

基于负数阶矩的正值 Alpha 稳定分布的参数估计

孙增国, 韩崇昭

(西安交通大学电子与信息工程学院, 710049, 西安)

摘要: 为了精确估计正值 alpha 稳定(PAS)分布的参数, 基于负数阶矩理论, 提出了比值估计、对数矩估计和迭代对数矩估计 3 种参数估计方法. 比值估计直接利用特定阶次的负数阶矩的比值来估计未知参数, 对数矩估计利用 PAS 分布的对数变换及其负数阶矩的 Taylor 展式从而获得解析的估计形式, 迭代对数矩估计通过样本分段迭代估计未知参数. 与传统的估计方法相比, 所提出的 3 种估计方法可以获得更高的估计精度, 并且对数矩估计具有较低的计算复杂度. Monte Carlo 仿真实验表明, 当独立运行次数为 100、样本总数为 5 000 时, 比值估计的估计精度可以达到 99.8%, 对数矩估计的估计精度可以达到 99.95%, 迭代对数矩估计的估计精度可以达到 99.94%.

关键词: 负数阶矩; 正值 alpha 稳定分布; Monte Carlo 仿真; 对数矩估计; 迭代对数矩估计

中图分类号: TN911 **文献标识码:** A **文章编号:** 0253-987X(2007)06-0645-05

Positive Alpha-Stable Distributed Parameter Estimation Based on Negative-Order Moments

Sun Zengguo, Han Chongzhao

(School of Electronics and Information Engineering, Xi'an Jiaotong University, Xi'an 710049, China)

Abstract: In order to accurately estimate the positive alpha-stable (PAS) distributed parameters, three kinds of parameter estimate methods, namely ratio, logarithmic moment and iterative logarithmic moment estimate are proposed based on negative-order moments. Ratio of the negative-order moments is directly used for the ratio estimation, and the logarithmic transformation of PAS distribution and Taylor series of the negative-order moment are used to obtain an analytical form of logarithmic moment estimation. Through dividing all samples into different segments, the iterative logarithmic moment estimation is obtained according to iterative computation of data segments. Compared to the conventional estimates, these three estimates achieve much higher estimation accuracy, and the logarithmic moment estimator has lower computing complexity. Monte Carlo simulation results demonstrate that the estimation accuracy of these three estimations can reach 99.8%, 99.95%, and 99.94% respectively when the independent running times are 100 and the number of all samples is 5 000.

Keywords: negative-order moment; positive alpha-stable distribution; Monte Carlo simulation; logarithmic moment estimation; iterative logarithmic moment estimation

正值 alpha 稳定(Positive Alpha-Stable, PAS)分布是一类特殊的稳定分布, 它的概率密度函数仅分布在正半轴上, 它能够有效表示非负信号和噪声的拖尾特征, 因此常用于非负物理量的建模中^[1-2].

在使用 PAS 分布来描述信号的过程中, 参数估计是一个关键问题, 但由于 PAS 分布没有解析形式的密度函数和特征函数, 因此常用于稳定分布参数估计的最大似然法和样本特征函数法很难精确估计 PAS

分布的未知参数^[1]. 为了正确估计出 PAS 分布的未知参数, 基于负数阶矩理论, 本文给出了比值估计、对数矩估计和迭代对数矩估计 3 种参数估计方法. 比值估计利用特定阶次的负数阶矩的比值来估计未知参数, 对数矩估计利用 PAS 分布的对数变换及其负数阶矩的 Taylor 展式来获得解析的估计形式, 迭代对数矩估计通过样本分段迭代估计未知参数. Monte Carlo 仿真试验证明, 这 3 种估计方法均可以获得较高的估计精度, 同时对数矩估计具有较低的计算复杂度.

