的某些性质。

1 损失函数的性质

定理1 $L(x_{n+1,1},x_{n+1,2},\cdots,x_{n+1,n+1})\leqslant L(x_{n1},x_{n2},\cdots,x_{nn})$。

本文1990年2月9日收到。

?1994-2014 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

证 因

$$L(x_{n1},x_{n2},\cdots,x_{nn})-L(x_{n1},x_{n2},\cdots,x_{nn},x_{nn}+1)$$

$$=\frac{1}{DX}\int_{x_{nn}+\frac{1}{2}}^{+\infty}[(x-x_{nn})^2-(x-x_{nn}-1)^2]\varphi(x)\mathrm{d}x$$

$$=\frac{2}{DX}\int_{x_{nn}+\frac{1}{2}}^{+\infty}(x-x_{nn}-\tfrac{1}{2})\varphi(x)\mathrm{d}x\geqslant 0,$$

而

$$L(x_{n+1,1},x_{n+1,2},\cdots,x_{n+1,n+1})\leqslant L(x_{n1},x_{n2},\cdots,x_{nn},x_{nn}+1),$$

即证。

定理2 $0<L(x_{n1},x_{n2},\cdots,x_{nn})\leqslant 1$。

证 因

$$L(x_1)=\frac{1}{DX}\int_{-\infty}^{+\infty}(x-x_1)^2\varphi(x)\mathrm{d}x$$

当 $x_1=EX$ 时取得最小值，所以 $x_{11}=EX$，并且 $L(x_{11})=1$。再由定理1即证。

定理3 $L(x_{n1},x_{n2},\cdots,x_{nn})\to 0\quad(n\to\infty)$。

证 对 ε，取

$$x_i=\{2\left(i-\left[\frac{n+1}{2}\right]\right)-1\}\sqrt{\frac{\varepsilon DX}{3}}\quad(i=1,2,\cdots,n),$$

有

$$L(x_1,x_2,\cdots,x_n)$$

$$=\frac{1}{DX}\left[\int_{-\infty}^{(x_1+x_2)/2}(x-x_1)^2\varphi(x)\mathrm{d}x+\sum_{i=2}^{n-1}\int_{(x_{i-1}+x_i)/2}^{(x_i+x_{i+1})/2}(x-x_i)^2\varphi(x)\mathrm{d}x\right.$$

$$\left.+\int_{(x_{n-1}+x_n)/2}^{+\infty}(x-x_n)^2\varphi(x)\mathrm{d}x\right],$$

而

$$\int_{x_1}^{(x_1+x_2)/2}(x-x_1)^2\varphi(x)\mathrm{d}x\leqslant\frac{\varepsilon DX}{3}\int_{x_1}^{(x_1+x_2)/2}\varphi(x)\mathrm{d}x,$$

$$\int_{(x_{i-1}+x_i)/2}^{(x_i+x_{i+1})/2}(x-x_i)^2\varphi(x)\mathrm{d}x\leqslant\frac{\varepsilon DX}{3}\int_{(x_{i-1}+x_i)/2}^{(x_i+x_{i+1})/2}\varphi(x)\mathrm{d}x\,(i=2,3,\cdots,n-1),$$

$$\int_{(x_{n-1}+x_n)/2}^{x_n}(x-x_n)^2\varphi(x)\mathrm{d}x\leqslant\frac{\varepsilon DX}{3}\int_{(x_{n-1}+x_n)/2}^{x_n}\varphi(x)\mathrm{d}x,$$

因 DX 存在，对 ε，有 M，当 $x_n>M$（这时 $x_1<-M$），即

?1994-2014 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

$$n>M\sqrt{\frac{3}{\varepsilon DX}}+2\text{时},$$

$$\int_{x_n}^{+\infty}(x-EX)^2\varphi(x)\mathrm{d}x<\frac{\varepsilon DX}{3},\quad \int_{-\infty}^{x_1}(x-EX)^2\varphi(x)\mathrm{d}x<\frac{\varepsilon DX}{3}.$$

又当$x_n>|EX|$(这时$x_1<-|EX|$)，即

$$n>|EX|\sqrt{\frac{3}{\varepsilon DX}}+2\text{时},$$

$$\int_{x_n}^{+\infty}(x-x_n)^2\varphi(x)\mathrm{d}x<\int_{x_n}^{+\infty}(x-EX)^2\varphi(x)\mathrm{d}x,$$

$$\int_{-\infty}^{x_1}(x-x_1)^2\varphi(x)\mathrm{d}x<\int_{-\infty}^{x_1}(x-EX)^2\varphi(x)\mathrm{d}x.$$

于是，对ε，取

$$N=\max\left\{\left[M\sqrt{\frac{3}{\varepsilon DX}}+2\right],\ \left[|EX|\sqrt{\frac{3}{\varepsilon DX}}+2\right]\right\},$$

