第11卷 第3期　郑 州 工 学 院 学 报　Vol.11 No.3
1990年 9 月　Journal of ZhengZhou Institute of Technology　Jun. 1990

分布参数u、σ的矩法估计量的改进*

张新育

(郑州工学院数力系)

摘　要: 本文改进了分布参数的矩法估计量$\hat{u}$、$\hat{\sigma}^2$. 证明了新的估计量$\hat{u}_{pq}$, $\hat{\sigma}^2_{pq}$的强相合性和其它大样本性质，并给出了他们在异常值检测方面的应用。

关键词: 矩法估计量，强相合性，异常值检测

在处理细尾极值分布样本的异常值问题时，首先会遇到异常值给分布的参数估计带来不良影响，以至使异常值的剔除出现遮蔽现象[4]。在利用剩余样本点构造统计量进行推断时，又遇到了确定统计量分布的难题。本文改进了u，σ^2的矩法估计量并使其强相合性条件有一定的减弱，这一估计量在细尾极值分布样本的异常值处理中有用。另外新估计量保留了原矩法估计量的渐近分布（若存在的话），并给出了正态总体下的具体结果，这一性质在大样本统计推断中有用。

1　改进的分布参数u、σ矩法估计量

1．1　分布参数矩法估计量的改进

分布函数F具有性质：$F(X)=G(\frac{X-u}{\sigma})$，其中标准分布函数G与u、$\sigma$无关，是已知确定的，u，$\sigma$是F(X)的参数，$\sigma>0$，其典型例子是正态分布N（u，$\sigma^2$)。我们先给出u，$\sigma^2$的矩法估计量。

定理1：设X_1，X_2，…，X_n iid$\sim F(X)=G(\frac{X-u}{\sigma})$，$E_GX$，$E_GX^2$存在，则u，$\sigma^2$的矩法估计量分别为：$\hat{u}=\overline{X}-u_1\hat{\sigma}$，$\hat{\sigma}^2=\frac{1}{n\sigma_1^2}\sum_{i=1}^{n}(X_i-\overline{X})^2$，$\hat{\sigma}=\sqrt{\hat{\sigma}^2}$，其中$u_1=E_GX_1$，$\sigma_1^2=D_GX_1$。

推论1：当E_GX，E_GX^2存在时，$\hat{u}$，$\hat{\sigma}^2$是u，σ^2的强相合估计量。(由强大数定律得

* 收稿日期：1989.09.04

到）。

推论2：当 $u_1=E_GX=0$，$\sigma_1^2=D_GX=1$ 时，$\hat{u}=\overline{X}$，$\hat{\sigma}^2=\frac{1}{n}\sum_{i=1}^{n}(X_i-\overline{X})^2$。

下面将顺序统计量 $X_{(1)}$，$X_{(2)}$，…，$X_{(n)}$ 中两端部分剔除，用剩余值构造两个新的估计量。

定义：设 X_1，X_2，…，X_n iid $\sim F(X)=G(\frac{X-u}{\sigma})$，$X_{(1)}<X_{(2)}<\cdots<X_{(n)}$ 为其顺序统计量，对 $0<p<q<1$，$u_{1pq}=\int_{G_p}^{G_q}x\mathrm{d}G(X)$，$u_{2pq}=\int_{G_p}^{G_q}x^2\mathrm{d}G(X)$。存在，其中 G_p，G_q 是G的下侧分位点，则 u，σ^2 的新的估计量为：

$$\hat{u}_{pq}=\frac{1}{q-p}\,\frac{1}{n}\sum_{i=[np]+1}^{[nq]}X_{(i)}-\frac{u_{1pq}}{q-p}\hat{\sigma}_{pq}$$

$$\hat{\sigma}_{pq}^2=\frac{(q-p)^2}{(q-p)u_{2pq}-u_{1pq}^2}\left[\frac{1}{q-p}\frac{1}{n}\sum_{i=[np]+1}^{[nq]}X_{(i)}^2-\left(\frac{1}{q-p}\frac{1}{n}\sum_{i=[np]+1}^{[nq]}X_{(i)}\right)^2\right]$$

1.2 u_{pq}，$\hat{\sigma}_{pq}^2$ 的强相合性

引理1：设 $X\sim F(X)=G(\frac{X-u}{\sigma})$，$u_{1pq}$，$u_{2pq}$ 存在，则：

$$\int_{F_p}^{F_q}X\mathrm{d}F(X)=u_{1pq}\sigma+(q-p)u^2$$

$$\int_{F_p}^{F_q}X^2\mathrm{d}F(X)=u_{2pq}\sigma^2+2u\sigma u_{1pq}+u^2(q-p)$$

其中：$F_p=\sigma G_p+u$，$F_q=\sigma G_q+u$。

证明：由 $X\sim F(X)=G(\frac{X-u}{\sigma})$，$Y=\frac{X-u}{\sigma}\sim G(y)$。$u_{1pq}$，$u_{2pq}$ 存在，则：

$$\int_{F_p}^{F_q}X\mathrm{d}F(X)=\int_{F_p}^{F_q}X\mathrm{d}G(\frac{X-u}{\sigma})=\int_{G_p}^{G_q}(\sigma y+u)\mathrm{d}G(y)=u_{1pq}\sigma+(q-p)u$$