1 比值估计

PAS 分布没有统一的解析形式的概率密度函数, 它的特征函数如下

$$\varphi(t) = \exp(-\gamma |t|^\alpha [1 + \text{isgn}(t)\tan(\alpha\pi/2)]) \quad (1)$$

式中: $0 < \alpha < 1$ 是特征指数, 它表示 PAS 分布的拖尾程度; $\gamma > 0$ 是尺度参数, 它表示 PAS 分布的分散程度; $\text{sgn}(\cdot)$ 表示符号函数^[1]. 如果随机变量 X 服从 PAS 分布, 则它的负数阶矩可写为

$$E(X^p) = \frac{\gamma^{p/\alpha} \Gamma(-p/\alpha) [1 + \tan^2(\pi\alpha/2)]^{p/2\alpha}}{\alpha \Gamma(-p)} \quad p < 0 \quad (2)$$

式中: $E(\cdot)$ 表示数学期望; p 表示负数阶次; $\Gamma(\cdot)$ 表示 Gamma 函数^[2]. 显然, α 可由如下比值得到

$$\frac{E(X^{2p})}{E^2(X^p)} = \frac{\Gamma(-p)}{\Gamma(1/2 - p)} \frac{\alpha^{2^{2p-2p/\alpha}} \Gamma(1/2 - p/\alpha)}{\Gamma(-p/\alpha)} \quad (3)$$

而 γ 可直接由式(2)获得^[3].

比值估计中的 $E(X^p)$ 和 $E(X^{2p})$ 可由样本计算近似得出, 一旦样本给定, 估计结果就仅由阶次 p

决定. 如果选择 $p = -0.5$, 经过适当化简, 可得 α 的估计($\hat{\alpha}$)如下

$$\frac{E(X^{-1})}{E^2(X^{-0.5})} = \frac{\hat{\alpha} \pi \Gamma(1/\hat{\alpha})}{\Gamma^2(1/2\hat{\alpha})} \quad (4)$$

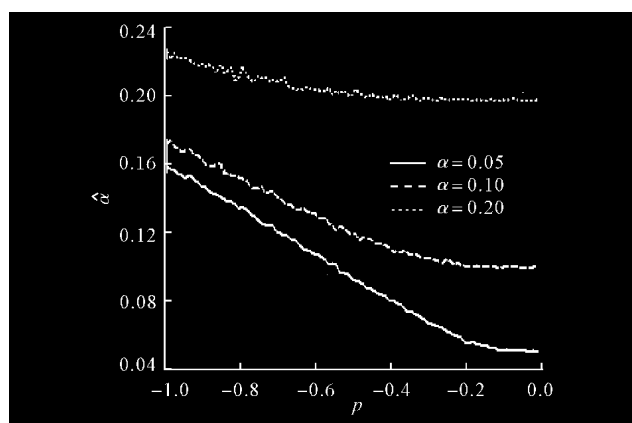
式(4)就是 Pierce 提出的估计式, 它仅是比值估计通式(3)的特例^[2].

表 1 给出了在不同阶次和样本数量条件下比值估计 $\hat{\alpha}$ 、 $\hat{\gamma}$ 的 Monte Carlo 仿真结果, 服从 PAS 分布的样本由文献[4]所提方法产生, 每组实验均独立运行 100 次. 由表 1 可见, 阶次 p 取 -0.1 时的估计性能明显优于 p 取 -0.5 , 即比值估计结果受到 p 值选取的影响很大. 图 1 给出了估计值随阶次 p 的变化关系, 为方便讨论, 这里限定 p 在 $(-1, 0)$ 上取值. 由图 1 可见, 随着 p 的增加, 估计值逐步接近真值, 估计偏差逐渐减小. 另外可以看出, 越是较小的 α 真值, 比值估计结果受到 p 值选取的影响就越大. 图 2 给出了估计值随 α 真值的变化关系, 可见 α 真值越小, 参数估计值对 p 值的选取就越敏感. 对于较小的 α 真值, 只有选择较大的 p 值才可能获得较高的估计精度.

