

Part I
Risk Methods for Bioinformatics

Generalized Information Criteria for the Best Logit Model

Christos P. Kitsos and Thomas L. Toulías

Abstract In this paper the γ -order Generalized Fisher's entropy type Information measure (γ -GFI) is adopted as a criterion for the selection of the best Logit model. Thus the appropriate Relative Risk model can be evaluated through an algorithm. The case of the entropy power is also discussed as such a criterion. Analysis of a real breast cancer data set is conducted to demonstrate the proposed algorithm, while algorithm's realizations, through MATLAB scripts, are cited in Appendix.

Keywords Fisher's entropy measure · Logit model · Relative Risk · Breast Cancer

1 Introduction

The two main lines of thought are adopted as far as the Fisher's information measure is concerned: The parametric approach and the entropy power [2]. In this paper we shall use the generalized form of the Fisher's entropy type information measure, as developed in Sect. 2, as well as a generalized form of the usual normal distribution. In Sect. 3 the binary response model is related to the developed theory of Sect. 2. An algorithm is proposed to choose the best binary response model for evaluating the Relative Risk. As a binary response case that demonstrates the algorithm the breast cancer problem is studied, see [16] among others.

The pioneering work of Jaynes [10] on the maximum entropy principle in Statistical Thermodynamics led to the adoption of this principle to other fields of interest. In the following, a compact form of various parametric and non-parametric information measures is discussed.

C.P. Kitsos (✉) · T.L. Toulías
Technological Educational Institute of Athens, Ag. Spyridonos
and Palikaridi Str., 12210 Egaleo, Athens, Greece
e-mail: xkitsos@teiath.gr

T.L. Toulías
e-mail: t.toulías@teiath.gr

© Springer International Publishing Switzerland 2015
C.P. Kitsos et al. (eds.), *Theory and Practice of Risk Assessment*,
Springer Proceedings in Mathematics & Statistics 136,
DOI 10.1007/978-3-319-18029-8_1

3

Let X be a random variable with probability density $f(x; \theta)$, with $\theta \in \Theta$ being the parameter vector from the parameter space $\Theta \subseteq \mathbb{R}^p$. Let

$$U(\theta) := \frac{\partial}{\partial \theta} \log f(x; \theta), \quad \theta \in \Theta \subseteq \mathbb{R}^p,$$

be the parametric score function. Then, the parametric information measure $I(\theta)$ can be defined as

$$I(\theta) := g(E[h(U(\theta))]), \quad \theta \in \Theta \subseteq \mathbb{R}^p,$$

where g and h being defined as real-valued functions, and $E[\cdot]$ denoting the expected value operator with respect to the parameter θ . For the univariate case, if $g := \text{id}$, the Fisher's information measure $I_F(\theta)$ is defined when $h(U) := U^2$, and the Vajda's information measure $I_V(\theta)$ when $h(U) := |U|^\lambda$, $\lambda \geq 1$. When $g(A) := A^k$, the Mathai's information measure $I_M(\theta)$ is obtained when $h(U) := |U|^\lambda$ and $k = 1/\lambda$, $\lambda \geq 1$, while the Boeke's information measure $I_B(\theta)$ is defined with $h(U) := |U|^{\lambda/(\lambda-1)}$ and $k = \lambda - 1$, $1 \neq \lambda > 0$. That is,

$$I(\theta) = \begin{cases} I_F(\theta), & g := \text{id}, \quad h(U) := U^2, \\ I_V(\theta), & g := \text{id}, \quad h(U) := |U|^\lambda, \quad \lambda \geq 1, \\ I_M(\theta), & g(A) := A^k, \quad h(U) := |U|^\lambda, \quad k = 1/\lambda, \quad \lambda \geq 1, \\ I_B(\theta), & g(A) := A^k, \quad h(U) := |U|^{\frac{\lambda}{\lambda-1}}, \quad k = \lambda - 1, \quad \lambda \in \mathbb{R}_+ \setminus 1. \end{cases} \quad (1)$$

Some of the merits of Fisher's information measure, $I_F(\theta)$, it remains invariant under orthogonal transformation, provides the well known Cramer-Rao lower bound, and plays an important role in optimal experimental design theory, see [8, 12, 23]. Therefore, $I(\theta)$ as in (1), defines the \mathfrak{F}_1 family of information measures.

