Higher Education
Australian Mathematical Sciences Institute

Maximum Leave-One-Out Likelihood Estimator

Nicholas Katada
Ray Kawai
University of Sydney

Australian Government
Department of Education

1 Introduction

The classical Maximum Likelihood Estimator for a location parameter is limited in that it cannot be applied to unbounded densities. The problem of estimating the location for an unbounded density has been previously approached by Ibragimov and Khasminskii in [3] and [4], using Bayesian estimation to attain a consistent estimator.

An alternative approach is taken in this work, where the problem is addressed through a modification of the estimator by omitting a single term in the likelihood function, yielding the *leave-one-out likelihood estimator*.

It can be shown that under quite reasonable conditions the leave-one-out likelihood estimator converges to the true value of the location parameter. In fact, a bound on the convergence rate has been established, and this work aims to investigate the asymptotic behaviour at this bound in order to determine whether this optimal rate is attained.

2 The Estimator

Suppose $X_1, ..., X_n$ are iid observations distributed by a density $f(x - \delta_0)$ that is differentiable everywhere except δ_0. Assume that

- $f(x) = p(x)|x|^{\alpha}$, for $\alpha \in (-1, 0)$, where p has bounded derivative on $\mathbb{R} \setminus \{0\}$ and, for some $\varepsilon_0 > 0$, is non-zero and continuous on either $[-\varepsilon_0, 0]$, or $[0, \varepsilon_0]$.
- There exists $b > 0$ such that $f(x) = O(|x|^{-b-1})$ when $|x| \to \infty$.
- For some $\varepsilon > 0$, the incomplete Fisher information $F_f(\varepsilon) = \mathbb{E}[f'^2/f^2(X) \mid |X| > \varepsilon]$ is finite.

We now introduce the estimator. First denote

$$k(\delta) = \text{argmin}_{k \in \{1,\ldots,n\}} |X_k - \delta|$$

.

Then define out likelihood function

$$l(\delta) = \prod_{i=1, i \neq k(\delta)}^{n} f_\delta(x_i),$$

and let our estimator $\hat{\delta_n}$ for the location parameter δ_0 be the maximiser of $l(\delta)$.

Theorem 1. *Suppose f satisfies the above requirements. Then for any $\beta \in (0, 1/(1+\alpha))$:*

$$n^\beta(\hat{\delta_n} - \delta_0) \xrightarrow{\mathbb{P}} 0.$$

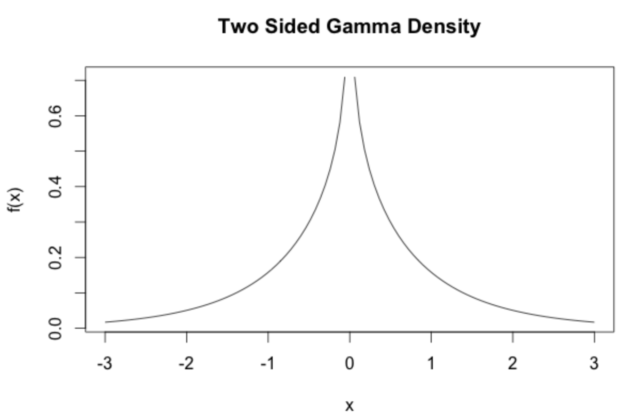
This result is proved in [1]. Here $1/(1+\alpha)$ is the critical value, and what would be the optimal rate if it were attained. Whether this is the case is not clear. To investigate this, we conduct simulations of the estimator.

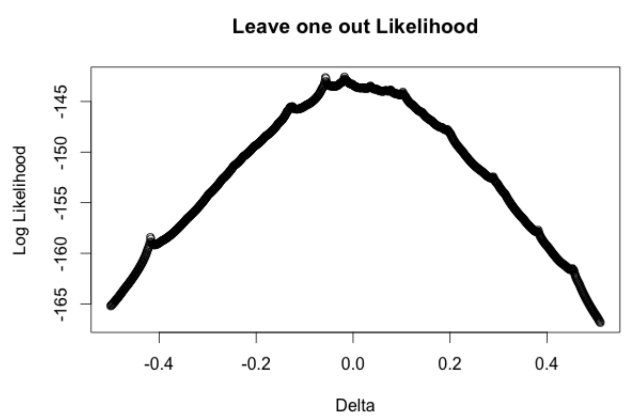
3 Numeric Simulations

The density used for out simulations will be the 'reflected' Gamma distribution, that is, a Gamma random variable that admits positive and negative values, given by density

$$f(x) = \frac{1}{2} \frac{|x-\delta|^{a-1}}{\Gamma(a)} e^{-|x-\delta|}, \ x \in \mathbb{R}$$

Here, the shape parameter is $a \in (0, 1)$, and scale parameter is 1. We can generate a sample from this density and then numerically compute the Leave-one-out likelihood.



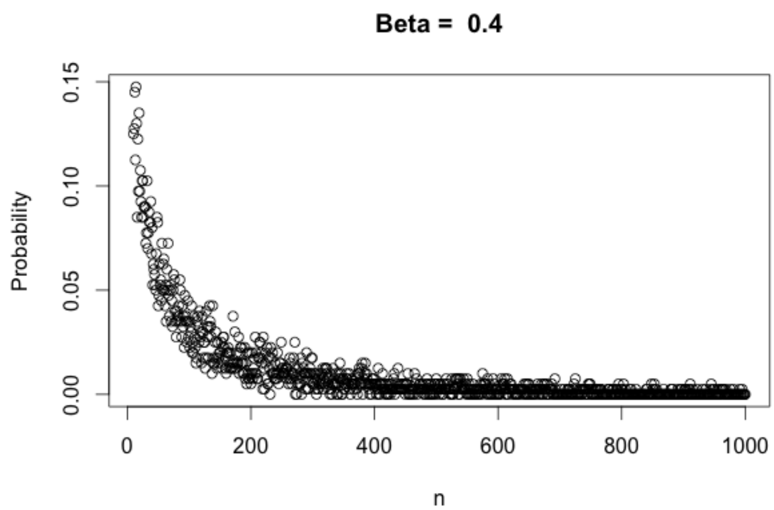


Note the spikes in the likelihood function: these coincide with changes to the observation being excluded.

To look into the statement $n^\beta(\hat{\delta_n}-\delta_0) \xrightarrow{\mathbb{P}} 0$, we are interested in the limiting value of $\mathbb{P}(|\hat{\delta_n}-\delta_0| > n^{-\beta}\varepsilon)$. To this end, we take a fixed value ε, and generate 400 samples of size n. We then estimate $\mathbb{P}(|\hat{\delta_n}-\delta_0| > n^{-\beta}\varepsilon)$ with the empirical probability (proportion). This is repeated for values of n from 10 to 1000.

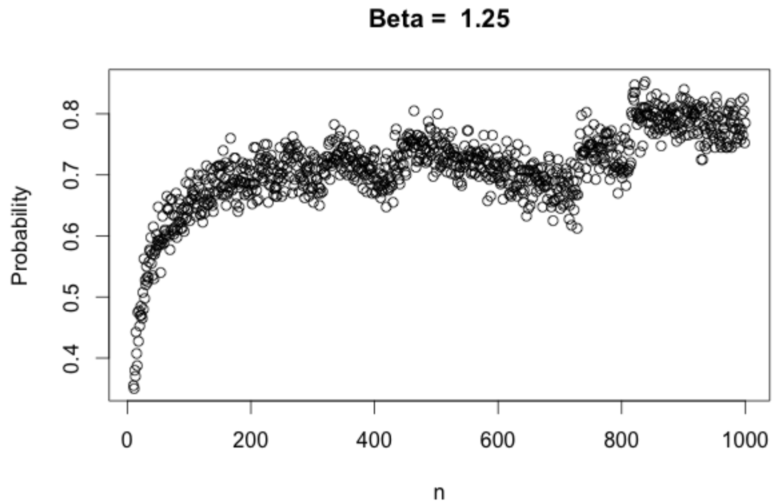
In our samples, we have taken the Gamma shape parameter $a = 0.8$, which means $\alpha = a - 1 = -0.2$, and our critical value is $1/(1+\alpha) = 1.25$

First consider the a case where $\beta = 0.4 < 1/(1+\alpha)$, fixing $\varepsilon = 1$.



Here, the observed empirical probability decisively converges to zero, which is what would be expected with the given result.

Of more interest is the case of the critical point. Here we take $\beta = 1/(1+\alpha)$. It is seen below that even taking a relatively large value for ε, the probability still diverges from zero.



This suggests that the optimal rate of $1/(1+\alpha)$ is not attained in this case.

4 Concluding Remarks

The initial direction for this project was an attempt to prove that the optimal rate of $n^{\frac{1}{1+\alpha}}$ was attained. When this approach did not yield results, we turned to Monte Carlo simulation of the estimator for the 'two sided Gamma' case, and established numerical results that suggest the optimal of convergence rate was not attained.

However, the results of these simulations are far from rigorous; finding a proof of the optimal rate or lack thereof remains an open problem. It is possible a modification of the proof presented in [1] could lead to a proof of this.

Another future direction of this research could be the application to Variance Gamma processes. As the time increments decrease, the distribution of the increments takes the form of an unbounded density, to which the estimator can be applied [2].

References

[1] Podgorski K., Wallin J. (2013) Maximizing leave-one-out likelihood for the location parameter of unbounded densities, *Annals of the Institute of Statistical Mathematics*, in press. doi:10.1007/s10463-013-0437-6.

[2] Kawai, R (2014) On the Likelihood Function of Small Time Variance Gamma Levy Processes, *Statistics*, doi: 10.1080/02331888.2014.918980

[3] Ibragimov, I.A., Khasminskii, R.Z. (1981a) Asymptotic behaviour of statistical estimates of the shift parameter for samples with unbounded density. *Journal of Mathematical Sciences*,16,1035-1041 (translation from Russian, the original date of publication: 1976).

[4] Ibragimov, I.A., Khasminskii, R.Z. (1981b) Statistical Estimation, Asymptotic theory, Applications of Mathematics, vol 16. Springer, New York (translation from Russian, the original date of publication, 1979)