Bayes' estimators of generalized entropies

Bayes' estimators of generalized entropies

D Holste[†], I Große[‡] and H Herzel[§][‖]

[†] Institute of Physics, Humboldt-University Berlin, Invalidenstrasse 110, D-10115 Berlin, Germany

[‡] Center for Polymer Studies and Department of Physics, Boston University, Boston, MA 02215, USA

[§] Institute for Theoretical Biology, Humboldt-University Berlin, Invalidenstrasse 43, D-10115 Berlin, Germany

Received 6 November 1997

Abstract. The order-q Tsallis (H_q) and Rényi entropy (K_q) receive broad applications in the statistical analysis of complex phenomena. A generic problem arises, however, when these entropies need to be estimated from observed data. The finite size of data sets can lead to serious systematic and statistical errors in numerical estimates. In this paper, we focus upon the problem of estimating generalized entropies from finite samples and derive the Bayes estimator of the order-q Tsallis entropy, including the order-1 (i.e. the Shannon) entropy, under the assumption of a uniform prior probability density. The Bayes estimator yields, in general, the smallest mean-quadratic deviation from the true parameter as compared with any other estimator. Exploiting the functional relationship between H_q and K_q, we use the Bayes estimator of H_q to estimate the Rényi entropy K_q. We compare these novel estimators with the frequency-count estimators for H_q and K_q. We find by numerical simulations that the Bayes estimator reduces statistical errors of order-q entropy estimates for Bernoulli as well as for higher-order Markov processes derived from the complete genome of the prokaryote *Haemophilus influenzae*.

1. Introduction

Building on the works of Shannon [1] and Khinchin [2], generalized entropies have witnessed an increasing interest in their application to characterize complex behaviour in models and real systems. As the Shannon entropy is formally defined as an average value, the idea underlying a generalization is to replace the average logarithms by an average of powers. Then this gives rise to the order-q Tsallis entropy H_q [3, 4] or, similarly, the Rényi entropy K_q [5]. The external parameter q applies to describe inhomogeneous structures of the probability distribution and hence the associated process under consideration. From both order-q entropies, H_q and K_q, the Shannon entropy is obtained in the limit $q \to 1$. Applications of order-q entropies occur in a variety of fields of sciences such as, e.g. nonlinear dynamical systems [6–10], statistical thermodynamics [11–16], classical mechanics [17], or evolutionary programming [18, 19].

Here we address the estimation of these entropies from a finite set of experimental data. Under the assumption of a stationary process generating the data, the data set is composed out of N data points chosen from M possible different outcomes. The problem that arises when entropies are to be estimated from these finite data sets is that the probabilities are *a priori* unknown. Naively replacing these probabilities by the sampled relative frequencies

[‖] Author to whom correspondence should be addressed.

0305-4470/98/112551+16$19.50

produces large statistical and systematic deviations of estimates from the true value [20, 21]. This problem becomes serious when the number of data points N is in the order of magnitude of the number of different states M, which occurs in many practical applications, for example in the estimations of correlations and dimensions. In such cases, the choice of an estimator with small deviations from the true value becomes important. Several different estimators have thus been developed, mainly devoted to the estimation of the Shannon entropy [22–27]. Specific estimators for the Rényi entropy and for the dimensions associated to them have also been derived, as well as for upper bounds on entropy estimates [24, 27].

While one can, in principle, calculate the systematic errors arising from frequency counts, less fluctuating entropy estimates can only be obtained by employing a different entropy estimator. The estimator which possesses the optimal property to minimize the mean-quadratic deviation of the estimate from the true value, subject to a certain prior assumption, is customarily referred to as the Bayes estimator. In this work, we derive the Bayes estimator of the Tsallis entropy and discuss its statistical properties. We then exploit this Bayes estimator to measure the Rényi entropy.

This paper is organized as follows. In section 2 we introduce the generalized entropy concept, starting from the definition of the canonical (Shannon) entropy, which we extend to non-logarithmic (order-q) averages. In section 3 we turn to the problem of estimating functions of probability distributions from a finite, discrete data set and introduce the Bayes estimator of a statistic. In section 4 of this work, we derive the Bayes estimator of the order-q Tsallis entropy under the prior assumption of a uniform prior probability density. We discuss properties of this estimator of H_q and contrast the result obtained with the frequency-count estimator. Using the functional relationship connecting the order-q Tsallis entropy, H_q, with the Rényi entropy, K_q, we propose a method on how to extract K_q from a data set in section 5. In section 6 we apply the Bayes estimators to numerically compute order-2 entropies of Markov processes with zero- and five-step memories. Our concluding remarks are given in section 7. Consigned to appendix A we present some analytic results about systematic errors (i.e. the bias) of the entropy estimators, which complete this work.

2. Generalized entropies

This section is aimed at introducing the notation used throughout this work as well as giving the definitions of the order-q Tsallis entropy, H_q, and the Rényi entropy, K_q. We then review some basic properties of these entropies, which will finally allow us the derivation of an indirect Bayes estimator of the Rényi entropy in section 5.

Consider a random variable A that can take on M different discrete values a_i, $i = 1, \ldots, M$, with an associated probability vector $\boldsymbol{p} \equiv \{p_1, \ldots, p_M\}$ with components $p_i \equiv p(a_i)$. The probabilities satisfy the two constraints $0 \leqslant p_i \leqslant 1$ and $\sum_{i=1}^{M} p_i = 1$. It is customary to refer to the set of all possible outcomes as the alphabet $\mathcal{A}$ with cardinality M. Then the Shannon entropy of A is defined as

$$H(A) = -\sum_{i=1}^{M} p_i \log_2 p_i \equiv -\langle \log_2 p_i \rangle. \tag{1}$$

Since the base of the logarithm is chosen to be 2, the Shannon entropy is measured in units of bits. One distinctive property of H, which is not shared by the generalized entropies, is worth mentioning: the entropy of a composite event can be given as the sum of the marginal and the conditional entropy.

By equation (1), events having either a particularly high or low occurrence do not contribute much to the Shannon entropy. In order to weight particular regions of the probability vector $\boldsymbol{p}$, one can consider the following partition function:

$$Z_q(A) = \sum_{i=1}^{M} p_i^q \equiv \langle p_i^{q-1} \rangle. \tag{2}$$

In contrast to equation (1), the average logarithm is now replaced by an average of powers of q. Clearly, a change of the order q will change the relative weights of how the event i contributes to the sum. Therefore, varying the parameter q allows us to monitor the inhomogeneous structure of the distribution $\boldsymbol{p}$: the larger q, the more heavily the larger probabilities enter into Z_q, and vice versa. Obviously, Z_0 equals the number of events i with non-vanishing probability, and Z_1 introduces normalization. Then the order-q Tsallis entropy is defined as

$$H_q(A) = \frac{1}{\ln 2} \frac{Z_q(A) - 1}{1 - q} \equiv \frac{1}{\ln 2} \frac{\langle p_i^{q-1} - 1 \rangle}{1 - q}. \tag{3}$$

Since the prefactor is chosen to be $1/\ln 2$, the Tsallis entropy is measured in units of bits. This can be seen by considering the limit $q \to 1$: we easily verify that $\lim_{q \to 1} H_q = H$ holds.

The order-q entropy due to Rényi is given by

$$K_q(A) = \frac{1}{1 - q} \log_2 Z_q(A) \equiv -\log_2 \langle p_i^{q-1} \rangle^{1/(q-1)}. \tag{4}$$

Here the argument of the logarithm is the generalized average of the numbers p_i. By equation (3), the relationship connecting both order-q entropies reads as

$$K_q(A) = \frac{1}{1 - q} \log_2 [1 + (1 - q) \ln 2 \; H_q(A)]. \tag{5}$$

From equation (5) we see that for fixed q, K_q and H_q are monotonic functions of one another and that $\lim_{q \to 1} K_q = H$ holds.