There are two main problems in applications concerning Fisher's $I_F(\theta)$ measure: is the measure singular or ill-conditioned? (see also [23]).

The elements of the non-parametric family \mathfrak{F}_2 of information measures are defined, through two given distributions $f_i = \frac{dP_i}{d\mu}$, $P_i \ll \mu$, $i = 1, 2$ with μ a σ -finite measure, that is

$$\mathfrak{F}_2 := \left\{ I(f_1, f_2) : I(f_1, f_2) := g \left(\int h(f_1, f_2) d\mu \right) \right\}. \quad (2)$$

Some known information measures (i.m.), or divergences, such as the Kullback-Leibler i.m. $I_{KL}(f_1, f_2)$, the Vajda's i.m. $I_V(f_1, f_2)$, the Kagan i.m. $I_K(f_1, f_2)$, the Csiszar i.m. $I_C(f_1, f_2)$, the Matusita i.m. $I_M(f_1, f_2)$, as well as the Rényi's divergence $I_R(f_1, f_2)$, are defined as follows

$$I(f_1, f_2) = \begin{cases} I_{KL}(f_1, f_2), & g(A) := A, & h(f_1, f_2) := f_1 \log(f_1/f_2), \\ I_V(f_1, f_2), & g(A) := A, & h(f_1, f_2) := f_1 |1 - (f_2/f_1)|^\lambda, \lambda \geq 1, \\ I_K(f_1, f_2), & g(A) := A, & h(f_1, f_2) := f_1 |1 - (f_2/f_1)|^2, \\ I_C(f_1, f_2), & g(A) := A, & h(f_1, f_2) := f_2 \phi(f_1/f_2), \phi \text{ convex}, \\ I_M(f_1, f_2), & g(A) := \sqrt{A}, & h(f_1, f_2) := (\sqrt{f_1} - \sqrt{f_2})^2, \\ I_R(f_1, f_2), & g(A) := \frac{\log A}{1-\lambda}, & h(f_1, f_2) := f_1^\lambda f_2^{1-\lambda}, 1 \neq \lambda > 0. \end{cases} \quad (3)$$

We can obtain the corresponding parametric information measures from the non-parametric ones, through the following general scheme, [6, 22]:

$$I(\theta) = \lim_{\Delta\theta \rightarrow 0} \inf_{\Delta\theta} \left\{ \frac{1}{\Delta\theta^2} I(f(x; \theta), f(x; \theta + \Delta\theta)) \right\}. \quad (4)$$

Then, for the univariate case and under certain regularity conditions [2], it can be proved that the parametric K–L information measure $I_{KL}(\theta)$, i.e. (4) with $I(\cdot, \cdot)$ being the K–L measure $I_{KL}(\cdot, \cdot)$ as in (3), is the half of the Fisher's $I_F(\theta)$ as in in (1), and $2/\lambda$ of Reyni's $I_R(\theta)$, or

$$I_{KL}(\theta) = \frac{1}{2} I_F(\theta) = \frac{2}{\lambda} I_R(\theta).$$

The two afore mentioned families of information measures, \mathfrak{F}_1 and \mathfrak{F}_2 are considered for the parametric case. Some of these measures attract special interest in ecological studies, see [1]. In the next Section we present the entropy type information measures.

2 Entropy Type Information Measures

Now as far as the entropy type information measures are concerned, notice that the well known Fisher's entropy type information measure $J(X)$ of a p -variate random variable is given by

$$J(X) = \int_{\mathbb{R}^p} [\nabla f(x)] [\nabla \log f(x)] dx = \int_{\mathbb{R}^p} f(x) \|\nabla \log f(x)\|^2 dx. \quad (5)$$

Recall that the Shannon entropy H of a r.v. X is defined as, [2],

$$H(X) := - \int_{\mathbb{R}^p} f(x) \log f(x) dx, \quad (6)$$

while the corresponding entropy power is given by

$$\mathbf{N}(X) = \nu e^{\frac{2}{p}\mathbf{H}(X)}, \quad (7)$$

with $\nu = (2\pi e)^{-1}$, see [2] for details.

