影响曲线与稳健估计

曾焰

版本 1.0，最后修订于 2017-11-05

摘要

孙山泽 [5] 第八章的内容摘要。

1 影响曲线

总体分布 $F(x)$ 的特征数 γ 可以看作是分布 $F(x)$ 的泛函 $T(F)$。例如，期望值

$$\mu = T_1(F) = \int_{-\infty}^{\infty} t dF(t),$$

中位数

$$\gamma = T_2(F) = \inf\left\{x : F(x) \geq \frac{1}{2}\right\}$$

若以样本经验分布函数 $F_n(x)$ 替代 $F(x)$，则构成一个统计量，如

$$T_1(F_n) = \int_{-\infty}^{\infty} t dF_n(t) = \frac{1}{n}\sum_{i=1}^{n} x_i = \bar{x},$$

$$T_2(F_n) = \inf\left\{x : F_n(x) \geq \frac{1}{2}\right\} = \mathrm{median}(x_1, \cdots, x_n)$$

绝大多数情况在适当的正则条件下，这些统计量服从渐近正态分布，即 $\sqrt{n}(T(F_n) - T(F))$ 有渐近正态分布 $N(0, V(F))$，这里 $V(F)$ 为 $T(F_n)$ 的渐近方差。

定义 1.1. *设 $T(\cdot)$ 是定义在一个分布函数 $F(x)$ 的集合上的泛函，x 是任意实数，$\delta_x(t)$ 是概率为 1 取值 x 的单点分布函数，$T(\cdot)$ 的***影响函数***定义为*

$$IC_{T,F}(x) = \lim_{\varepsilon \to 0}\left[\frac{T((1-\varepsilon)F(\cdot) + \varepsilon\delta_x(\cdot)) - T(F(\cdot))}{\varepsilon}\right]$$

$IC_{T,F}(x)$ 曲线反映了总体分布 $F(x)$ 在 x 点分布概率有一点增加时，$T(\cdot)$ 的变化。对均值统计量 $T_1(\cdot)$，有

$$IC_{T_1,F}(x) = x - \mu,$$

其中 $\mu=\int_{-\infty}^{\infty}tdF(t)$ 为 $F(x)$ 的期望值。对中位数统计量 $T_2(\cdot)$，有

$$IC_{T_2,F}(x)=\begin{cases}-\frac{1}{2f\left(F^{-1}\left(\frac{1}{2}\right)\right)}, & x<F^{-1}\left(\frac{1}{2}\right),\\ 0, & x=F^{-1}\left(\frac{1}{2}\right),\\ \frac{1}{2f\left(F^{-1}\left(\frac{1}{2}\right)\right)}, & x\;F^{-1}\left(\frac{1}{2}\right),\end{cases}$$

其中 $f(t)=F'(t)$ 为密度函数。

可以看出当 x 取远离中心值的极端值，其概率有一点增加时，样本均值比样本中位数的变化要大。这说明样本中位数比样本均值对极端的样本观测值有更好的抗御能力，具有较好的稳健性。但是对于正态分布总体，中位数远不如样本均值有效。人们总是希望寻求稳健性较好，同时对最常遇到的正态等特殊形式的分布又效率较高的估计量。

A. A. Filippova 在论文 [1] 中用泰勒展开证明了，在一般情况下有

$$\sqrt{n}\left[T(F_n)-T(F)-\int_{-\infty}^{\infty}IC_{T,T}(t)dF_n(t)\right]\xrightarrow{P}0,$$

$T(F_n)$ 的渐近方差为

$$V(F)=\int_{-\infty}^{\infty}[IC_{T,F}(t)]^2dF(t)$$

2 次序统计量的线性组合估计

许多常用的稳健估计是样本次序统计量的线性组合。以 $X_{(1)}\leq\cdots\leq X_{(n)}$ 记样本 X_1，$\cdots$，X_n 的次序统计量，

$$T=\sum_{i=1}^{n}a_iX_{(i)},$$

其中 $\sum_{i=1}^{n}a_i=1$。当取 $a_i=\frac{1}{n}$ $(i=1,\cdots,n)$ 时，即为样本均值。在上一节我们已经看到对样本的极端观察值，均值是非常敏感的。为使其不太敏感，一个非常简单的方法是对样本观测值进行“修剪”，删去那些最极端的观测值。通常规定一个删去的比例 α，用泛函的形式表示，修剪均值可写成

$$T_\alpha(F)=\frac{1}{1-2\alpha}\int_{\alpha}^{1-\alpha}F^{-1}(t)dt=\frac{1}{1-2\alpha}\int_{F^{-1}(\alpha)}^{F^{-1}(1-\alpha)}tdF(t)$$

对应的估计量（称为**修剪均值**或**截尾均值**）即为

$$T_\alpha(F_n)=\frac{bX_{[\alpha n]+1}+X_{[\alpha n]+2}+\cdots+X_{n-[\alpha n]-1}+bX_{n-[\alpha n]}}{n(1-2\alpha)},$$

