Applied Mathematical Sciences, Vol. 4, 2010, no. 41, 2013 - 2019

Exponential Family and Taneja's Entropy

A. Beitollahi

Department of Statistics
Islamic Azad University - Roudehen Branch
Tehran, Iran
arman_beitollahi@yahoo.com

P. Azhdari

Department of Statistics
Islamic Azad University - North Tehran Branch
Tehran, Iran

Abstract

Menendez [3] derived Shaonnon's entropy formula for exponential family. In this paper we proposed Taneja's entropy formula for exponential family. Notice that this relation is the generalization of Menendez work. Also we will obtain proper Taneja's entropy formulas for Bernoulli, Geometry, Gamma, Beta and Normal distributions. Finally we present Taneja's entropy formula for multivariate normal distribution.

Keywords: Taneja's entropy, Regular exponential family, Multivariate Normal distribution

1 Introduction

Let $(\chi, \beta_\chi, P_\theta, \theta \in \Theta)$ be a statistical space where Θ is an open subset of R^M. We consider that there exist p.d.f. $f_\theta(x)$ for the distribution P_θ with respect to a σ-finite measure μ. In 1975 Taneja [6] introduced the generalized entropy as follows:

$$H_T(\theta) = -2^{r-1} \int_\chi f_\theta^r(x) \log f_\theta(x) d\mu(x)$$

which by taking $r = 1$, Shannon's entropy [5] is obtained. Salicru et al.(1993) [4] defined (h, φ)-entropy for $f_\theta(x)$ as follows:

$$H_\varphi^h(f_\theta(x)) = h(\int_\chi \varphi(f_\theta(x)) d\mu(x))$$

Where either $\varphi : [0, \infty) \to \mathbb{R}$ is concave and $h : \mathbb{R} \to \mathbb{R}$ is an increasing concave or φ is convex and h is a decreasing concave. Further we assume that h and φ are in $\mathcal{C}^3$ (functions with continues third derivatives). Taneja's entropy is defined for special case of $\varphi(x) = x^r \log x$ and $h(x) = -2^{r-1}x$.
The exponential family of k-parameter distribution $\{f_\theta(x) : \theta \in \Theta \subseteq \mathbb{R}^k\}$ belongs to the family of k-parameter distribution with suitable situation if we can write $f_\theta(x)$ as:

$$f_\theta(x) = \exp\{\sum_{j=1}^{k} T_j(x)\theta_j - b(\theta) - R(x)\} \qquad (1)$$

with support $S = \{x : f_\theta(x) > 0\}$ where:
a) S is a subset of $\mathbb{R}$ which is independent of $\theta_1, \theta_2, \ldots, \theta_k$.
b) The set Θ of the points $\theta = (\theta_1, \theta_2, \ldots, \theta_k)$ is called natural parameter space and convex.
c) T_j is a real valued function with continuous derivative on S and $R(x)$ is a real valued function and continuous on S.
d) b is a real valued function on Θ which $b(\theta)$ is a function as below:

$$b(\theta) = \ln(\int_{\chi} \exp\{\sum_{j=1}^{k} T_j(x)\theta_j - R(x)\} d\mu(x)) \qquad (2)$$

If $R(x) = 0$, the exponential family is called regular exponential family. The functions defined by :

$$\int_{\chi} f(x) \exp\{\sum_{j=1}^{k} T_j(x)\theta_j - R(x)\} d\mu(x)$$

is continuous and has all derivatives with respect to θ_j and the derivatives can be obtained by derivation under the integral sign.
Here suppose that

$$I_{ij}(\theta) = cov[\frac{\partial}{\partial\theta_i} \log f_\theta(x), \frac{\partial}{\partial\theta_j} \log f_\theta(x)]$$

and we assume that $I_F(\theta)$ is positive definite. We show that:

$$I_F(\theta) = (\frac{\partial^2 b(\theta)}{\partial\theta_i \partial\theta_j})_{i,j=1,2,\ldots,k}$$

Since $\frac{\partial}{\partial\theta_i} \log f_\theta(x) = T_i - \frac{\partial}{\partial\theta_i} b(\theta)$ then, $I_{ij}(\theta) = cov(T_i, T_j)$. On the other hand

$$\int_{x} \exp\{\sum_{j=1}^{k} T_j(x)\theta_j - b(\theta) - R(x)\} d\mu(x) = 1$$

Next we derivative with respect to θ_j

$$\int_x (T_j(x) - \frac{\partial}{\partial\theta_j} b(\theta)) f_\theta(x) d\mu(x) = 0$$

It means $E(T_j(X)) = \frac{\partial}{\partial\theta_j} b(\theta)$, by derivation with respect to θ_k again we obtain $cov(T_j, T_k) = \frac{\partial^2}{\partial\theta_j \partial\theta_k} b(\theta)$ finally

$$I_F(\theta) = (\frac{\partial^2 b(\theta)}{\partial\theta_i \partial\theta_j})_{i,j=1,2,\ldots,k}$$

2 Taneja's entropy in exponential family

In this section we calculate Taneja's entropy for regular exponential k-parameter families and introduce some of their specialities.

