American Journal of Scientific Research
ISSN 1450-223X Issue 55 (2012), pp. 70-75

http://www.eurojournals.com/ajsr.htm

Estimation of Single Distributions Parameter by T.O.M with Exponential Families

Jubran Abdul Ameer K.
Department of Statistics and Informations
College of Computer Science and Mathematics
University of Al-Qadisiya, Diwaniya, Iraq
E-mail: jubalaa98@yahoo.com

Abstract

We discuss the T.O.M (Term Omission Method) to estimate the distribution parameters that belong to exponential family with one single distribution parameter, and compare it with chosen method like MLE (Maximum Likelihood Estimation).

Keywords: Term Omission Method, Exponential Family, Maximum Likelihood Estimator

1. Introduction

Estimators that based on T.O.M[5] deals with many distributions with discrete or continuous random variable that have exponential family, and for simplifying we take some distributions as an examples like Poisson, Normal and other distributions.

2. Exponential Family

We note that there are many definitions of such type of that representation of exponential families. In this research, we will recall the regular type of exponential class.

3. Definitions

Def. (1): A family of continuous (discrete) random variables is called an exponential family if the probability density functions (probability mass functions) can be expressed in the form

$$f_X(x/\theta) = h(y)c(\theta)\exp\left(\sum_{i=1}^{k}\theta_i t_i(x)\right) \text{ , } x = 0,1,2,... \tag{1}$$

for x in the common domain of the $f_X(x/\theta)$, $\theta\in R^k$.

Obviously h and c are non-negative functions. The $t_i(x)$ are real-valued functions of the observations.[1]

Def. (2): Let ϑ be an interval on the real line. Let $\{f(x;\theta): \theta\in\vartheta\}$ be a family of pdf's (or pmf's). We assume that the set $\{\underline{x} : f(\underline{x};\theta) > 0\}$ is independent of θ, where $\underline{x} = (x_1, x_2,..., x_n)$. We say that the family $\{f(x;\theta): \theta\in\vartheta\}$ is a one-parameter exponential family if there exist real-valued functions Q(θ) and D(θ) on ϑ and Borel-measurable functions T($\underline{X}$) and S($\underline{X}$) on R^n such that

$$f(\underline{x};\theta) = \exp\left(Q(\theta)T(\underline{x}) + D(\theta) + S(\underline{x})\right) \tag{2}$$

if we write $f(\underline{x};\theta)$ as

$$f(\underline{x};\eta) = h(\underline{x})c(\eta)\exp(\eta T(\underline{x})) \tag{3}$$

where $h(\underline{x}) = \exp(S(\underline{x}))$, $\eta = Q(\theta)$, and $c(\eta) = \exp(D(Q^{-1}(\eta)))$, then we call this the exponential family in canonical form for a natural parameter η.

Def. (3): Let $\underline{\vartheta} \subseteq R^k$ be a k-dimensional interval. Let {f(x; $\underline{\theta}$): $\underline{\theta} \in \underline{\vartheta}$} be a family of pdf's (or pmf's). We assume that the set $\{\underline{x} : f(\underline{x};\underline{\theta}) > 0\}$ is independent of $\underline{\theta}$, where $\underline{x} = (x_1, x_2, ..., x_n)$. We say that the family {f(x; $\underline{\theta}$): $\underline{\theta} \in \underline{\vartheta}$} is a k-parameter exponential family if there exist real-valued functions $Q_1(\underline{\theta}),\ldots, Q_k(\underline{\theta})$ and $D(\underline{\theta})$ on $\underline{\vartheta}$ and Borel-measurable functions $T_1(\underline{X}),\ldots,T_k(\underline{X})$ and $S(\underline{X})$ on R^n such that:[2]

$$f(\underline{x};\underline{\theta}) = \exp\left(\sum_{i=1}^{k} Q_i(\underline{\theta})T_i(\underline{x}) + D(\underline{\theta}) + S(\underline{x})\right) \tag{4}$$

Def. (4): Exponential family is a class of distributions that all share the following form:

$$P(y/\eta) = h(y)\exp\{\eta^T T(y) - A(\eta)\} \tag{5}$$

- Is the natural parameter. For a given distribution η specifies all the parameters needed for that distribution.
- *T(y)* is the sufficient statistic of the data (in many cases T(y) = y, in which case the distribution is said to be in canonical form and η is referred to as the canonical parameter).
- A(η) is the log-partition function which ensures that p(y/η) remains a probability distribution.
- h(y) is the non-negative base measure (in many cases it is equal to 1).