同理：$\int_{F_p}^{F_q}X^2\mathrm{d}F(X)=u_{2pq}\sigma^2+2u\sigma u_{1pq}+(q-p)u^2$

引理2：设 $X_n\sim F_n(X)$，$X\sim F(X)$，h(x)连续，$\int|h(x)|\mathrm{d}F_n(x)$ 一致可积，$F_n\to F$，则 $\int h(x)\mathrm{d}F_n(x)\to\int h(x)\mathrm{d}F(x)$。（见[1]，P137(Vil)(d)）。

引理3：X_1，X_2，…，X_n iid $\sim F(X)=G(\frac{X-u}{\sigma})$，$X_{(1)}<X_{(2)}<\cdots<X_{(n)}$ 为其顺序统计量，G为严格上升、连续，h连续，则：

$$\frac{1}{n}\sum_{i=[np]+1}^{[nq]}h(X_{(i)})\to\int_{F_p}^{F_q}h(x)\mathrm{d}F(x)\qquad \text{a.s}\quad 0<p<q<1$$

证明：

$$\frac{1}{n}\sum_{i=[np]+1}^{[nq]}h(X_{(i)})=\sum_{\substack{i=[np]+1,\cdots,[nq]\\ F_p<X_{(i)}<F_q}}h(X_{(i)})\frac{1}{n}+\sum_{\substack{i=[np]+1,\cdots,[nq]\\ X_{(i)}<F_p}}h(X_{(i)})\frac{1}{n}+\sum_{\substack{i=[np]+1,\cdots,[nq]\\ X_{(i)}>F_q}}h(X_{(i)})\frac{1}{n}$$

① 由格列汶科定理：

$$|F_n(X_{([np]+1)})-F(X_{([np]+1)})| \xrightarrow{n\to\infty} 0,\ a.s$$

又：$F_n(X_{([np]+1)})=\dfrac{[np]+1}{n}\to P$, a.s

即：$|F_n(X_{([np]+1)})-P| \xrightarrow{n\to\infty} 0$, a.s

所以：$|F(X_{([np]+1)})-P|\to 0$, a.s

又F严格上升、连续，故F^{-1}存在且连续，故：

$$|F^{-1}[F(X_{([np]+1)})]-F^{-1}(P)|\to 0$$

即：$(X_{([np]+1)}\to F^{-1}(P)=F_p$, a.s

故：$\sum\limits_{\substack{i=[np]+1,\cdots,[nq]\\ X_{(i)}<F_p}} h(X_{(i)})\dfrac{1}{n}\to 0$, a.s

同理：$\sum\limits_{\substack{i=[np]+1,\cdots,[nq]\\ X_{(i)}>F_q}} h(X_{(i)})\dfrac{1}{n}\to 0$, a.s

② $\sum\limits_{\substack{i=[np]+1,\cdots,[nq]\\ F_p\leqslant X_{(i)}\leqslant F_q}} h(X_{(i)})\dfrac{1}{n}=\int_{F_p}^{F_q}h(x)dF_n(x)$，$F_n(x)$是$X_1$，$X_2$，…，$X_n$的经验分布函数，$F_n \xrightarrow{a.s} F\ (n\to\infty)$。

$h(x)$连续，$|h(x)|$在$[F_p, F_q]$上一致连续，必有界。

设：$|h(x)|\leqslant M$，$x\in[F_p, F_q]$，

则：$0\leqslant\int_{F_p}^{F_q}|h(x)|dF_n(x)\leqslant\dfrac{[nq]-[np]}{n}M<M$，

故$\int_{F_p}^{F_q}|h(x)|dF_n(x)$一致可积。

由引理2知：$\int_{F_p}^{F_q}h(x)dF_n(x)\to\int_{F_p}^{F_q}h(x)dF(x)$。

综合(1)和(2)知· $\dfrac{1}{n}\sum\limits_{i=[np]+1}^{[nq]}h(X_{(i)})\to\int_{F_p}^{F_q}h(x)dF(x)$, a.s

定理2：$\hat{u}_{pq} \xrightarrow{a.s} u$，$\hat{\sigma}^2_{pq} \xrightarrow{a.s} \sigma^2$，$(n\to\infty)$G连续，严格上升。

