Cumhuriyet Üniversitesi Fen Fakültesi
Fen Bilimleri Dergisi (CFD), Cilt:36, No: 3 Özel Sayı (2015)
ISSN: 1300-1949

Cumhuriyet University Faculty of Science
Science Journal (CSJ), Vol. 36, No: 3 Special Issue (2015)
ISSN: 1300-1949

Estimation based on minimum distance between empirical and theoretical distribution functions

Marjan JALILIAN[1*], Maryam SHARIFDOUST[1]

[1] *Department of Statistics & Mathematics, Islamic Azad University, Khomeinishar Branch, Isfahan, Iran*

Received: 01.02.2015; Accepted: 05.05.2015

Abstract. Distribution function, other than goodness-of-fit test, is also used in point estimation approach, especially in distributions with closed-form distribution functions. In this paper, the goal is to estimate parameters with minimum distance between empirical and theoretical distribution functions by means of Hellinger and Jeffrey distance measures. Monte Carlo simulation for estimating parameters of generalized Pareto distribution with three parameters represents acceptable results. This distribution does not have any close form moment for obtaining MME and also the estimation of parameters by MLE needs some numerical ways. Finally, the introduced methods is implemented to real data and compared to classical estimators.

Keywords: Method of Moment, Maximum Likelihood, Hellinger Distance, Jeffrey Distance, Monte Carlo

1. INTRODUCTION

There are various methods to obtain point estimators, such as method of moments estimation (MME), maximum likelihood estimation (MLE), least square estimation, bayesian method, minimum chi square and minimum distance. To choose the best estimator, there are several criteria that retaining the domain of unknown parameter, unbiaseness, efficiency and consistency are of the most important.

Using classic estimation methods such as MME and MLE requires obtaining preliminary moments and maximizing the likelihood function, respectively, and weakness of these methods shows up with appearance of multi parametric distributions, distributions without closed-form moments, or which requires numerical solutions for attaining estimators. The estimation of parameters by minimum distance between theoretical and estimation of parameters by transformation, which is mostly used in testing hypothesis, was first officially used by Beran (1977) in point estimation. He used the difference between theoretical and empirical density function and showed that the point estimators obtained by Hellinger distance act robust on the presence of outliers, which usually results unreasonable estimates. The statistic used in these tests is called distance measure, which must get minimized. Recently, in addition to distance between theoretical and empirical distribution functions, distance between other distribution characteristics has been of interest. Sim and Ong (2010) considered some estimators based on the generalization of Hellinger and some other distances between probability generating function (pgf) and the empirical probabiliey generating function (epgf) for negative binomial (NB) distribution that are robust estimators. Also Sharifdoust et al. (2014) studied the behavior of pgf based Hellinger and Jeffrey distances for NB and some distribution in presence and without presence of outliers and compared the results with other estimators including Hellinger MME and MLE [3, 4].

In this paper, we are going to use this method for some distributions with closed-form distribution function in which MLE does not have closed-form or classical estimators do not show

*Corresponding author. *Email address: Marjan.jalilian@iauksh.ac.ir*

Special Issue: The Second National Conference on Applied Research in Science and Technology

good properties. These distributions do not have any close form moment for obtaining MME and also the estimation of parameters by MLE needs some numerical ways. To discuss optimal conditions of the estimator, generalized Pareto distribution is used in simulation. Therefore, we try to estimate the parameters of generalized Pareto distribution by minimizing the distance between theoretical and empirical distribution functions. In other words, estimator, which results the minimum distance between distribution of data and theoretical distribution, will be chosen. Among the available statistical distances, we consider Hellinger and Jeffrey robust distances [2, 3]. To compare suggested estimators, we use biasness and mean squared error.

So, for this purpose, in the second section, we introduce the suggested distances between two functions, that are named Jeffery and Hellinger distances. In the third section, we represent the estimating method and in the fourth section, to evaluate the proposed estimating method and to compare it with classic estimators, we apply Monte Carlo simulation to three-parameter generalized Pareto distribution. In the fifth section, we fit the three-parameter generalized Pareto distribution to real data and estimate parameters. And finally in the sixth section, a brief review of the results will be discussed.

