Open Journal of Statistics, **2012, 2, 415-419**
http://dx.doi.org/10.4236/ojs.2012.24050 Published Online October 2012 (http://www.SciRP.org/journal/ojs)

Maximum Entropy and Maximum Likelihood Estimation for the Three-Parameter Kappa Distribution

Bungon Kumphon
Department of Mathematics, Faculty of Sciences, Mahasarakham University, Mahasarakham, Thailand
Email: bungon.k@msu.ac.th

Received June 27, 2012; revised July 29, 2012; accepted August 11, 2012

ABSTRACT

The two statistical principles of maximum entropy and maximum likelihood are investigated for the three-parameter kappa distribution. These two methods become equivalent in the discrete case with $x,\beta>0$ where $0<\alpha=1/(2k+1)\leq 1, k=0,1,2,\cdots$, for the maximum entropy method.

Keywords: Maximum Entropy; Maximum Likelihood; Kappa Distribution; Lagrange Multiplier

1. Introduction

Statistical entropy deals with a measure of uncertainty or disorder associated with a probability distribution. The principle of maximum entropy (ME) is a tool for inference under uncertainty [1,2]. This approach produces the most suitable probability distribution given the available information as seeks the probability distribution that maximizes the information entropy subject to the information constraints, typically via the method of Lagrange multipliers. More precisely, the result is a probability distribution that is consistent with the known constraints expressed in terms of averages or expected values of one or more quantities, but is otherwise as unbiased as possible—*i.e*. one obtains the least-biased estimate possible on the given information, maximally noncommittal with regard to missing information.

A family of positively skewed distributions known as kappa distributions introduced by Mielke [3] and Mielke and Johnson [4], is very popular for analyzing precipitation data (cf. Park *et al*. [5], Kysely and Picek [6], Dupuis and Winchester [7]). Various methods of estimation for this type of data include the L-moment, Moment, and Maximum Likelihood (ML) techniques. Many research papers have shown that the ML is too sensitive to extreme values, especially for small samples although but it may be satisfactory for large samples, and the final estimate is not always a global maximum because it can depend upon the starting values. The ME can remove this ambiguity, as various authors have shown—e.g. Hradil and Rehacek [8], and Papalexious and Koutsoyiannis [9]. Singh and Deng [10] considered the ME method for the four-parameter kappa distributions, which include the three-parameter kappa distribution (K3D) introduced by Meilke [3]. In this study, we investigate the theoretical background for parameter estimation by the ME method in the K3D case. The limitation of its performance compared to the ML method is also discussed.

2. Three-Parameter Kappa Distribution

Let a random variable be denoted by *X*. The distribution function of the three-parameter kappa distribution (K3D) is

$$f(x)=\frac{\alpha}{\beta}\left[\alpha+\left(\frac{x-\mu}{\beta}\right)^{\alpha}\right]^{-\left(\frac{\alpha+1}{\alpha}\right)}\quad x,\alpha,\beta>0 \tag{1}$$

where μ,β,α denote the location, scale and shape parameters respectively (Park *et al*. [5]), and the corresponding cumulative distribution function of the K3D is

$$F(x)=\left(\frac{x-\mu}{\beta}\right)\left[\alpha+\left(\frac{x-\mu}{\beta}\right)^{\alpha}\right]^{-\frac{1}{\alpha}},\quad x>0. \tag{2}$$

It is notable that the K3D distribution function, Equation (1), involves adding the location parameter μ to the two-parameter kappa distribution (K2D), in contrast to Meilke [3] where only a new shape parameter α is introduced.

3. The Entropy Framework

3.1. Entropy Measure and the Principle of Maximum Entropy

The concept of entropy was originally developed by

Ludwig Boltzmann in statistical mechanics. A famous and well justified measure is the Boltzmann-Gibbs-Shannon (BGS) entropy

$$S = -\int_0^\infty f(x) \ln f(x) \mathrm{d}x \tag{3}$$

for a continuous non-negative random variable *X*, where $f(x)$ is the probability density function of *X*. The given information used in the principle of maximum entropy (ME) is expressed as a set of constraints representing expectations of functions $g_j(X)$—*i.e.*

$$E\left[g_j(X)\right] = \int_0^\infty g_j(x) f(x) \mathrm{d}x = c_j,\ j = 1, 2, \cdots, n \tag{4}$$