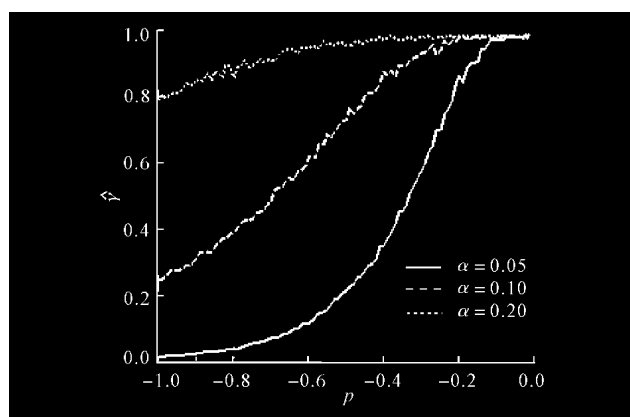
表 1 比值估计的仿真结果 ($\alpha=0.1, \gamma=1$)

样本数量	$\hat{\alpha}$			$\hat{\gamma}$		
	$p=-0.5$	$p=-0.3$	$p=-0.1$	$p=-0.5$	$p=-0.3$	$p=-0.1$
1 000	0.119 1	0.106 9	0.100 2	0.754 7	0.931 5	1.003 7
	(0.013 4)	(0.009 4)	(0.002 8)	(0.123 0)	(0.102 9)	(0.036 7)
2 000	0.113 6	0.102 5	0.099 9	0.815 5	0.976 5	0.999 7
	(0.0119)	(0.007 6)	(0.002 0)	(0.130 4)	(0.086 4)	(0.025 2)
5 000	0.109 6	0.102 5	0.100 2	0.868 0	0.975 7	1.000 8
	(0.009 5)	(0.005 6)	(0.001 2)	(0.127 5)	(0.063 5)	(0.014 9)

注: 括号中的数值是多次实验结果的标准差, 以下各表相同.

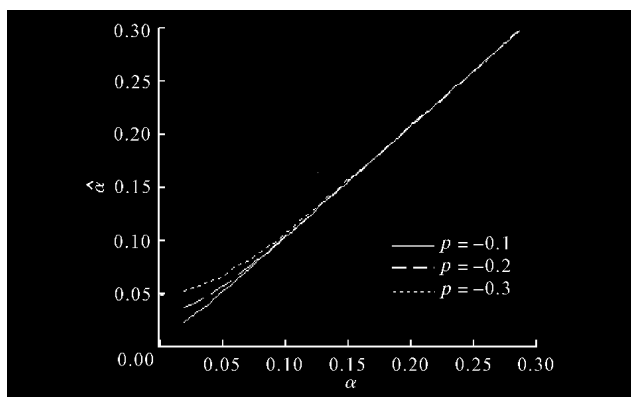


(a) 特征指数估计值

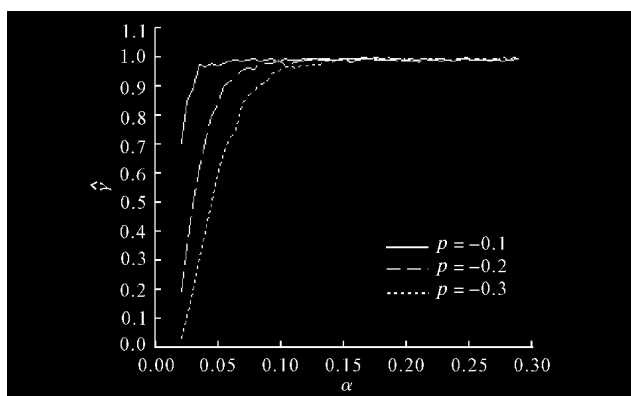


(b) 尺度参数估计值

图 1 比值估计的参数估计值随阶次的变化关系 (独立运行 100 次, 样本数为 1 000, $\gamma=1$)



(a)特征指数估计值



(b)尺度参数估计值

图2 比值估计的参数估计值随 α 真值的变化关系
(独立运行100次,样本数为1 000, $\gamma=1$)

面两式中的均值和方差用经过对数变换后的样本均值和样本方差代替,就可以方便地估计出 α 和 γ .和比值估计相比,对数矩估计完全由样本决定,并且具有较为简单的形式,便于计算.

表2给出了对数矩估计的 Monte Carlo 仿真结果,每组实验均独立运行100次.由表2可见,不论 α 真值取为多少,这种方法均能获得较高精度的参数估计结果.图3给出了估计结果随样本数量的变化关系.由图3可见,由于样本产生的随机性,估计值出现明显的振荡,但总体上随着样本数量的增加而逼近真值.与估计值相比,估计标准差的变化曲线要光滑得多,并且随着样本数量的增加逐步降低,即较高精度的估计结果需要较多的样本数量.

