Applied Methods of Statistical Analysis. Statistical Computation and Simulation

Proceedings of the International Workshop

18-20 September 2019

Novosibirsk

2019

Optimal index estimation of log-gamma distribution

Dimitris N. Politis[1], Vyacheslav A. Vasiliev[2] and Sergey E. Vorobeychikov[2]
University of California, San Diego, USA[1],
National Research Tomsk State University, Tomsk, Russia[2]
e-mail: dpolitis@ucsd.edu, vas@mail.tsu.ru, sev@mail.tsu.ru

Abstract

The problem of optimal estimation of the heavy tail index is revisited from the point of view of truncated estimation. A class of these estimators is introduced having guaranteed accuracy based on a sample of fixed size [7]. The optimality of considered log-gamma index estimators in the sense of a special type risk function is established. The considered risk function makes possible to optimize not only the asymptotic variances of the estimators, as well as used for estimation of sample size. Optimization of the parameters of log-gamma distribution is presented. Simulation results confirm theoretical one's.

Keywords: Optimal parameter estimation, heavy tails, log-gamma distribution, optimal convergence rate.

Introduction

This paper presents results of optimality for the parameter estimators of log-gamma distribution, introduced in [7]. Some general properties of parameter estimators are used only and are such that the considered class of estimators is sufficiently wide.

In this paper, we use the risk function of a special type which is a linear combination of mean-square deviation of parameter estimators and sample size. The requirement of both good parameter estimation quality and reasonable duration of observations is formulated as a risk efficiency problem. The risk function of similar structure was proposed in [1], see also references therein. The criterion is given by a certain loss function and optimization is performed based on it.

Further the loss and risk functions of the type proposed in [1] were used in, e.g., [8, 9] for optimization of interpolators and predictors of a scalar AR(1) process with unknown parameters. Similar optimization problem of the sequential parameter estimator of AR(1) was considered in [3]. There was considered a risk function defined on the basis of squared estimation error of sequential estimator of the dynamic parameter.

Later the results of those papers were refined and extended to other stochastic models. In particular, this approach was applied to construction of optimal adaptive predictors of the stochastic processes related with discrete and continuous-time dynamical systems, see, e.g, [16, 2]. The proposed procedures are based on the so-called truncated estimators which have been developed in order to estimate ratio type functionals from a wide class by dependent observations and by samples of fixed size so that they had guaranteed accuracy in the sense of the L_{2m}-norm, $m \geq 1$. Examples of parameter estimation problems of discrete and continuous time systems on a time interval of a fixed length are considered.

The truncated estimators may keep asymptotic properties of the estimators they are based upon. Another approaches do not guarantee prescribed estimation accuracy when using samples of non-random finite size and lead up to complicated analytical problems in adaptive procedures. Applications of truncated estimators with the said quality makes possible to optimize the risk function which is a linear combination of sample mean of mean-square deviation of predictors and sample size.

Results of non-asymptotic non-parametric problems can be found also in [5, 6] among others. In particular, they have investigated non-asymptotic properties of the regression and density function kernel-type estimators.

It should be noted that first truncated parameter estimation method was applied for construction of adaptive optimal predictors of VAR(1) in [10]. Then this method was applied to more complicated stochastic systems. Among the processes considered are stable multivariate discrete time AR(1), ARMA(1,1) and RCA(1), as well as continuous time diffusion and time delayed processes, see, e.g., [2]. The proposed procedure is shown to be asymptotically risk efficient as the cost of prediction error tends to infinity.

1 Log-gamma density function

Consider the parameter estimation problem based on i.i.d. observations $X_1, \ldots, X_n$ with the log-gamma density function

$$f(x) = C_f x^{-(\gamma+1)} \log^{\beta-1} x, \quad x \geq 1.$$

Our main aim is to prove the optimality of the truncated estimators β_n, γ_n and θ_n of the parameters β, γ and θ presented in [7] in the sense of the risk function considered above.

To define the truncated estimators we introduce, similar to [7] for some given $a > 0$ the functional

$$\Phi(a) = E \log^a X_1.$$

Using the definition of $f(x)$ according to [7] we have

$$\Phi(a) = \frac{\gamma}{\beta + a} \Phi(a+1).$$

Analogously for a given $b \neq a$,

$$\Phi(b) = \frac{\gamma}{\beta + b} \Phi(b+1).$$

Thus

$$\beta\Phi(a) - \gamma\Phi(a+1) = -a\Phi(a),$$
$$\beta\Phi(b) - \gamma\Phi(b+1) = -b\Phi(b)$$