ME distributions emerge by maximizing the selected form of entropy, subject to Equation (4) and the obvious additional constraint

$$\int_0^\infty f(x) \mathrm{d}x = 1. \tag{5}$$

As precisely mentioned, the maximization is usually accomplished via the method of Lagrange multipliers, such that the general solution form of the ME distributions from maximizing the BGS entropy Equation (3) (Levine and Tribus, [11]) is

$$f(x) = \exp\left[-\lambda_0 - \sum_{j=1}^{n} \lambda_j g_j(x)\right], \tag{6}$$

where $\lambda_j, j = 1, 2, \cdots, n$, are the Lagrange multipliers linked to the constraints in Equation (4) and λ_0 is the multiplier linked to the additional constraint Equation (5).

3.2. Justification of the Constraints

Samples are drawn from positively skew or heavy-tailed distributions, located on the right far from the mean. Statistically, such values are considered to be outliers and consequently strongly influence the sample moments. The logarithm function is applied to the data set to eliminate the influence of extreme values. The maximum entropy distribution is uniquely defined by the chosen constraints, which normally contain information from observations or theoretical considerations. Thus in geophysical applications for example, important prior characteristics of the underlying distribution should be preserved—e.g. a J-shaped, Bell-shape or heavy-tailed distribution. The constraints should also be chosen based on the suitability of the resulting distribution in regard to the empirical evidence. More details on appropriate constraints are discussed in [11]. In this study, we choose a single constraint to express the features of the distribution given the empirical evidence.

3.3. The Estimation of Maximum Entropy

There are four steps in the ME method to estimate the objective distribution—viz.

1) Specification of appropriate constraints;
2) Construction of the Lagrange multipliers;
3) Derivation of the entropy function of the distribution; and
4) Derivation of the relation between the Lagrange multiplier and the constraints.

<u>Step 1</u> *Specification of Appropriate Constraints.*

Taking the natural logarithm, from (1) we have

$$\ln f(x) = \ln\frac{\alpha}{\beta} - \left(\frac{\alpha+1}{\alpha}\right)\ln\left[\alpha + \left(\frac{x-\mu}{\beta}\right)^{\alpha}\right] \tag{7}$$

To establish the entropy as expressed in Equation (3), multiply Equation (7) with $-f(x)$ and integrate over the entries space $(0, \infty)$ to obtain

$$S = -\int_0^\infty \ln\frac{\alpha}{\beta} - \left(\frac{\alpha+1}{\alpha}\right)\ln\left[\alpha + \left(\frac{x-\mu}{\beta}\right)^{\alpha}\right] f(x) \mathrm{d}x \tag{8}$$

which is to be subject to the constraints

$$\int_0^\infty \ln\left[\alpha + \left(\frac{x-\mu}{\beta}\right)^{\alpha}\right] f(x) \mathrm{d}x = E\left[\ln\left[\alpha + \left(\frac{x-\mu}{\beta}\right)^{\alpha}\right]\right] \tag{9}$$

$$\int_0^\infty f(x) \mathrm{d}x = 1. \tag{10}$$

<u>Step 2</u> *Construction of the Lagrange Multipliers.*

From Equation (6)

$$f(x) = \exp\left[-\lambda_0 - \lambda_1 \ln\left[\alpha + \left(\frac{x-\mu}{\beta}\right)^{\alpha}\right]\right], \tag{11}$$

where λ_0, λ_1 are the Lagrange multipliers. Substituting Equation (11) into Equation (10) we have

$$\int_0^\infty \exp\left[-\lambda_0 - \lambda_1 \ln\left[\alpha + \left(\frac{x-\mu}{\beta}\right)^{\alpha}\right]\right] \mathrm{d}x = 1, \tag{12}$$

such that

$$\begin{aligned}
\exp(\lambda_0) &= \int_0^\infty \exp\left[-\lambda_1 \ln\left[\alpha + \left(\frac{x-\mu}{\beta}\right)^{\alpha}\right]\right] \mathrm{d}x \\
&= \int_0^\infty \left[\alpha + \left(\frac{x-\mu}{\beta}\right)^{\alpha}\right]^{-\lambda_1} \mathrm{d}x = \frac{\beta}{\alpha}\int_0^\infty z^{-\lambda_1}(z-\alpha)^{(1/\alpha)-1} \mathrm{d}z \\
&= \frac{\beta}{\alpha}\left(\alpha^{(1/\alpha)-1}\right)\int_0^\infty (u\alpha)^{-\lambda_1}(u-1)^{(1/\alpha)-1} \alpha \mathrm{d}u \\
&= (-1)^{(1/\alpha)-1} \beta\alpha^{(1/\alpha)-1-\lambda_1}\int_0^\infty u^{-\lambda_1}(1-u)^{(1/\alpha)-1} \mathrm{d}u
\end{aligned} \tag{13}$$