2. DISTANCE BETWEEN FUNCTIONS

Mathematical distances are quantities used for calculating between two mathematical objects. These distances are usually meter and not necessarily symmetric [4]. In probability theory, f-divergent is a function that measure the difference between two probability distributions. This divergence was first introduced and studied by Morimoto (1963). Hellinger's distance was first introduced by Hellinger (1909) to determine the similarity between two probability distributions. This distance is a special form of f-divergent in the condition $\mathrm{f} = \left(\sqrt{\mathrm{t}} - 1\right)^2$. If P and Q are probability measures which are strictly continuous with respect to probability measure λ, the second exponent of Hellinger distance between P and Q will have the form

$$H^2(P,Q) = \frac{1}{2}\int \left(\sqrt{\frac{dP}{d\lambda}} - \sqrt{\frac{dQ}{d\lambda}}\right)^2 d\lambda \tag{1}$$

where $\frac{dP}{d\lambda}$ and $\frac{dQ}{d\lambda}$ are Radon-Nikodin derives of P and Q, respectively. Since this definition does not depend on λ, thus Hellinger distance between P and Q will not be changed when λ is replaced by another probability measure in which P and Q are also strictly continuous. So the distance can be represented in following form:

$$H^2(P,Q) = \frac{1}{2}\int \left(\sqrt{dP} - \sqrt{dQ}\right)^2 \tag{2}$$

The Kulback-Leibler distance is an asymmetric measure of difference between two probability distribution P and Q and determines a measure of missing information when is used for approximating. In fact, P shows the real distribution of data or observations and Q is an approximation of P. This distance, which is a special case of f-divergent at the condition $\mathrm{f} = tlnt$, has been of interest in many researches of engineering and reliability theory such as Moreno et al. (2004) and Seghouane and Amari (2007) [7, 8]. Let P and Q be probability measures on set Ω and P be strictly continuous with respect to Q. Then Kullback-Leibler distance will have the following form:

$$D_{KL}(P\|Q) = \int_{\Omega} ln\left(\frac{dP}{dQ}\right) dP \tag{3}$$

where $\frac{dP}{dQ}$ is the Radon-Nikodin derive of P according to Q. The big problem in using Kullbak-Leibler distance is its asymmetric form. Jeffrey distance is a symmetric version of Kullback-Leibler distance; If P and Q are discrete distribution functions, this distance will be determined as below [9]:

$$D_j(P,Q) = D_{KL}(P\|Q) + D_{KL}(Q\|P) = \sum_i (p(i) - q(i)) ln \frac{p(i)}{q(i)} \tag{4}$$

The Jeffrey distance, which is not determined in p(i)=0 and q(i)=0, has many usages like classification of multimedia data [7].

3. ESTIMATING WITH METHOD OF MINIMUM DISTANCE BETWEEN DISTRIBUTION AND SAMPLE CHARACTERISTICS

In many distributions, probability distribution function has a simpler form than other distribution characteristics like probability generating function or density function. Therefore, we can use distribution function instead of other characteristics. In this paper we use Hellinger and Jeffrey distances based on distribution function, which are as followed:

$$\mathrm{H}\big(\mathrm{F_X(x)}, \mathrm{F_n(x)}\big) = \int \left(\sqrt{\mathrm{F_X(x)}} - \sqrt{\mathrm{d_n(x)}}\right)^2 \mathrm{dx} \tag{5}$$

$$\mathrm{J}\big(\mathrm{F_X(x)}, \mathrm{F_n(x)}\big) = \int (\mathrm{F_X(x)} - \mathrm{F_n(x)}) \log \frac{\mathrm{F_X(x)}}{\mathrm{F_n(x)}} \mathrm{dx} \tag{6}$$

where $\mathrm{F_X(x)}$ and $\mathrm{F_n(x)}$ are empirical distribution functions. The empirical distribution function of n observation $\mathrm{X_1, X_2, \ldots, X_n}$ will be obtained from equation below:

$$\mathrm{F_n(t)} = \frac{\#\, of\ members\ \le \mathrm{t}}{\mathrm{n}} = \frac{1}{\mathrm{n}} \sum_{\mathrm{i=1}}^{\mathrm{n}} \mathrm{I\{x_i \le t\}} \tag{7}$$

where *I(A)* is the indicator function of event A. Estimators obtained by minimizing equations (5) and (6) are called MHDE and MJDE, respectively.

To evaluate the goodness of the method of estimation, we use this strategy that as smaller the quantity of the distances, the estimation of the parameters are more suitable with this data.. In fact, this appropriateness shows that model parameters are well-estimated. Thus we can say that parameter values, which distances in equations (5) and (6) are minimized, are best estimates for parameters.

Since the classic MLE sometimes need to be solved numerically and is not even unique or domain-retaining in many cases, are also MME does not have acceptable results in small samples, and many distributions does not have moment generating functions with closed form, using the distance between distribution characteristics or even distribution function itself can be so helpful.