and the solution of this system has the form

$$\beta = \frac{b\Phi(b)\Phi(a+1) - a\Phi(a)\Phi(b+1)}{\Delta_{a,b}},$$

$$\gamma = \frac{(b-a)\Phi(a)\Phi(b)}{\Delta_{a,b}},$$

as well as

$$\theta = \frac{\Delta_{a,b}}{(b-a)\Phi(a)\Phi(b)},$$

where

$$\Delta_{a,b} = \Phi(a)\Phi(b+1) - \Phi(b)\Phi(a+1).$$

Now we define the empirical functional estimator

$$\Phi_n(a) = \frac{1}{n}\sum_{k=1}^{n} \log^a X_k$$

of $\Phi(a)$ and the truncated estimators β_n, γ_n and θ_n (see also [7]) as follows

$$\beta_n = \frac{b\Phi_n(b)\Phi_n(a+1) - a\Phi_n(a)\Phi_n(b+1)}{\Delta_{a,b}(n)} \cdot \chi(|\Delta_{a,b}(n)| \geq \log^{-1} n), \quad (1)$$

$$\gamma_n = \frac{(b-a)\Phi_n(a)\Phi_n(b)}{\Delta_{a,b}(n)} \cdot \chi(|\Delta_{a,b}(n)| \geq \log^{-1} n), \quad (2)$$

$$\theta_n = \frac{\Delta_{a,b}(n)}{(b-a)\Phi_n(a)\Phi_n(b)} \cdot \chi(|(b-a)\Phi_n(a)\Phi_n(b)| \geq \log^{-1} n), \quad (3)$$

where

$$\Delta_{a,b}(n) = \Phi_n(a)\Phi_n(b+1) - \Phi_n(b)\Phi_n(a+1).$$

From [7] it follows that the asymptotic normality property, defined in [7] is fulfilled for the estimators γ_n, θ_n and β_n with the rate $\alpha_n = \sqrt{n}$ and the asymptotic variance of $n \cdot \gamma_n$ is defined by equations

$$\sigma_\gamma^2 = (b-a)^2\Delta_{a,b}^{-2} \cdot \sigma_1^2 + 2(b-a)^2\Delta_{a,b}^{-3} \cdot \sigma_2 + (b-a)^2\Delta_{a,b}^{-4} \cdot \sigma_3^2, \quad (4)$$

where

$$\sigma_1^2 = \Phi^2(a)\Phi(2b) + \Phi(2a)\Phi^2(b) + 2\Phi(a)\Phi(b)\Phi(a+b) - 4\Phi^2(a)\Phi^2(b),$$

$$\sigma_2 = -\Phi(a)\Phi(b+1)\Phi(a+b) - \Phi^2(a)\Phi(2b+1) + \Phi(a)\Phi(a+1)\Phi(2b) - \Phi(2a)\Phi(b)\Phi(b+1)$$
$$+\Phi(a+1)\Phi(b)\Phi(a+b) + \Phi^2(b)\Phi(2a+1) + 4\Phi^2(a)\Phi(b)\Phi(b+1) - 4\Phi(a)\Phi(a+1)\Phi^2(b),$$

$$\sigma_3^2 = \Phi^2(a)\Phi^2(b) \cdot \{\Phi(2a)\Phi^2(b+1) - 4\Phi^2(a)\Phi^2(b+1) + \Phi^2(a)\Phi(2(b+1))$$
$$+2\Phi(a)\Phi(b+1)\Phi(a+b+1) + \Phi^2(a+1)\Phi(2b) + \Phi^2(b)\Phi(2(a+1)) - 4\Phi^2(a+1)\Phi^2(b)$$
$$+2\Phi(a+1)\Phi(b)\Phi(a+b+1) - 2\Phi(a+1)\Phi(b+1)\Phi(a+b) - 2\Phi(a)\Phi(a+1)\Phi(2b+1)$$
$$+8\Phi(a)\Phi(b)\Phi(a+1)\Phi(b+1) - 2\Phi(2a+1)\Phi(b)\Phi(b+1) - 2\Phi(a)\Phi(b)\Phi(a+b+2)\}.$$

Consider the case of known β. The parameter Gamma can be represented in the form

$$\gamma = (\beta + a)\frac{\Phi(a)}{\Phi(a+1)},$$

the estimator is defined as

$$\gamma_n = (\beta + a)\frac{\Phi_n(a)}{\Phi_n(a+1)} \cdot \chi(\Phi_n(a+1) \geq \log^{-1} n)$$

and its asymptotic variance is equal to

$$\sigma_\gamma^2 = (\beta + a)^2 \frac{\Phi(2a)\Phi(a+1) + \Phi^2(a)\Phi(a+1)\Phi(2(a+1)) - 2\Phi(a)\Phi(2a+1)}{\Phi^3(a+1)} \qquad (5)$$

Consider the optimization procedure of the parameter estimation of log-gamma distribution.