On setting $z=\alpha+\left(\frac{x-\mu}{\beta}\right)^{\alpha}$ such that $dx\frac{\beta}{\alpha}(z-\alpha)^{(1/\alpha)-1}dz$, and $u=\frac{z}{\alpha}$ such that $dz=\alpha du$.

Since $\exp(\lambda_0)>0$ from Equation (13) we require $(1/\alpha)-1=2k, k=0,1,2,\cdots,$ implying $0<\alpha=1/(2k+1)\le 1, k=0,1,2,\cdots$.

Consequently,

$$\begin{aligned}\exp(\lambda_0)&=\beta\alpha^{(1/\alpha)-1-\lambda_1}\int_0^{\infty}u^{1-\lambda_1-1}(1-u)^{(1/\alpha)-1}du\\&=\beta\alpha^{(1/\alpha)-1-\lambda_1}\frac{\Gamma(1-\lambda_1)\Gamma\left(\frac{1}{\alpha}\right)}{\Gamma\left(1-\lambda_1+\frac{1}{\alpha}\right)}.\end{aligned}\tag{14}$$

Then on taking logarithms we have

$$\begin{aligned}\lambda_0&=\ln\beta+((1/\alpha)-1-\lambda_1)\ln\alpha+\ln\Gamma(1-\lambda_1)\\&+\ln\Gamma\left(\frac{1}{\alpha}\right)-\ln\Gamma\left(1-\lambda_1+\frac{1}{\alpha}\right).\end{aligned}$$

<u>Step 3</u> *Derivation of the Entropy Function of the Distribution.*

Substituting Equation (14) into Equation (11) gives

$$f(x)=\frac{\Gamma\left(1-\lambda_1+\frac{1}{\alpha}\right)}{\Gamma(1-\lambda_1)\Gamma\left(\frac{1}{\alpha}\right)}\frac{\left[\alpha+\left(\frac{x-\mu}{\beta}\right)^{\alpha}\right]^{-\lambda_1}}{\beta\alpha^{(1/\alpha)-1-\lambda_1}}$$

and again taking the natural logarithms, we have

$$\begin{aligned}\ln f(x)&=\ln\Gamma\left(1-\lambda_1+\frac{1}{\alpha}\right)-\ln\Gamma(1-\lambda_1)\\&-\ln\Gamma\left(\frac{1}{\alpha}\right)-\ln\beta-\left(\frac{1}{\alpha}-1-\lambda_1\right)\ln\alpha\\&-\lambda_1\ln\left[\alpha+\left(\frac{x-\mu}{\beta}\right)^{\alpha}\right],\end{aligned}$$

and hence from the definition of entropy, Equation (3),

$$\begin{aligned}S&=-\ln\Gamma\left(1-\lambda_1+\frac{1}{\alpha}\right)+\ln\Gamma(1-\lambda_1)\\&+\ln\Gamma\left(\frac{1}{\alpha}\right)+\ln\beta+((1/\alpha)-1-\lambda_1)\ln\alpha\\&+\lambda_1E\left[\ln\left[\alpha+\left(\frac{x-\mu}{\beta}\right)^{\alpha}\right]\right].\end{aligned}\tag{15}$$

<u>Step 4</u> *Derivation of the Relation between the Lagrange Multipliers and Constraints.*

Let $a=\frac{1}{\alpha}$ such that $da=-\frac{1}{\alpha^2}d\alpha$, $b=1-\lambda_1$ such that $db=-d\lambda_1$, and $c=1-\lambda_1+\frac{1}{\alpha}$ such that $\frac{\partial c}{\partial\lambda_1}=-1$ and $\frac{\partial c}{\partial\alpha}=-\frac{1}{\alpha^2}$.

Since $\frac{\partial}{\partial t}\ln\Gamma(t)=\psi(t)$ is the digamma function, it follows that

$$\frac{\partial}{\partial\alpha}\ln\Gamma\left(\frac{1}{\alpha}\right)=-\frac{1}{\alpha^2}\psi(a),$$

$$\frac{\partial}{\partial\lambda_1}\ln\Gamma(1-\lambda_1)=-\psi(b),$$

$$\frac{\partial}{\partial\alpha}\ln\Gamma\left(1-\lambda_1+\frac{1}{\alpha}\right)=-\frac{1}{\alpha^2}\psi(c),$$

and

$$\frac{\partial}{\partial\lambda_1}\ln\Gamma\left(1-\lambda_1+\frac{1}{\alpha}\right)=-\psi(c).$$

There are four parameters in Equation (15)—viz. μ,α,β and λ_1. To maximize Equation (15), we need to set the following partial derivative to zero:

$$\begin{aligned}\frac{\partial S}{\partial\alpha}&=-\frac{\partial}{\partial\alpha}\ln\Gamma\left(1-\lambda_1+\frac{1}{\alpha}\right)+\frac{\partial}{\partial\alpha}\ln\Gamma\left(\frac{1}{\alpha}\right)\\&+\left(\frac{1}{\alpha}-1-\lambda_1\right)\frac{1}{\alpha}-\left(\frac{1}{\alpha^2}\right)\ln\alpha\\&+\lambda_1E\left[\frac{1}{\alpha+\left(\frac{x-\mu}{\beta}\right)^{\alpha}}\left(1+\left(\frac{x-\mu}{\beta}\right)^{\alpha}\ln\left(\frac{x-\mu}{\beta}\right)\right)\right]\\&=\frac{1}{\alpha^2}(\psi(c)-\ln\alpha)-\frac{1}{\alpha^2}\psi(a)+\left(\frac{1}{\alpha}-1-\lambda_1\right)\frac{1}{\alpha}\\&+\lambda_1E\left[\frac{1}{\alpha+\left(\frac{x-\mu}{\beta}\right)^{\alpha}}\left(1+\left(\frac{x-\mu}{\beta}\right)^{\alpha}\ln\left(\frac{x-\mu}{\beta}\right)\right)\right],\end{aligned}\tag{16}$$

$$\begin{aligned}\frac{\partial S}{\partial\lambda_1}&=-\frac{\partial}{\partial\lambda_1}\ln\Gamma\left(1-\lambda_1+\frac{1}{\alpha}\right)+\frac{\partial}{\partial\lambda_1}\ln\Gamma(1-\lambda_1)\\&-\ln\alpha+E\left[\ln\left(1+\left(\frac{x-\mu}{\beta}\right)^{\alpha}\right)\right]\\&=\psi(c)-\psi(b)-\ln\alpha+E\left[\ln\left[\alpha+\left(\frac{x-\mu}{\beta}\right)^{\alpha}\right]\right]\\&=0\end{aligned}$$

$$\frac{1}{\alpha^2}\left(\psi(c)-\ln\alpha\right)=\frac{1}{\alpha^2}\psi(b)-\frac{1}{\alpha^2}E\left[\ln\left(1+\left(\frac{x-\mu}{\beta}\right)^{\alpha}\right)\right], \tag{17}$$

$$\frac{\partial S}{\partial\mu}=-\lambda_1\alpha E\left[\frac{\left(\frac{x-\mu}{\beta}\right)^{\alpha-1}}{\alpha+\left(\frac{x-\mu}{\beta}\right)^{\alpha}}\right]=-(\alpha+1)E\left[\frac{\left(\frac{x-\mu}{\beta}\right)^{\alpha-1}}{\alpha+\left(\frac{x-\mu}{\beta}\right)^{\alpha}}\right], \tag{18}$$

and

$$\frac{\partial S}{\partial\beta}=\frac{1}{\beta}-\frac{\lambda_1\alpha}{\beta}E\left[\frac{\left(\frac{x-\mu}{\beta}\right)^{\alpha}}{\alpha+\left(\frac{x-\mu}{\beta}\right)^{\alpha}}\right]=\frac{1}{\beta}-\frac{\alpha+1}{\beta}E\left[\frac{\left(\frac{x-\mu}{\beta}\right)^{\alpha}}{\alpha+\left(\frac{x-\mu}{\beta}\right)^{\alpha}}\right] \tag{19}$$

Assuming $\lambda_1=\frac{\alpha+1}{\alpha}$, Equations (16) and (17) yield

$$\begin{aligned}\frac{\partial S}{\partial\alpha}&=\frac{\psi(b)}{\alpha^2}-\frac{1}{\alpha^2}E\left[\ln\left(1+\left(\frac{x-\mu}{\beta}\right)^{\alpha}\right)\right]-\frac{\psi(a)}{\alpha^2}-\frac{2}{\alpha}+\left(\frac{\alpha+1}{\alpha}\right)E\left[\frac{1}{\alpha+\left(\frac{x-\mu}{\beta}\right)^{\alpha}}\left(1+\left(\frac{x-\mu}{\beta}\right)^{\alpha}\ln\left(\frac{x-\mu}{\beta}\right)\right)\right]\\&=C-\frac{1}{\alpha^2}E\left[\ln\left(1+\left(\frac{x-\mu}{\beta}\right)^{\alpha}\right)\right]+\left(\frac{\alpha+1}{\alpha}\right)E\left[\frac{1}{\alpha+\left(\frac{x-\mu}{\beta}\right)^{\alpha}}\left(1+\left(\frac{x-\mu}{\beta}\right)^{\alpha}\ln\left(\frac{x-\mu}{\beta}\right)\right)\right]\end{aligned} \tag{20}$$

where $C=\frac{1}{\alpha^2}\left(\psi(b)-\psi(a)\right)-\frac{2}{\alpha}$ is a constant.

The parameter estimation for the K3D (*i.e.* of μ,β,α) from the ME for $\beta>0$ and $0<\alpha=1/(2k+1)\le 1$ $k=0,1,2,\cdots,$ are obtained from Equations (18)-(20), respectively. By the definition of expectation of random variable $E(Y)=\sum_{\text{all}\,x}y\cdot P(Y=y)$ assume that $P(Y=y)$ equal to 1, thus Equation (18),

$$E\left[\frac{\left(\frac{x-\mu}{\beta}\right)^{\alpha-1}}{\alpha+\left(\frac{x-\mu}{\beta}\right)^{\alpha}}\right]\approx\sum_{i=1}^{n}\left[\frac{\left(\frac{x_i-\mu}{\beta}\right)^{\alpha-1}}{\alpha+\left(\frac{x_i-\mu}{\beta}\right)^{\alpha}}\right], \tag{21}$$

and apply this assumption to Equations (19) and (20).

4. The Maximum Likelihood Estimation

From Equation (1), the log-likelihood function can be written as

$$\begin{aligned}&\ln L(\mu,\alpha,\beta)\\&=n\ln\frac{\alpha}{\beta}-\left(\frac{\alpha+1}{\alpha}\right)\sum_{i=1}^{n}\ln\left[\alpha+\left(\frac{x_i-\mu}{\beta}\right)^{\alpha}\right]\end{aligned} \tag{22}$$

where x_i is the i-th value of the random variable X and n is a sample size. Multiply with –1 and differentiating Equation (22) partially with respect to each parameter, we obtain the MLE by equating each of the following partial derivatives to zero:

$$\frac{\partial}{\partial\mu}\ln L=-(\alpha+1)\sum_{i=1}^{n}\left[\frac{\left(\frac{x_i-\mu}{\beta}\right)^{\alpha-1}}{\alpha+\left(\frac{x_i-\mu}{\beta}\right)^{\alpha}}\right] \tag{23}$$

$$\frac{\partial}{\partial\beta}\ln L=\frac{n}{\beta}-\left(\frac{\alpha+1}{\beta}\right)\sum_{i=1}^{n}\left[\frac{\left(\frac{x_i-\mu}{\beta}\right)^{\alpha}}{\alpha+\left(\frac{x_i-\mu}{\beta}\right)^{\alpha}}\right] \tag{24}$$

$$\frac{\partial}{\partial\alpha}\ln L=-\frac{n}{\alpha}-\frac{1}{\alpha^2}\sum_{i=1}^{n}\ln\left[\alpha+\left(\frac{x_i-\mu}{\beta}\right)^{\alpha}\right]+\left(\frac{\alpha+1}{\alpha}\right)\sum_{i=1}^{n}\left[\frac{1}{\alpha+\left(\frac{x_i-\mu}{\beta}\right)^{\alpha}}\left(1+\left(\frac{x_i-\mu}{\beta}\right)^{\alpha}\ln\left(\frac{x_i-\mu}{\beta}\right)\right)\right] \tag{25}$$