证明：由引理1和引理3，分别取$h(x)=x$，$h(x)=x^2$，

得：
$$\frac{1}{n}\sum_{i=[np]+1}^{[nq]}h(X_{(i)}) \xrightarrow{a.s} \int_{F_p}^{F_q}XdF(X)=\sigma u_{1pq}+u^2(q-p) \tag{1}$$

$$\frac{1}{n}\sum_{i=[np]+1}^{[nq]}(X^2_{(i)}) \xrightarrow{a.s} \int_{F_p}^{F_q}X^2dF(X)=\sigma^2u_{2pq}+2u\sigma u_{1pq}+u^2(q-p) \tag{2}$$

由：$\frac{(2)}{q-p}-\left(\frac{(1)}{q-p}\right)^2$

得：$$\sigma^2=\frac{(q-p)^2}{(q-p)u_{2pq}-u_{1pq}^2}\left\{\frac{\int_{F_p}^{F_q}X^2dF(x)}{q-p}-\left[\frac{\int_{F_p}^{F_q}XdF(x)}{q-p}\right]^2\right\} \quad (3)$$

由(1)和(3)得：

$$u=\frac{\int_{F_p}^{F_q}XdF(x)}{q-p}-\frac{u_{1pq}}{q-p}\sigma$$

故：$$\hat{u}_{pq}=\frac{1}{q-p}\frac{1}{n}\sum_{i=[np]+1}^{[nq]}X_{(i)}-\frac{u_{1pq}}{q-p}\hat{\sigma}_{pq}\to u,\quad a.s$$

$$\sigma_{pq}^2=\frac{(q-p)^2}{(q-p)u_{2pq}-u_{1pq}^2}\left\{\frac{1}{q-p}\frac{1}{n}\sum_{i=[np]+1}^{[nq]}X_{(i)}-\left(\frac{1}{q-p}\frac{1}{n}\sum_{i=[np]+1}^{[nq]}X_{(i)}\right)\right\}\to\sigma^2,\quad a.s$$

对比定理 1 的推论和定理 2 知 $\hat{u}_{pq}$，$\hat{\sigma}_{pq}^2$ 的强相合性条件比 $\hat{u}$，$\hat{\sigma}^2$ 有些地方得到了减弱，如 G 为哥西分布时，矩法估计量不存在更不具备相合性，而 $\hat{u}_{pq}$，$\hat{\sigma}_{pq}^2$ 分别是 u，σ^2 的强相合估计。$0<p<q<1$.

1．3　估计量 $\hat{u}_{pq}$，$\hat{\sigma}_{pq}^2$ 的渐近分布

引理 4：$\{X_n, Y_n\}$，$n=1, 2, \cdots$，是一列 R.V 变量偶，且：

$$|X_n-Y_n|\xrightarrow{P}0,\quad Y_n\xrightarrow{L}Y,$$

则：$X_n\xrightarrow{L}Y$　（见[1]，P319(ix)）。

定理 5：当 E_GX，E_GX^2 存在时，

若：$L(\hat{u})\to F_u$，　$L(\hat{\sigma}^2)\to F_{\sigma^2}$

则：$L(\hat{u}_{pq})\to F_u$，　$L(\hat{\sigma}_{pq}^2)\to F_{\sigma^2}$

引理 5：在正态总体下 $\hat{u}$，$\hat{\sigma}^2$ 是 u，σ^2 的极大似然估计，

σ^2 已知时，$L(\sqrt{n}\frac{\hat{u}-u}{\sigma})\to N(0, 1)$；

u 已知时，$L(\sqrt{\frac{n}{2}}\frac{\hat{\sigma}^2-\sigma^2}{\sigma^2})\to N(0, 1)$，（见[2]，P2-20，例 2.3.6）

定理 4：在正态总体下，

σ^2 已知时：$L(\sqrt{n}\frac{\hat{u}_{pq}-u}{\sigma})\to N(0, 1)$；

u 已知时，$L(\sqrt{\frac{n}{2}}\frac{\hat{\sigma}_{pq}^2-\sigma^2}{\sigma^2})\to N(0, 1)$　$(0<p<q<1)$。

在大样本推断中定理 3、定理 4 很有用，因为可以用剔除了异常值的统计量$\hat{u}_{pq}$，$\hat{\sigma}^2_{pq}$直接推断，使推断更精确。在构造新的估计量时可以根据需要截取多段顺序统计量，从而得到更一般的结论，限于篇幅不再给出。

1.4 若干常用系数

对$0<p<q<1$，估计量$\hat{u}_{pq}=W_{pq}-u_{1pq}\hat{\sigma}^2_{pq}$，$\hat{\sigma}^2_{pq}=C_{pq}(M_{pq}-W^2_{pq})$。