Note that since η contains all the parameters needed for a particular distribution in its original form, we can express it with respect to the mean parameter θ:[3]

$$P(y/\theta) = h(y)\exp\{\eta(\theta)T(y) - A(\eta(\theta))\} \tag{6}$$

Def. (5): (Regular Exponential Family): Consider a one-parameter family $\{f(x;\theta) : \theta \in \Omega)$ of probability density functions, where Ω is the interval set $\Omega = \{\theta : \gamma < \theta < \delta\}$, where γ and δ are known constants, and where[4]

$$f(x;\theta) = \begin{cases} e^{p(\theta)k(x)+s(x)+q(\theta)} & \text{a < x < b} \\ 0 & \text{o.w} \end{cases} \tag{7}$$

The form (7) is said to be a member of the exponential class of probability density functions of the continuous type, if the following conditions satisfy:

1) Neither (a) nor (b) depends upon θ. increasingly
2) p(θ) is a nontrivial continuous function of θ.
3) Each of k′(x)≠0 and s(x) is a continuous function of x.

and the following conditions with discrete random variable X_i:

1) The set $\{x : x = a_1, a_2, ...\}$ does not depend upon θ.
2) p(θ) is a nontrivial continuous function of θ.
3) k(x) is a nontrivial function of x.

4. Examples of Distributions that belongs to Exponential Family

We will take here some distributions as an example, with discrete and continuous random variables, to how can write the p.d.f as an exponential class form.

Example 1: Let X be a discrete random variable of Poisson distribution with parameter θ, 0<θ<∞, with p.d.f

$$f(x;\theta)=\frac{\theta^x e^{-\theta}}{x!} \qquad \mathrm{x}=0,1,2,...$$

we can rewrite the p.d.f of this distribution represented by exponential class as follows:

$$f(x;\theta)=\exp[x(\ln\theta)-\ln(x!)-\theta] \qquad \mathrm{x}=0,1,2,...$$

where $k(x)=x$, $\mathrm{p}(\theta)=\ln(\theta)$, $\mathrm{s(x)}=-\ln[\mathrm{x}!]$, $\mathrm{q}(\theta)=-\theta$ with satisfying conditions.

Example 2: The family $\{f(x;\theta);0<\theta<\infty\}$, which each member of the family with a single parameter f(x;θ) is N(0,θ), can represents a regular case of the exponential class of the continuous type because

$$f(x;\theta)=\frac{1}{\sqrt{2\pi\theta}}e^{-\frac{x^2}{2\theta}} \qquad -\infty<\mathrm{x}<\infty$$

can written in the exponential form

$$f(x;\theta)=\exp\left(-\frac{1}{2\theta}x^2-\ln\sqrt{2\pi\theta}\right) \qquad -\infty<\mathrm{x}<\infty$$

where $k(x)=x^2$, $\mathrm{p}(\theta)=\frac{-1}{2\theta}$, $\mathrm{s(x)}=0$, $\mathrm{q}(\theta)=-\ln\sqrt{2\pi\theta}$ with satisfying conditions.

Example 3: The family $\{f(x;\theta);0<\theta<\infty\}$, which each member of the family with a single parameter f(x;θ) is N(θ,1), can represents a regular case of the exponential class of the continuous type because

$$f(x;\theta)=\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}(x-\theta)^2} \qquad -\infty<\mathrm{x}<\infty$$

can written in the exponential form

$$f(x;\theta)=\exp\left(\theta x-\frac{1}{2}x^2-\frac{1}{2}\theta^2-\ln\sqrt{2\pi}\right) \qquad -\infty<\mathrm{x}<\infty$$

where $k(x)=x$, $\mathrm{p}(\theta)=\theta$, $\mathrm{s(x)}=-\frac{1}{2}\mathrm{x}^2$, $\mathrm{q}(\theta)=-\frac{1}{2}\theta^2-\ln\sqrt{2\pi}$ with satisfying conditions.