当$n>N$时，

$$L(x_1,x_2,\cdots,x_n)$$

$$<\frac{1}{DX}\left[\int_{-\infty}^{x_1}(x-EX)^2\varphi(x)\mathrm{d}x+\frac{\varepsilon DX}{3}\int_{x_1}^{x_n}\varphi(x)\mathrm{d}x+\int_{x_n}^{+\infty}(x-EX)^2\varphi(x)\mathrm{d}x\right]$$

$$<\frac{1}{DX}\left(\frac{\varepsilon DX}{3}+\frac{\varepsilon DX}{3}+\frac{\varepsilon DX}{3}\right)=\varepsilon.$$

而

$$0<L(x_{n1},x_{n2},\cdots,x_{nn})\leqslant L(x_1,x_2,\cdots,x_n),$$

所以

$$L(x_{n1},x_{n2},\cdots,x_{nn})\to 0(n\to\infty).$$

2 Xn与X之间的统计关系

$L(x_1,x_2,\cdots,x_n)$在$x_{n1},x_{n2},\cdots,x_{nn}$取得最小值，$x_{n1},x_{n2},\cdots,x_{nn}$必满足

$$\frac{\partial L}{\partial x_1}=0,\frac{\partial L}{\partial x_2}=0,\cdots,\frac{\partial L}{\partial x_n}=0,$$

即

$$\int_{-\infty}^{(x_{n1}+x_{n2})/2}x\varphi(x)\mathrm{d}x=x_{n1}p_{n1},$$

$$\int_{(x_{n,i-1}+x_{ni})/2}^{(x_{ni}+x_{n,i+1})/2}x\varphi(x)\mathrm{d}x=x_{ni}p_{ni}\ (i=2,3,\cdots,n-1),\qquad(1)$$

?1994-2014 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

$$\int_{(x_{n,n-1}+x_{nn})/2}^{+\infty} x\varphi(x)\mathrm{d}x = x_{nn}p_{nn}.$$

定理 4　$EX_n = EX$, $EX_n^2 = EXX_n$,

一般
$$EX_n^k = EXX_n^{k-1} \quad (k=1,2,\cdots).$$

证　$EXX_n^{k-1} = EX[f_n(X)]^{k-1} = \int_{-\infty}^{+\infty} x[f_n(x)]^{k-1}\varphi(x)\mathrm{d}x$

$$= \int_{-\infty}^{(x_{n1}+x_{n2})/2} x x_{n1}^{k-1}\varphi(x)\,\mathrm{d}x + \sum_{i=2}^{n-1}\int_{(x_{n,i-1}+x_{ni})/2}^{(x_{ni}+x_{n,i+1})/2} x x_{ni}^{k-1}\varphi(x)\mathrm{d}x$$

$$+ \int_{(x_{n,n-1}+x_{nn})/2}^{+\infty} x x_{nn}^{k-1}\varphi(x)\mathrm{d}x,$$

由(1),

$$EXX_n^{k-1} = \sum_{i=1}^{n} x_{ni}^k p_{ni} = EX_n^k.$$

定理 5　$DX_n = [1 - L(x_{n1}, x_{n2}, \cdots, x_{nn})]DX$.

证　$L(x_{n1}, x_{n2}, \cdots, x_{nn})DX$

$$= \int_{-\infty}^{(x_{n1}+x_{n2})/2} (x-x_{n1})^2\varphi(x)\,\mathrm{d}x + \sum_{i=2}^{n-1}\int_{(x_{n,i-1}+x_{ni})/2}^{(x_{ni}+x_{n,i+1})/2} (x-x_{ni})^2\varphi(x)\mathrm{d}x$$

$$+ \int_{(x_{n,n-1}+x_{nn})/2}^{+\infty} (x-x_{nn})^2\varphi(x)\mathrm{d}x$$

$$= \int_{-\infty}^{+\infty} [x - f_n(x)]^2\varphi(x)\mathrm{d}x = E[X - f_n(X)]^2 = E(X - X_n)^2, \qquad (2)$$

由定理 4,

$$L(x_{n1}, x_{n2}, \cdots, x_{nn})DX = EX^2 + EX_n^2 - 2EXX_n$$
$$= EX^2 + EX_n^2 - 2EX_n^2 = EX^2 - EX_n^2$$
$$= [EX^2 - (EX)^2] - [EX_n^2 - (EX_n)^2] = DX - DX_n,$$ 即证。

由定理1、3和5可知

$$DX_n \leqslant DX_{n+1},\quad DX_n \to DX \quad (n\to\infty).$$

定理 6　$\gamma_{X,X_n} = [1 - L(x_{n1}, x_{n2}, \cdots, x_{nn})]^{\frac{1}{2}}$.