其中:

$$W_{pq}=\frac{1}{q-p}\frac{1}{n}\sum_{i=[np]+1}^{[nq]}X_{(i)},\quad M_{pq}=\frac{1}{q-p}\frac{1}{n}\sum_{i=[np]+1}^{[nq]}X^2_{(i)},\quad C_{pq}=\frac{(q-p)^2}{(q-p)u_{2pq}0u^2_{1pq}}$$

当 G 为 N(0, 1)，指数分布 P(1)及哥西分布时有如下系数表:

	N (0, 1)		P(1)		哥西分布	
p	1/4	1/8	0	0	1/4	1/8
q	3/4	7/8	3/4	7/8	3/4	7/8
u_{1pq}	0	0	0.4034	0.6151	0	0
u_{2pq}	0.07147	0.2762	0.3264	0.6896	0.1366	0.6619
C_{pq}	6.9959	2.7154	6.8540	3.4020	3.6603	1.1331

2 $\hat{u}_{pq}$，$\hat{\sigma}^2_{pq}$在细尾分布异常值检测中的应用

定理 4: X_1，X_2，…，X_n是来自 F(x)的样本，$F(X)=G(\frac{X-u}{\sigma})$，$X_{(1)}<X_{(2)}<\cdots<X_{(n)}$为其顺序统计量，G 细尾严格上升，且有密度函数 g(x)，$\exists\,\delta_1<\delta_2$，使$\int_{\delta_2}^{\delta_1}x\,dG(X)$，$\int_{\delta_2}^{\delta_1}x^2dG(X)$存在，则当 n 较大时有:

① $P\{\frac{X_{(i)}-\hat{u}_{pq}}{\hat{\sigma}_{pq}}<l_\alpha\}=1-\alpha$;

② $P\{\frac{\hat{u}_{pq}-X_{(i)}}{\hat{\sigma}_{pq}}<k_\alpha\}=1-\alpha$。

其中: $l_\alpha\approx k_\alpha\approx G^{-1}(\sqrt[n]{1-\alpha})$，$G(\delta_1)\leqslant p<q\leqslant G(\delta_2)$。

证明: 设Y_1，Y_2，…，Y_n idd～R(0, 1)，$Y_{(1)}<Y_{(2)}<\cdots<Y_{(n)}$为其顺序统计量，则$G(\frac{X_{(n)}-\hat{u}_{pq}}{\hat{\sigma}_{pq}})$，$G(\frac{\hat{u}_{pq}-X_{(1)}}{\hat{\sigma}_{pq}})$分别近似与$Y_{(n)}$，$1-Y_{(n)}$同分布，又$1-Y_{(n)}$与$Y_{(n)}$同分布，故有: $G(\frac{\hat{u}_{pq}-X_{(1)}}{\hat{\sigma}_{pq}})$，$G(\frac{X_{(n)}-\hat{u}_{pq}}{\hat{\sigma}_{pq}})$与$Y_{(n)}$近似同分布。其中$X_{(1)}$，$X_{(n)}$看作 R.V。又$Y_{(n)}\sim P(x)=nX^{n-1}(0<X<1)$，故$l_\alpha\approx k_\alpha\approx G^{-1}(\sqrt[n]{1-\alpha})$。

注1：一般有$\int_{\delta_1}^{\delta_2} dG(x) \approx 1$。

注2：适当选择p、q可以使检测减弱遮蔽现象。

注3：当u、σ^2出现已知时，以真值代替估计量。

注4：定理4是常见方法的改进，在常见方法中也可直接使用$\hat{u}_{pq}$，$\hat{\sigma}^2_{pq}$得出新结果[4]。

致谢：本文作者感谢韩芝隆、刘江国两位老师的帮助。

参　考　文　献

〔1〕C.R.劳著．线性统计推断及其应用．科学出版社

〔2〕成平，陈希儒著．参数估计．上海科技出版社，1985

〔3〕茆诗松，王静龙编．数理统计教程．华东师大出版社

〔4〕《异常值处理》．华东师大数理统计系研究生讲义，1989

〔5〕朱宏．用样本分位数方法同时检测正态样本多个异常值．数理统计与应用概率．Vol.4，No.1，1989

〔6〕中国科学院数学所概率统计室编．常用数理统计表．科学出版社，1974

An Improvement of the Moment Method Estimation for the Parameters u and σ of the Distribution

Zhang Xinyu

(ZhengZhou Institute of Technology)

Abstract: This paper improves the moment method estimation $\hat{u}$ and $\hat{\sigma}^2$ for the parameters of the extremal, distribution. It prove that the new estimation $\hat{u}_{pq}$ and $\hat{\sigma}^2_{pq}$ have strongly consi-stency and other great sampling prepertirs. It give the application of them in rejecting outliers.

Keywords: moment method estimation, strongly consistency, rejecting outliers