Define for an estimator γ_n of parameter γ the loss function

$$L_n = A(\gamma_n - \gamma)^2 + n.$$

Parameter A stands for a cost of mean square quality of the estimator γ_n of parameter γ and n is a sample size. We suppose that the cost of observations is included in the definition A (see, for comparison, [1]).

The corresponding risk function $R_n = EL_n$ has the form

$$R_n = AE(\gamma_n - \gamma)^2 + n$$

and we solve the optimization problem

$$R_n \to \min_n \qquad (6)$$

Consider two cases.

– Case of known asymptotic variance σ^2 of γ_n.

Thus the principal term of the risk function has the form

$$R_n = \frac{A\sigma^2}{n} + n.$$

For A large enough the optimal sample size is equal to

$$n_A^0 = \sqrt{A\sigma^2}, \qquad (7)$$

as well as the corresponding principal term of the risk function $R_{n_A^0}$

$$R_A^0 := \frac{A\sigma^2}{n_A^0} + n_A^0 = 2\sqrt{A\sigma^2}. \qquad (8)$$

As follows the problem is solved if the number σ^2 is known.

– Case of unknown σ^2.

First define the estimator σ_n^2 of the variance σ^2 as

$$\sigma_n^2 = \frac{\Phi_n(2a)\Phi_n(a+1) + \Phi_n^2(a)\Phi^n(a+1)\Phi_n(2(a+1)) - 2\Phi_n(a)\Phi_n(2a+1)}{\Phi_n^3(a+1)} \quad (9)$$

$$\cdot(b+a)^2\chi(\Phi_n(a+1) \geq \log^{-1} n).$$

Since (7) is directly involved in the expression (8) for R_A^0, the optimal sample size cannot be obtained as before. Similarly to Konev and Lai (1995), Sriram (1988), Sriram and Iaci (2014) and Kusainov and Vasiliev (2014), one uses the stopping time N_A as an estimator of n_A^0 replacing σ^2 in its definition with the estimator σ_n^2

$$N_A = \inf\{n \geq n_A : \ n \geq A^{1/2}\overline{\sigma}_A\}, \quad (10)$$

where $\overline{\sigma}_A = \min\{\sigma_{n_A}, \log A\}$, $\sigma_n = \sqrt{\sigma_n^2}$. We use here in comparison with mentioned above papers the estimator $\overline{\sigma}_A$ instead of σ_n to simplify the proofs. At the same time all results remain true.

It should be noted that for A large enough the following property is fulfilled

$$E(\overline{\sigma}_A^2 - \sigma^2)^{2p} \leq 2r_{n_A}(p), \quad (11)$$

where $r_n(p)$ is some deterministic sequence such that

$$A \cdot r_{n_A}(p) = o(1) \quad \text{as} \quad A \to \infty.$$

Indeed, for, e.g., $\log^2 A - \sigma^2 \geq 1$, using the Chebyshev inequality we have

$$E(\overline{\sigma}_A^2 - \sigma^2)^{2p} = E(\sigma_{n_A}^2 - \sigma^2)^{2p}\chi(\sigma_{n_A} \leq \log A) + (\log^2 A - \sigma^2)P(\sigma_{n_A} > \log A)$$

$$\leq r_{n_A}(p) + (\log^2 A - \sigma^2)\frac{E(\sigma_{n_A}^2 - \sigma^2)^{2p}}{(\log^2 A - \sigma^2)^{2p}} \leq 2r_{n_A}(p).$$

We prove the asymptotic equivalence of N_A and n_A^o in the almost surely and mean senses (see Theorem 1 below) and the optimality of the adaptive estimation procedure in the sense of equivalence of the obviously modified risk

$$\overline{R}_A := EL_{N_A} = AE(\gamma_{N_A} - \gamma)^2 + EN_A. \quad (12)$$

Theorem 1. *The observation numbers (10) and (7) and corresponding risk functions (12) and (8) are asymptotically equivalent in the following sense: as* $A \to \infty$

$$\frac{N_A}{n_A^o} \to 1 \quad a.s., \quad (13)$$

$$\frac{EN_A}{n_A^o} \to 1, \quad (14)$$

$$\frac{\overline{R}_A}{R_A^o} \to 1. \quad (15)$$

2 Simulation results

To illustrate the theoretical properties of the optimal adaptive procedure we give some numerical results for log-gamma distribution. We obtained the estimators σ_n^2 of the variance of parameter estimators. The results for different values of n are presented in Fig. 1. The horizontal line shows the asymptotic value of σ^2.

