Iranian Int. J. Sci. **5**(2), 2004, p.267-279

Testing Statistical Hypothesis via Shannon's Entropy in Exponential Families

E., Pasha[1], M., Khodabin[2]*G.R., Mohtashami Borzadran[3]
[1]*Department of Mathematics, Teacher Training University, Tehran, Iran*
[2]*PHD-Student in Islamic Azad University, Sciences and Researches branch , Tehran, Iran*
[3]*Department of Mathematics and Statistics, University of Birjand, Birjand, Iran*
(received: 5/12/2003 ; accepted: 9/6/2004)

Abstract

In this paper we present an entropy characterization of general exponential model and use entropy of regular model to construct a testing of hypothesis for parameters of some common distributions such as normal, exponential, gamma and beta. Furthermore we use these concepts and methods to construct interval estimators for $H(\theta)$ and for θ if $H(\theta)$ is one to one, where $H(\theta)$ is Shannon's entropy of X with density function $f_\theta(x)$ or probability mass function $P_\theta(X=x)$.

Keywords: *Shannon's entropy, Asymptotic distribution, Exponential Family, Testing hypotheses, Confidence coefficient, Confidence interval.*

1. Introduction

A mathematical theory of hypothesis testing in which tests are derived as solutions of clearly stated optimum problems was developed by Neyman and Pearson in the 1930, and since then has been considerably extended. For review of the classic method of hypothesis testing see Lehmann, E.L (1997) and (1983) and George Casella and Roger L.Berger (1990). Consider hypothesis testing structure where we have a null hypothesis H_o and an alternative hypothesis H_1. Let $\hat{\theta}$ be the MLE of the parameter θ and θ_o the true value of parameter under H_o. Application of information theory and Shannon's entropy in testing statistical hypothesis is quite new, and recently some papers in this subject have been published, (Menendez, 1999; Cover and Thomas, 1991), Pasha *et al.,* (2004)). Menendez (1999) published a paper on Shannon's entropy in exponential family and discussed some properties of them. He introduced an asymptotic distribution of $\left(H(\hat{\theta}) - H(\theta_0)\right)$. This article is organized as follows. In the following section we discuss the expression of Shannon's entropy and its asymptotic distribution. In section 3, we present testing hypothesis by Shannon's entropy in the exponential family as well as some examples. In section 4, hypothesis testing for parameters of gamma distribution is introduced. In section 5, we do hypothesis testing for parameters of beta distribution. In section 6, we obtain confidence interval for $H(\theta)$ and for θ if $H(\theta)$ be one to one. Finally the conclusion of paper is given in section 7.

2. Shannon's entropy and asymptotic distribution of it

In this section, we obtain an easy expression for Shannon's entropy and the asymptotic distribution of it in the exponential family when the parameter θ is replaced by its corresponding MLE.

Let $\left(\chi, \beta_\chi, P_\theta\right)_{\theta\in\Theta}$ be a family of Probability Spaces, where Θ is an open convex subset of R^K.

Consider the exponential family:

$$f_\theta(x) = \exp\left\{\sum_{j=1}^{k} T_j(x)\theta_j - b(\theta) - R(x)\right\} \qquad (1.2)\ ,$$

If in (1.2) R(x)=0, then the family is regular which has been considered by Menendez (1999).

Lemma1. Let X be a random variable with distribution of the form (1.2) then the expression of Shannon's entropy is given by:

$$H(\theta) = -\sum_{j=1}^{K} \theta_j \tau_j(\theta) + b(\theta) + E(R(X)),$$

where $\tau_j(\theta) = E(T_j(X)) = \dfrac{\partial b(\theta)}{\partial \theta_j}$ and $\tau(\theta) = \left(\dfrac{\partial b(\theta)}{\partial \theta_1},\ \ldots,\ \dfrac{\partial b(\theta)}{\partial \theta_K}\right)$.

It's easy to see that:

$$b(\theta) = Ln\left(\int_x \exp\left\{\sum_{j=1}^{k} T_j(x)\theta_j - R(x)\right\} d\mu(x)\right).$$

Proof: We know that:

$$H(\theta) = -\int_x f_\theta(x)\left[\sum_{j=1}^{K} T_j(x)\theta_j - b(\theta) - R(x)\right] d\mu(x)$$

$$= -\sum_{j=1}^{K} \theta_j \int_x T_j(x) f_\theta(x) d\mu(x) + b(\theta) + \int_x R(x) f_\theta(x) d\mu(x)$$

Then:

$$H(\theta) = -\sum_{j=1}^{K} \theta_j \tau_j(\theta) + b(\theta) + E(R(X)).$$

In the following we use Shannon's entropy for regular exponential model i.e

$$H(\theta) = -\sum_{j=1}^{k} \theta_i \tau_j(\theta) + b(\theta).$$

Remark1. If $y_n \xrightarrow{L} y$, $A_n \xrightarrow{p} a$, $B_n \xrightarrow{p} b$ then $A_n + B_n y_n \xrightarrow{L} a + by$, where $\xrightarrow{p}$ and $\xrightarrow{L}$ denote the convergence in probability and convergence in distribution respectively.