Theorem 1: Let $f_\theta(x)$ be a density of the form (1) with $R(x) = 0$,then :

$$H_T(\theta) = -2^{r-1}[e^{-rb(\theta)+b(r\theta)}]\{\sum_{j=1}^{k} \theta_j (\frac{\partial b(r\theta)}{\theta_j}) - b(\theta)\}$$

Proof: We know

$$H_T(\theta) = -2^{r-1} \int_\chi f_\theta^r(x) [\sum_{j=1}^{k} T_j(x)\theta_j - b(\theta)] d\mu(x) \qquad (3)$$

Therefore we need to calculate $\int_\chi f_\theta^r(x) d\mu(x)$ and then $\int_\chi f_\theta^r(x) T_j(x) d\mu(x)$ in the relation (3).
Since

$$\int_\chi e^{\sum_j T_j(x)(r\theta_j) - b(r\theta)} d\mu(x) = 1$$

then

$$\int_\chi f_\theta^r(x) d\mu(x) = \int_\chi e^{\sum_j T_j(x)(r\theta_j) - rb(\theta)} d\mu(x)$$

$$= e^{-rb(\theta)+b(r\theta)} \int_\chi e^{\sum_j T_j(x)(r\theta_j) - b(r\theta)} d\mu(x) = e^{-rb(\theta)+b(r\theta)} \qquad (4)$$

If we obtain partial derivative with respect to θ_j, we have:

$$\int_\chi [rT_j(x) - r\frac{\partial b(\theta)}{\partial\theta_j}] e^{\sum_j T_j(x)(r\theta_j) - b(r\theta)} d\mu(x) = [-r\frac{\partial b(\theta)}{\partial\theta_j} + r\frac{\partial b(r\theta)}{\partial\theta_j}] e^{-rb(\theta)+b(r\theta)}$$

then we omit coefficient r and simplify the relation so we have

$$\int_\chi f_\theta^r(x) T_j(x) d\mu(x) - \frac{\partial b(\theta)}{\partial\theta_j} \int_\chi f_\theta^r(x) d\mu = -\frac{\partial b(\theta)}{\partial\theta_j} e^{-rb(\theta)+b(r\theta)} + \frac{\partial b(r\theta)}{\partial\theta_j} e^{-rb(\theta)+b(r\theta)}.$$

Now by use of (4) in above relation we have

$$\int_{\chi} f_\theta^r(x) T_j(x) d\mu(x) = \frac{\partial b(r\theta)}{\partial \theta_j} e^{-rb(\theta)+b(r\theta)} \quad (5)$$

Finally (4) ,(5) complete the proof of theorem.
Pasha et. al. [2] obtained the formula of divergence measure by use of Taneja's entropy in exponential family.

1) Bernoulli distribution
We know in this distribution exponential form is

$$f_\theta(x) = \exp\{\theta x - \ln(\frac{1+e^{-\theta}}{e^{-\theta}})\}$$

where

$$\theta = \ln(\frac{p}{1-p}), \quad b(\theta) = \ln(\frac{1+e^{-\theta}}{e^{-\theta}}) = \ln(e^\theta + 1),$$

therefore Taneja's entropy will be

$$H_T(\theta) = -2^{r-1}[\frac{1+e^{r\theta}}{(1+e^\theta)^r}][\theta r \frac{e^{r\theta}}{e^{r\theta}+1} - \ln(1+e^\theta)]$$

2) Geometry distribution
In this distribution exponential form is

$$f_\theta(x) = \exp\{x\theta - \ln(\frac{e^\theta}{1-e^\theta})\}$$

where

$$\theta = \ln(1-p) \ , \quad b(\theta) = \ln(\frac{e^\theta}{1-e^\theta}),$$

therefore Taneja's entropy will be

$$H_T(\theta) = -2^{r-1}[\frac{(1-e^\theta)^r}{1-e^{r\theta}}][\frac{r\theta}{1-e^{r\theta}} - \ln(\frac{e^\theta}{1-e^\theta})].$$

