Hindawi Publishing Corporation
Research Letters in Signal Processing
Volume 2008, Article ID 790607, 5 pages
doi:10.1155/2008/790607

Research Letter

Generalized Cumulative Residual Entropy for Distributions with Unrestricted Supports

Noomane Drissi, Thierry Chonavel, and Jean Marc Boucher

Lab-STICC (CNRS FRE 3167), Institut Telecom, Telecom Bretagne, Technopole Brest Iroise, CS 83818, 29238 Brest Cédex, France

Correspondence should be addressed to Noomane Drissi, noomane.drissi@telecom-bretagne.eu

Received 6 April 2008; Accepted 19 June 2008

Recommended by Andreas Jakobsson

We consider the cumulative residual entropy (CRE) a recently introduced measure of entropy. While in previous works distributions with positive support are considered, we generalize the definition of CRE to the case of distributions with general support. We show that several interesting properties of the earlier CRE remain valid and supply further properties and insight to problems such as maximum CRE power moment problems. In addition, we show that this generalized CRE can be used as an alternative to differential entropy to derive information-based optimization criteria for system identification purpose.

1. INTRODUCTION

The concept of entropy is important for studies in many areas of engineering such as thermodynamics, mechanics, or digital communications. An early definition of a measure of the entropy is the Shannon entropy [1, 2]. In Shannon's approach, discrete values and absolutely continuous distributions are treated in a somewhat different way through entropy and differential entropy, respectively. Considering the complementary cumulative distribution function (CCDF) instead of the probability density function in the definition of differential entropy leads to a new entropy measure named cumulative residual entropy (CRE) [3, 4]. In [3, 4], CRE is defined as

$$\mathcal{E}(X) = -\int_{\mathbb{R}_+^n} P(|X| > u) \log P(|X| > u)\, du, \qquad (1)$$

where n is the dimension of the random vector X. Clearly, this formula is valid both for a discrete or an absolutely continuous random variable (RV), or with both a discrete and an absolutely continuous part, because it resorts to the CCDF of $|X|$. In addition, unlike Shannon differential entropy it is always positive, while preserving many interesting properties of Shannon entropy. The concept of CRE has found nice interpretations and applications in the fields of reliability (see [5] where the concept of dynamic CRE is introduced) and images alignment [3].

Shannon entropy can be seen as a particular case of exponential entropy, when entropy order tends to 1. Thus, following the work in [4], a modified version of the exponential entropy, where PDF is replaced by CCDF, has been introduced in [6], leading to new entropy-type measures, called survival entropies.

However, both Rao et al.'s CRE and its exponential entropy generalization by Zografos and Nadarajah lead to entropy-type definitions that assume either positive valued RVs or apply to $|X|$ otherwise. Although the positive case is of great interest for many applications, CRE and exponential entropies entail difficulties when working with RVs with supports that are not restricted to positive values.

In this paper, we show that for an RV X, (1) remains a valid expression when $P(|X| > u)$ is replaced by $P(X > u)$ and integration is performed over $\mathbb{R}^n$, without further hypothesis than in [4]. In addition, some desirable properties are enabled by this CRE definition extension. We also complete the power moment constrained maximum CRE distributions problem that was adressed in [7], for classes of distributions that have lower-unbounded supports. Finally, we illustrate the potential superiority of the proposed generalized CRE (GCRE) against differential entropy in mutual information-based estimation problems.

The paper is organized as follows. Section 2 introduces the GCRE definition. Some properties of GCRE are discussed in Section 3. In Section 4, we introduce cumulative entropy rate and mutual information rate. Section 5 deals with maximum GCRE distributions. With a view to illustrate the potentiality of GCRE, in Section 6, we show on a simple example a possible benefit of GCRE for systems identification.

2. GENERALIZED CUMULATIVE RESIDUAL ENTROPY (GCRE)

We will denote by $F_X^c(x)$ the complementary cumulative distribution function (survival function) of a multivariate RV $X = [X_1, \ldots, X_n]^T$ of dimension n: $F_X^c(X) = P(X > x) = P(X_i > x_i,\ i = 1, \ldots, n)$. We denote by $H_C(X)$ the GCRE of X that we define by

$$H_C(X) = -\int_{\mathbb{R}^n} F_X^c(u) \log F_X^c(u) du. \tag{2}$$

Clearly, like the CRE, the GCRE is a positive and concave function of F_X^c. In addition, existence of GCRE can be established without further assumption upon distribution than those assumed for the CRE in [4].

Theorem 1. *$H_C(X) < \infty$ if for some $p > n$, $E[|X|^p] < \infty$.*

Proof. First let us remark that from the proof of the existence of CRE in [4], it is sufficient to prove the result when X is a scalar RV, that is $n = 1$, and for $p > 1$. Then, letting $p^{-1} < \alpha < 1$, we use the following inequality:

$$\begin{aligned} -F_X^c(x) \log F_X^c(x) \le{}& \frac{e^{-1}}{1-\alpha}\left[F_X^c(x)\right]^{\alpha} 1\mathrm{I}_{]0,\infty[}(x) \\ &+ \left(1 - F_X^c(x)\right) 1\mathrm{I}_{]-\infty,0]}(x), \end{aligned} \tag{3}$$

where $1\mathrm{I}_A(x) = 1$ if $x \in A$ and $1\mathrm{I}_A(x) = 0$ otherwise. The existence of $\int_{\mathbb{R}} (e^{-1}/(1-\alpha))[F_X^c(x)]^{\alpha} 1\mathrm{I}_{[0,\infty[}(x)dx$ can be proven just as in [4]. Now, letting $u = -x$, we have

$$\begin{aligned} &\left(1 - F_X^c(x)\right) 1\mathrm{I}_{]-\infty,0]}(x) \\ &\quad = F_X(x) 1\mathrm{I}_{]-\infty,0]}(x) \\ &\quad \le 1\mathrm{I}_{[-1,0[}(t) + F_X(x) 1\mathrm{I}_{]-\infty,-1[}(x) \\ &\quad \le 1\mathrm{I}_{]0,1]}(u) + F_X(-u) 1\mathrm{I}_{]1,\infty]}(u) \\ &\quad \le 1\mathrm{I}_{]0,1]}(u) + F_{-X}(u) 1\mathrm{I}_{]1,\infty]}(u) \\ &\quad \le 1\mathrm{I}_{]0,1]}(u) + F_{|-X|}^c(u) 1\mathrm{I}_{]1,\infty]}(u) \\ &\quad \le 1\mathrm{I}_{]0,1]}(u) + u^{-p} E[|X|^p] 1\mathrm{I}_{]1,\infty]}(u). \end{aligned} \tag{4}$$

Thus,

$$\begin{aligned} &\int_{\mathbb{R}} \left(1 - F_X^c(x)\right) 1\mathrm{I}_{]-\infty,0]}(x) dx \\ &\quad \le \int_{\mathbb{R}} \left(1\mathrm{I}_{]0,1]}(u) + u^{-p} E[|X|^p] 1\mathrm{I}_{]1,\infty]}(u)\right) du \\ &\quad \le 1 + \int_1^{\infty} u^{-p} E[|X|^p] du \\ &\quad < \infty. \end{aligned} \tag{5}$$

Finally, putting all pieces together one finally proves convergence of right-hand side of (2). □

3. A FEW PROPERTIES OF GCRE

Let us now exhibit a few more interesting properties of the GCRE. First, it is easy to check that like Shannon entropy the GCRE remains constant with respect to variable translation:

$$\forall a \in \mathbb{R}^n, \quad H_C(X + a) = H_C(X). \tag{6}$$

In the same way, it is clear that

$$\forall a \in \mathbb{R}_+, \quad H_C(aX) = aH_C(X). \tag{7}$$

When $a < 0$, we do not have such a nice property. However, let us consider the important particular case where the distribution of X has a symmetry of the form

$$\exists \mu,\ \forall x, \quad F_X^c(\mu - x) = 1 - F_X^c(\mu + x). \tag{8}$$

In this case, we get the following result.

Theorem 2. *For an RV X that satisfies symmetry property* (8), *one has*

$$\forall a \in \mathbb{R}, \quad H_C(aX) = |a| H_C(X). \tag{9}$$

Proof. Since it is clear that for all $a \in \mathbb{R}_+$, $H_C(aX) = aH_C(X)$, we just have to check that $H_C(-X) = H_C(X)$, which can be established as follows:

$$\begin{aligned} -H_C(-X) &= \int_{\mathbb{R}} F_{-X}^c(x) \log F_{-X}^c(x) dx \\ &= \int_{\mathbb{R}} F_{-X}^c(x - \mu) \log F_{-X}^c(x - \mu) dx \\ &= \int_{\mathbb{R}} F_X(-x + \mu) \log F_X(-x + \mu) dx \\ &= \int_{\mathbb{R}} \left(1 - F_X(x + \mu)\right) \log\left(1 - F_X(x + \mu)\right) dx \\ &= \int_{\mathbb{R}} F_X^c(x + \mu) \log F_X^c(x + \mu) dx \\ &= \int_{\mathbb{R}} F_X^c(x) \log F_X^c(x) dx \\ &= -H_C(X). \end{aligned} \tag{10}$$

□

When the entries of vector X are independent, it has been shown in [4] that if the X_i are nonnegative, then

$$H_C(X) = \sum_i \left(\Pi_{j \ne i} \mathbb{E}[X_j]\right) H_C(X_i). \tag{11}$$

However, this formula does not extend to RVs with distributions carried by $\mathbb{R}^n$ because $F^C(X)$ can be integrated over $\mathbb{R}_+^n$ in general but never over $\mathbb{R}^n$. However, if the X_is are independent and have lower bounded supports with respective lower bounds $m_1, \dots, m_n$,

$$\begin{aligned} H_C(X) &= \int_{\Pi_i]m_i,\infty]} -F_X^c(x)\log F_X^c(x)dx \\ &= \sum_i \big(\Pi_{j\neq i}(\mathbb{E}[X_j]-m_j)\big)H_C(X_i), \end{aligned} \tag{12}$$

because

$$\begin{aligned} \int_{m_i}^{\infty} F_{X_i}^c(u)du &= \big[uF_{X_i}^c(u)\big]_{m_i}^{\infty} + \int_{m_i}^{\infty} uP_{X_i}(du) \\ &= -m_i + \mathbb{E}[X_i]. \end{aligned} \tag{13}$$

Conditional GCRE definition is a direct extension of the definition of conditional CRE: the conditional GCRE of X knowing that Y is equal to y is defined by

$$\begin{aligned} &H_C(X \mid Y = y) \\ &\quad = \int_{\mathbb{R}^n} -P(X > x \mid Y = y)\log P(X > x \mid Y = y)dx. \end{aligned} \tag{14}$$

We recall an important result from [4] that states that conditioning reduces the entropy.

Theorem 3. *For any X and Y,*

$$H_C(X \mid Y) \leq H_C(X) \tag{15}$$

equality holds if and only if X is independent of Y.

As a consequence, if $X \to Y \to Z$ is a Markov chain, we have the data processing inequality for GCRE:

$$H_C(Z \mid X, Y) \leq H_C(Z \mid X). \tag{16}$$

4. ENTROPY AND MUTUAL INFORMATION RATES

4.1. Entropy rate

The GCRE of a stochastic process $\{X_i\}$ is defined by

$$H_C(X) = \lim_{n\to\infty} H_C(X_n \mid X_{n-1}, X_{n-2}, \dots, X_1) \tag{17}$$

when the limit exists.

Theorem 4. *For stationary processes, the limit exists.*

Proof. Consider

$$\begin{aligned} H_C(X_{n+1} \mid X_n, \dots, X_1) &\leq H_C(X_{n+1} \mid X_n, \dots, X_2) \\ &\leq H_C(X_n \mid X_{n-1}, \dots, X_1). \end{aligned} \tag{18}$$

The first line follows from the fact that conditioning reduces entropy and the second follows from the stationarity (see [2] for the equivalent proof in the case of Shannon entropy). □

4.2. Mutual information

Let X and Y be two RVs. We define the cumulative mutual information between X and Y as follows:

$$I_C(X; Y) = H_C(X) - H_C(X \mid Y). \tag{19}$$

Theorem 5. *I_C is nonnegative and it vanishes if and only if X and Y are independent.*

Proof. It is clear that I_C is nonnegative because of Theorem 3. □

For a random vector $\mathbf{X} = (X_1, X_2, \dots, X_n)$ of size n, mutual information is defined by

$$I_C(\mathbf{X}) = \sum_{i=1}^{n} H_C(X_i) - H_C(X_n \mid X_{n-1}, \dots, X_1). \tag{20}$$

In the case of stochastic processes $\{X_i\}$, we have $H_C(X) = \lim_{n\to\infty} H_C(X_n|X_{n-1}, \dots, X_1)$ and the limit exists for stationary processes. Then the mutual information rate for $\{X_i\}$ is defined as

$$I_C(X) = \lim_{T\to\infty} \frac{1}{T}\sum_{t=1}^{T} H_C(X_t) - H_C(X), \tag{21}$$

where $H_C(X_t)$ is the marginal GCRE of the process X.

5. MAXIMUM GCRE DISTRIBUTIONS

In this section, we only consider the case of one-dimensional RVs ($n = 1$). Maximum entropy principle is useful in many scientific areas and most important distributions can be derived from it [8]. The maximum CRE distribution has been studied in [7]. For an RV X with a symmetric CCDF in the sense of (8), we are looking for the maximum GCRE distribution, that is, the CCDF that solves the following moment problem:

$$\begin{gathered} \max_{F^c} H_C(F^c) \\ \int_{\mathbb{R}} r_i(x)p(x)dx = c_i, \quad i = 1, \dots, m, \end{gathered} \tag{22}$$

where $p(x) = -(d/dx)F^c(x)$, $(r_i)_{i=1,m}$, and $(c_i)_{i=1,m}$ are fixed C^1 real valued functions and real coefficients, respectively. The solution of this problem is supplied by the following result.

Theorem 6. *When the symmetry property* (8) *holds, the solution of problem* (22)*, when it can be reached, is of the form*

$$F^c(x) = \frac{1}{1+\exp\big[\sum_1^m \lambda_i r_i'(x-\mu)\big]} \quad (x \geq \mu). \tag{23}$$

Proof.

$$
\begin{aligned}
-H_C(X) &= \int_{-\infty}^{\mu} F^c(x)\log\left(F^c(x)\right)dx \\
&\quad + \int_{\mu}^{\infty} F^c(x)\log\left(F^c(x)\right)dx \\
&= \int_{\mathbb{R}_-} F^c(x+\mu)\log\left(F^c(x+\mu)\right)dx \\
&\quad + \int_{\mathbb{R}_+} F^c(x+\mu)\log\left(F^c(x+\mu)\right)dx \\
&= \int_{\mathbb{R}_-} \left(1-F^c(\mu-x)\right)\log\left(1-F^c(\mu-x)\right)dx \\
&\quad + \int_{\mathbb{R}_+} F^c(x+\mu)\log\left(F^c(x+\mu)\right)dx \\
&= \int_{\mathbb{R}_+} \left[\left(1-F^c(\mu+x)\right)\log\left(1-F^c(\mu+x)\right)\right. \\
&\qquad \left. + F^c(\mu+x)\log\left(F^c(\mu+x)\right)\right]dx.
\end{aligned}
\tag{24}
$$

Let us define $\widetilde{f}$ by

$$
\begin{aligned}
&\widetilde{f}\left(F^c(x), -p\right) \\
&\quad = -\left(1-F^c(\mu+x)\right)\log\left(1-F^c(\mu+x)\right) \\
&\qquad - F^c(\mu+x)\log\left(F^c(\mu+x)\right) + \sum_1^m \lambda_i\left(-p(x)\right)r_i(x).
\end{aligned}
\tag{25}
$$

Then, since $(F^c(x))' = -p(x)$, the Euler-Lagrange equation [9] states that the solution F^c of problem (22) is a solution of equation

$$\frac{d}{dx}\widetilde{f}_{-p}(x) = \widetilde{f}_{F^c}(x), \tag{26}$$

where $\widetilde{f}_u$ is the partial derivative of $\widetilde{f}$ with respect to component u. From (25), we get

$$
\begin{aligned}
\frac{d}{dx}\widetilde{f}_{-p}(x) &= \sum_1^m \lambda_i r_i'(x), \\
\widetilde{f}_{F^c}(x) &= \log\left(1-F^c(\mu+x)\right) - \log F^c(\mu+x).
\end{aligned}
\tag{27}
$$

Then,

$$
\begin{aligned}
\log\frac{1-F^c(\mu+x)}{F^c(\mu+x)} &= \sum_1^m \lambda_i r_i'(x), \\
F^c(x) &= \frac{1}{1+\exp\left(\sum_1^m \lambda_i r_i'(x-\mu)\right)},
\end{aligned}
\tag{28}
$$

for $x \in [\mu, \infty[$. □

5.1. Example

We set the constraints $\mathbb{E}[X] = \mu$ and $\mathbb{E}[X^2] = \sigma^2$. Then the maximum GCRE symmetric solution for the CCDF of X is given by

$$F_X^c(x) = \frac{1}{1+\exp\left(\lambda_1 + 2\lambda_2 x\right)}, \tag{29}$$

for $x > 0$, which is the CCDF of a logistic distribution. The moment constraints lead to $\lambda_1 + 2\lambda_2 x = (\sigma\sqrt{3}/\pi)(x-\mu)$. The corresponding PDF is defined on $\mathbb{R}$ by

$$p_X(x) = \frac{\sigma\sqrt{3}}{\pi}\frac{\exp\left(-(\sigma\sqrt{3}/\pi)(x-\mu)\right)}{\left(1+\exp\left(-(\sigma\sqrt{3}/\pi)(x-\mu)\right)\right)^2}. \tag{30}$$

5.2. Positive random variables

It has been shown in [7] that the maximum CRE (i.e., the maximum GCRE under additional nonnegative constraint) distribution has CCDF in the form

$$F^C(x) = \exp\left(-\sum_{i=1}^m \lambda_i r_i'(x)\right), \tag{31}$$

for $x \in [0, \infty[$. In [7], this result is derived from the log-sum inequality, but of course it can also be derived from the Euler-Lagrange equation along the same lines as in the proof of Theorem 6.

With a positive support constraint and under first and second moment constraints, it comes that the optimum CCDF is of the form $F^c(x) = \exp(-\lambda_1 - 2\lambda_2 x)$, for $x > 0$. Thus the solution, if it exists, is an exponential distribution. In fact, the first and second power moment constraints must be such that $\mathbb{E}[X^2] = 2(\mathbb{E}[X])^2$, otherwise the problem has no exact solution.

6. SIMULATION RESULTS

With a view to emphasize the potential practical interest of GCRE, we consider a simple system identification problem. Here, we consider an $MA(1)$ process, denoted by $Y = (Y_n)_{n\in\mathbb{Z}}$, generated by a white noise $X = (X_n)_{n\in\mathbb{Z}}$ and corrupted by a white noise N:

$$Y_n = X_n - aX_{n-1} + N_n. \tag{32}$$

The model input X and output Y are observed and the system model ($MA(1)$) is assumed to be known. We want to estimate the coefficient a without prior knowledge upon the distributions of X and Y. Thus, we resort to mutual information (MI) to estimate a as the coefficient α such that RVs $\widehat{Y}_n^\alpha = X_n - \alpha X_{n-1}$ and Y_n show the highest dependence. Shannon MI between $\widehat{Y}_n^\alpha$ and Y_n is given by $f_S(\alpha) = I_S(\widehat{Y}_n^\alpha, Y_n) = H_S(\widehat{Y}_n^\alpha) - H_S(\widehat{Y}_n | Y_n)$, where H_S is Shannon differential entropy. Similarly, for GCRE, MI will be defined as $f_C(\alpha) = I_C(\widehat{Y}_n^\alpha, Y) = H_C(\widehat{Y}_n^\alpha) - H_C(\widehat{Y}_n^\alpha | Y_n)$. We compare estimation performance for a by maximizing both $f_S(\alpha)$ and $f_C(\alpha)$. Since true values of $f_S(\alpha)$ and $f_C(\alpha)$ are not available, they are estimated from empirical distributions of $(Y_n, \widehat{Y}_n^\alpha)$.

For simulations, we have chosen X Gaussian and N with a Laplace distribution: $p_N(x) = (\lambda/2)\exp(-\lambda|x|)$. We consider an experiment with $a = 0.5$ and noise variance equal to 0.2. Estimation is carried out from observation of $(X_n, Y_n)_{n=1,400}$. Here, optimization of MIs is realized on a fixed regular grid of 200 points over interval [0,1]. Estimation performance is calculated from 200 successive

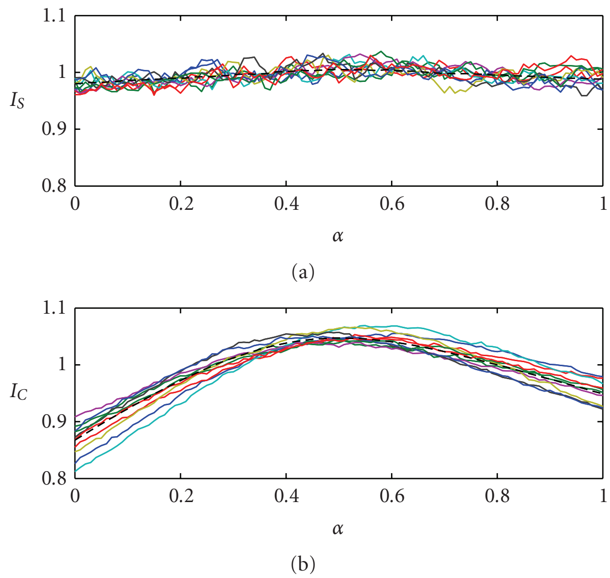

FIGURE 1: Ten estimates of (a) $f_S(\alpha)$ and (b) $f_C(\alpha)$. Dotted line: mean estimate averaged from 200 realizations.

experiments. Estimation of a from Shannon MI leads to bias and standard deviation that are equal to 0.032 and 0.18, respectively, while they are equal to 0.004 and 0.06, respectively, for GCRE MI.

More important, we see on Figure 1(a) that Shannon MI estimates are much more irregular than GCRE MI (Figure 1(b)) estimates because of smoothing brought by density integration in the calculation of CCDF. This difference is important since the use of an iterative local optimization technique would have failed in general to find Shannon's estimated MI global optimum, because of its many local maxima.

Of course, this drawback can be partly solved by kernel smoothing of the empirical distribution of $(Y_n, \widehat{Y}_n^\alpha)$, for instance by using the method proposed in [10]. However, we have checked that, for the above example, very strong smoothing is necessary and then bias and variance performance remain worse than with GCRE MI estimator.

7. CONCLUSION

We have shown that the concept of cumulative residual entropy (CRE) introduced in [3, 4] can be extended to distributions with general supports. Generalized CRE (GCRE) shares many nice features of CRE. We also pointed out specific properties of GCRE such as its maximum, moment constrained, and distribution and we have illustrated practical interest of GCRE by showing how it can be used in system identification procedures.

REFERENCES

[1] C. E. Shannon, "A mathematical theory of communication," *Bell System Technical Journal*, vol. 27, pp. 379–423, 1948.

[2] J. M. Cover and J. H. Thomas, *Elements of Information Theory*, Wiley Interscience, New York, NY, USA, 2006.

[3] F. Wang, B. C. Vemuri, M. Rao, and Y. Chen, "A new & robust information theoretic measure and its application to image alignment," in *Proceedings of the 18th International Conference on Information Processing in Medical Imaging (IPMI '03)*, vol. 2732 of *Lecture Notes in Computer Science*, pp. 388–400, Springer, Ambleside, UK, July 2003.

[4] M. Rao, Y. Chen, B. C. Vemuri, and F. Wang, "Cumulative residual entropy: a new measure of information," *IEEE Transactions on Information Theory*, vol. 50, no. 6, pp. 1220–1228, 2004.

[5] M. Asadi and Y. Zohrevand, "On the dynamic cumulative residual entropy," *Journal of Statistical Planning and Inference*, vol. 137, no. 6, pp. 1931–1941, 2007.

[6] K. Zografos and S. Nadarajah, "Survival exponential entropies," *IEEE Transactions on Information Theory*, vol. 51, no. 3, pp. 1239–1246, 2005.

[7] M. Rao, "More on a new concept of entropy and information," *Journal of Theoretical Probability*, vol. 18, no. 4, pp. 967–981, 2005.

[8] J. N. Kapur, *Maximum Entropy Methods in Science and Engineering*, John Wiley & Sons, New York, NY, USA, 1989.

[9] J. L. Troutman, *Variational Calculus and Optimal Control, Optimization with Elementary Convexity*, Springer, New York, NY, USA, 2nd edition, 1996.

[10] D. T. Pham, "Fast algorithm for estimating mutual information, entropies and score functions," in *Proceedings of the 4th International Symposium on Independent Component Analysis and Blind Signal Separation (ICA '03)*, pp. 17–22, Nara, Japan, April 2003.