Remark2. Let $X_1,...,X_n$ be a random sample from $f_\theta(x)$, under some assumptions, any consistent sequence $\hat{\theta}_n = \hat{\theta}(X_1,...,X_n)$ for roots of the likelihood equation satisfies $\sqrt{n}(\hat{\theta}_n - \theta) \xrightarrow{L} N\left(0, I_F^{-1}(\theta)\right)$ and $\sqrt{n}(\hat{\theta}_{jn} - \theta_j) \xrightarrow{L} N\left(0, (I_F^{-1}(\theta))_{jj}\right)$. For details see Lehmann, E.L (1983).

Remark3. Let $\sqrt{n}(T_n - \theta) \xrightarrow{L} N(0, \tau^2)$ and there exist $f'(\theta) \neq 0$ then:

$$\sqrt{n}\left(f(T_n) - f(\theta)\right) \xrightarrow{L} N\left(0, \tau^2 [f'(\theta)]^2\right).$$

Lemma2. Let $I_F(\theta) = \left(\dfrac{\partial^2 b(\theta)}{\partial\theta_i \partial\theta_j}\right)_{i,j=1,2,3,...,k}$ be Fisher information matrix and $T' = (t_1, t_2, ..., t_k)$ where $t_i = \left(\dfrac{\partial H(\theta)}{\partial\theta_i}\right)_{\theta=\theta_o}$ then:

$$\sqrt{n}\left(H(\hat{\theta}) - H(\theta_0)\right) \overset{L}{\to} N(0, T' I_F^{-1}(\theta_o) T),$$

Proof: Consider a Taylor's expansion of $H(\hat{\theta})$ around θ by:

$$H(\hat{\theta}) = H(\theta) + \sum_{i=1}^{k} \frac{\partial H(\theta)}{\partial \theta_i}\left(\hat{\theta}_i - \theta_i\right) + R_n = H(\theta) + T'(\hat{\theta} - \theta) + R_n \qquad (2.2),$$

where $R_n = \frac{1}{2}\sum_{j=1}^{d}\sum_{i=1}^{d}\left[\frac{\partial^2 H(\theta^*)}{\partial\theta_j \partial\theta_i}(\hat{\theta}_j - \theta_j)(\hat{\theta}_i - \theta_i)\right]$ for $\left\|\theta - \theta^*\right\| < \left\|\theta - \hat{\theta}\right\|$.

Then from (2.2) we have:

$$\sqrt{n}(H(\hat{\theta}) - H(\theta)) = \sqrt{n}T'(\hat{\theta} - \theta) + \sqrt{n}R_n .$$

Also we know $\sqrt{n}T'(\hat{\theta} - \theta_0) \xrightarrow{L} N\left(0, T'[I_F(\theta_o)]^{-1}T\right)$ and $\sqrt{n}R_n \xrightarrow{p} 0$, then it is immediate from above remarks that: $\sqrt{n}\left(H(\hat{\theta}) - H(\theta_0)\right) \xrightarrow{L} N\left(0, T'[I_F(\theta_o)]^{-1}T\right)$, provided that $T'[I_F(\theta_o)]^{-1}T > 0$.

Example1. Shannon's entropy for gamma distribution

The canonical probability density function of gamma distribution is given by:

$$f(x) = \exp\left\{\theta_1 x + \theta_2 Lnx - \left(Ln\Gamma(\theta_2 + 1)\right) + (\theta_2 + 1)Ln(-\frac{1}{\theta_1})\right\},$$

where $\theta_1 = -\frac{1}{\beta}$ and $\theta_2 = \alpha - 1$ and $b(\theta_1, \theta_2) = Ln\Gamma(\theta_2 + 1) - (\theta_2 + 1)Ln(-\theta_1)$.