5. T.O.M. with Exponential Family

T.O.M[6] can used to estimate the value of parameter, below we will derivative the T.O.M of distributions (with one parameter) that have exponential family as follows:

For sample with size n having the p.d.f f(x;θ), and for any two values x_i and x_{i+1}, where $1\leq i\leq n-1$,

$\mathrm{x_i}$ $y_i=e^{k(x_i)p(\theta)+q(\theta)+s(x_i)}$

$\mathrm{x_{i+1}}$ $y_j=e^{k(x_{i+1})p(\theta)+q(\theta)+s(x_{i+1})}$

by taking the natural logarithm to y_i , y_{i+1} we have

x_i $k(x_i)p(\theta)+q(\theta)+s(x_i)$

x_{i+1} $k(x_{i+1})p(\theta)+q(\theta)+s(x_{i+1})$

and by subtract the last result y_{i+1} from y_i we obtain

$$p(\theta)[k(x_{i+1})-k(x_i)]+[s(x_{i+1})-s(x_i)]$$

and again by subtract $[s(x_{i+1})-s(x_i)]$ from the final amount we have

$$p(\theta)[k(x_{i+1})-k(x_i)]$$

Finally, by divided this amount over $[k(x_{i+1})-k(x_i)]$ we have the function of θ, p(θ).

Therefore, we can define the $p^i(\theta)$ as follows:

$$p^i(\theta)=\frac{[\ln(y_{i+1})-\ln(y_i)]-[s(x_{i+1})-s(x_i)]}{[k(x_{i+1})-k(x_i)]} \qquad (8)$$

where $p^i(\theta)$ represent to the values that we have from previous steps of T.O.M, ∀i=1,2,…,n-1. Thus from eq. (8) we can educe values of θ^i from $p^i(\theta)$.

Therefore the estimation of θ can found now using the least square error with the following equation:

$$\hat{\theta} = Min\left(\sum_{m=1}^{n}\left[f(x_m,\theta^i) - y_m\right]\right)^2 \qquad i = 1,2,...,n-1$$

where $f(x_m,\theta^i)$ is the value of function $f(x_m)$ on θ^i, and $y_m = y(x_m)$ is the observed value on x_m.

Example 4: In example (1) of exponential class of Poisson distribution we can obtain the following:

$$f(x;\theta) = \exp\left[x(\ln\theta) - \ln(x!) - \theta\right] \qquad x = 0,1,2,... \tag{9}$$

where $k(x) = x$, $p(\theta) = \ln(\theta)$, $s(x) = -\ln[x!]$, $q(\theta) = -\theta$

therefore by using (8) we can write

$$p^i(\theta) = \frac{\left[\ln(y_{i+1}) - \ln(y_i)\right] - \left[-\ln[x!_{i+1}] + \ln[x!_i]\right]}{\left[x_{i+1} - x_i\right]}$$

or, simply

$$p^i(\theta) = \frac{\ln\left(\frac{y_{i+1}}{y_i} . \frac{x!_{i+1}}{x!_i}\right)}{x_{i+1} - x_i}, \quad i=1,2,\ldots,n-1$$

Example 5: In example (2) of exponential class of Normal dist. N(0,θ) we can obtain the following:

$$f(x;\theta) = \exp\left(-\frac{1}{2\theta}x^2 - \ln\sqrt{2\pi\theta}\right) \qquad -\infty < x < \infty$$

$$k(x) = x^2 \ , \ p(\theta) = \frac{-1}{2\theta} \ , \ s(x) = 0 \ , \ q(\theta) = -\ln\sqrt{2\pi\theta}$$

therefore by using (8) we can write

$$p^i(\theta) = \frac{\left[\ln(y_{i+1}) - \ln(y_i)\right]}{\left[x^2_{i+1} - x^2_i\right]}, \quad i=1,2,\ldots,n-1$$

Example 6: In example (3) of exponential class of Normal dist. N(θ,1) we can obtain the following:

$$f(x;\theta) = \exp\left(\theta x - \frac{1}{2}x^2 - \frac{1}{2}\theta^2 - \ln\sqrt{2\pi}\right) \qquad -\infty < x < \infty$$

where $k(x) = x$, $p(\theta) = \theta$, $s(x) = -\frac{1}{2}x^2$, $q(\theta) = -\frac{1}{2}\theta^2 - \ln\sqrt{2\pi}$

therefore by using (8) we can write

$$p^i(\theta) = \frac{\left[\ln(y_{i+1}) - \ln(y_i) - \left(-\frac{1}{2}x^2_{i+1} + \frac{1}{2}x^2_i\right)\right]}{\left[x_{i+1} - x_i\right]}$$

or

$$p^i(\theta) = \frac{\left[\left(\ln(y_{i+1}) + \frac{1}{2}x^2_{i+1}\right) - \left(\ln(y_i) + \frac{1}{2}x^2_i\right)\right]}{\left[x_{i+1} - x_i\right]}$$