4. SIMULATION

In this section, the aim is to discuss and compare the estimating methods of minimum distance between theoretical and experimental distribution functions and classic estimators; MLE and MME. For this purpose, we use generalized Pareto distribution with three parameters (GPD3). As we know, this distribution is a schema for many social-economical phenomena. This distribution, which is compatible with skew phenomena, is very important in studying life time of creatures and reliability theory [11]. The GPD3 contains a big family of heavy-tailed distributions, exponential distribution family and a subclass of Beta distributions and distributions with bounded support. Distribution function and density function of this distribution is represented in equations (8) and (9), respectively.

$$F_X(x) = \begin{cases} 1 - \left(1 - \frac{a(x-c)}{b}\right)^{\frac{1}{a}} & ; \quad a \neq 0 \\ 1 - exp\left\{-\frac{x-c}{b}\right\} & ; \quad a = 0 \end{cases} \tag{8}$$

$$f_X(x) = \begin{cases} \frac{1}{b}\left[1 - \frac{a(x-c)}{b}\right]^{\frac{1}{a}-1} & ; \quad a \neq 0 \\ \frac{1}{b} exp\left\{-\frac{x-c}{b}\right\} & ; \quad a = 0 \end{cases} \tag{9}$$

where a is shape parameter, b scale parameter, and c is location parameter. Afterward, for different sample sizes 10, 20, 30, 50, 100, 200, and 500 from GPD3 distribution, simulation is done, parameters are estimated with by the classic ML and MM and the proposed methods MHD and MJD for each sample size, and biasness and MSE are obtained and calculated in Table 1.

In cases, which complex integrals cannot be solved with common integrating methods, we use numerical methods to approximate the integrals.

As it can be seen on Table 1, bias value of MME decreases as sample size increases. Although this decrement is very negligible, it holds for all three parameters of distribution. MSE value is also invariant as sample size increases.

On the other hand, we see that MLE also decreases for different sample sizes and has an invert relevance with sample size. But for the threshold variable, bias value tends to zero as sample size get bigger. MSE is almost invariant for shape and scale parameters and tends to zero for threshold variable. Proposed estimator act more suitable that classic ones since bias and MSE values are smaller that values of classic methods. Also two methods MHDE and MJDE are very similar, their values of bias and MSE for all parameters are the same and the only difference is MHDE needs more time for calculation than MJDE. Thus, as a result, MHDE and MJDE are at least as good as MLE and MME.

Table1. Parameter BIAS and MSE under MME, MLE, MJDE and MHDE for simulation of different sizes fromGPD3.

MJDE			MHDE			MLE			MME				
c	b	a	c	b	a	c	b	a	c	b	a		
0.03349968	0.1076689	0.1068301	0.03751284	0.1260546	0.1332235	0/09014941	0/128702	0/06310695	-0/00834771	0/1005578	0/100008	BIAS	n=10
0.00486586	0.4293261	0.6110634	0.006065481	0.5072224	0.681196	0/01545574	0/02723238	/0051166730	0/01256924	0/4003866	0/5864917	MSE	
0.06510454	0.0777186	0.2017821	0.08820269	0.1017605	0.04852102	0/04852102	-0/0659769	0/03691643	-0/00340995	0/05783779	0/06640465	BIAS	n=20
0.001354892	0.2177829	0.3705034	0.001686035	0.2917875	0.4252346	/0043965310	/0060196020	/0019159650	0/005586149	0/1913062	0/3257667	MSE	
0.01225811	0.06543971	0.06603979	0.01441773	0.09132504	0.08269046	0/03397863	-0/03935429	0/02590559	0/000228846-	0/04674474	0/05620145	BIAS	n=30
0.000715612	0.1504346	0.2217627	0.00081602	0.2274305	0.2798173	/0021975750	/0025705930	/0009838280	0/003470717	0/1047413	0/1845918	MSE	
0.006837838	0.05605263	0.06120184	0.009295387	0.05976424	0.08035477	0/02024252	-0/02225964	0/01581443	0/000376698	0/03391863	0/04131192	BIAS	n=50
0.000231432	0.1188617	0.1457336	0.000348476	0.1843261	0.206875	/0007763150	/0008138250	/0003770090	0/002058114	0/05615484	0/09954428	MSE	
0.003631739	0.03413785	0.02523923	0.005087456	0.06002055	0.04360259	0/01057447	-0/0108733	/0081538320	-0/00356948	0/1078131	0/01267217	BIAS	n= 100
7.36E –0.5	0.08388194	0.09494305	0.000104675	0.1520189	0.1624977	/0002201680	/0001974550	/0001055840	0/001047012	0/02445725	0/04519335	MSE	
0.001635924	0.02404	0.01211024	0.002583457	0.05249535	0.03681133	0.005314222	-0.005097321	0.003929863	-0.000198754	0.00490901	0.005089851	BIAS	n= 200
1.71E – 0.5	0.0696732	0.06808303	2.62E – 0.5	0.1439366	0.1290109	5.85E –0.5	4.63E −0.5	2.65E −0.5	0.000463979	0.0117486	0.02108669	MSE	
0.000554769	0.02813257	0.01145132	0.001041942	0.05931511	0.02565333	0.00197699	−0.001836863	0.001450578	-0.000264417	0.000183656	0.00019510-9	BIAS	n= 500
2/ 11E-06	0.07828963	0.05394872	4.37E – 0.6	0.1545947	0.1184875	7.52E −0.6	5.66E −0.6	3.49E −0.6	0.000192197	0.00478933	0.008813122	MSE	

*Corresponding author. *Email address: Marjan.jalilian@iauksh.ac.ir*

Special Issue: The Second National Conference on Applied Research in Science and Technology

5. RESULTS OF REAL DATA

In this section, we consider the river flooding data, fit the GP3 distribution to them, and estimate the parameters with classic and proposed methods and compare results.

River flooding data have been used by Dumonceaux and Antel (1973). This data, which is mostly about maximum level of river flooding in Pennsylvania state, has been recorded in a four-year period from 1890 to 1969 with million cube feet per minute [13]. The data is listed in Table 2.

Table 2. The river flooding data

0.6540	0.6130	0.3150	0.4490	0.2970
0.4020	0.3790	0.4230	0.3790	0.3235
0.2690	0.740	0.4180	0.4120	0.4940
0.4160	0.3380	0.3920	0.4840	0.2650

This data is skew to right with skew coefficient 1.156, thus the GP3 distribution can be a good candidate for fitting. Now we estimate parameters of model with classic and proposed methods. Estimated values of river flooding data by means of different methods is represented in Table 3.

Table 3. The different estimation of parameters for river flooding data

MJDE	MHDE	MLE	MME	
8.891724×10^{5}	8.211710×10^{5}	2.42913×10^{-1}	0.24291381	a
8.687275×10^{4}	1.835673×10^{5}	9.051908×10^{5}	0.18977162	b
3.6869×10^{-2}	1.0×10^{-6}	0.265	-0.01544028	c

6. CONCLUSION

In this paper, we use Hellinger and Jeffrey distances between theoretical and empirical distribution functions to obtain the estimates of parameters that classic methods cannot have a closed-form distribution function. To evaluate the functionality of proposed methods, we use generalized Pareto distribution with three parameters. Simulation results shows suitable function of proposed methods in comparison to classic ones. Finally, proposed and classic methods were applied for a skew to right set of data.

REFRENCES

[1] Beran, R. (1977), Minimum Hellinger distance estimates for parametric models, Annals of Statistics, 5, 445–463.

[2] Sim, S. Z. and Ong, S. H. (2010), Parameter Estimation for Discrete Distributions by Generalized Hellinger-Type Divergence Based on Probability Generating Function, *Communications in Statistics - Simulation and Computation*, 39, 305– 314.

[3] Sharifdoust, M, Chung, M. N, and Ong, S. H. (2014), Probability generating function based Jeffrey's divergence for statistical inference, Accepted in *Communications in Statistics -Simulation and Computation".*

[4] Dodge, Y. and Institute, I. S. (2003), The Oxford Dictionary of Statistical Terms, Oxford University Press.

[5] Morimoto, T. (1963), Markov processes and the H-theorem, Journal of the Physical Society of Japan, 18, 328-331.

*Corresponding author. *Email address: Marjan.jalilian@iauksh.ac.ir*

Special Issue: The Second National Conference on Applied Research in Science and Technology

[6] Hellinger, E. (1909), NeueBegründung der TheoriequadratischerFormen von unendlichvielenVeränderlichen,, Journal für die reine und angewandteMathematik (in German), 136, 210-271.

[7] Moreno, P., Ho, P., and Vasconcelos, N. (2004), A Kullback-Leibler divergence based kernel for SVM classification in multimedia applications, Advances in Neural Information Processing Systems Vambridge: MIT Press, 16, 1385-1392.

[8] Seghouane, A. and Amari, S. (2007), The AIC criterion and symmetrizing the Kullback-Leibler divergence. IEEE Transactions on Neural Networks, 18, 97-106.

[9] Jeffreys, H. (1946), An invariant form for the prior probability in estimation problems, Proceeding of the Royal Society A: Mathematical, physics science, 186, 453-461.

[10] Michel, M. D. and Elena, D. (2013), Encyclopedia of Distances, Springer.

[11] Norman, L. J. and Kotz, S. (1970), Distribution in statistics continuous univariate, Honghton Mifflin Company.

[12] Dumonceaux, R. and Antle, C. E. (1973), Discrimination between the lognormal and the Weibull distributions, Technometrics, 15, 923-926.