Let us summarize the following features of order-q entropies.

(i) $H_q \geqslant 0$ and $K_q \geqslant 0$. For given M, the global maxima (minima) are attained at $p_i = 1/M \; \forall i$ for $q > 0$ ($q < 0$). In particular, we have that $K_q^{\max} = H^{\max}$.

(ii) H_q and K_q are monotonically decreasing functions of q for arbitrary probability vectors $\boldsymbol{p}$: $H_q \geqslant H_{q'}$ and $K_q \geqslant K_{q'}$ for $q < q'$.

(iii) $H_q(A)$ is a concave (convex) function of the probabilities given $q > 0$ ($q < 0$). The curvature dependence of K_q upon q and $\boldsymbol{p}$ is non-trivial [4]. Yet the following two inequalities hold: K_q is a convex (concave) function of p_i for $q < 0$ ($0 < q \leqslant 1$).

(iv) Considering two subsets, $\mathcal{A}$ and $\mathcal{B}$, then $K_q(A, B)$ obeys additivity for independent random variables, whereas $H_q(A, B)$ is pseudo-additive. That is, we find $H_q(A, B) = H_q(A) + H_q(B) + (1 - q) H_q(A) H_q(B)$. Furthermore, $H_q(A, B)$ generalizes the Shannon-additivity to the order q (see, e.g. [1] or [4] for a definition and discussion).

By the above properties, the whole set of order-q entropies (which generalize the Shannon entropy) provides us with a whole spectrum of entropies, in which $q = 1$ is singled-out by the property of composite events. In the light of the fact that K_q is indeed additive but, in general, not a concave (convex) function of the probabilities p_i on the entire simplex, it is remarkable that via the nonlinear transformation (5) we are able to switch between two types of entropies of order q, either having the property of additivity or of well-defined concavity (convexity).

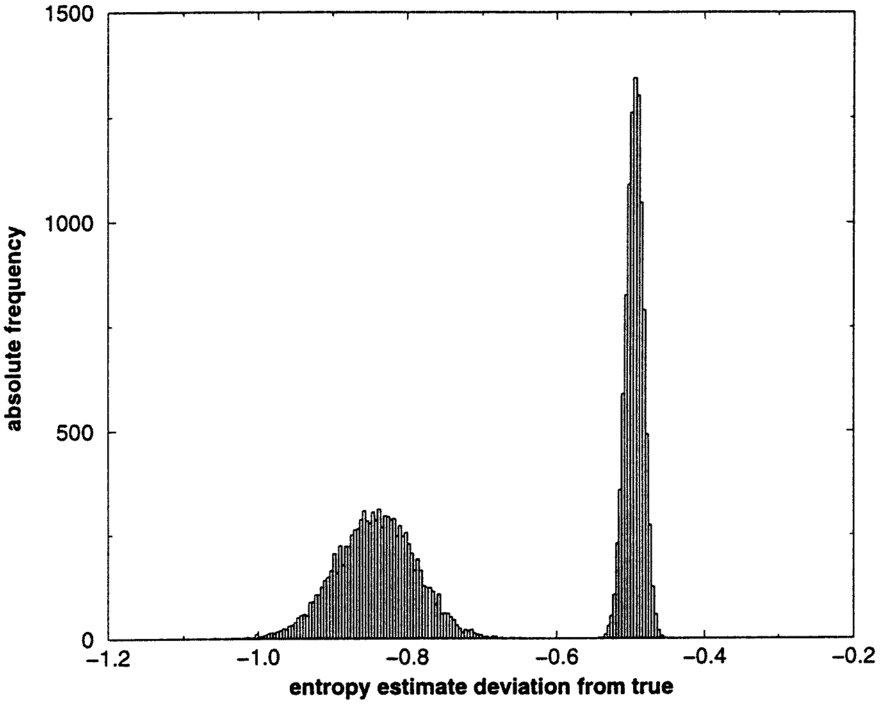

Figure 1. Comparison of the Bayes and the frequency-count estimator of the Shannon entropy, $\widehat{H_1}$ and $\bar{H}_1$ respectively. We generate an ensemble of 10 000 sequences, each of which composed of $N = 250$ data points chosen from an alphabet with cardinality $M = 256$. The 256 possible outcomes were samples from a uniform distribution, $p_i = 1/M$. From each such sequence, the entropy is estimated by the Bayes estimator and the frequency-count estimator. This figure displays the corresponding histograms of $\widehat{H_1}$ (right) and $\bar{H}_1$ (left), measured in units of bits per symbol. It can be seen that the variance of $\widehat{H_1}$ is about one order of magnitude smaller in comparison with the variance of $\bar{H}_1$. Note that the significant negative bias can, in principle, be approximated by length correction formulae. Therefore, it is the smaller variance of $\widehat{H_1}$ that makes this estimator superior.

3. Bayes' estimators

The generic problem in estimating entropies from finite realizations is that the probabilities p_i remain hidden to the observer. For a given ensemble of data sets one estimator can on average come close to the true value for some probability distribution, while a change in the distribution may favour another estimator. In particular, if the cardinality of $\mathcal{A}$ is in the order of the number of data points, then fluctuations increase and estimates usually become significantly biased. By bias we denote the deviation of the expectation value of the estimator from the true value. Thus, the problem in estimating functions of probability distributions is twofold: we seek to construct an estimator whose estimates both fluctuate with the smallest possible variance and are least biased.

The Bayes estimator has the optimal property of minimizing the mean-quadratic deviation [28–30]. This feature is illustrated in figure 1, by displaying the distribution of estimates of the order-1 Tsallis (i.e. the Shannon) entropy for two estimators: the Bayes estimator $\widehat{H_1}$ and the frequency-count estimator $\bar{H}_1$ with $M = 256$ and $N = 250$. For this investigation, the Bayes estimator is given in (17). Defining the relative frequencies to be $f_i = N_i/N$, where N_i is the number of observations of the symbol i, the frequency-count estimator reads as $\bar{H}_1 = -\sum_{i=1}^{M} f_i \log_2 f_i$. Inspecting the width of the variances of our estimates reveals the superiority of the Bayes estimator: fluctuations of its estimates are significantly suppressed, as compared with the fluctuations of the frequency-count estimates. However, let us point out yet another feature, namely that there is still a substantial bias affecting both estimates. An approach to approximate (and hence correct) the entropy bias of $\widehat{H_q}$ will be given in appendix A.