Introducing an extra parameter, say δ , Kitsos and Tavoularis in [15] defined the Generalized (entropy type) Fisher's Information measure (δ -GFI), J_δ as follows:

$$J_\delta(X) := \int_{\mathbb{R}^p} f(x) \|\nabla \log f(x)\|^\delta dx. \quad (8)$$

For parameter value $\delta = 2$ we get the known Fisher's information, i.e. $J_2(X) = J(X)$. The extension of the entropy power, the Generalized Entropy Power (δ -GEP) is defined for $\delta \in \mathbb{R} \setminus [0, 1]$, as

$$\mathbf{N}_\delta(X) := \nu_\delta e^{\frac{\delta}{p}\mathbf{H}(X)}, \quad (9)$$

where

$$\nu_\delta := \left(\frac{\delta-1}{\delta e}\right)^{\delta-1} \pi^{-\delta/2} \left[\frac{\Gamma(\frac{p}{2} + 1)}{\Gamma(p\frac{\delta-1}{\delta} + 1)} \right]^{\delta/p}, \quad \delta \in \mathbb{R} \setminus [0, 1], \quad (10)$$

see [15]. Trivially, when $\delta = 2$, (9) is reduced to the known entropy power $\mathbf{N}(X)$, i.e. $\mathbf{N}_2(X) = \mathbf{N}(X)$ as $\nu_2 = \nu$. Moreover, it can be shown [15] that

$$J_\delta(X) \mathbf{N}_\delta(X) \geq p, \quad (11)$$

with p being the number of the involved parameters, i.e. $\Theta \subseteq \mathbb{R}^p$. Therefore, $J_\delta(X) \approx p/\mathbf{N}_\delta(X)$.

The above extensions give rise to a generalization of the multivariate normal distribution. This new distribution plays the same role as the classical normal distribution for the Fisher's entropy type information, and we shall call it the γ -order Generalized Normal Distribution (γ -GND).

Recall the p -dimensional random variable X_γ is said to follow the γ -GND, denoted by $X_\gamma \sim \mathcal{N}_\gamma^p(\mu, \Sigma)$, with mean vector $\mu \in \mathbb{R}^{p \times 1}$ and positive definite scale matrix $\Sigma \in \mathbb{R}^{p \times p}$, when the density function, f_{X_γ} , is of the form, see [15, 17],

$$f_{X_\gamma}(x; \mu, \Sigma) := C_\gamma^p(\Sigma) \exp \left\{ -\frac{\gamma-1}{\gamma} Q(x)^{\frac{\gamma}{2(\gamma-1)}} \right\}, \quad x \in \mathbb{R}^{p \times 1}, \quad (12)$$

with quadratic form $Q(x) := (x - \mu)^T \Sigma^{-1} (x - \mu)$ and the normality factor $C_\gamma^p(\Sigma)$ defined as

$$C_\gamma^p(\Sigma) := \pi^{-p/2} \frac{\Gamma(\frac{p}{2} + 1)}{\Gamma\left(p\frac{\gamma-1}{\gamma} + 1\right)} \left(\frac{\gamma-1}{\gamma}\right)^{p\frac{\gamma-1}{\gamma}} |\det \Sigma|^{-1/2}. \quad (13)$$

Notice that the 2–GND coincides with the usual multivariate (elliptically contoured) Normal distribution, i.e. $\mathcal{N}_2^p(\mu, \Sigma) = \mathcal{N}^p(\mu, \Sigma)$, while the 1–GND and the $(\pm\infty)$ –GND reduced in limit to the multivariate (elliptically contoured) Laplace and Uniform distributions respectively, i.e. $\mathcal{N}_1^p(\mu, \Sigma) = \mathcal{U}^p(\mu, \Sigma)$ and $\mathcal{N}_{\pm\infty}^p(\mu, \Sigma) = \mathcal{L}^p(\mu, \Sigma)$. Moreover, for dimensions $p = 1, 2$ the 0–GND is reduced to the degenerate Dirac distribution, i.e. $\mathcal{N}_0^p(\mu, \Sigma) = \mathcal{D}^p(\mu)$. See [21] for details.