Furthermore we know that:

$$\Psi(x) = \frac{\partial \Gamma(x)}{\partial x} = \sum_{n=0}^{\infty}\left(\frac{1}{n+1} - \frac{1}{n+x}\right) - 0.57721566...$$

Then the Shannon's entropy is given by:

$$H(\theta) = (\theta_2 + 1) - \theta_2\left(\frac{\Psi(\theta_2 + 1)}{\Gamma(\theta_2 + 1)}\right) + Ln\Gamma(\theta_2 + 1) - Ln(-\theta_1)$$

Example2. Shannon's entropy for beta distribution

The canonical pdf of beta distribution is given by

$$f(x) = \exp\left\{\begin{array}{l}\theta_1\Gamma(x) + \theta_2\Gamma(1-x) - \\ \left[Ln\Gamma(\theta_1 + 1) + Ln\Gamma(\theta_2 + 1) - Ln\Gamma(\theta_1 + \theta_2 + 2)\right]\end{array}\right\} \qquad 0 < x < 1$$

where $\theta_1 = \alpha - 1$ and $\theta_2 = \beta - 1$ and:

$$b(\theta_1, \theta_2) = Ln\Gamma(\theta_1 + 1) + Ln\Gamma(\theta_2 + 1) - Ln\Gamma(\theta_1 + \theta_2 + 2).$$

Then the Shannon's entropy is given by:

$$H(\theta_1,\theta_2) = -\theta_1 \frac{\Psi_{\theta_1}(\theta_1+1)}{\Gamma(\theta_1+1)} + \theta_1 \frac{\Psi_{\theta_1}(\theta_1+\theta_2+2)}{\Gamma(\theta_1+\theta_2+2)} - \theta_2 \frac{\Psi_{\theta_2}(\theta_2+1)}{\Gamma(\theta_2+1)}$$

$$+\theta_2 \frac{\Psi_{\theta_2}(\theta_1+\theta_2+2)}{\Gamma(\theta_1+\theta_2+2)} + Ln \frac{\Gamma(\theta_1+1)\Gamma(\theta_2+1)}{\Gamma(\theta_1+\theta_2+2)}$$

Where: $$\Psi_a(a+b) = \frac{d}{da}\Gamma(a+b).$$

3. Testing Hypothesis

In this section we use the asymptotic distribution of Shannon′s entropy for regular exponential model to testing statistical hypothesis and comparison has been made with classical tests. Consider:

$$H_o : \theta = \theta_o \quad versus \quad H_1 : \theta \neq \theta_o ,$$

which is equivalent to

$$H_o : H(\theta) = H(\theta_o) \quad versus \quad H_1 : H(\theta) \neq H(\theta_o),$$

via lemma 2 we apply

$$Z_n = \sqrt{n}\frac{H(\hat{\theta}) - H(\theta_o)}{\sigma(\theta_o)}, \text{ where } \sigma^2(\theta_o) = T'I_F^{-1}(\theta_o)T ,$$

which has asymptotically a standard normal distribution under H_o for sufficiently large n , therefore H_o is rejected at the level α if $|Z_n| > Z_{\frac{\alpha}{2}}$.

Corollary1. This test is asymptotically size α and the power of it tends to 1 as $n \to \infty$.

Proof: If H_o is true, then $H(\theta) = H(\theta_o)$ and $Z_n \xrightarrow{L} Z \sim N(0,1)$, then the type ($I$) error probability is $p(|Z_n| > Z_{\frac{\alpha}{2}}) \to p(|Z| > Z_{\frac{\alpha}{2}}) = \alpha$, and this is an asymptotically size α test. Now consider an alternative parameter value $H(\theta) \neq H(\theta_o)$,

$$Z_n = \sqrt{n}\frac{H(\hat{\theta}) - H(\theta_o)}{\sigma(\theta_o)} = \sqrt{n}\frac{H(\hat{\theta}) - H(\theta)}{\sigma(\theta_o)} + \sqrt{n}\frac{H(\theta) - H(\theta_o)}{\sigma(\theta_o)} ,$$

$$\sqrt{n}\frac{H(\hat{\theta}) - H(\theta)}{\sigma(\theta_o)} \xrightarrow{L} N(0,1) \quad and \quad \frac{\sigma(\theta_o)}{\sqrt{n}} \xrightarrow{p} 0 \quad then$$

$$\sqrt{n}\,\frac{H(\theta)-H(\theta_o)}{\sigma(\theta_o)} \xrightarrow{p} \begin{cases} +\infty & if \quad H(\theta)-H(\theta_o)>0 \\ & or \\ -\infty & if \quad H(\theta)-H(\theta_o)<0 \end{cases},$$

Thus: $Z_n \xrightarrow{p} \pm\infty$ and:

power = p_θ (reject H_o)= $p(|Z_n| > Z_{0.5\alpha}) \to 1 \quad as \quad n \to \infty$.

Example3. The following sample of size 30 was simulated from standard normal distribution:

1, 0.2, 5, -0.6, 0.6, 2.6, 0.4, -2, 0.2, -5, -1, -0.4, 1, -3.2, 0.4, -2.4, 0, 1, 3, -0.8, -1.2, 1.4, 1.8, 2.8, 1.6, -2, -1.8, -3.8, 2, -1

To test $H_o : \sigma^2 = 4 \quad versus \quad H_1 : \sigma^2 \neq 4$, we have

$\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n} X_i^2 = 4.52$, $H(\theta) = \frac{1}{2}\ln(\frac{\pi e}{\theta})$, $H(\hat{\theta}) = H(\frac{1}{2\hat{\sigma}^2}) = 2.172$ and $H(\theta_o) = 2.112$, then:

$$\sigma(\theta_o) = \sqrt{T' I_F^{-1}(\theta_o) T} = \sqrt{\frac{1}{2}} = 0.707,$$

and the value of corresponding statistic is equal to:

$$Z_n = \sqrt{n}\,\frac{H(\hat{\theta})-H(\theta_o)}{\sigma(\theta_o)} = \sqrt{30}\,\frac{2.172-2.112}{0.707} = 0.468 \ngtr Z_{0.025} = 1.96,$$

then H_o isn't rejected and the corresponding p-value is:

$$p_v = p_{H_o}(|Z| > 0.468) = 2p_{H_o}(Z > 0.468) = 0.6384.$$

The value of classic statistic is equal to:

$$\chi^2 = \frac{(n-1)S^2}{\sigma_o^2} = \frac{135.4}{4} = 33.85 \quad \notin \left[\chi^2_{0.975,29}, \chi^2_{0.025,29}\right] = [16.047, 45.7],$$

then H_o isn't rejected and the corresponding p-value is:

$$p_v = p_{H_o}(\chi^2_{29} > 33.85) = 0.25$$

Example4. The following data are a random sample of size 15 simulated from an exponential distribution with mean 5.

3.24, 4.29, 1.17, 5.46, 3.38, 0.49, 1.23, 1.01, 0.58, 1.29, 2.64, 0.99, 2.76, 6.18, 2.97

We consider the hypothesis testing problem:

$$H_o : \beta = 1 \quad versus \quad H_1 : \beta = 5.$$

The MLE for β is $\hat{\beta} = \frac{1}{n}\sum_{i=1}^{n} X_i = \frac{38.33}{15} = 2.56$, also we have $H(\theta) = 1 - \ln\theta$, then $H(\hat{\theta}) = 1.921$ and $H(\theta_o) = 1$ and $\sigma(\theta_o) = \sqrt{T' I_F^{-1}(\theta_o) T} = 1$, so the value of corresponding statistic is

$$Z_n = \sqrt{n}\,\frac{H(\hat{\theta}) - H(\theta_o)}{\sigma(\theta_o)} = \sqrt{15}(1.921 - 1) = 3.567 > Z_{0.05} = 1.64,$$

then H_o is rejected and the corresponding p-value is

$$p_v = p_{H_o}(|Z| > 3.567) = 2p_{H_o}(Z > 3.567) \approx 0.0026.$$

In the classic form, by Neyman and Pearson lemma there exist a UMP test which is given by

$$\phi(x) = \begin{cases} 1 & if \quad \sum_{i=1}^{15} X_i \geq 21.88 \\ \\ 0 & if \quad \sum_{i=1}^{15} X_i < 21.88 \end{cases},$$

so by $\sum_{i=1}^{15} X_i = 38.33$ H_o is rejected and the corresponding p-value is :

$$p_v = p_{H_o}(\sum_{i=1}^{15} X_i > 38.33) = p_{H_o}(2\sum_{i=1}^{15} X_i > 76.66) < 0.005$$

4 - Hypothesis testing for parameters of gamma distribution

Let $X_1,\ldots,X_n$ be a random sample from gamma distribution with parameters α and β. We find $\hat{\alpha}$ numerically from $\frac{\Gamma'(\hat{\alpha})}{\Gamma(\hat{\alpha})} + \ln\frac{\overline{X}}{\hat{\alpha}} - \frac{1}{n}\sum_{i=1}^{n}\ln X_i = 0$ and $\hat{\beta} = \frac{\overline{X}}{\hat{\alpha}}$ to find MLE for α and β. Furthermore let $\theta = (\alpha, \beta)$, $\theta_o = (a, b)$, $\hat{\theta} = (\hat{\alpha}, \hat{\beta})$. Testing $H_o : \theta = \theta_o$ *versus* $H_1 : \theta \neq \theta_o$ is equivalent to testing $H_o : H(\theta) = H(\theta_o)$ *versus* $H_1 : H(\theta) \neq H(\theta_o)$. Via lemm2 and example1 we have

$$H(\hat{\theta}) = (\hat{\theta}_2 + 1) - \hat{\theta}_2(\frac{\Psi(\hat{\theta}_2 + 1)}{\Gamma(\hat{\theta}_2 + 1)}) + Ln\Gamma(\hat{\theta}_2 + 1) - Ln(-\hat{\theta}_1) = \hat{\alpha} - (\hat{\alpha} - 1)\frac{\Psi(\hat{\alpha})}{\Gamma(\hat{\alpha})} - Ln\frac{\Gamma(\hat{\alpha})}{\hat{\beta}}$$

and

$$H(\theta_o) = a - (a-1)\frac{\Psi(a)}{\Gamma(a)} - Ln\frac{\Gamma(a)}{b},$$

then the information matrix and array T is given by:

$$I_F(\theta_o) = \begin{bmatrix} ab^2 & b \\ b & \dfrac{\Psi'(a)\Gamma(a) - \Psi^2(a)}{\Gamma^2(a)} \end{bmatrix}$$

and

$$T = \left[t_1 = \frac{\partial H(\theta)}{\partial \theta_1}\Big|_{\theta_1=\theta_o} = b \quad t_2 = \frac{\partial H(\theta)}{\partial \theta_2}\Big|_{\theta_2=\theta_o} = 1 + \frac{\Psi'(a)\Gamma(a) - \Psi^2(a)}{\Gamma^2(a)} \right]$$ respectively.

Example5. Menendez (1999) gave an example for this distribution, here we give another example. The following data are a random sample of size 50 simulated from gamma (3.5,5) distribution.

19.93	29.4	12.02	6.83	8.92	56.12	23.53	42.06	13.29	16.19
9.77	18.46	7.67	9.57	38.07	18.15	5.59	28.2	39.18	7.49
9.88	8.5	12.93	3.95	6.92	21.72	11.26	14.13	21.76	6.77
22.38	24.69	47.24	8.36	13.31	12.76	24.65	43.53	15.2	25.19
6.85	13.82	15.91	9.94	19.76	24.15	27.3	9.61	17.89	23.42

We consider the hypothesis testing

$$H_o : \theta = (3.5,5) \quad \textit{versus} \quad H_1 : \theta \neq (3.5,5).$$

Note: We have written computer programs using MATHCAD2000 package to obtain the results.
From above note and example1we have:

$\hat{\alpha}$	$\hat{\beta}$	$\overline{X}$	$\sum_{i=1}^{50} \ln X_i$	$H(\hat{\theta})$	$H(\theta_o)$	$\sigma(\theta_o)$
2.873	6.503	18.684	137.216	2.515	1.151	24.73

The value of corresponding statistic is:

$$Z_n = \sqrt{n}\,\frac{H(\hat{\theta}) - H(\theta_o)}{\sigma(\theta_o)} = \sqrt{50}\,\frac{2.515 - 1.151}{4.97} = 1.94 \ngtr Z_{0.025} = 1.96,$$

then H_o isn't rejected and the corresponding p-value is:

$$p_v = p_{H_o}(|Z| > 1.94) = 2p_{H_o}(Z > 1.94) = 0.0524$$

5 - Hypothesis testing for parameters of beta distribution

In this section we use asymptotic distribution of Shannon's entropy for regular exponential model to testing statistical hypothesis for parameters of beta distributions. Let $X_1,...,X_n$ be a random sample of beta (α,β) distribution.

We solve: $$\frac{\Psi_\alpha(\alpha+\beta)}{\Gamma(\alpha+\beta)}-\frac{\Psi(\alpha)}{\Gamma(\alpha)}+\frac{1}{n}\sum_{i=1}^{n}\ln X_i=0$$

and $$\frac{\Psi_\beta(\alpha+\beta)}{\Gamma(\alpha+\beta)}-\frac{\Psi(\beta)}{\Gamma(\beta)}+\frac{1}{n}\sum_{i=1}^{n}\ln(1-X_i)=0$$

numerically to find MLE of α and β. Let $\theta=(\alpha,\beta)$, $\theta_o=(a,b)$, $\hat{\theta}=(\hat{\alpha},\hat{\beta})$. Consider: $H_o:\theta=\theta_o$ *versus* $H_1:\theta\neq\theta_o$, which is equivalent to: $H_o:H(\theta)=H(\theta_o)$ *versus* $H_1:H(\theta)\neq H(\theta_o)$, From example2 we have:

$$I_{11}=\frac{\partial^2 b(\theta)}{\partial\theta_1^{\,2}}\Big|_{\theta=\theta_o}=\frac{\Psi_a'(a)\Gamma(a)-\Psi_a^{\,2}(a)}{\Gamma^2(a)}-\frac{\Psi_a'(a+b)\Gamma(a+b)-\Psi_a^{\,2}(a+b)}{\Gamma^2(a+b)}$$

$$I_{12}=I_{21}=\frac{\partial^2 b(\theta)}{\partial\theta_1\partial\theta_2}\Big|_{\theta=\theta_o}=-\frac{\Psi_b'(a+b)\Gamma(a+b)-\Psi_b^{\,2}(a+b)}{\Gamma^2(a+b)}$$

$$I_{22}=\frac{\partial^2 b(\theta)}{\partial\theta_2^{\,2}}\Big|_{\theta=\theta_o}=\frac{\Psi'(b)\Gamma(b)-\Psi^2(b)}{\Gamma^2(b)}-\frac{\Psi_b'(a+b)\Gamma(a+b)-\Psi_b^{\,2}(a+b)}{\Gamma^2(a+b)}$$

$$H(\hat{\theta})=(1-\hat{\alpha})\frac{\Psi(\hat{\alpha})}{\Gamma(\hat{\alpha})}+(1-\hat{\beta})\frac{\Psi(\hat{\beta})}{\Gamma(\hat{\beta})}+(\hat{\alpha}-1)\frac{\Psi_{\hat{\alpha}}(\hat{\alpha}+\hat{\beta})}{\Gamma(\hat{\alpha}+\hat{\beta})}+(\hat{\beta}-1)\frac{\Psi_{\hat{\beta}}(\hat{\alpha}+\hat{\beta})}{\Gamma(\hat{\alpha}+\hat{\beta})}+Ln\frac{\Gamma(\hat{\alpha})\Gamma(\hat{\beta})}{\Gamma(\hat{\alpha}+\hat{\beta})}$$

$$H(\theta_o)=(1-a)\frac{\Psi(a)}{\Gamma(a)}+(1-b)\frac{\Psi(b)}{\Gamma(b)}+(a-1)\frac{\Psi_a(a+b)}{\Gamma(a+b)}+(b-1)\frac{\Psi_b(a+b)}{\Gamma(a+b)}+Ln\frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}$$

Then the information matrix and array T are $I_F(\theta_o)=\begin{bmatrix} I_{11} & I_{12}\\ I_{21} & I_{22}\end{bmatrix}$ and $T=[t_1,t_2]$ respectively, where:

$$t_1 = -\frac{\Psi(a)}{\Gamma(a)} + (1-a)\frac{\Psi'(a)\Gamma(a) - \Psi^2(a)}{\Gamma^2(a)} + \frac{\Psi_a(a+b)}{\Gamma(a+b)} + (a-1)\frac{\Psi'_a(a+b)\Gamma(a+b) - {\Psi_a}^2(a+b)}{\Gamma^2(a+b)}$$
$$+ (b-1)\frac{\Psi'_{ba}(a+b)\Gamma(a+b) + \Psi_a(a+b)\Psi_b(a+b)}{\Gamma^2(a+b)} + \frac{\Gamma(b)\Psi(a)\Gamma(a+b) - \Psi_a(a+b)\Gamma(a)\Gamma(b)}{\Gamma(a)\Gamma(b)\Gamma(a+b)}$$

$$t_2 = -\frac{\Psi(b)}{\Gamma(b)} + (1-b)\frac{\Psi'(b)\Gamma(b) - \Psi^2(b)}{\Gamma^2(b)} + \frac{\Psi_b(a+b)}{\Gamma(a+b)} + (b-1)\frac{\Psi'_b(a+b)\Gamma(a+b) - {\Psi_b}^2(a+b)}{\Gamma^2(a+b)}$$
$$+ (a-1)\frac{\Psi'_{ab}(a+b)\Gamma(a+b) + \Psi_a(a+b)\Psi_b(a+b)}{\Gamma^2(a+b)} + \frac{\Gamma(a)\Psi(b)\Gamma(a+b) - \Psi_b(a+b)\Gamma(a)\Gamma(b)}{\Gamma(a)\Gamma(b)\Gamma(a+b)}$$

Example6. The following data are a random sample of size 40 simulated from beta (3,4) distribution.

0.02	0.21	0.35	0.26	0.7	0.3	0.33	0.05	0.29	0.12
0.2	0.43	0.76	0.68	0.96	0.14	0.21	0.17	0.73	0.73
0.03	0.46	0.59	0.12	0.35	0.36	0.41	0.35	0.14	0.65
0.33	0.64	0.24	0.44	0.24	0.06	0.45	0.7	0.68	0.66

We consider the hypothesis testing

$$H_o : \theta = (1.5,2.2) \quad \textit{versus} \quad H_1 : \theta \neq (1.5,2.2)$$

Note: We have written computer programs using MATHCAD2000 package to obtain the results.

From above note and example2 we have $I_F(\theta_o) = \begin{bmatrix} 0.6247 & -0.31 \\ -0.31 & 0.2629 \end{bmatrix}$ and

$\hat{\alpha}$	$\hat{\beta}$	$\sum_{i=1}^{40} \ln(1-X_i)$	$\sum_{i=1}^{40} \ln X_i$	$H(\hat{\theta})$	$H(\theta_o)$	$\sigma(\theta_o)$
1.217	1.892	-24.568	-49.316	-0.113	-0.139	154.861

Then the value of corresponding statistic is:

$$Z_n = \sqrt{n}\,\frac{H(\hat{\theta}) - H(\theta_o)}{\sigma(\theta_o)} = \sqrt{40}\,\frac{-0.113 + 0.139}{12.444} = 0.014 \quad \textit{then} \quad |Z_n| \not> Z_{0.025} = 1.96 \,,$$

H_0 isn't rejected and the corresponding p-value is:

$$p_v = p_{H_o}(|Z_n| > 0.014) = 2p_{H_o}(Z_n > 0.014) = 0.984$$

6 - Confidence Interval

In this section we use asymptotic distribution of Shannon's entropy for regular exponential model to present a method of finding interval estimators for entropy $H(\theta)$ and for θ if entropy $H(\theta)$ is one to one.

Consider hypothesis testing $H_o : \theta = \theta_o$ *versus* $H_1 : \theta \neq \theta_o$ which is equivalent to:

$$H_o : H(\theta) = H(\theta_o) \quad versus \quad H_1 : H(\theta) \neq H(\theta_o)$$

Via lemma2 we have:

$p\left(|Z_n| = \left| \sqrt{n}\, \frac{H(\hat{\theta}) - H(\theta_o)}{\sigma(\theta_o)} \right| \leq Z_{\frac{\alpha}{2}} \right) = 1 - \frac{\alpha}{2}$, then the asymptotic confidence interval with confidence coefficient $(1-\alpha)$ for $H(\theta)$ is equal:

$$\left[H(\hat{\theta}) - Z_{\frac{\alpha}{2}} \frac{\sigma(\theta_o)}{\sqrt{n}} \quad , \quad H(\hat{\theta}) + Z_{\frac{\alpha}{2}} \frac{\sigma(\theta_o)}{\sqrt{n}} \right],$$

with $L = 2Z_{\frac{\alpha}{2}} \frac{\sigma(\theta_o)}{\sqrt{n}}$. Furthermore the minimum sample size giving an error at most ε at a confidence level $(1-\alpha)$ is $n = \left[\frac{1}{\varepsilon^2} \sigma^2(\theta_o) Z^2_{\frac{\alpha}{2}} \right] + 1$. If $H(\theta)$ is one to one then the asymptotic confidence interval with confidence coefficient $(1-\alpha)$ for θ is equal to:

$$\left[H^{-1}\left(H(\hat{\theta}) - Z_{\frac{\alpha}{2}} \frac{\sigma(\theta_o)}{\sqrt{n}} \right) \quad , \quad H^{-1}\left(H(\hat{\theta}) + Z_{\frac{\alpha}{2}} \frac{\sigma(\theta_o)}{\sqrt{n}} \right) \right]$$

In the similar way we can construct the asymptotic confidence interval with confidence coefficient $(1-\alpha)$ for $H(\theta_1) - H(\theta_2)$ as:

$$\left[\left(H(\hat{\theta}_1) - H(\hat{\theta}_2) \right) - Z_{\frac{\alpha}{2}} \sqrt{\frac{\sigma_1^2(\hat{\theta}_o)}{n_1} + \frac{\sigma_2^2(\hat{\theta}_o)}{n_2}} \quad , \quad \left(H(\hat{\theta}_1) - H(\hat{\theta}_2) \right) + Z_{\frac{\alpha}{2}} \sqrt{\frac{\sigma_1^2(\hat{\theta}_o)}{n_1} + \frac{\sigma_2^2(\hat{\theta}_o)}{n_2}} \right]$$

where $\sigma_j^2(\theta_o) = T_j' I_F^{-1}(\theta_o) T_j$ and $\hat{\theta}_o$ is MLE θ under H_o.

Furthermore for $n = n_1 = n_2$ the minimum sample size giving an error at most ε at a confidence level $(1-\alpha)$ is $n = \left[\frac{1}{\varepsilon^2} (\sigma_1^2(\hat{\theta}_o) + \sigma_2^2(\theta_o)) Z^2_{\frac{\alpha}{2}} \right] + 1$.

In example4 we had, $\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n} X_i^2 = 4.52$, $H(\hat{\theta}) = 2.172$, $H(\theta_o) = 2.112$ and $\sigma(\theta_o) = 0.129$, so asymptotic confidence interval with confidence coefficient %95 for $H(\theta)$ is $[1.919 \quad 2.425]$, then:

$$1.919 \le \ln(\frac{\pi e}{\theta}) \le 2.425 \qquad and \qquad 0.067 < \theta = \frac{1}{2\sigma^2} < 0.183$$

then $2.73 \le \sigma^2 \le 7.465$ with $L = 7.465 - 2.73 = 4.735$.

The classic confidence interval with confidence coefficient %95 for σ^2 is:

$$\frac{(n-1)S^2}{\chi^2_{0.025}} < \sigma^2 < \frac{(n-1)S^2}{\chi^2_{0.975}} \qquad then \qquad \frac{135.4}{45.7} = 2.96 < \sigma^2 < \frac{135.4}{16.047} = 8.44$$

with $L = 8.44 - 2.96 = 5.48$.

Example7. The following data are two independent random sample of size 10 and 15 were simulated from normal (0,9) and normal (0,25) distributions.

Sample from N(0,9)	0.3 , -1.5 , -3.6 , -2.4 , -4.5 , 1.5 , 2.1 , 2.4 , 0.9 , 4.5
Sample from N(0,25)	-10, 6 , -5.5 , 2 , -1 , 1 , 1.5 , 3 , -1.5 , -10.5 , -2 , -0.5 , -3.5 , 5 , 7

In this example we have

$$H(\theta) = \frac{1}{2}\ln(\frac{\pi e}{\theta}) \qquad , \hat{\theta}_o = \frac{1}{n_1 + n_2}\left(\sum_{i-1}^{n_1} X_{i1}^2 + \sum_{i-1}^{n_2} X_{i2}^2\right) = 18.452$$

$$t_K = -\frac{1}{2\theta_o} \quad , \quad I_F^j(\theta_o) = \frac{1}{2\theta_o^2} \quad , \quad \sigma_j^2(\theta_o) = \frac{1}{2} \quad , \quad \hat{\sigma}_j^2 = \frac{1}{n_j}\sum_{i-1}^{n_j} X_{ij}^2$$

The results for this example is summarized in the following table:

	N	$\hat{\theta}$	$H(\hat{\theta})$	$\sigma(\hat{\theta}_O)$
Sample(1)	10	0.067	2.425	0.5
Sample (2)	15	0.019	3.054	0.5

Then the asymptotic confidence interval with confidence coefficient %95 for $H(\theta_1) - H(\theta_2)$ is:

$$\left[(2.425 - 3.054) - 1.96\sqrt{\frac{0.5}{10} + \frac{0.5}{15}} \quad , \quad (2.425 - 3.054) + 1.96\sqrt{\frac{0.5}{10} + \frac{0.5}{15}}\right] = [-1.1948 \quad -0.0632]$$

7 - Conclusion

Some characterization results based on Shannon's entropy, hypothesis testing, confidence interval, by Shannon's entropy and comparison with other classical tests are discussed in the paper. The main contribution of this paper is application of Shannon's entropy in hypothesis testing. Furthermore we have considered some examples to throw some light on this new theory. As we expected the method of asymptotic distribution of $\left(H(\hat{\theta})-H(\hat{\theta}_0)\right)$ is more powerful than methods of classical form. An analytic proof of it seems to be difficult and for further research we appeal to simulation methods for justifications.

References

Lehmann, E.L (1997) *Testing statistical hypotheses*, Verlog New York Inc.

Lehmann, E.L. (1983) *Theory of point estimation*, John Wiley New York.

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

Cover,T.M., and Thomas, J.A. (1991) *Elements of information theory*, John Wiley New York

Casella, G., and Berger, R.L. (1990) *Statistical inference*, Brooks/cole, Publishing Company.

Pasha, E., and khodabin, M., and mohtashami borzadaran, G.R. (2004) *Entropy in exponential families*, journal of science, Azad university, science-branch, Iran-Tehran, to appear.