4. The Tsallis entropy estimator

In this section we focus upon the first task stated in the preceding section, by deriving the Bayes estimator of the generalized Tsallis entropy H_q. The total number of symbols available in a sample for the estimation is given by $N = \sum_{i=1}^{M} N_i$. Let further $P(\boldsymbol{N}|\boldsymbol{p}) = N![\prod_{i=1}^{M} p_i^{N_i}/N_i!]$ be the underlying conditional probability distribution to obtain the (multinomially distributed) observable vector $\boldsymbol{N}$ with components N_i. Finally, $Q(\boldsymbol{p})$ denotes the prior probability density of the probability vector $\boldsymbol{p}$. It satisfies the constraint $\int_{\mathcal{S}} \mathrm{d}\boldsymbol{p}\, Q(\boldsymbol{p}) = 1$ where the integration extends over the whole simplex $\mathcal{S} \equiv \{\boldsymbol{p}|\forall i\ \ p_i \geqslant 0, \sum_{i=1}^{M} p_i = 1\}$. Then the Bayes estimator of H_q reads as

$$\widehat{H_q}(\boldsymbol{N}) = \frac{1}{W(\boldsymbol{N})} \int_{\mathcal{S}} \mathrm{d}\boldsymbol{p}\, H_q(\boldsymbol{p}) P(\boldsymbol{N}|\boldsymbol{p}) Q(\boldsymbol{p}) \tag{6}$$

where the normalization constant is given by

$$W(\boldsymbol{N}) = \int_{\mathcal{S}} \mathrm{d}\boldsymbol{p}\, P(\boldsymbol{N}|\boldsymbol{p}) Q(\boldsymbol{p}). \tag{7}$$

According to Bayes' theorem, $P(\boldsymbol{p}|\boldsymbol{N}) = P(\boldsymbol{N}|\boldsymbol{p})Q(\boldsymbol{p})/Q(\boldsymbol{N})$. Thus, equation (6) is equivalent to $\widehat{H_q}(\boldsymbol{N}) = \int_{\mathcal{S}} \mathrm{d}\boldsymbol{p}\ H_q(\boldsymbol{p}) P(\boldsymbol{p}|\boldsymbol{N})$, which is the average of $H_q(\boldsymbol{p})$ over the posterior distribution $P(\boldsymbol{p}|\boldsymbol{N})$.

In what follows, we will derive the Bayes estimator of H_q under the assumption of a uniform prior probability density $Q(\boldsymbol{p}) = \text{constant}$. That is to say, we regard all possible probability vectors $\boldsymbol{p} \in \mathcal{S}$ to be relevant.

If we write down the Bayes estimator of H_q as

$$\widehat{H_q}(\boldsymbol{N}) = \frac{1}{\ln 2} \frac{1}{1-q} [\widehat{Z_q}(\boldsymbol{N}) - 1] \qquad \text{with } \widehat{Z_q}(\boldsymbol{N}) = \frac{1}{W(\boldsymbol{N})} \int_{\mathcal{S}} \mathrm{d}\boldsymbol{p} \sum_{i=1}^{M} p_i^q P(\boldsymbol{N}|\boldsymbol{p}) \tag{8}$$

then it can be seen that the derivation of $\widehat{H_q}$ reduces to the derivation of the Bayes estimator of the partition function Z_q. The normalization constant W and the quantity W' (see later) will be evaluated in appendix B. Interchanging the integral with the finite sum, $\widehat{Z_q}$ may be cast into the form

$$\widehat{Z_q}(\boldsymbol{N}) = \frac{\Gamma(N+M)}{\prod_{j=1}^{M} \Gamma(N_j+1)} \times \left\{ \sum_{i=1}^{M} \int_{\mathcal{S}} \prod_{j=1}^{M} \mathrm{d}p_j\, p_j^{(N_j + \delta_{ij} q)} \right\}. \tag{9}$$

Integrating over $M-1$ of the M components, we obtain

$$\widehat{Z_q}(\boldsymbol{N}) = \frac{\Gamma(N+M)}{\prod_{j=1}^{M} \Gamma(N_j+1)} \times \left\{ \sum_{i=1}^{M} \int_{p_i=0}^{1} \mathrm{d}p_i\, W'(p_i; \boldsymbol{N}) p_i^q \right\}. \tag{10}$$

Evaluating the remaining integral, we arrive at

$$\widehat{Z_q}(\boldsymbol{N}) = \frac{\Gamma(N+M)}{\Gamma(N+M+q)} \times \left[\sum_{i=1}^{M} \frac{\Gamma(N_i+1+q)}{\Gamma(N_i+1)}\right] \tag{11}$$

and thus, composing all above expressions, we eventually obtain

$$\widehat{H_q}(\boldsymbol{N}) = \frac{1}{\ln 2}\frac{1}{1-q}\left\{\frac{\Gamma(N+M)}{\Gamma(N+M+q)} \times \left[\sum_{i=1}^{M} \frac{\Gamma(N_i+1+q)}{\Gamma(N_i+1)}\right] - 1\right\}. \tag{12}$$

Expression (12) constitutes a central result of this work: the Bayes estimator of the Tsallis entropy of order q. To illustrate the differences between the fluctuations of the Bayes estimator and the frequency-count estimator of H_q, in the following we will simplify expression (12) for the special cases $q = 1$ and $q = 2$. The motivation for this parameter choice stems from the following. We recall that H_q is indeed a generalization of H, providing upper and lower bounds for the Shannon entropy. As such, we wish to make contact with the Bayes estimator $\hat{H}$ for the Shannon entropy. This is realized in the limit $q \to 1$. The second example, $q = 2$, plays an important role in the statistical analysis of nonlinear dynamical systems. Here $q = 2$ gives rise to quantities such as the correlation dimension and the second-order Kolmogorov entropy (see, e.g. [31] and references therein) as well as a generalization of the mutual information which preserves positivity [32].

To obtain $\widehat{H_1}$, we introduce the auxiliary function

$$F(q) = \left[\sum_{i=1}^{M} \frac{\Gamma(N_i+1+q]}{\Gamma(N_i+1)}\right] \Big/ \Gamma(N+M+q). \tag{13}$$

This will become useful due to the necessary consideration of the limit $q \to 1$, since expression (12) is not defined otherwise. Introducing $\alpha_q = N_i + 1 + q$ and $\beta_q = N + M + q$, we may write at the limit point

$$\widehat{H_1}(\boldsymbol{N}) = \lim_{q\to 1} \hat{H}_q(\boldsymbol{N}) = -\frac{\Gamma(\beta_0)}{\ln 2}\frac{\partial F(q)}{\partial q}\bigg|_{q=1} \tag{14}$$

where

$$\frac{\partial F(q)}{\partial q} = \sum_{i=1}^{M}\left\{\frac{\Gamma(\alpha_q)}{\Gamma(\alpha_0)\Gamma(\beta_q)}(\psi^{(1)}(\alpha_q) - \psi^{(1)}(\beta_q))\right\}. \tag{15}$$

Here $\psi^{(1)}(z) = \mathrm{d}\ln\Gamma(z)/\mathrm{d}z$ is the Digamma function. Since α_1 and β_1 are integers, we may express $\psi^{(1)}(z)$ in terms of the finite harmonic sum $\psi^{(1)}(z) = \sum_{l=1}^{z-1} 1/l - E_c$, with $E_c = \lim_{R\to\infty}(\sum_{r=1}^{R} 1/r - \ln R)$ being Euler's constant. Inserting this expression into equation (15), we obtain

$$\frac{\partial F(1)}{\partial q} = -\sum_{i=1}^{M}\left\{\frac{\Gamma(\alpha_1)}{\Gamma(\alpha_0)\Gamma(\beta_1)}\left(\sum_{l=\alpha_1}^{\beta_0}\frac{1}{l}\right)\right\}. \tag{16}$$

Hence we arrive at

$$\widehat{H_1}(\boldsymbol{N}) = \frac{1}{\ln 2}\left[\sum_{i=1}^{M}\frac{N_i+1}{N+M}\left(\sum_{l=N_i+2}^{N+M}\frac{1}{l}\right)\right]. \tag{17}$$

Equation (17) defines the Bayes estimator of the order-1 Tsallis entropy under a uniform prior probability density. Comparing the above expression with results derived in [29, 33], we verify the consistency of expression (12) in the limit $q \to 1$. That is, the Bayes estimator of the order-1 Tsallis entropy is identical to the Bayes estimator of the Shannon entropy: $\widehat{H_1} \equiv \hat{H}$.

We now turn to the case $q = 2$. From equation (11) we can read off the Bayes estimator of p_i^q to be

$$\widehat{p_i^q} = \frac{\Gamma(N+M)}{\Gamma(N+M+q)} \times \frac{\Gamma(N_i+1+q)}{\Gamma(N_i+1)}. \tag{18}$$

Thus, we write down $\widehat{H_2}$ in the form

$$\widehat{H_2}(\boldsymbol{N}) = \frac{1}{\ln 2}\left(1 - \sum_{i=1}^{M} \widehat{p_i^2}\right) \qquad \text{with } \widehat{p_i^2} = \left(\frac{N_i+1}{N+M}\right)\left(\frac{N_i+2}{N+M+1}\right). \tag{19}$$

In general, we find the following characteristics of the Bayes estimator to be noteworthy.

(i) $\widehat{H_q}$ is defined in the parameter range $q \in (-1, \infty)$. Apparently, cases of particular interest (and simplicity) are given when q takes on integer values $n \in \mathcal{N}$ (set of non-negative integer numbers) which allow one to replace Gamma functions by factorials. Similar simple expressions can also be obtained when $q = (n+1)/2$.

(ii) Given $q = n$, then equation (18) factorizes into a product of n terms, which takes on the following singled-out form:

$$\widehat{p_i^n} = \left(\frac{N_i+1}{N+M}\right)\left(\frac{N_i+2}{N+M+1}\right)\left(\frac{N_i+3}{N+M+2}\right)\cdots\left(\frac{N_i+n}{N+M-1+n}\right).$$

As we have shown above, $\widehat{H_q}_{|q=1}$ includes the Bayes estimator of the Shannon entropy. Setting now $n = 1$, we furthermore reobtain Laplace's (successor rule) estimator (see, e.g. [27]): $\widehat{p_i^n}_{|n=1} = (N_i+1)/(N+M)$. Moreover, for $q = n$ the asymptotic approach $\widehat{p_i^n} \to {f_i}^n$ is realized by allowing $N \to \infty$, i.e. $\widehat{H_n}$ converges towards the frequency-count estimator of H_n. Thus, the Bayes estimator is consistent.

(iii) We note that the Bayes estimator of H_q is not equal to the estimator obtained by inserting the Bayes estimator of the probability vector $\boldsymbol{p}$, i.e. $\widehat{H_q}(\boldsymbol{N}) \neq H_q(\hat{\boldsymbol{p}})$.

5. The Rényi entropy estimator

In this section, we consider the Bayes estimator of the Rényi entropy K_q. Substituting K_q for H_q in equation (6), the problem of deriving the estimator is the calculation of the integral

$$\widehat{K_q}(\boldsymbol{N}) = \frac{1}{1-q}\frac{1}{W(\boldsymbol{N})}\int_{\mathcal{S}} \mathrm{d}\boldsymbol{p}\, \log_2 Z_q(\boldsymbol{p}) P(\boldsymbol{N}|\boldsymbol{p}) Q(\boldsymbol{p}). \tag{20}$$

Even in the simple case of $M = 2$, finding the explicit analytical solution of the above integral turns out to be very complicated. In appendix C we will show that the Bayes estimator of the binary Rényi entropy (under the assumption of a uniform prior probability density) can be written as

$$\widehat{K_q}(N_1, N_2) = \frac{1}{\ln 2}\frac{1}{1-q}\left(I_q(N_1, N_2) - q\sum_{l=N_1}^{N}\frac{1}{l+1}\right) \tag{21}$$

for all $N_1 + N_2 = N$. In the above expression we have introduced the following notation:

$$I_q(N_1, N_2) = \frac{\Gamma(N+2)}{\Gamma(N_1+1)\Gamma(N_2+1)}\int_0^\infty \mathrm{d}x\, \frac{x^{N_2}}{(1+x)^{N+2}} \ln(1+x^q). \tag{22}$$

Although the integrand in the above integral is well defined and thus this integral exists for all q, we could not obtain a closed analytical expression for arbitrarily given N_1, N_2 and q. This does also hold for the case $M > 2$. So the explicit evaluation of equation (20) remains a challenge.

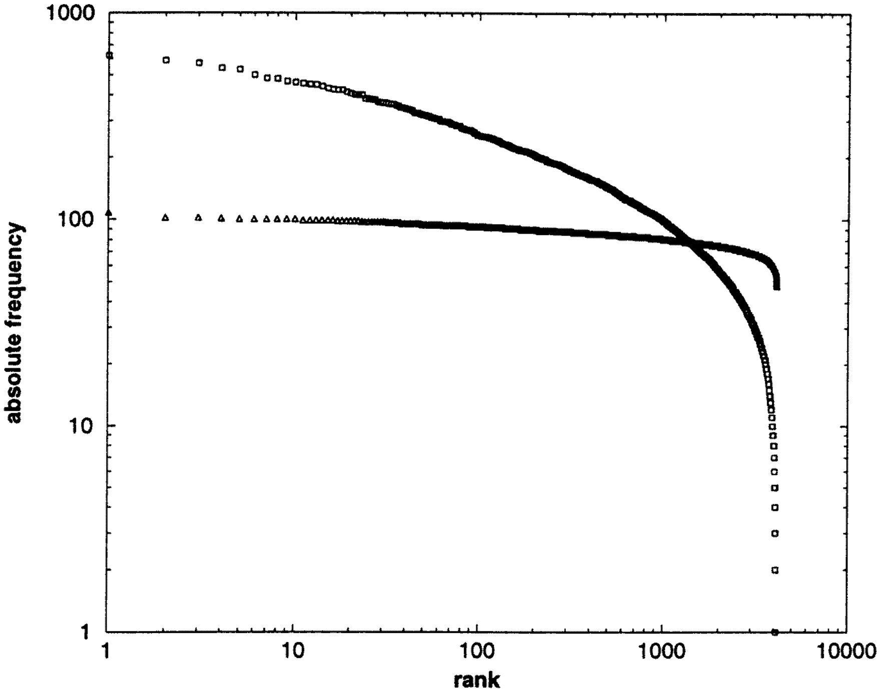

Figure 2. The rank-ordered hexamer distribution of the complete *Haemophilus influenzae* DNA sequence displayed as a double-logarithmic plot (□). For a comparison, the rank ordered hexamer distribution of a Bernoulli-sequence of same length has been included in the figure (△).

Although the equation in the binary case (see equation (22)) could be solved numerically to give $\widehat{K_q}$, we may seek another strategy which is of practical use also in the multi-variate case $M > 2$. We recall that H_q and K_q are intimately related to each other via equation (5). Therefore, a natural way to estimate K_q would be to estimate H_q and then use relation (5) to compute K_q of corresponding order. Hence, we may write down the (indirect) Bayes estimator $\widetilde{K_q}$ (see equation (11)) in the form†

$$\widetilde{K_q}(\boldsymbol{N}) = \frac{1}{1-q}\log_2\left\{\frac{\Gamma(N+M)}{\Gamma(N+M+q)} \times \left[\sum_{i=1}^{M}\frac{\Gamma(N_i+1+q)}{\Gamma(N_i+1)}\right]\right\}. \tag{23}$$

Since $\lim_{q\to 1}\widetilde{K_q} = \lim_{q\to 1}\widehat{H_q}$, the limit $\widetilde{K_1} = \hat{H}$ holds and we again reobtain the Bayes estimator of the Shannon entropy. The motivation to proceed in this way is led by the fact that both $\widehat{H_q}$ and $\widetilde{K_q}$ can be understood as entropies computed from the Bayes estimator of the partition function Z_q. As such, we gain a significant reduction of the entropy variance due to $\widehat{Z_q}$.