Proposition 1 *The Shannon entropy of a random variable $X_\gamma \sim \mathcal{N}_\gamma^p(\mu, \Sigma)$, is of the form*

$$H(X_\gamma) = p \frac{\gamma-1}{\gamma} - \log C_\gamma^p(\Sigma). \quad (14)$$

Proof Consider the p.d.f. f_{X_γ} as in (12). From the definition (6) we have that the Shannon entropy of X is

$$H(X_\gamma) = -\log C_\gamma^p(\Sigma) + C_\gamma^p(\Sigma) \frac{\gamma-1}{\gamma} \int_{\mathbb{R}^p} Q(x)^{\frac{\gamma}{2(\gamma-1)}} \exp \left\{ -\frac{\gamma-1}{\gamma} Q(x)^{\frac{\gamma}{2(\gamma-1)}} \right\} dx.$$

Applying the linear transformation $z = (x - \mu)^T \Sigma^{-1/2}$ with $dx = d(x - \mu) = \sqrt{|\det \Sigma|} dz$, the $H(X_\gamma)$ above is reduced to

$$H(X_\gamma) = -\log C_\gamma^p(\Sigma) + C_\gamma^p(\mathbb{I}_p) \frac{\gamma-1}{\gamma} \int_{\mathbb{R}^p} \|z\|^{\frac{\gamma}{\gamma-1}} \exp \left\{ -\frac{\gamma-1}{\gamma} \|z\|^{\frac{\gamma}{\gamma-1}} \right\} dz,$$

where \mathbb{I}_p denotes the $p \times p$ identity matrix. Switching to hyperspherical coordinates, we get

$$H(X_\gamma) = -\log C_\gamma^p(\Sigma) + C_\gamma^p(\mathbb{I}_p) \frac{\gamma-1}{\gamma} \omega_{p-1} \int_{\mathbb{R}_+} \rho^{\frac{\gamma}{\gamma-1}} \exp \left\{ -\frac{\gamma-1}{\gamma} \rho^{\frac{\gamma}{\gamma-1}} \right\} \rho^{p-1} d\rho,$$

where $\omega_{p-1} := 2\pi^{p/2} / \Gamma(\frac{p}{2})$ is the volume of the $(p-1)$ -sphere. Applying the variable change $du := d(\frac{\gamma-1}{\gamma} \rho^{\gamma/(\gamma-1)}) = \rho^{1/(\gamma-1)} d\rho$ we obtain successively

$$\begin{aligned} H(X_\gamma) &= -\log C_\gamma^p(\Sigma) + C_\gamma^p(\mathbb{I}_p) \omega_{p-1} \int_{\mathbb{R}_+} u e^{-u} \rho^{\frac{(p-1)(\gamma-1)-1}{\gamma-1}} du \\ &= -\log C_\gamma^p(\Sigma) + p \frac{\gamma-1}{\gamma} \Gamma(p \frac{\gamma-1}{\gamma}) C_\gamma^p(\mathbb{I}_p) \omega_{p-1}. \end{aligned}$$

Finally, by substitution of the volume ω_{p-1} and the normalizing factors $C_\gamma^p(\Sigma)$ and $C_\gamma^p(\mathbb{I}_p)$ and as in (13), relation (14) is obtained.

Example 1 Substituting (14) into (9), the δ -GEP is then

$$N_\delta(X_\gamma) = \left(\frac{\delta-1}{e\delta}\right)^{\delta-1} \left(\frac{e\gamma}{\gamma-1}\right)^{\delta\frac{\gamma-1}{\gamma}} \left[\frac{\Gamma\left(p\frac{\gamma-1}{\gamma} + 1\right)}{\Gamma\left(p\frac{\delta-1}{\delta} + 1\right)} \right]^{\delta/p} |\det \Sigma|^{\frac{\delta}{2p}}. \quad (15)$$

Moreover, the generalized Fisher's entropy type information measure $J_\delta(X_\gamma)$ with X_γ spherically contoured, i.e. $X_\gamma \sim \mathcal{N}_\gamma(\mu, \sigma^2 \mathbb{I}_p)$, is given by the formula, [20],

$$J_\delta(X_\gamma) = \left(\frac{\gamma}{\gamma-1}\right)^{\frac{\delta}{\gamma}} \frac{\Gamma\left(\frac{\delta+p(\gamma-1)}{\gamma}\right)}{\sigma^\delta \Gamma\left(p\frac{\gamma-1}{\gamma}\right)}. \quad (16)$$