证　由定理 4,

$$\mathrm{cov}(X, X_n) = EXX_n - EXEX_n = EX_n^2 - (EX_n)^2 = DX_n,$$

由定理 5,

?1994-2014 China Academic Journal Electronic Publishing House. All rights reserved.　http://www.cnki.net

$$\gamma_{X,\ X_n}=\frac{\operatorname{cov}(X,X_n)}{\sqrt{DX\ DX_n}}=\sqrt{\frac{DX_n}{DX}}=[1-L(x_{n1},x_{n2},\cdots,x_{nn})]^{\frac{1}{2}}.$$

由定理1、3和6可知

$$\gamma_{X,\ X_n}\leqslant\gamma_{X,X_{n+1}},\quad \gamma_{X,X_n}\to 1\ (n\to\infty).$$

定理 7 $E(X_n|X=x)=f_n(x)$，$E(X|X_n=x_{ni})=x_{ni}(i=1,2,\cdots,n)$。

证 前式是显然的，今证后式。

当 $i=2,3,\cdots,n-1$ 时，因在条件 $X_n=x_{ni}$ 下，X 的条件分布函数为

$$F(x|x_{ni})=P\{X<x|X_n=x_{ni}\}=\frac{P\{X<x,\ X_n=x_{ni}\}}{P\{X_n=x_{ni}\}}$$

$$=\frac{1}{p_{ni}}P\left\{X<x,\frac{x_{n,i-1}+x_{ni}}{2}<X\leqslant\frac{x_{ni}+x_{n,i+1}}{2}\right\}$$

$$=\begin{cases}0, & x\leqslant\dfrac{x_{n,i-1}+x_{ni}}{2};\\ \dfrac{1}{p_{ni}}\displaystyle\int_{(x_{n,i-1}+x_{ni})/2}^{x}\varphi(x)\mathrm{d}x, & \dfrac{x_{n,i-1}+x_{ni}}{2}<x\leqslant\dfrac{x_{ni}+x_{n,i+1}}{2};\\ 1, & \dfrac{x_{ni}+x_{n,i+1}}{2}<x,\end{cases}$$

于是在条件 $X_n=x_{ni}$ 下，X 的条件密度函数为

$$\varphi(x|x_{ni})=\begin{cases}\dfrac{\varphi(x)}{p_{ni}}, & \dfrac{x_{n,i-1}+x_{ni}}{2}<x\leqslant\dfrac{x_{ni}+x_{n,i+1}}{2};\\ 0, & \text{其它},\end{cases}$$

就有

$$E(X|X_n=x_{ni})=\int_{-\infty}^{+\infty}x\,\varphi(x|x_{ni})\mathrm{d}x=\int_{(x_{n,i-1}+x_{ni})/2}^{(x_{ni}+x_{n,i+1})/2}x\frac{\varphi(x)}{p_{ni}}\mathrm{d}x=x_{ni}.$$

同理可证

$$E(X|X_n=x_{n1})=x_{n1},\quad E(X|X_n=x_{nn})=x_{nn}.$$

由定理 7 可知 X_n 关于 X 的回归是 $x-x_n$ 平面上方程为 $x_n=f_n(x)$ 的阶梯曲线，而 X 关于 X_n 的回归是 $x-x_n$ 平面上坐标为 $(x_{ni},\ x_{ni})$，$i=1,2,\cdots,n$ 的点组成的点集[4]。

定理 8 设 X 的分布函数为 $F(x)$，X_n 的分布函数为 $F_n(x)$，则 $F_n(x)\to F(x)(n\to\infty)$。

证 由(2)和定理3，

$$E(X_n-X)^2=L(x_{n1},x_{n2},\cdots,x_{nn})DX\to 0\ (n\to\infty),$$

即 X_n 均方收敛于 X，从而 X_n 依分布收敛于 X[5]。

?1994-2014 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

参　考　文　献

1 方开泰，贺曙东.在正态总体中如何选择给定数目的代表点的问题.应用数学学报，1984；3

2 张立民.如何在母体中选取给定数目的代表点问题.数理统计与应用概率，1987；2

3 费荣昌.在Pearson分布族总体中选取代表点的问题.无锡轻工业学院学报，1990；1

4 费史.概率论及数理统计.上海科学技术出版社，1962

5 复旦大学.概率论　第一册.人民教育出版社，1979

Statistical Relationship Between the Representative Point and the Population

Fei Rongchang

(Dept. of Fund. Coures)

Abstract In this paper, statistical relationship between the representative point and the population is studied, and several properties of lose function are obtained.

Keywords population; representative point; lose function, statistical relationship

?1994-2014 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net