† Please note that we distinguish the indirect from the direct Bayes estimator by a tilde.

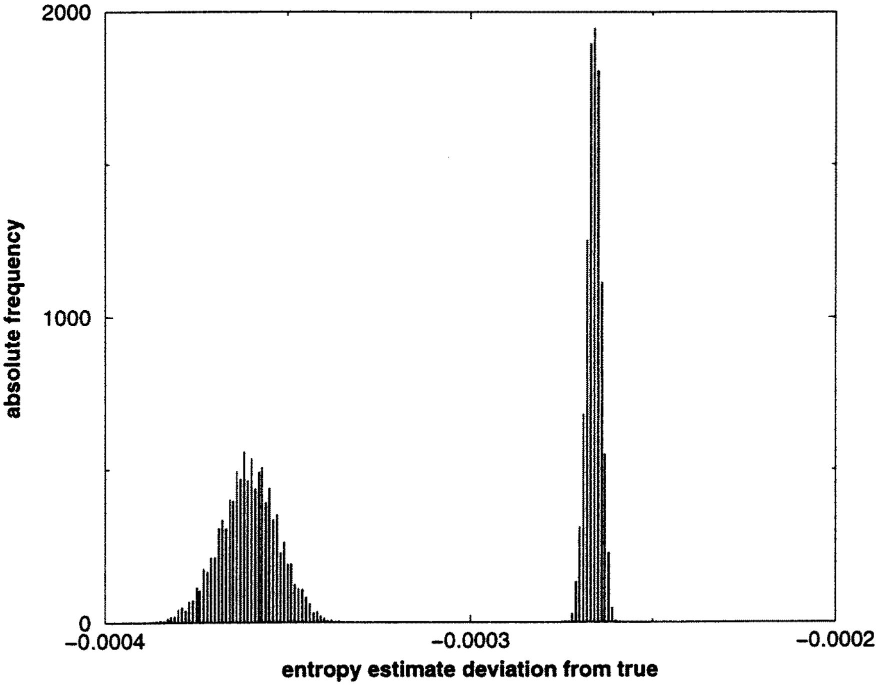

Figure 3. Comparison of the entropy estimators $\widehat{H_2}$ (right) and $\bar{H}_2$ (left) with $M = 4096$, $N = 4000$ and equidistributed $p_i = 1/M$. We observe the small width of the variance of the Bayes estimator $\widehat{H_2}$ as compared with the frequency-count estimator $\bar{H}_2$. Equation (A3) predicts the entropy bias with $\Delta\widehat{H_2} = -2.66 \times 10^{-4}$ (bits/symbol), in good agreement to the observed value. According to [34], the bias of $\bar{H}_2$ can be approximated to be $\Delta\bar{H}_1 = -0.36 \times 10^{-3}$ (bits/symbol), which is also in good agreement with the observed value. In samples where N is in the order of magnitude of M, the reliability of the Bayes estimator is significantly higher than the reliability of the frequency-count estimator.

6. Numerical tests

In this section we compare the variances of the direct and indirect Bayes estimators, $\widehat{H_q}$ and $\widetilde{K_q}$, with the variances of the frequency-count estimators, $\bar{H}_q$ and $\bar{K}_q$. To investigate and contrast the performance of the two different estimators we choose m-step memory Markov processes belonging to the following cases. (a) Generated by a process with no memory, i.e. $m = 0$, and (b) generated by a process with memory $m = 5$. In (a) we choose a process with equidistributed probabilities (henceforth denoted as Bernoulli process), whereas in the latter case we use the fifth-order transition probabilities taken from the complete 1830 240 nucleotides long *Haemophilus influenzae* DNA sequence [35] to generate a Markov chain with fifth-order memory. Figure 2 shows the rank-ordered statistics obtained from the above DNA sequence and from a sequence of same length derived from a Bernoulli process. It can be seen that the DNA sequence is far more inhomogeneous than the realization of the Bernoulli process. The derived rank-order frequencies might count as a typical example representing hexamer distributions in (prokaryotic) DNA. The entropy analysis of biosequences has received applications in order to distinguish between coding and non-coding DNA [36], to detect repeated nucleotide sequences [37], and to characterize

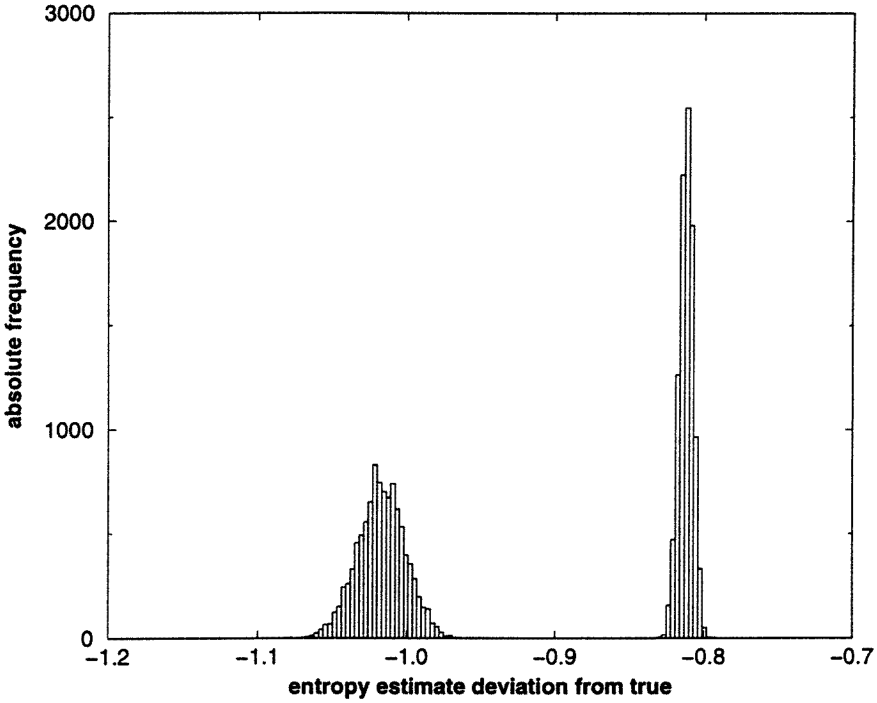

Figure 4. Comparison of the entropy estimators $\widetilde{K}_2$ (right) and $\bar{K}_2$ (left) with $M = 4096$, $N = 4000$ and equidistributed $p_i = 1/M$. We observe that fluctuations of the Bayes estimator $\widetilde{K}_2$ are strongly suppressed in comparison with the frequency-count estimator $\bar{K}_2$. Equation (A5) predicts the entropy bias with $\Delta\widetilde{K}_2 = -0.81$ (bits/symbol), in good agreement to the observed value. According to [34], the bias of $\bar{K}_2$ can be approximated to be $\Delta\bar{K}_2 = -1.02$ (bits/symbol), which is also in good agreement with the observed value.

protein sequences [23, 38]. A prerequisite to the application of generalized entropies in biosequence analysis are reliable estimators. Therefore we consider a probability vector derived from a DNA sequence to test the performence of the Bayes estimators, given by expressions (12) and (23), versus the frequency-counts estimators, which are obtained by defining $\bar{Z}_q \to \sum_{i=1}^{M} f_i^q$ with $f_i = N_i/N$.

Since we are particularly interested in the case where the size of the sequence length is in the order of magnitude of the cardinality of the alphabet, $M = 4^6$, we perform our numerical simulations with $N_{(a)} = 4 \times 10^3$ and $N_{(b)} = 8 \times 10^3$. Then, according to the probability vector $\boldsymbol{p} \equiv (p_1, \ldots, p_{4096})$, a sequence S is randomly generated from which we estimate the entropy values. In both cases we can also compute the theoretical hexamer entropies (since we take the relative frequencies obtained from the DNA sequence as probabilities by definition). Hence, the difference between the estimated and theoretical values defines a random variable, which we define as ‘entropy estimate deviation from true’. Generating an ensemble of 10 000 sequences and estimating the entropies from each sequence, we obtain the histograms displayed in figures 3–5. These studies demonstrate the merit of the Bayes order-2 entropy estimators as compared with the frequency-count estimators. Indeed, the variances of the Bayes estimates are significantly smaller than the variances of the frequency-count estimates for both Markov processes with memory $m = 0$ and $m = 5$. In repeated simulations with different sequence lengths and different values of q ranging from

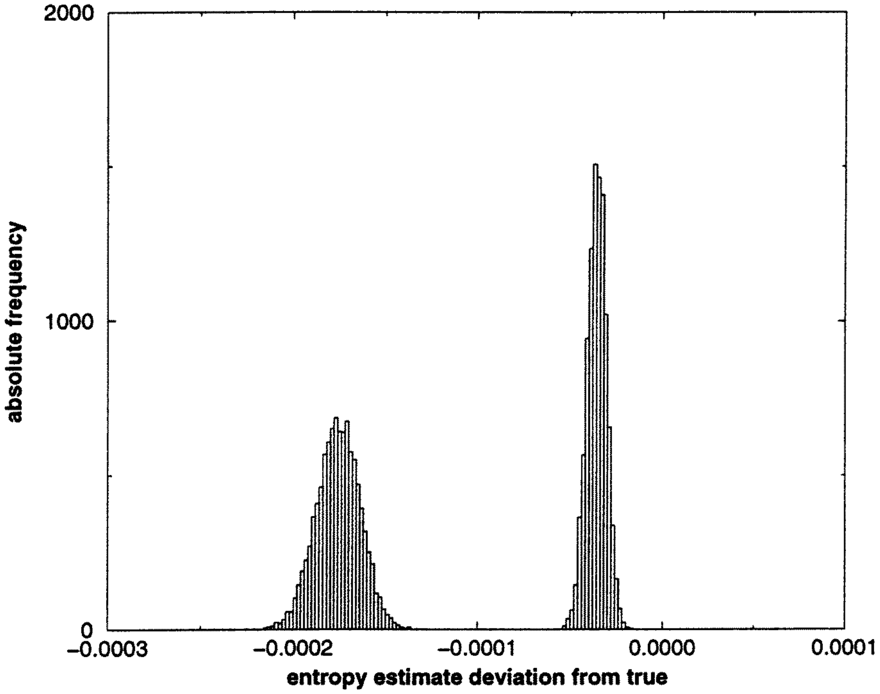

Figure 5. Comparison of the entropy estimators $\widehat{H_2}$ (right) and $\bar{H}_2$ (left) with $M = 4096$, $N = 8000$ and p_i derived from the *H influenzae* DNA sequence. We observe the smaller variance of the Bayes estimator $\widehat{H_2}$ in comparison with the frequency-count estimator $\bar{H}_2$. Equation (A3) predicts the entropy bias with $\Delta\widehat{H_2} = -0.38 \times 10^{-4}$ (bits/symbol) and, according to [34], the bias of $\bar{H}_2$ can be approximated to be $\Delta\bar{H}_1 = -0.18 \times 10^{-3}$ (bits/symbol).

-1 to 50 we could observe similar results: the Bayes estimator of H_q and K_q produces significantly smaller variances than the frequency-count estimator.

As analytical calculations and numerical simulations reveal, the Bayes estimator of Z_q (and hence for H_q and K_q) is biased. As we will show in appendix A, this bias can be approximated within $\mathcal{O}(1/N)$, by using a straightforward analytical approach.

7. Summary and conclusions

In this paper we derived the direct Bayes estimator $\widehat{H_q}$ of the order-q Tsallis entropy and the indirect Bayes estimators $\widetilde{K_q}$ of order-q Rényi entropy of a finite, discrete data set.

Our approach for deriving the Bayes estimators of H_q and K_q was motivated by the requirement to estimate generalized entropies from realizations where the total sample size N available may only be in the order of magnitude of the cardinality M. The central result of this work, namely the Bayes estimator of the Tsallis entropy H_q, is stated in expression (12). As we could not arrive at a closed form expression of the direct Bayes estimator of the Rényi entropy, we proposed an indirect Bayes estimator by the transformation formula which connects the Tsallis with the Rényi entropy. In fact, both estimators, $\widehat{H_q}$ and $\widetilde{K_q}$, are based on the Bayes estimator of the partition function Z_q, which may be exploited to estimate related quantities such as generalized dimensions or order-q Kolmogorov entropies. In the case of $q = (n+1)/2$, $n \in \mathcal{N}$, these estimators are easy to implement for numerical purposes.

A comparative study of the accuracy by which both the Bayes and the frequency-count estimators extract the order-2 entropies of m-step memory Markov chains demonstrates the strength of the Bayes estimator. Over the whole parameter range $q \in (-1, \infty)$ the Bayes estimator outperforms the frequency-count estimator by a significantly smaller variance of its estimates. This makes the Bayes estimator appropriate to measure generalized entropies in a sample, whose size N may be as small as the cardinality M of the alphabet.

The Bayes estimators $\widehat{H_q}$ and $\widetilde{K_q}$ have been derived under the assumption of a uniform prior probability density. Clearly, the specific choice of an assumption for the prior probability density is application dependent. Given no other constraint except $\boldsymbol{p} \in \mathcal{S}$, we assumed a constant prior probability density over the simplex. Note that this does not mean that the probabilities p_i are equidistributed, but rather that all probability vectors $\boldsymbol{p}$ on the simplex $\mathcal{S}$ are equiprobable. Nevertheless, the numerical simulations demonstrate that for the probability vectors considered is this work, which are by no means equidistributed on the simplex, the Bayes estimator with $Q(\boldsymbol{p}) =$ constant leads to variances which are significantly smaller in comparison with the variances of the frequency-counts estimators of generalized entropies of order q.

Acknowledgments

We are grateful to Lev B Levitin, W Ebeling, O Weiss and M Stemmler for stimulating discussions. This work was supported by the National Institutes of Health and the Deutsche Forschungsgemeinschaft.

Appendix A. Finite-size effects

This appendix is devoted to asymptotic length corrections of the entropy bias of $\widehat{H_q}$ and $\widetilde{K_q}$. As shown by numerical simulations in sections 4 and 5, although the variance is significantly small both Bayes estimators produce biased entropy estimates. It is a general feature that many estimators, in particular those minimizing the variance, share this property of being biased. Consequently, the systematic deviation of the expectation value of the estimated entropies from the true entropy value, namely the bias, has to be calculated and taken into account in order to correct the bias of the observed estimates. Explicitly,

$$\Delta\widehat{H_q} = \mathrm{E}\widehat{H_q}(\boldsymbol{N}) - H_q(\boldsymbol{p}) = \frac{1}{\ln 2}\frac{1}{1-q}\left(\sum_{i=1}^{M}\Delta\widehat{p_i^q}\right) \tag{A1}$$

defines the bias of the estimator $\widehat{H_q}$. Here by E we denote the expectation value with respect to the multinomial distribution: $\mathrm{E}(\cdot) = \sum_{(N_1,\ldots,N_M)} P(\boldsymbol{N}|\boldsymbol{p})(\cdot)\delta(\sum_{i=1}^{M} N_i - N)$. Clearly, an unbiased statistic satisfies $\Delta(\cdot) = 0$.