3) Gamma distribution
In this family we have :

$$f_\theta(x) = \exp\{\theta_1 x + \theta_2 \ln x - (\ln \Gamma(\theta_2+1) + (\theta_2+1)\ln(-\frac{1}{\theta_1}))\}$$

Where:

$$\theta_1 = -\frac{1}{\beta} \quad \text{and} \quad \theta_2 = \alpha - 1$$

and

$$b(\theta) = \ln\Gamma(\theta_2+1) - (\theta_2+1)\ln(-\theta)$$

Then the Taneja's entropy is :

$$H_T(\theta) = -2^{r-1}[e^{-rb(\theta)+b(r\theta)}]\{-(r\theta_2+1)+\theta_2(\frac{\Gamma'_{\theta_2}(r\theta_2+1)}{\Gamma(r\theta_2+1)} - r\ln(-r\theta_1)) - b(\theta)\},$$

where:

$$\Gamma'_x(ax+b) = \frac{d\Gamma(ax+b)}{dx}$$

and

$$[e^{-rb(\theta)+b(r\theta)}] = (\frac{(-\theta_1)^{\theta_2+1}}{\Gamma(\theta_2+1)})^r(\frac{\Gamma(r\theta_2+1)}{(-r\theta_1)^{r\theta_2+1}})$$

4) Beta distribution

In this family we have :

$$f_\theta(x) = \exp\{\theta_1\ln(x) + \theta_2\ln(1-x) - b(\theta)\},$$

where :

$$\theta_1 = \alpha - 1 \quad \text{and} \quad \theta_2 = \beta - 1$$

and

$$b(\theta) = b(\theta_1,\theta_2) = \ln\Gamma(\theta_1+1) + \ln\Gamma(\theta_2+1) - \ln\Gamma(\theta_1+\theta_2+2)$$

Then the Taneja's entropy is :

$$H_T(\theta) = -2^{r-1}[e^{-rb(\theta)+b(r\theta)}]\{\theta_1\frac{\partial b(r\theta)}{\partial\theta_1} + \theta_2\frac{\partial b(r\theta)}{\partial\theta_2} - b(\theta)\}$$

where:

$$[e^{-rb(\theta)+b(r\theta)}] = (\frac{\Gamma(\theta_1+\theta_2+2)}{\Gamma(\theta_1+1)\Gamma(\theta_2+1)})^r(\frac{\Gamma(r\theta_1+1)\Gamma(r\theta_2+1)}{\Gamma(r\theta_1+r\theta_2+2)})$$

5) Normal distribution with mean μ and variance σ^2

In this family we have :

$$f_\theta(x) = \exp\{\theta_1 x + \theta_2 x^2 - \frac{1}{2}\ln(-\frac{\pi}{\theta_2}) + \frac{\theta_1^2}{4\theta_2}$$

where :

$$\theta_1 = \frac{\mu}{\sigma^2} \quad , \quad \theta_2 = -\frac{1}{2\sigma^2} \quad \text{and} \quad b(\theta) = \frac{1}{2}\ln(-\frac{\pi}{\theta_2}) - \frac{\theta_1^2}{4\theta_2},$$

then the Taneja's entropy is :

$$H_T(\theta) = -2^{r-1}(\frac{-\theta_2}{\pi})^{\frac{r}{2}}(\sqrt{\frac{-\pi}{r\theta_2}})[-\frac{1}{2} - \frac{1}{2}\ln(\frac{\pi}{-\theta_2}) + \frac{\theta_1^2}{4\theta_2} - \frac{r\theta_1^2}{4\theta_2^2}]$$

6) d-variate Normal distribution

We know probability density function of multivariate normal distribution is the same as below:

$$f_\theta(x) = \frac{1}{(2\pi)^{\frac{d}{2}}|\Sigma|^{\frac{1}{2}}} e^{-\frac{1}{2}(x-\mu)^t\Sigma^{-1}(x-\mu)}$$

where $\mu = (\mu_1, \ldots, \mu_d)$ is the vector of mean and Σ is the matrix of variance covariance.

Theorem 2: Let $\theta = (\mu, \Sigma)$ and also $f_\theta(x)$ be probability density function of multivariate normal distribution, then Taneja's entropy is:

$$H_T(\theta) = [2^{r-2}d\log 2\pi + 2^{r-2}\log|\Sigma| + 2^{r-2}(\frac{d}{r})]A$$

where

$$A = \frac{(2\pi)^{-\frac{d}{2}(r-1)}r^{-\frac{d}{2}}}{|\Sigma|^{\frac{r-1}{2}}}.$$

proof: By Taneja's entropy definition and probability density function of d-variate normal distribution, we have:

$$H_T(\theta) = -2^{r-1}\int_\chi f_\theta^r(x)\log f_\theta(x)d\mu(x)$$

$$\log f_\theta(x) = -\frac{d}{2}\log 2\pi - \frac{1}{2}\log|\Sigma| - \frac{1}{2}(x-\mu)^t\Sigma^{-1}(x-\mu)$$

therefore

$$H_T(\theta) = -2^{r-1}\{[-\frac{d}{2}\log 2\pi - \frac{1}{2}\log|\Sigma|]\int f_\theta^r dx - \int (x-\mu)^t\Sigma^{-1}(x-\mu)f_\theta^r(x)dx\}$$

it means that:

$$H_T(\theta) = \{2^{r-2}d\log 2\pi + 2^{r-2}\log|\Sigma|\}\int f_\theta^r(x)dx +$$
$$2^{r-2}\int (x-\mu)^t\Sigma^{-1}(x-\mu)f_\theta^r(x)dx \qquad (6)$$

but the first integral in above relation is equal to:

$$\int f_\theta^r(x)dx = \int\{\frac{1}{(2\pi)^{\frac{d}{2}}|\Sigma|^{\frac{1}{2}}}e^{-\frac{1}{2}(x-\mu)^t\Sigma^{-1}(x-\mu)}\}dx$$

$$= \frac{1}{(2\pi)^{\frac{rd}{2}}|\Sigma|^{\frac{r}{2}}} \int e^{-\frac{1}{2}(x-\mu)^t\Sigma^{-1}(x-\mu)}dx$$

$$= (2\pi)^{-\frac{d}{2}(r-1)} \frac{r^{-\frac{d}{2}}}{|\Sigma|^{\frac{r-1}{2}}} \qquad (7)$$

and also the second integral in above relation is equal to:
$\int (x-\mu)^t\Sigma^{-1}(x-\mu)f_\theta^r(x)dx$

$$= \frac{1}{r}\int [(x-\mu)^t(\frac{\Sigma}{r})^{-1}(x-\mu)]\frac{1}{(2\pi)^{\frac{rd}{2}}|\Sigma|^{\frac{r}{2}}}e^{-\frac{1}{2}(x-\mu)^t(\frac{\Sigma}{r})^{-1}(x-\mu)}dx$$

$$= \frac{1}{r}(2\pi)^{-\frac{d}{2}(r-1)}\frac{r^{-\frac{d}{2}}}{|\Sigma|^{\frac{r-1}{2}}}\int \frac{1}{(2\pi)^{\frac{d}{2}}|\frac{\Sigma}{r}|^{\frac{1}{2}}}(x-\mu)^t(\frac{\Sigma}{r})^{-1}(x-\mu)e^{-\frac{1}{2}(x-\mu)^t(\frac{\Sigma}{r})^{-1}(x-\mu)}dx$$

On the other hand we know $(x-\mu)^t\Sigma^{-1}(x-\mu)$ has $\chi^2(d)$ distribution [2]. The last integral in the above relation is equal to mathematical expectation of $(x-\mu)^t(\frac{\Sigma}{r})^{-1}(x-\mu)$ in d-variate normal distribution with mean vector μ and variance covariance matrix $\frac{\Sigma}{r}$. Therefore the above integral is equal to d. Finally we have:

$$\int (x-\mu)^t\Sigma^{-1}(x-\mu)f_\theta^r(x)dx = \frac{d}{r}(2\pi)^{-\frac{d}{2}(r-1)}\frac{r^{-\frac{d}{2}}}{|\Sigma|^{\frac{r-1}{2}}} \qquad (8)$$

by (6), (7) and (8) the proof of theorem is complete.

References

[1] R. V. Hogg. and A. T. Criag., *An introduction to mathematical statistics, 4th ed,* The Macmillan company, N. Y. (1978).

[2] E. Pasha, A. Beitollahi and M. Khodabin, Divergence measure and testing statistical hypothesis, Pakistan Journal of Statistics. 23(3) (2007) 205-220.

[3] M.L. Menendez, Shannon's entropy in exponential families, statistical applications, Applied Mathematics Letters,13(1), (2000), 37-42.

[4] M. Salicru, M. L. Menendez, D. Morales and L. Pardo, Asymptotic distribution of (h, φ)-entropies, Commun. In Stat.(Theory and Methods). 27(7), (1993), 2015-2031.

[5] C. E. Shannon, A mathematical theory of communication, Bell. Syst. Tech. J, 27, (1948), 379 -423.

[6] I. J. Taneja, A study of generalized measures in information theory, Ph. D. Thesis. University of Delhi, (1975).

Received: November, 2009