By Equation (21), a comparison of the equations of the ME and the MLE immediately reveals that Equation (18) is equivalent to Equation (23), Equation (19) to Equation (24) and Equation (20) to Equation (25), where $x,\beta>0$ and $0<\alpha=1/(2k+1)\le 1 \quad k=0,1,2,\cdots$. Consequently, the two methods become equivalent for discrete random variables.

5. Conclusion

A positive skewness distribution, the three-parameter kappa distribution, is considered. Parameter estimation by the maximum likelihood method requires a certain cutoff in the parameter space or a best starting value, for otherwise the solution may appear under-determined instead of a unique answer (there can exist a concave set). The principle of maximum entropy is another tool to address this problem under constraints that show the characteristic of the distribution given the empirical evidence, using the method of Lagrange multipliers. For $0<\alpha=1/(2k+1)\le 1 \quad k=0,1,2,\cdots$, and $x,\beta>0$ the principle of maximum entropy method is equivalent to the maximum likelihood method for the discrete case.

REFERENCES

[1] E. T. Jaynes, "Information Theory and Statistical Mechanics," *Physical Review*, Vol. 106, No. 4, 1957, p. 620. doi:10.1103/PhysRev.106.620

[2] E. T. Jaynes, "Prior Probabilities," *IEEE Transactions on Systems Science and Cybernetics*, Vol. 3, No. 4, 1968, pp. 227-241. doi:10.1109/TSSC.1968.300117

[3] P. W. Mielke, "Another Family of Distributions for Describing and Analyzing Precipitation Data," *Journal of Applied Meteorology*, Vol. 12, No. 2, 1973, pp. 275-280. doi:10.1175/1520-0450(1973)012<0275:AFODFD>2.0.CO;2

[4] P. W. Mielke and E. R. Johnson, "Three-Parameter Kappa Distribution Maximum Likelihood Estimates and Likelihood Ratio Tests," *Monthly Weather Review*, Vol. 101, No. 9, 1973, pp. 701-707. doi:10.1175/1520-0493(1973)101<0701:TKDMLE>2.3.CO;2

[5] J. S. Park, S. C. Seo and T. Y. Kim, "A Kappa Distribution with a Hydrological Application," *Stochastic Environmental Research and Risk Assessment*, Vol. 23, No. 5, 2009, pp. 579-586. doi:10.1007/s00477-008-0243-5

[6] J. Kysely and J. Picek, "Probability Estimates of Heavy Precipitation Events in a Flood-Prone Central-European Region with Enhanced Influence of Mediterranean Cyclones," *Advances in Geosciences*, Vol. 12, 2007, pp. 43-50. doi:10.5194/adgeo-12-43-2007

[7] D. J. Dupuis and C. Winchester, "More on the Four-Parameter Kappa Distribution," *Journal of Statistical Computation and Simulation*, Vol. 7, No. 2, 2001, pp. 99-113. doi:10.1080/00949650108812137

[8] Z. Hradil and J. Rehacek, "Likelihood and Entropy for Statistical Inversion," *Journal of Physics*: *Conference Series*, Vol. 36, 2006, pp. 55-59.

[9] S. M. Papalexious and D. Koutsoyiannis, "Entropy Base Derivation of Probability Distributions: A Case Study to Daily Rainfall," *Advances in Water Resources*, 2012, in Press. doi:10.1016/j.advwatres.2011.11.007

[10] V. P. Singh and Z. Q. Deng, "Entropy-Based Parameter Estimation for Kappa Distribution," *Journal of Hydrologic Engineering*, Vol. 8, No. 2, 2003, pp. 81-92. doi:10.1061/(ASCE)1084-0699(2003)8:2(81)

[11] R. D. Levine and M. Tribus, "The Maximum Entropy Formalism," MIT Press, Cambridge, 1978.