The problem encountered in deriving the bias of entropy estimators is that it is difficult to obtain a closed form expression. However, in this case one may still obtain an approximation to the exact bias, for example, by expanding a power-series around the true values of $\widehat{p_i^q}$ and applying E to each individual term within this series. The underlying idea exploits the fact that any probability distribution can, in principle, be extensively described by all of its moments. For the Bayes estimators derived in this work, this applies to $q = n$, $n > 1$. Expanding the exact entropy bias as a series in terms of $(1/N)^d$ with $d = 1, 2, \ldots$, we arrive at a $d = 1$ approximation by Taylor-expanding the entropies in powers of $(f_i - p_i)^m$, $m \in \mathcal{N}$, and truncating this series after the quadratic term.

As principles of this technique have been discussed in detail, for example in [21, 34], we will not elaborate on this in further detail here, but only present the final results of the $\mathcal{O}(1/N)$ approximation of the entropy bias for the case $q = 2$. Since

$$\Delta \widehat{p_i^q}_{|q=2} = [2(2Np_i + 1) - (2N + M)Mp_i^2]/(N + M)^2 \qquad \text{(A2)}$$

we obtain the entropy bias of the order-2 Tsallis entropy as

$$\Delta \widehat{H_2} = -\frac{1}{\ln 2}\frac{(2N + M)(2 - MZ_2)}{(N + M)^2} \qquad \text{(A3)}$$

(note that the approximation is exact for the case $q = 2$). In order to obtain order-2 Tsallis entropy estimates that are unbiased in $\mathcal{O}(1/N)$, we define the estimator

$$\widehat{H_2}^{(d=1)} = \widehat{H_2} + \frac{1}{\ln 2}\frac{(2N + M)(2 - M\widehat{Z_2})}{(N + M)^2}. \qquad \text{(A4)}$$

For the Bayes estimator of the Renyi entropy it is more difficult to calculate the entropy bias. Given equidistributed states, we find that $\mathrm{E}\log(\cdot) \approx \log \mathrm{E}(\cdot)$ holds, and thus we can obtain an approximation to the bias of $\widetilde{K_2}$, which reads as:

$$\Delta \widetilde{K_2} = -\log_2\left[\frac{N^2 + 2(2N + M)/Z_2}{(N + M)^2}\right] + \mathcal{O}\left(\frac{1}{N^2}\right). \qquad \text{(A5)}$$

Hence, in analogy to equation (A4), we obtain Rényi entropy estimates $\widetilde{K_2}^{(1)}$ that are unbiased in $\mathcal{O}(1/N)$. For non-Bernoulli processes fluctuations increase, which render the above approximation to be, in general, no longer reliable. In this case, unbiased estimates of K_2 (in the order of $\mathcal{O}(1/N)$) may be obtained by a transformation of the unbiased Tsallis entropy $\widehat{H_2}^{(1)}$.

According to the correction terms (see expressions (A3) and (A5)) the systematic error depends on the individual probability components p_i as well as the cardinality of the alphabet M. Since the simulation performed in this investigation are not aimed at a detailed analysis of finite-size effects but rather a study of the variances of the Bayes entropy estimator versus the frequency-count estimator, we insert the theoretical values of p_i in the above correction terms, i.e. we set $\hat{p}_i = p_i$. In any attempt to estimate these quantities from a sample of data points, it is crucial to the entropy bias by which method we estimate the unknown variables p_i (see, e.g. [39]). A study of the quantification of the order-q entropy bias, using asymptotic length corrections, deserves further investigations and will be undertaken in forthcoming work.

Appendix B. Calculation of the normalization constant W

Under the assumption of a stationary, independent distributed sample of data points, the conditional probability density to observe a sample with occupation numbers $\{N_1, \ldots, N_M\}$ is given by the multinomial distribution $P(\boldsymbol{N}|\boldsymbol{p}) = C_{\boldsymbol{N}} \prod_{i=1}^{M} p_i^{N_i}$. Here the multinomial coefficient reads as $C_{\boldsymbol{N}} = N!/\prod_{i=1}^{M} N_i!$, and the size of the sample is $N = \sum_{i=1}^{M} N_i$.

We define $W'(\boldsymbol{N}) = W(\boldsymbol{N})/C_{\boldsymbol{N}}$. Then, with a uniform prior probability density, the reduced normalization constant reads as

$$W'(\boldsymbol{N}) = \frac{1}{C_{\boldsymbol{N}}}\int_{\mathcal{S}} \mathrm{d}\boldsymbol{p}\, P(\boldsymbol{N}|\boldsymbol{p})Q(\boldsymbol{p}) = \int_{\mathcal{S}} \prod_{j=1}^{M} \mathrm{d}p_j\, p_j^{N_j}. \qquad \text{(B1)}$$

Introducing the auxiliary variable $k_j = 1 - \sum_{l=1}^{j} p_l$, the explicit integral takes on the form

$$W'(\boldsymbol{N}) = \int_{p_1=0}^{1} \mathrm{d}p_1\, p_1^{N_1} \int_{p_2=0}^{k_1} \mathrm{d}p_2\, p_2^{N_2} \ldots \int_{p_{M-1}=0}^{k_{M-2}} \mathrm{d}p_{M-1}\, p_{M-1}^{N_{M-1}} (k_{M-2} - p_{M-1})^{N_M}. \tag{B2}$$

In the above expression, all integrals are of the type $\int \mathrm{d}u\, u^a (\xi - u)^b$. Changing co-ordinates $u = \xi v$, these integrals can be rewritten in terms of ordinary Beta-functions

$$\int_0^{\xi} \mathrm{d}u\, u^a (\xi - u)^b = \xi^{a+b+1} \mathrm{B}(a+1, b+1) \tag{B3}$$

for all positive real numbers a and b, and $\mathrm{B}(a, b) = \Gamma(a)\Gamma(b)/\Gamma(a+b)$. The relation $\Gamma(n+1) = n!$ holds for $n \in \mathcal{N}$.

Using relation (B3), we may integrate equation (B2) over p_{M-1} to obtain

$$W'(\boldsymbol{N}) = \mathrm{B}(N_{M-1}+1, N_M+1) \times \left\{ \int_{p_1=0}^{1} \mathrm{d}p_1\, p_1^{N_1} \int_{p_2=0}^{k_1} \mathrm{d}p_2\, p_2^{N_2} \ldots \right.$$
$$\left. \ldots \int_{p_{M-2}=0}^{k_{M-3}} \mathrm{d}p_{M-2}\, p_{M-2}^{N_{M-2}} (k_{M-3} - p_{M-2})^{(N_{M-1}+N_M+1)} \right\}.$$

Completing the iteration for all but the integration over p_1, this yields

$$W'(\boldsymbol{N}) = \prod_{m=2}^{M-1} \mathrm{B}\left(N_m + 1, \sum_{j=m}^{M-1} N_{j+1} + (M-m) \right)$$
$$\times \left\{ \int_{p_1=0}^{1} \mathrm{d}p_1\, p_1^{N_1} (1-p_1)^{(\sum_{j=2}^{M} N_j + (M-2))} \right\}.$$

Expressing the Beta-functions in terms of Gamma-functions, we obtain $W'(p_1; \boldsymbol{N})$ in the form

$$W'(p_1; \boldsymbol{N}) = p_1^{N_1} \frac{\prod_{j=2}^{M} \Gamma(N_j+1)}{\Gamma(\sum_{j=2}^{M} N_j + M - 1)} (1-p_1)^{(\sum_{j=2}^{M} N_j + (M-2))}. \tag{B4}$$

Inspecting the above expression, we realize that equation (B4) can, in fact, be readily written down for a general ith component:

$$W'(p_i; \boldsymbol{N}) = p_i^{N_i} \frac{\prod_{\substack{j=1 \\ j \neq i}}^{M} \Gamma(N_j+1)}{\Gamma(\sum_{j=1}^{M} (1-\delta_{ij}) N_j + M - 1)} (1-p_i)^{(\sum_{j=1}^{M} (1-\delta_{ij}) N_j + (M-2))}. \tag{B5}$$

Integrating (B5) over p_i, we arrive at the normalization constant

$$W(\boldsymbol{N}) = C_{\boldsymbol{N}} \int_{p_i=0}^{1} \mathrm{d}p_i\, W'(p_i; \boldsymbol{N}) = \frac{\Gamma(N+1)}{\Gamma(N+M)}. \tag{B6}$$

Appendix C. Bayes' estimator of the binary Renyi entropy K_q

Under the assumption of a uniform prior probability density, $Q(\boldsymbol{p}) =$ constant, the Bayes estimator of the binary Rényi entropy of order q can be written as

$$\widehat{K_q}(N_1, N_2) = \frac{1}{1-q} \frac{1}{W'(N_1, N_2)} \int_0^1 \mathrm{d}p\, p^{N_1} (1-p)^{N_2} \log_2[p^q + (1-p)^q] \tag{C1}$$

for all $N_1 + N_2 = N$. Using the normalization constant (B6), we rewrite equation (C1) in the form

$$\widehat{K_q}(N_1, N_2) = \frac{1}{\ln 2}\frac{1}{1-q}\frac{\Gamma(N+2)}{\Gamma(N_1+1)\Gamma(N_2+1)} \times \left\{ q\int_0^1 \mathrm{d}p\, p^{N_1}(1-p)^{N_2}\ln p \right.$$
$$\left. + \int_0^1 \mathrm{d}p\, p^N \left(\frac{1-p}{p}\right)^{N_2} \ln\left[1 + \left(\frac{1-p}{p}\right)^q\right]\right\}. \tag{C2}$$

The first term on the right-hand side of (C2) can be calculated to become

$$\begin{aligned} q\int_0^1 \mathrm{d}p\ \ p^{N_1}(1-p)^{N_2}\ln p &\equiv q\frac{\partial}{\partial N_1}\left(\int_0^1 \mathrm{d}p\, p^{N_1}(1-p)^{N_2}\right) \\ &= q\frac{\partial}{\partial N_1}\mathrm{B}(N_1+1, N_2+1) \\ &= -q\mathrm{B}(N_1+1, N_2+1)\left(\sum_{l=N_1}^{N}\frac{1}{l+1}\right). \end{aligned} \tag{C3}$$

In the remaining term in equation (C2), we change the co-ordinate $x = (1-p)/p$ and thus arrive at

$$\widehat{K_q}(N_1, N_2) = \frac{1}{\ln 2}\frac{1}{1-q}\left(I_q(N_1, N_2) - q\sum_{l=N_1}^{N}\frac{1}{l+1}\right) \tag{C4}$$

where we define

$$I_q(N_1, N_2) = \frac{\Gamma(N+2)}{\Gamma(N_1+1)\Gamma(N_2+1)}\left\{\int_0^\infty \mathrm{d}x\, x^{N_2}[1+x]^{-(N+2)}\ln(1+x^q)\right\}. \tag{C5}$$

References

[1] Shannon C E 1948 *Bell Syst. Tech. J.* **27** 379

[2] Khinchin A I 1957 *Mathematical Foundations of Information Theory* (New York: Dover)

[3] Tsallis C 1988 *J. Stat. Phys.* **52** 479

[4] Curado E M F and Tsallis C 1991 *J. Phys. A: Math. Gen.* **24** L69

[5] Rényi A 1970 *Probability Theory* (Amsterdam: North-Holland)

[6] Grassberger P and Procaccia I 1983 *Physica* **9D** 189

[7] Halsey T C, Jensen M H, Kadanoff L P, Procaccia I and Shraiman B I 1986 *Phys. Rev.* A **33** 1141

[8] Kurths J and Herzel H 1987 *Physica* **25D** 167

[9] Grassberger P, Schreiber T and Schaffrath C 1991 *Int. J. Bifurcation Chaos Appl. Sci. Eng.* **1** 521

[10] Beck C and Schlögl F 1993 *Thermodynamics of Chaotic Systems* (Cambridge: Cambridge University Press)

[11] Chame A and de Mello E V L 1994 *J. Phys. A: Math. Gen.* **27** 3363

[12] Ramshaw J D 1995 *Phys. Lett.* A **198** 119
Ramshaw J D 1995 *Phys. Lett.* A **198** 122

[13] Tsallis C, Levy S V F, Souza A M C and Maynard R 1995 *Phys. Rev. Lett.* **75** 3589

[14] Lucena L S, da Silva L R and Tsallis C 1995 *Phys. Rev.* E **51** 6247

[15] Boghosian B M 1996 *Phys. Rev.* E **53** 475

[16] Chame A and de Mello E V L 1997 *Phys. Lett.* A **228** 159

[17] Plastino A R, Plastino A and Tsallis C 1994 *J. Phys. A: Math. Gen.* **27** 5707

[18] Stariolo D A and Tsallis C 1995 *Annual Reviews of Computational Physics* vol 1, ed D Stauffer (Singapore: World Scientific)

[19] Penna T J P 1995 *Phys. Rev.* E **51** 34

[20] Basharin G P 1959 *Theor. Prob. Appl.* **4** 361

[21] Harris B 1975 *Colloquia Mathematica Societatis János Bolyai (Keszthely)* **16** 323

[22] Levitin L B and Reingold R 1978 An improved estimate for the entropy of a discrete random variable *Annu. Conf. of the Israel Statistical Association (Tel Aviv)*

[23] Herzel H 1988 *Syst. Anal. Modelling Simul.* **5** 435

[24] Grassberger P 1988 *Phys. Lett.* **128A** 369
[25] Herzel H, Schmitt A O and Ebeling W 1994 *Chaos Solitons Fractals* **4** 97
[26] Ebeling W, Pöschel T and Albrecht K-F 1995 *Int. J. Bifurcation Chaos Appl. Sci. Eng.* **5** 51
[27] Schürmann T and Grassberger P 1996 *Chaos* **6** 414
[28] Berger O 1985 *Statistical Decision Theory and Bayesian Analysis* (New York: Springer)
[29] Wolpert D H and Wolf D R 1995 *Phys. Rev.* E **52** 6841
[30] Große I 1995 Statistical analysis of biosequences *Thesis* Humboldt-University, Berlin
[31] Ruelle D 1989 *Chaotic Evolution and Strange Attractors* (Cambridge: Cambridge University Press)
[32] Pompe B 1993 *J. Stat. Phys.* **73** 587
[33] Große I 1996 Estimating entropies from finite samples *Dynamik, Evolution, Strukturen* ed J A Freund (Berlin: Köster)
[34] Holste D 1997 Entropy estimators and their limitations: the statistical analysis of symbol sets *Thesis* Humboldt-University, Berlin
[35] Fleischmann R *et al* 1995 *Science* **269** 496
[36] Fickett J W and Tung C-S 1992 *Nucleic Acids Res.* **24** 6441
[37] Herzel H, Ebeling W and Schmitt A O 1994 *Phys. Rev.* E **6** 5061
[38] Strait B J and Dewey T G 1996 *Biophys. J.* **71** 148
[39] Schmitt A O, Herzel H and Ebeling W 1993 *Europhys. Lett.* **23** 303