其中 $[\alpha n]$ 表示 αn 的整数部分，$b=1+[\alpha n]-\alpha n$ 是一个加权系数。也就是说，如果 αn 是整数，则修剪均值是对样本观测值依大小从两端各删去 αn 个观测后的平均值。若 αn 不是整数，则从两端各删去 $[\alpha n]$ 个观测外，对剩余的观测值中最大值和最小值再给一个删去权数 $\alpha n-[\alpha n]$ 进行加权平均。特别 $\alpha=0$ 时，修剪均值就是通常的样本均值，$\alpha=0.5$ 时，修剪均值对应的就是样本中位数，$\alpha=0.25$ 对应的修剪均值一般称为中段均值（midmean）。常用的 α 取值为 0.05 或 0.10。

α-修剪均值的影响曲线为

$$IC(x)=\begin{cases}\frac{1}{1-2\alpha}(F^{-1}(\alpha)-C(\alpha)), & x<F^{-1}(\alpha),\\ \frac{1}{1-2\alpha}(x-C(\alpha)), & F^{-1}(\alpha)\le x\le F^{-1}(1-\alpha),\\ \frac{1}{1-2\alpha}(F^{-1}(1-\alpha)-C(\alpha)), & x>F^{-1}(1-\alpha),\end{cases}$$

其中 $C(\alpha)=\int_{\alpha}^{1-\alpha}F^{-1}(t)dt+\alpha(F^{-1}(\alpha)+F^{-1}(1-\alpha))$。$T_\alpha(F_n)$ 的渐近方差为

$$V(F)=\frac{1}{(1-2\alpha)^2}\left\{\int_{F^{-1}(\alpha)}^{F^{-1}(1-\alpha)}(t-C(\alpha))^2dF(t)+\alpha\left[(F^{-1}(\alpha)-C(\alpha))^2+(F^{-1}(1-\alpha)-C(\alpha))^2\right]\right\}$$

另一类样本次序统计量线性组合形式的稳健估计为适当选择的线性组合。这一类估计量写成泛函形式为

$$T(F)=\sum_{i=1}^{m}\beta_iF^{-1}(t_i),$$

其中 $\sum_{i=1}^m\beta_i=1$，$0<t_1<\cdots<t_m<1$。这类估计量的影响曲线为

$$IC(x)=\sum\frac{\beta_i}{f(F^{-1}(t_i))}=C(F),$$

上式中求和号 $\sum$ 是对满足 $i:F^{-1}(t_i)<x$ 条件的 i 值求和，$C(F)$ 为常数，使得 $IC(x)$ 满足 $\int_{-\infty}^{\infty}IC(x)dF(x)=0$。这类估计的渐近方差可用下式求出：

$$V(F)=\int_{-\infty}^{\infty}[IC(x)]^2dF(x)$$

3 M 估计

设随机变量 X_1，$\cdots$，X_n 相互独立同分布，是来自总体 $F(x-\theta)$ 的样本，$F(x-\theta)$ 关于 θ 对称。位置参数 θ 的传统的最小二乘估计方法，是选择 $\theta=\hat\theta$，使

$$Q=\sum_{i=1}^{n}(X_i-\theta)^2$$

达到最小值。可通过正规方程

$$\frac{\partial Q}{\partial\theta}=-2\sum_{i=1}^{n}(X_i-\theta)=0$$

解出 $\hat\theta$。这里采用了平方 $(X_i-\theta)^2$ 作为 X_i 与 θ 的距离。这种获得估计的方法很易推广到一般的距离函数 $\rho(t)$：我们可以选择 $\theta=\hat\theta$，使

$$Q=\sum_{i=1}^{n}\rho(X_i-\theta)$$

达到最小值。当 $\rho(t)$ 对全部 t 有有限导数时，$\hat\theta$ 应满足方程

$$\sum_{i=1}^{n}\rho'(X_i-\theta)=0$$

也有一些重要的距离 $\rho(t)$，并非对一切 t 都有导数。如 $\rho(t)=|t|$ 在 $t=0$ 不存在导数。要选 $\theta=\hat{\theta}$，使得

$$Q=\sum_{i=1}^{n}|X_i-\theta|$$

达到最小值。可以证明，这样的 $\hat{\theta}$ 是样本中位数。在这种情形，$\hat{\theta}$ 可以看成是满足下列方程的一个解

$$\sum_{i=1}^{n}\psi(X_i-\theta)=0,$$

其中函数

$$\psi(t)=\begin{cases}-1, & t<0,\\ 0, & t=0,\\ 1, & t>0\end{cases}$$

但这一方程的解不一定唯一。解决这一问题的方法是采用如下定义。

定义 3.1. 设 $\psi(\cdot)$ 是满足下列性质的函数：

(i) $\psi(t)$ 是在全直线上定义的一个非降、非常数的函数。

(ii) 对一切 t，有 $\psi(t)=-\psi(-t)$。

由 $\psi(\cdot)$ 产生的下列估计称为M **估计**：

$$\hat{\theta}(X_1,\cdots,X_n)=\frac{\theta^*+\theta^{**}}{2},$$

其中

$$\theta^*(x_1,\cdots,x_n)=\sup\left\{\theta:\sum_{i=1}^{n}\psi(x_i-\theta)>0\right\},$$

$$\theta^{**}(x_1,\cdots,x_n)=\inf\left\{\theta:\sum_{i=1}^{n}\psi(x_i-\theta)<0\right\}$$

可以证明，一定存在有限的 θ^* 和 θ^{**}。当 $\psi(t)$ 连续且严格递增时，M 估计 $\hat{\theta}$ 可从下式解出

$$\sum_{i=1}^{n}\psi(X_i-\theta)=0$$

如果总体分布除未知位置参数 θ 外，还有尺度参数 σ，则构造 M 估计需要 $\psi(\cdot)$ 和另一个函数 $\eta(\cdot)$，从方程组

$$\begin{cases}\sum_{i=1}^{n}\psi\left(\frac{X_i-\theta}{\sigma}\right)=0,\\ \sum_{i=1}^{n}\eta\left(\frac{X_i-\theta}{\sigma}\right)=0\end{cases}$$

解出 θ 和 σ。这是一个很广泛的估计类，参数估计中常用的极大似然估计也在其中。若总体分布有密度函数

$$\frac{1}{\sigma}f\left(\frac{t-\theta}{\sigma}\right),$$

可以证明，当取

$$\psi(t)=-\frac{f'(t)}{f(t)},\ \eta(t)=-\frac{tf'(t)}{f(t)}-1$$

时，对应的 M 估计即为极大似然估计。

P. J. Huber 在 1964 年提出一类 M 估计，他取

$$\psi(t)=\begin{cases}-k, & t<-k,\\ t, & -k\le t\le k,\\ k, & t>k.\end{cases}$$

而取

$$\eta(t)=\psi'(t)-\beta(k),$$

其中 $\beta(k)=\int_{-\infty}^{\infty}\psi^2(t)d\Phi(t)$，$\Phi(t)$ 为标准正态分布函数。解方程组

$$\begin{cases}\sum_{i=1}^n \psi\left(\frac{x_i-\theta}{\sigma}\right)=0,\\ \sum_{i=1}^n \eta\left(\frac{x_i-\theta}{\sigma}\right)=0\end{cases}$$

一般需要通过迭代解出 θ 和 σ，初始值可以取 θ_0 为样本中位数，σ_0 取 $(x_{(n)}-x_{(1)})/1.35$，$x_{(n)}-x_{(1)}$ 为样本差程。Huber 所取的 $\psi(\cdot)$ 对应的距离为

$$\rho(t)=\begin{cases}\frac{1}{2}t^2, & |t|\le k,\\ k|t|-\frac{1}{2}k^2, & |t|>k.\end{cases}$$

这个函数在中部采用了平方距离，而在两端采用了绝对值距离，因此这一类 M 估计倾向于像取平均一样处理中间大小的样本观测值，而像取中位数一样不重视两边极端的样本观测值。

如果假定总体分布密度为 $\frac{1}{\sigma}f\left(\frac{x-\theta}{\sigma}\right)$，位置参数 θ 和尺度参数 σ 的上述这些 M 估计写成泛函 $T(F)$ 和 $S(F)$。则在 $F(x)$ 对称，$\psi(\cdot)$ 为奇函数，$\eta(\cdot)$ 为偶函数时，可求出 $T(F)$ 和 $S(F)$ 的影响曲线为

$$IC_{T,F}(x)=\left[S(F)/\int_{-\infty}^{\infty}\psi'\left(\frac{t}{S(F)}\right)dF(t)\right]\psi\left(\frac{x}{S(F)}\right),$$

$$IC_{S,F}(x)=\left[S(F)/\int_{-\infty}^{\infty}\eta'\left(\frac{t}{S(F)}\right)\frac{t}{S(F)}dF(t)\right]\eta\left(\frac{x}{S(F)}\right)$$

可以看到，对于对称分布 $F(\cdot)$，上面提到的 Huber 的 M 估计当取 $\alpha=F(-kS(F))$ 时，其 $T(F)$ 的影响曲线与 α-修剪平均的影响曲线相同。

扩展阅读： Wilcox [2] 提供了关于稳健统计学的非技术性阐述。Wilcox [4] 将经典统计学理论与稳健统计学理论进行了对比阐述，并提供了详细的算法描述。Wilcox [3] 则是以 R 为脚本语言的算法手册。

参考文献

[1] A. A. Filippova. “Mises theorem on the asymptotic behavior of functionals of empirical distribution functions and its statistical applicaitons”, *Theory of probability and its applications*, vol. 7 (1962), 24-57.

[2] Rand R. Wilcox. *Fundamentals of Modern Statistical Methods: Substantially Improving Power and Accuracy*, 2nd ed., Springer, 2010.

[3] Rand R. Wilcox. *Introduction to Robust Estimation and Hypothesis Testing*, 4th ed., Academic Press, 2016.

[4] Rand R. Wilcox. *Modern Statistics for the Social and Behavioral Sciences: A Practical Introduction*, 2nd ed., Chapman and Hall/CRC, 2017.

[5] 孙山泽：《非参数统计讲义》。北京：北京大学出版社，2000.4。