6. Results

Below we have tables for comparison between two methods (T.O.M , MLE[7]) for the distributions in this research and choose the best method according to MSE (Mean Square Error) with equation below:

$$MSE = \frac{\sum_{i=1}^{n}\left[f(x_i;\theta) - f(x_i;\hat{\theta})\right]^2}{n}$$

where $f(x_i;\theta), f(x_i;\hat{\theta})$ represent to the probability function of variable x_i, θ, $\hat{\theta}$ parameter and estimating parameter respectively.

Table 1: The estimate values of parameter that founding by MLE and T.O.M for Poisson distribution.

θ	N	T.O.M	MLE
0.4	20	0.4	0.5
	40	0.384615	0.475
	100	0.38806	0.42
	1000	0.411552	0.428
1.1	20	1.166667	1.3
	40	1.076923	1.125
	100	1.125	1.01
	1000	1.10119	1.084
3	20	3	3.05
	40	3	2.825
	100	2.909091	3.17
	1000	3	2.987
7.1	20	7	6.45
	40	7	6.825
	100	6.857143	7.5
	1000	7.086614	7.048

Table 2: The estimate values of parameter that founding by MLE and T.O.M for Normal distribution with mean 0 and variance θ.

θ	N	T.O.M	MLE
0.4	20	0.400687	0.3485652
	40	0.399523	0.3074637
	100	0.399914	0.3216095
	1000	0.400065	0.3199984
1.1	20	1.135039	0.6645947
	40	1.105634	0.7128143
	100	1.099953	0.7883161
	1000	1.100218	0.8048009
3	20	3.041161	1.421986
	40	3.000293	1.426634
	100	2.995623	1.510042
	1000	3.000064	1.563115
7.1	20	7.62959	3.266826
	40	7.317262	3.506112
	100	7.060247	2.820753
	1000	7.096881	3.218951

Table 3: The estimate values of parameter that founding by MLE and T.O.M for Normal distribution with mean θ and variance 1.

θ	N	T.O.M	MLE
0.4	20	0.398992	0.6907246
	40	0.396855	0.4490734
	100	0.399628	0.4859558
	1000	0.400036	0.3855798

Table 3: The estimate values of parameter that founding by MLE and T.O.M for Normal distribution with mean θ and variance 1. - continued

1.1	20	1.101443	1.860171
	40	1.099408	1.42258
	100	1.100216	1.153501
	1000	1.099968	1.090096
3	20	2.996207	2.976046
	40	2.999867	3.103523
	100	3.000096	3.126839
	1000	2.999979	3.028492
7.1	20	7.103125	7.085547
	40	7.102167	7.334194
	100	7.100064	7.143281
	1000	7.100005	7.112599

7. Discussion

In this research and from the pervious tables we can see the following results:

1) The preference of the T.O.M with other method using MSE in all samples,
2) Approximation of T.O.M when sample large.
3) We have an exact estimation when sample exact fitting the distribution.

References

[1] Exponential Families of Random Variables, 2008
[2] Jurgen, S., 1999. "Mathematical Statistics I", Utah State University, p: 120.
[3] Andreas Vlachos, 2010, "Notes on exponential family distributions and generalized linear models".
[4] Robert V. Hogg and Allen T. Craig, 1978. "Introduction to Mathematical Statistics", 4th Edition, p: 357.
[5] Labban, J.A., 2005. "Modified and Simplified Method for Finite Difference (Divided)". Al-Qadisiya Journal for Science, Vol. 10, No. 2, pp. 244-251.
[6] Labban, J.A., 2009. "Approximation of Scale Parameter of Inverted Gamma Distribution by TOM Modified". Accepted for Publishing
[7] Ziegler, Andreas, 2011. "Generalized Estimating Equations", Springer, p. 37-43.