表2 对数矩估计的仿真结果($\gamma=1$)

样本 数量	$\hat{\alpha}$			$\hat{\gamma}$		
	$\alpha=0.1$	$\alpha=0.5$	$\alpha=0.9$	$\alpha=0.1$	$\alpha=0.5$	$\alpha=0.9$
1 000	0.100 1 (0.003 5)	0.502 0 (0.013 5)	0.900 2 (0.012 3)	1.003 9 (0.034 0)	1.002 0 (0.032 9)	0.995 7 (0.100 6)
2 000	0.099 9 (0.002 5)	0.500 2 (0.009 4)	0.902 1 (0.010 0)	1.000 6 (0.026 4)	1.000 8 (0.022 8)	0.982 3 (0.081 0)
5 000	0.100 0 (0.001 5)	0.501 0 (0.006 1)	0.900 0 (0.006 5)	0.999 5 (0.014 9)	0.996 4 (0.014 2)	0.999 6 (0.053 5)

2 对数矩估计

比值估计性能严重依赖阶次 p 的选择,一旦阶次选择不当,估计性能就会显著下降,因此希望获得不依赖阶次的参数估计方法^[5].如果随机变量 X 服从PAS分布,可以把 $E(X^p)$ 写为 $E(e^{p \ln X})$,并定义新的随机变量 $Y=\ln X$,利用 Taylor 级数可得

$$E(e^{pY}) = \sum_{k=0}^{\infty} E(Y^k) \frac{p^k}{k!} = C(p, \alpha) \gamma^{p/\alpha} \quad (5)$$

式中: $C(p, \alpha) = \frac{\Gamma(-p/\alpha) [1 + \tan^2(\pi\alpha/2)]^{p/2\alpha}}{\alpha \Gamma(-p)}$.由式(5)可得

$$E(Y^k) = \left. \frac{d^k (C(p, \alpha) \gamma^{p/\alpha})}{dp^k} \right|_{p=0} \quad (6)$$

经过适当化简可得

$$E(Y) = C_e \left(\frac{1}{\alpha} - 1 \right) + \frac{\ln \gamma}{\alpha} + \frac{\ln(\sec^2(\pi\alpha/2))}{2\alpha} \quad (7)$$

$$D(Y) = \frac{\pi^2(1-\alpha^2)}{6\alpha^2} \quad (8)$$

式中: C_e 代表欧拉常数; $D(\cdot)$ 表示方差.如果把上

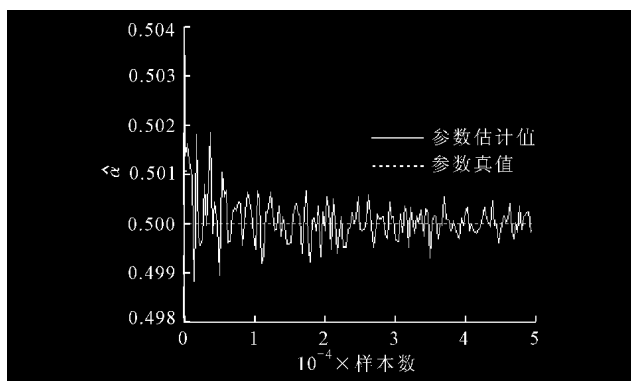
3 迭代对数矩估计

对数矩估计必须要用大量数据样本才能获得较高精度的估计结果,并且参数估计是通过所有样本的一次计算完成的.迭代对数矩估计的基本思想^[6]是:把经过对数变换的所有数据样本分为不重叠的 B 个数据段,每个数据段均含有 N 个样本点,并用 $\hat{\alpha}(k)$ 和 $\hat{\gamma}(k)$ 表示由前 k ($1 \leq k \leq B$)个数据段得到的参数估计值,它们可以由 $\hat{\alpha}(k-1)$ 和 $\hat{\gamma}(k-1)$ 得到.假定 $M_{1,k}$ 表示由前 k 个数据段得到的样本均值,则

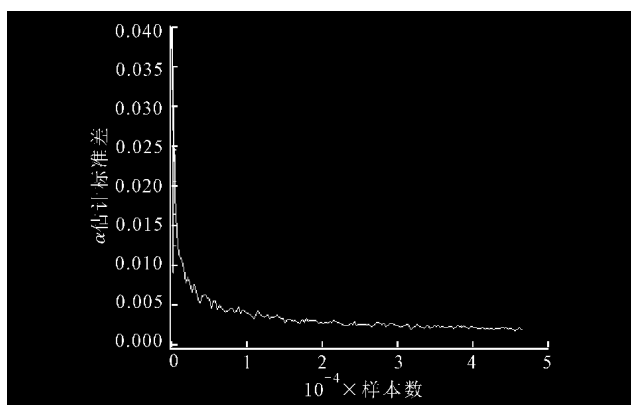
$$M_{1,k} = \frac{k-1}{k} M_{1,k-1} + \frac{1}{k} M_k \quad (9)$$

式中: M_k 表示由第 k 个数据段获得的样本均值.类似地,假定 $V_{1,k}$ 表示由前 k 个数据段获得的样本方差,则

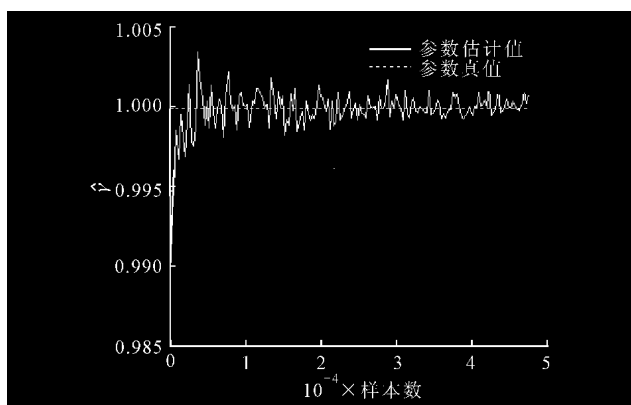
$$V_{1,k} = \frac{(k-1)N-1}{kN-1} V_{1,k-1} + \frac{N-1}{kN-1} V_k + \frac{(k-1)N}{k(kN-1)} (M_{1,k-1} - M_k)^2 \quad (10)$$



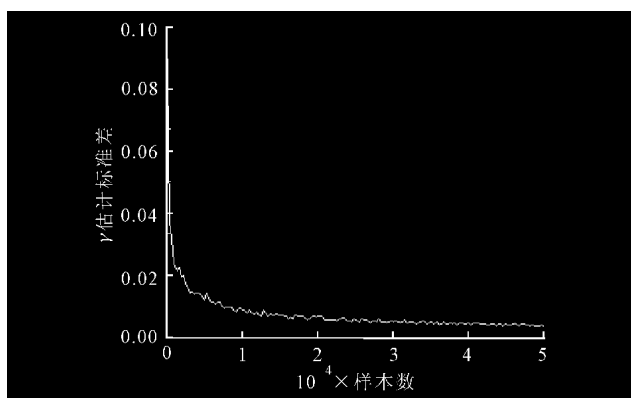
(a)特征指数估计值



(b)特征指数估计标准差



(c)尺度参数估计值



(d)尺度参数估计标准差

图3 对数矩估计的性能随样本数量的变化关系
(独立运行100次, $\alpha=0.5, \gamma=1$)

式中: V_k 表示由第 k 个数据段获得的样本方差. 如果把上面两式中的样本均值 $M_{1,k}$ 和 $M_{1,k-1}$ 以及样本方差 $V_{1,k}$ 和 $V_{1,k-1}$ 分别用式(7)、式(8)中的 $E(Y)$ 和 $D(Y)$ 代替, 经过适当化简, 可得迭代对数矩估计如下

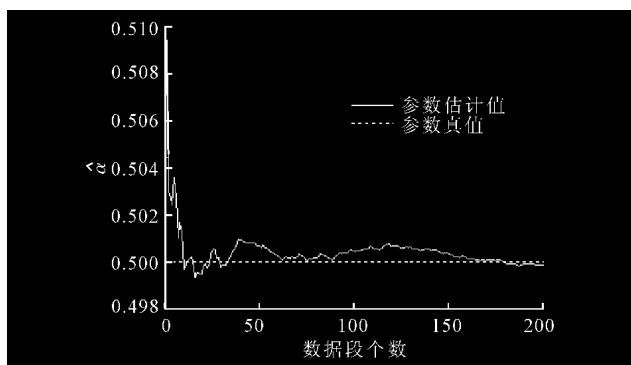
$$\frac{\pi^2[1-\hat{\alpha}^2(k)]}{6\hat{\alpha}^2(k)} = \frac{(k-1)N-1}{kN-1} \frac{\pi^2[1-\hat{\alpha}^2(k-1)]}{6\hat{\alpha}^2(k-1)} + \frac{N-1}{kN-1} V_k + \frac{(k-1)N}{k(kN-1)} \left\{ C_e \left[\frac{1}{\hat{\alpha}(k-1)} - 1 \right] + \frac{\ln \hat{\gamma}(k-1)}{\hat{\alpha}(k-1)} + \frac{\ln(\sec^2(\pi \hat{\alpha}(k-1)/2))}{2\hat{\alpha}(k-1)} - M_k \right\}^2 \quad (11)$$

$$\frac{\ln \hat{\gamma}(k)}{\hat{\alpha}(k)} = \frac{k-1}{k} \left\{ C_e \left[\frac{1}{\hat{\alpha}(k-1)} - 1 \right] + \frac{\ln \hat{\gamma}(k-1)}{\hat{\alpha}(k-1)} + \frac{\ln(\sec^2(\pi \hat{\alpha}(k-1)/2))}{2\hat{\alpha}(k-1)} \right\} + \frac{M_k}{k} - C_e \left[\frac{1}{\hat{\alpha}(k)} - 1 \right] - \frac{\ln(\sec^2(\pi \hat{\alpha}(k)/2))}{2\hat{\alpha}(k)} \quad (12)$$

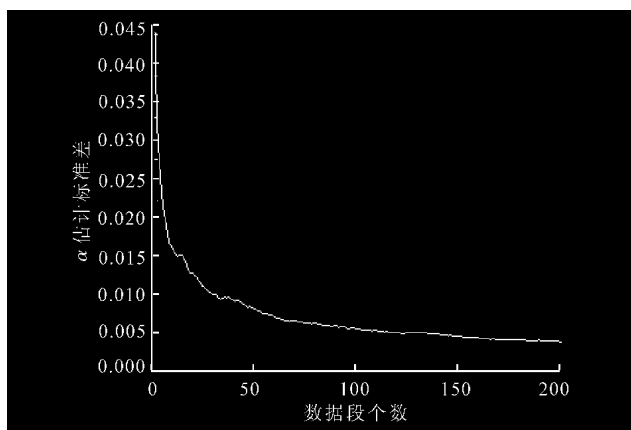
表3给出了迭代对数矩估计的 Monte Carlo 仿真结果, 每组实验均独立运行100次. 由表3可见, 在每次迭代使用较少样本的条件下, 迭代对数矩估计通过数据段的迭代同样可以获得较高的估计精度. 由表2和表3可知, 在样本总数相同的情况下, 迭代对数矩估计和非迭代的对数矩估计的性能接近, 但前者的计算量要小得多. 数据段个数的选择对迭代对数矩估计的性能有着直接影响, 图4给出了估计性能随着数据段个数的变化关系. 可见, 当数据段个数过小时, 参数估计值出现剧烈振荡, 但随着数据段个数的增加, 参数估计值很快就趋于稳定, 并且接近真值. 可以说, 迭代对数矩估计是稳定并且高效的估计方法.

表3 迭代对数矩估计的仿真结果 ($\alpha=0.5, \gamma=1$)

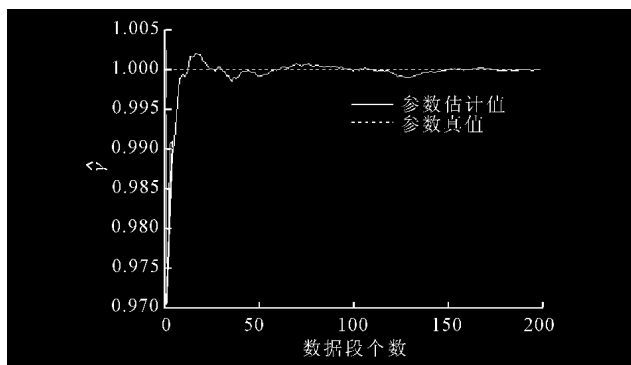
每段 样本数	$\hat{\alpha}$			$\hat{\gamma}$		
	$B=50$	$B=100$	$B=200$	$B=50$	$B=100$	$B=200$
50	0.499 6	0.499 1	0.500 6	0.999 4	1.001 4	1.000 5
	(0.009 1)	(0.007 5)	(0.004 6)	(0.020 0)	(0.013 8)	(0.010 7)
100	0.499 8	0.499 6	0.499 8	1.000 6	1.000 6	1.000 6
	(0.007 2)	(0.005 0)	(0.003 5)	(0.015 3)	(0.011 4)	(0.007 7)
200	0.499 9	0.500 0	0.500 0	0.999 9	0.999 5	1.000 2
	(0.004 1)	(0.003 2)	(0.002 3)	(0.010 3)	(0.007 6)	(0.004 7)



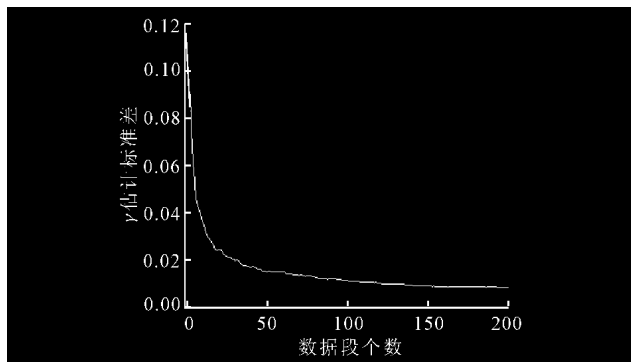
(a)特征指数估计值



(b)特征指数估计标准差



(c)尺度参数估计值



(d)尺度参数估计标准差

图4 迭代对数矩估计的估计性能随数据段个数的变化关系(独立运行100次,每段100个样本, $\alpha=0.5, \gamma=1$)

4 总结

本文把负数阶矩理论应用于PAS分布的参数估计,提出了比值估计、对数矩估计和迭代对数矩估计3种参数估计方法.比值估计利用特定阶次的负数阶矩的比值来估计未知参数,对数矩估计利用PAS分布的对数变换及其负数阶矩的Taylor展开从而获得解析的估计形式,迭代对数矩估计通过样本分段迭代估计未知参数.通过Monte Carlo仿真验证了这3种方法均具有较高的估计精度,同时对数矩估计还具有计算复杂度低的优点.在独立运行次数为100,同时样本总数为5000时,比值估计在负数阶次取为-0.1时可以达到99.8%的估计精度,对数矩估计可以达到99.95%的估计精度,迭代对数矩估计在样本分段数取为50时可以达到99.94%的估计精度.因此,比值估计、对数矩估计和迭代对数矩估计是PAS分布参数估计的有效方法.

参考文献:

- [1] Nikias C L, Shao Min. Signal processing with alpha-stable distribution and application [M]. New York: Wiley, 1995.
- [2] Pierce R D. Application of the positive alpha-stable distribution [C]//Proceedings of the IEEE Signal Processing Workshop on Higher-Order Statistics. Los Alamitos, USA: IEEE Computer Society, 1997: 420-424.
- [3] Kuruoglu E E, Zerubia J. Modeling SAR images with a generalization of Rayleigh distribution [J]. IEEE Transactions on Image Processing, 2004, 13(4): 527-533.
- [4] Szajnowski W J, Wynne J B. Simulation of dependent samples of symmetric alpha-stable clutter [J]. IEEE Transactions on Signal Processing Letters, 2001, 8(5): 151-152.
- [5] Kuruoglu E E. Density parameter estimation of skewed alpha-stable distributions [J]. IEEE Transactions on Signal Processing, 2001, 49(10): 2192-2201.
- [6] Ma X Y, Nikias C L. Parameter estimation and blind channel identification in impulsive signal environments [J]. IEEE Transactions on Signal Processing, 1995, 43(12): 2884-2897.

(编辑 刘杨)