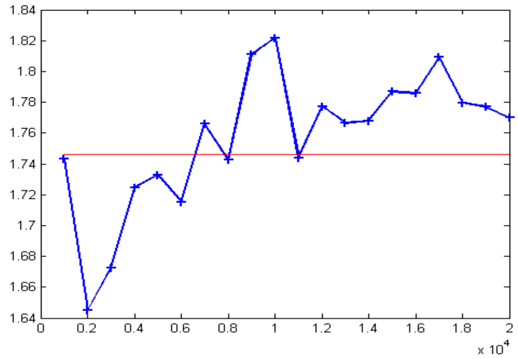

Figure 1: Log-gamma distribution. $\gamma = 1.666$

The quantities C_N and C_R are given in Fig. 2, 3 where $C_N = \frac{EN_A}{n_A^o}$, $C_R = \frac{\overline{R}_A}{R_A^o}$. Here n_A^0, R_A^0 are defined by (7, 8) and $N_A, \overline{R}_A$ – by (10, 12). Note that EN_A and $\overline{R}_A$ were computed as an empirical average over 1000 Monte Carlo replications of the experiment (for each value of A).

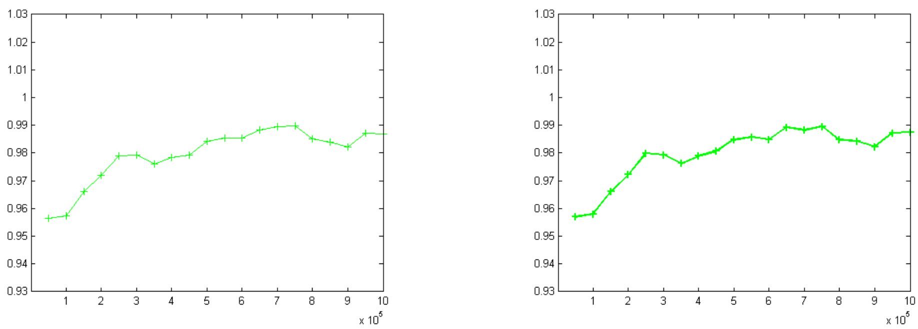

Figure 2: Log-gamma distribution. $\gamma = 1.6666$ C_N - left, C_R - right

The obtained numerical results are close to the theoretical properties of the proposed adaptive procedure.

Conclusion

The paper presents the method of optimal parameter estimation of log-gamma distribution. The truncated estimator is used to minimize the loss function which includes the weighed mean square deviation and the sample size. It is shown that the proposed procedure is asymptotically efficient.

The theoretical results are illustrated by numerical results which confirm the optimality properties.

Acknowledgements

This study was supported by The Ministry of Education and Science of the Russian Federation, Goszadanie No 2.3208.2017/4.6

References

[1] Chernoff H. (1972). *Sequential Analysis and Optimal Design*. SIAM, Philadelphia.

[2] Dogadova T.V., Kusainov M.I., Vasiliev V.A. (2017). Truncated estimation method and applications. *Serdica. Mathematical Journal Bulgarian Academy of Sciences Institute of Mathematics and Informatics*. Vol. **43**, pp. 221-266.

[3] Konev V.V., Lai T L. (1995). Estimators with Prescribed Precision in Stochastic Regression Models. sl Sequential Analysis. Vol. **14**, pp. 179-192.

[4] Kusainov M.I., Vasiliev V.A. (2015). On optimal adaptive prediction of multivariate autoregression. *Sequential Analysis*. Vol. **34**, pp. 211-234.

[5] Politis D.N., Vasiliev V.A. (2012). Sequential kernel estimation of a multivariate regression function. *Proceedings of the IX International Conference 'System Identification and Control Problems, SICPRO'12, Moscow, 30 January – 2 February, 2012, V. A. Trapeznikov Institute of Control Sciences*. pp. 996-1009.

[6] Politis D.N., Vasiliev V.A. (2013). Non-parametric sequential estimation of a regression function based on dependent observations. *Sequential Analysis: Design Methods and Applications*. Vol. **32**, pp. 243–266. http://www.tandfonline.com/doi/full/10.1080/07474946.2013.803398. DOI:10.1080/07474946.2013.803398.

[7] Politis D.N., Vasiliev V.A., Vorobeychikov S. E. (2018). Truncated Estimation of Ratio Statistics with Application to Heavy Tail Distributions. *Math. Methods Statist*. Vol. **27**, pp. 226-243.

[8] Sriram T.N. (1988). Sequential Estimation of the Autoregressive Parameter in a First Order Autoregressive Process. *Sequential Analysis*. Vol. **7**, pp. 53-74.

[9] Sriram T.N., Ross , Iaci . (2014). Sequential Estimation for Time Series Models. *Sequential Analysis*. Vol. **33**, pp. 136-157.

[10] Vasiliev V. (2014) A truncated estimation method with guaranteed accuracy. *Annals of the Institute of Statistical Mathematics*. Vol. **66**. pp. 141–163.