Example 2 For the usual entropy power of the γ -GND, i.e. for the second-GEP of the r.v. $X_\gamma \sim \mathcal{N}_\gamma(\mu, \Sigma)$, we have that

$$N(X_\gamma) = \frac{1}{2e} \left(\frac{e\gamma}{\gamma-1}\right)^{2\frac{\gamma-1}{\gamma}} \left[\frac{\Gamma\left(p\frac{\gamma-1}{\gamma} + 1\right)}{\Gamma\left(\frac{p}{2} + 1\right)} \right]^{2/p} |\det \Sigma|^{1/p}.$$

Note that for the limiting cases of X_1 (1-GND) and $X_{\pm\infty}$ ($\pm\infty$ -GND) we obtain the usual entropy power for the multivariate (and elliptically contoured) Uniform and Laplace distributions respectively, i.e.

$$N(X_1) = \lim_{\gamma \rightarrow 1^+} N(X_\gamma) = \frac{|\det \Sigma|^{1/p}}{2e\Gamma^{2/p}\left(\frac{p}{2} + 1\right)},$$

$$N(X_{\pm\infty}) = \lim_{\frac{\gamma-1}{\gamma} \rightarrow 1^+} N(X_\gamma) = 2^{\frac{2-p}{p}} e \left[\frac{(p-1)! \sqrt{|\det \Sigma|}}{\Gamma\left(\frac{p}{2}\right)} \right]^{2/p}.$$

Finally, for the r.v. X_2 we obtain the usual entropy power for the multivariate Normal, i.e. $N(X_2) = \sqrt[2]{|\det \Sigma|}$ (Fig. 1).

Table 1 provides an evaluation of $N_\delta(X_\gamma)$ with $X_\gamma \sim \mathcal{N}_\gamma^1(0, 1)$ for certain γ and $\delta \geq 1$ values.

Example 3 We make the following observations here. When $\delta = \gamma$, from (15), we have that

$$N_\gamma(X_\gamma) = |\det \Sigma|^{\frac{\gamma}{2p}}. \quad (17)$$

Thus, $N_0(X_0) = 1$, i.e. the 0-GEP of the Dirac distributed, $X_0 \sim \mathcal{D}(0)$ is 1 while $N_\delta(X_0) = +\infty$ for every defined $\delta \in \mathbb{R} \setminus [0, 1]$ as it is derived through (15).

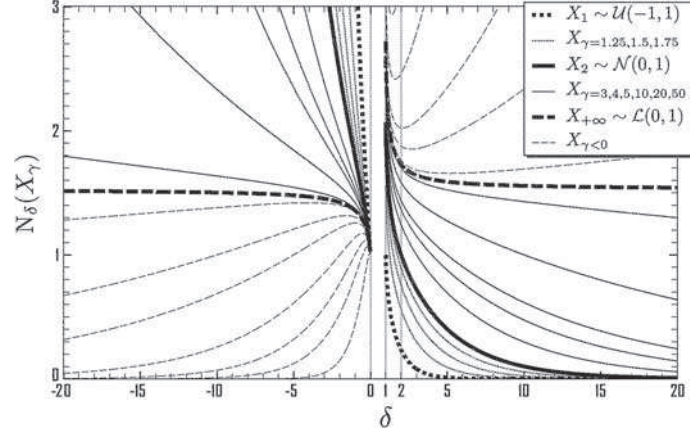


Fig. 1 Graphs of $N_\delta(X_\gamma)$ along δ for various γ values, where $X_\gamma \sim \mathcal{N}(0, \sigma^2)$ with $\sigma = 0.8, 1, 1.5$

Table 1 Evaluation of $N_\delta(X_\gamma)$ with $X_\gamma \sim \mathcal{N}_\gamma^1(0, 1)$ for various γ and $\delta \geq 1$ parameters

γ/δ	1	3/2	2	3	5	10	50	$+\infty$
-50	2.7412	1.8834	1.7598	1.684	1.6566	1.6912	2.3295	$+\infty$
-10	2.8309	1.9766	1.8769	1.8549	1.9461	2.3340	11.660	$+\infty$
-5	2.9393	2.0912	2.0233	2.0761	2.3482	3.3981	76.283	$+\infty$
-2	3.2430	2.4235	2.463	2.7884	3.8392	9.0834	10411.	$+\infty$
-1	3.6945	2.9469	3.1967	4.1230	7.3677	33.452	7.05e+6	$+\infty$
-1/10	8.3767	10.0610	16.434	48.057	441.49	1.2e+5	$\approx +\infty$	$+\infty$
1	1.0	0.4150	0.2342	0.0818	0.0107	7.06e-5	2.95e-22	0.0
3/2	1.7974	1.0	0.7566	0.4748	0.2008	0.0248	1.59e-9	0.0
2	2.0664	1.2327	1.0	0.7214	0.4032	0.1002	1.74e-6	0.0
3	2.3040	1.4513	1.2433	1.0	0.6950	0.2977	0.0004	0.0
5	2.4779	1.6187	1.438	1.2440	1.0	0.6163	0.0149	0.0
10	2.6009	1.7406	1.5842	1.4384	1.2739	1.0	0.1684	0.0
50	2.6952	1.8362	1.7012	1.6007	1.5223	1.4281	1.0	0.0
$\pm\infty$	2.7183	1.8598	1.7305	1.6422	1.5886	1.5552	1.5317	1.0

Moreover, $N_1(X_1) = |\det \Sigma|^{1/(2p)}$ while $N_1(X_{\pm\infty}) = e|\det \Sigma|^{1/(2p)}$. For the Laplace distributed $X_{\pm\infty}$, (17) implies

$$N_{\pm\infty}(X_{\pm\infty}) = \begin{cases} 0, & |\det \Sigma| < 1, \\ 1, & |\det \Sigma| = 1, \\ +\infty, & |\det \Sigma| > 1. \end{cases}$$

See Appendix 1 for the minimum/maximum analysis of the generalized entropy power N_δ of a γ -GND random variable.

3 Generalized Fisher's Information and Relative Risk

Consider a subject with attributes given by the input vector $X = (X_1, X_2, \dots, X_p)^T$. In risk analysis, the focus is on the parameter $p(x)$, i.e. the probability that this subject has a certain characteristic C , given that the input vector takes on the real vector value x , i.e. $X = x$, and measures the odds ratio or the Relative Risk (RR):

$$RR = \frac{p(x)}{1 - p(x)} \quad \text{with} \quad \log \frac{p(x)}{1 - p(x)} = x^T \beta, \quad (18)$$

where β being an appropriate vector of regression parameters, see also [11, 13].

Due to the Central Limit Theorem, the involved Binomial distribution $\mathcal{B}(n, P)$, $P = p(x)$, corresponding to the binary response model under investigation, approximated by the Normal distribution, i.e.

$$\mathcal{B}(n, P) \approx \mathcal{N}(nP, nP(1 - P)) = \mathcal{N}_2^1(nP, nP(1 - P)).$$

For the normally distributed $X \sim \mathcal{N}(\mu, \sigma^2) := \mathcal{N}(nP, nP(1 - P))$, the Shannon entropy is

$$H(X) = \frac{1}{2} + \log \sqrt{2\pi nP(1 - P)},$$

while the entropy power is

$$N(X) = \frac{1}{2\pi e} e^{2H(X)} = \frac{1}{2\pi e} e^{1 + \log\{2\pi nP(1 - P)\}} = nP(1 - P) := \sigma^2.$$

The generalized entropy power $N_\delta(X)$ introduced in (9), for this case is

$$N_\delta(X) = \left(\frac{\pi}{2}\right)^{\frac{\delta}{2}} e^{1 - \frac{\delta}{2}} \left(\frac{\delta - 1}{\delta}\right)^{\delta - 1} \Gamma^{-\delta} \left(\frac{\delta - 1}{\delta} + 1\right) [nP(1 - P)]^\delta. \quad (19)$$

Given the above discussion we propose the following algorithm for the examination of the optimum variable entering the logit model, based on the methodology of [11, 18], while for the maximum entropy see [22]. That is, for each k -risk variable model the maximum δ -GFI models (with respect to the parameter δ) is chosen, and among them, we obtain the one with the minimum γ -GFI value. The steps of the proposed algorithm for a bioassay, [7] are presented as follows: