Nuclear Physics B 383 (1992) 622–642
North-Holland

NUCLEAR
PHYSICS B

Large-deviation analysis of multiplicity fluctuations

J. Seixas and R. Vilela Mendes [1]

TH Division, CERN, CH-1211 Geneva 23, Switzerland

Received 12 December 1991
Accepted for publication 1 July 1992

In this paper we introduce and briefly discuss some of the basic results of large-deviation theory. We propose that large-deviation analysis be used as a tool complementary to the usual intermittency analysis because, as we show, systems exhibiting similar factorial moment behaviour may be quite different from the point of view of large-deviation properties. Large-deviation functions for some simple statistical models are computed and a method is developed to perform this analysis on experimental or simulated data.

1. Introduction

Clustering of many particles in small rapidity intervals has been observed in many multiparticle production experiments at high energies. Some of these events are exceptional both in the sense of the clustering in small intervals and in the sense of deviating from average values. For example, in the NA22 experiment, the average number of particles per event is around seven but there is an event with a multiplicity 25 and 10 particles clustered in a bin of 0.1 (in rapidity).

The clustering structure is suggestive of fractal behaviour of the type observed in fluctuations in turbulent fluids. This led Bialas and Peschanski [1] to make the pioneering suggestion that a similar (momentum-space) hierarchy of scales might exist in the multiparticle production process, leading to a fractal dependence on bin resolution. Some intuitive insight may be obtained from a parton fragmentation picture. However to decide on the dynamical origin of such a hierarchy of scales in the production process, requires a better understanding of the relation between QCD and the hadronization process. The geometrical clustering structure of the phenomenon is well characterised by the factorial moments of the distribution, and these have become a standard tool for the analysis of the data.

[1] Permanent address: Instituto de Física e Matemática, Av. Gama Pinto 2, 1699 Lisbon Codex, Portugal.

Correspondence to: Dr. R. Vilela Mendes, Instituto de Física e Matemática, Av. Gama Pinto 2, 1699 Lisbon Codex, Portugal.

Whenever large fluctuations are observed in a process there is however another feature of the data which should be adequately characterised. This is the deviation from the average for finite sample sizes. Trained to think in terms of gaussian or Poisson processes, many physicists live under the impression that the deviation from asymptotic average values, coming from the finiteness of the samples, is a minor and well-controlled effect. Think however, for a moment, about a set of data $\{\xi_n\}$ where each point ξ_n is generated by the multiplication of 1000 numbers each one being either zero or four with probability $\frac{1}{2}$. The expectation value $\langle \xi \rangle$ (or the asymptotic average $\lim_{n\to\infty}(1/n)\Sigma\xi_n$) of the data is 2^{1000}, a huge number. However if an experimental approach is taken, without prior knowledge of the nature of the process, the result might be quite different. Generating for example the "data" on a computer and analysing it according to well-established experimental standards, even with statistics unmatched by any current particle experiment, we bet that the result would be zero (and with negligible error bars).

Incidentally, if our multiplicatively generated data seems too removed from any physically relevant problem, just remember that a random multiplicative process is exactly one of the models used to interpret the fluctuations in multiparticle distributions.

Large-deviation theory is the set of results that studies deviations from the asymptotic averages when the sample is finite (as it always is in experimental physics). If we are interested in measuring an observable it is its expectation value or asymptotic average that we want to obtain. Large deviations from this value are, in this sense, exceptional events. What large-deviation theory tells us is how probable exceptional events are.

We should point out that the question of the clustering structure in particle distributions and the question of large deviations, may be independent issues. In the following sections we will indeed show how two models that display similar factorial moment dependence on the bin resolution, and therefore have similar clustering structure, have nevertheless different behaviour as far as large-deviation properties are concerned. We therefore suggest that the data analysis using the factorial moments should be complemented by a large-deviation analysis. Both should be useful in uncovering the dynamical origins of the multiplicity fluctuations.

The plan of the paper is the following: In sect. 2 we make a summary of the most useful results in large-deviation theory. Most of the results are used in the following sections and we thought that, rather than spreading throughout the paper the references to the theory, a systematic summary might be more useful for experimentalists and theoreticians that might want to use these tools. In sect. 3 we study, from the point of view of large deviations, the random cascading model (α-model) and a spin model. We have chosen these models because explicit calculations could be performed and their distinct large-deviation behaviour exhibited.

We have on purpose refrained from commenting and evaluating actual phenomenological particle production models, because we want to emphasise that the ideal situation would be to have the data itself analysed to extract its large-deviation behaviour. We indicate in sect. 4 a method to perform such an analysis.

2. A summary of large-deviation theory

Large-deviation theory is the study of the fluctuations away from the limit values given by the law of large numbers, that are observed in empirical statistics, due to the finiteness of the sample size.

In probability theory one consider a space Ω of elementary events ω, a probability measure P and a set $\mathscr{F}$ of measurable subsets of events. Consider now a sequence $(\Omega_n, \mathscr{F}_n, P_n)$ of probability spaces and a sequence of random vectors (vector-valued random functions) $\xi_n(\omega)$ $\{\omega \in \Omega_n\}$ with values in a metric space Y.

Assume that ξ_n tends to a limit point y_0 when $n \to \infty$ (law of large numbers). Define now the following sequence of probability measures Q_n on the Borel sets of Y,

$$Q_n\{A\} = P_n\{\omega\colon \xi_n(\omega) \in A\}, \qquad A \in \mathscr{B}(Y). \tag{2.1}$$

The existence of the limit $\xi_n \to y_0$ implies that $Q_n\{A\} \to 0$ if y_0 is not in the closure of A.

The sequence Q_n of probability measures is said to have a *large-deviation property* (LDP) if there is a sequence $a_n \to \infty$ $a_n \in \mathbb{N}$ and a function $I(y)\colon Y \to [0, \infty]$ such that

(a) $I(y)$ is lower semicontinuous [i.e. $y_n \to y \Rightarrow \liminf I(y_n) \geqslant I(y)$].

(b) $I(y)$ has compact level sets (i.e. $\{y \in Y\colon I(y) \leqslant L\}$ compact $\forall L \in \mathbb{R}$).

(c) $\lim_{n\to\infty} \sup 1/a_n \log Q_n\{F\} \leqslant -I(F)$, F closed.
$\lim_{n\to\infty} \inf 1/a_n \log Q_n\{G\} \geqslant -I(G)$, G open.

where $I(F) = \inf I(y)$, $y \in F$. These two limits are usually represented by the symbol $\asymp$, namely

$$Q_n(\mathrm{d}y) \asymp \exp[-a_n I(y)]\ \mathrm{d}y. \tag{2.2}$$

Furthermore if $I(\mathrm{int}\ A) = I(\mathrm{cl}\ A)$ then

$$\lim_{n\to\infty} \frac{1}{a_n} \log Q_n\{A\} = -I(A). \tag{2.3}$$

We now list a few large-deviation theorems that will be useful in the applications of the following sections.

Theorem 1. If the function I exists for a sequence $\{a_n\}$ then it is unique.

Theorem 2. If $Q_n(\mathrm{d}y) \asymp \exp[-a_n I(y)]\,\mathrm{d}y$ on Y, then $I(\mathrm{Y}) = 0$.

Theorem 3. [Obtaining the deviation function from the free energy (or pressure)] Let W_n be random vectors with scaling constants a_n. If the following limit exists,

$$c(t) = \lim_{n\to\infty} \frac{1}{a_n} \log \mathrm{E}^{P_n}\{\exp(t, W_n)\}, \qquad t \in \mathrm{Y}, \tag{2.4}$$

and if $c(t)$ is differentiable, then the deviation function in

$$Q_n(\mathrm{d}x) = P_n\{a_n^{-1} W_n \in \mathrm{d}x\} \asymp \exp[-a_n I(x)]\,\mathrm{d}x$$

is the Legendre transform

$$I(x) = \sup_{t\in\mathrm{Y}} \{tx - c(t)\}. \tag{2.5}$$

E^{P_n} denotes the expectation value in the probability space $(\Omega_n, \mathscr{F}_n, P_n)$ and $a_n^{-1} W_n$ plays the role of the random vector ξ_n in the definition of the large deviation above. The method we propose for the large-deviation analysis of the data is based on this result.

Theorem 4. (Laplace–Varadhan method.) If $Q_n(\mathrm{d}y) \asymp \exp[-a_n I(y)]\,\mathrm{d}y$ on Y and F is a function, F: $\mathrm{Y} \to \mathbb{R}$, then

$$\lim_{n\to\infty} \frac{1}{a_n} \log \int \exp[a_n F(y)]\; Q_n(\mathrm{d}y) = \sup_{y\in\mathrm{Y}} \{F(y) - I(y)\}.$$

Theorem 5. (Contraction principle.) Let ψ be a continuous mapping ψ: $\mathrm{Y} \to \mathrm{W}$. If $Q_n(\mathrm{d}y) \asymp \exp[-a_n I(y)]\,\mathrm{d}y$ on Y, then

$$Q_n \circ \psi^{-1}(\mathrm{d}z) \asymp \exp[-a_n J(z)]\;\mathrm{d}z \text{ on W},$$

where

$$J(z) = \inf\{I(y)\colon y \in \mathrm{Y},\, \psi(y) = z\}$$

and if z is not in the range of ψ then $J(z) = \infty$.

Theorem 6. Let $Q_n(\mathrm{d}y) \asymp \exp[-a_n I(y)]\,\mathrm{d}y$ on Y, F: $\mathrm{Y} \to \mathbb{R}$ a continuous bounded function and

$$Q_{n,F}\{\mathrm{A}\} = \int_{\mathrm{A}} \frac{\exp[a_n F(y)]}{\int_{\mathrm{Y}} \exp[a_n F(z)] Q_n(\mathrm{d}z)} Q_n(\mathrm{d}y).$$

Then

$$Q_{n,F}(\mathrm{d}y) \asymp \exp[-a_n I_F(y)]\ \mathrm{d}y,$$

with

$$I_F(y) = I(y) - F(y) - \inf_{z \in \mathrm{Y}} \{I(z) - F(z)\}.$$

This theorem is useful for example to obtain finite-volume Gibbs states, where the exponential factor is $\exp[-\beta H_n(\omega)]$ and the function $F(y)$ is obtained isolating the leading a_n term in the finite-volume energy $-H_n(\omega) = a_n\ F(\xi_n(\omega)) + o(a_n)$. This will be used in sect. 3 to obtain the level-2 large deviations in spin systems.

2.1. LEVELS OF DESCRIPTION OF STATISTICAL SYSTEMS AND LARGE-DEVIATION LEVELS

Let a statistical phenomenon be described by a sequence of values of a random variable X,

$$\ldots X_{-2} X_{-1} X_0 X_1 X_2 \ldots,$$

which takes values in a space Y. Y is called the *state space* and the space of sequences $\mathrm{Y}^{\mathbb{Z}}$ is called the *path space*.

The statistical properties of the phenomenon may be described at three different levels:

(1) By the expectation values of observables,

$$\langle g(X) \rangle = \lim_{n \to \infty} \frac{1}{n} \sum_{i=1}^{n} g(X_i); \tag{2.6}$$

(2) By probability measures on the state space Y,

$$\rho(\mathrm{d}y) = P\{X_i \in (y, y + \mathrm{d}y)\}; \tag{2.7}$$

(3) By probability measures on path space $\mathrm{Y}^{\mathbb{Z}}$. This may be constructed by defining the probabilities $P\{\Sigma\}$ on cylinder sets

$$\Sigma_k = \{\omega\colon X_i \in \Delta_i, \ldots, X_{i+k} \in \Delta_{i+k}\}. \tag{2.8}$$

The finite-sample versions of the expectation values, the probability on the state space and the probability on path space are

(a) The mean partial sum,

$$\frac{1}{n} S_n = \frac{1}{n} \sum_{i=1}^{n} g(X_i); \tag{2.9}$$

(b) The empirical measure,

$$L_n(\omega, \mathrm{A}) = \frac{1}{n} \sum_{i=1}^{n} \delta_{X_i}(\omega)(\mathrm{A}); \tag{2.10}$$

(A being a measurable set in Y.)

(c) The empirical process.

To define a measure in $\mathrm{Y}^{\mathbb{Z}}$ out of a finite sample, the sample is repeated infinitely many times in the future and the past generating an element $X(n, \omega)$ in $\mathrm{Y}^{\mathbb{Z}}$. Then the empirical process is

$$R_n(\omega, \Sigma) = \frac{1}{n} \sum_{i=0}^{n-1} \delta_{\sigma^i X(n,\omega)}\{\Sigma\}, \tag{2.11}$$

where σ is the shift operator and Σ a cylinder set in $\mathrm{Y}^{\mathbb{Z}}$.

The large deviations of mean partial sums are called *level-1 large deviations (L.D.)*. The large deviations of the empirical measure are called *level-2 L.D.* and the large deviations of the empirical process are called *level-3 L.D.*

2.2. EXAMPLES

2.2.1. Level-1 L.D.

$$Q_n(\mathrm{d}x) = P\{n^{-1}S_n \in (x, x+\mathrm{d}x)\} \asymp \exp[-nI(x)]\ \mathrm{d}x. \tag{2.12}$$

For a gaussian process,

$$\rho(\mathrm{d}x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp(-x^2/2\sigma^2)\ \mathrm{d}x, \qquad I(x) = \frac{1}{2\sigma^2}x^2, \tag{2.13}$$

and for

$$\rho = \tfrac{1}{2}\delta_1 + \tfrac{1}{2}\delta_{-1} \tag{2.14a}$$

the deviation function is

$$\begin{aligned} I(x) &= \tfrac{1}{2}(1-x)\log(1-x) + \tfrac{1}{2}(1+x)\log(1+x) \qquad |x| \leqslant 1 \\ &= \infty \qquad\qquad |x| > 1. \end{aligned} \tag{2.14b}$$

2.2.2. Level-2 L.D.

$$Q_n(\mathrm{d}\nu) = P\{\omega\colon L_n(\omega, \cdot) \in \mathrm{d}\nu\} \asymp \exp[-nI^{(2)}(\nu, \rho)]\ \mathrm{d}\nu. \tag{2.15}$$

For independent identically distributed (I.I.D.) random vectors,

$$I^{(2)}(\nu, \rho) = \int_Y \left(\log \frac{d\nu}{d\rho}(x) \right) \nu(dx) \quad \text{if } \nu \ll \rho$$
$$= \infty \qquad \text{otherwise,} \tag{2.16}$$

and if the state space Y is finite, $Y = \{1, 2, \ldots, r\}$,

$$I^{(2)}(\nu, \rho) = \sum_{i=1}^{r} \nu_i \log\left(\frac{\nu_i}{\rho_i} \right) \quad \text{if } \nu \ll \rho$$
$$= \infty \qquad \text{otherwise} \tag{2.17}$$

(the symbol $\ll$ means "absolutely continuous with respect to").

2.2.3. Level-3 L.D.

$$Q_n(dP) = P_\rho\{\omega\colon R_n(\omega, \cdot) \in dP\} \asymp \exp\left[-nI^{(3)}(P, P_\rho)\right] dP. \tag{2.18}$$

The finite-dimensional marginals of a probability measure in $Y^{\mathbb{Z}}$ are defined by restriction to cylinder sets pinned down at a finite number of points.

$$\Sigma_1 = \{\omega \in Y^{\mathbb{Z}}\colon X_1 = y_1\}, \tag{2.19a}$$

$$\Sigma_2 = \{\omega \in Y^{\mathbb{Z}}\colon X_1 = y_1,\ X_2 = y_2\}, \quad \text{etc.} \tag{2.19b}$$

Then for the empirical process the one-dimensional marginal

$$R_n(\omega, \Sigma_1) = L_n(\omega, \{y_1\}) \tag{2.20}$$

gives the empirical measure, the two-dimensional marginal

$$R_n(\omega, \Sigma_2) = M_n(\omega, \{y_1, y_2\}) \tag{2.21}$$

the empirical pair measure, etc.

Denote by $\pi_\alpha P$ the α-dimensional marginal of P. Then for independent identically distributed (I.I.D.) random vectors,

$$I^{(2)}(\pi_\alpha P, \pi_\alpha P_\rho) = \sum_\omega \pi_\alpha P\{\omega\} \log \frac{\pi_\alpha P\{\omega\}}{\pi_\alpha P_\rho\{\omega\}} \tag{2.22}$$

and

$$I^{(3)}(P, P_\rho) = \lim_{\alpha \to \infty} \frac{1}{\alpha} I^{(2)}(\pi_\alpha P, \pi_\alpha P_\rho). \tag{2.23}$$

Large-deviation theory was formulated in its modern form by Donsker and Varadhan and developed and applied in several domains by many authors. Here we have simply assembled a minimum set of results which we apply in the following sections and which might be useful for the large-deviation analysis of particle production data. For further results and applications we refer to the standard references [2–4].

3. Large deviation analysis of intermittency models

In this section we analyse the level-2 large deviations in two models which have been used to mimic the clustering effects observed in multiparticle production. The first is a random cascading model, the α-model [1], and the second a one-dimensional ferromagnetic spin model which has been shown [5] to display, for certain values of the parameters, a structure of exceptional events and factorial moments similar to some of the particle production data. The correspondence with the particle model in the spin case is made by assigning a certain number of spins to a bin and interpreting the number of spins that point upwards as the number of particles in the bin. Level-2 large deviations are deviations of the empirical measure away from the probability distribution of the number of particles per bin.

3.1. LARGE DEVIATIONS IN A RANDOM CASCADING MODEL (α-MODEL)

We consider each event to be generated by a cascade of s steps with λ branches, and for definiteness we take $\lambda = 2$. Up to a normalisation factor the particle number per bin Z (i.e. the average of the associated Poisson process) is given by the product of s random variables,

$$Z = W_1 W_2 \dots W_s, \tag{3.1}$$

with $\langle W \rangle = 1$. The distribution of the random variable W is

$$r(W) = u\delta(W - \beta_+) + (1-u)\delta(W - \beta_-). \tag{3.2}$$

Let β_+ and β_- be of the form $\beta_+ = p/\Gamma$ and $\beta_- = q/\Gamma$, p and q being prime numbers. Then Z can take $s+1$ different values,

$$Z \in \{ Z_r = \beta_+^r \beta_-^{s-r} / \Gamma^s;\ r = 0, 1, \dots, s \}$$

and the probability to find each one of these values is

$$\rho_r = u^r (1-u)^{s-r} \binom{s}{r}. \tag{3.3}$$

TABLE 1

r	0	1	2	3	4	5	6	7	8
Z_r/N	10.4	2.1	0.4	0.08	0.016	0.003	0.0006	0.00013	0.000026
ρ_r	$\frac{1}{256}$	$\frac{8}{256}$	$\frac{28}{256}$	$\frac{56}{256}$	$\frac{70}{256}$	$\frac{56}{256}$	$\frac{28}{256}$	$\frac{8}{256}$	$\frac{1}{256}$

Because the events are independent, the level-2 function ruling the large deviations of the empirical measure $\{\nu_r\}$ from the probability $\{\rho_r\}$ is

$$I^{(2)}(\nu, \rho) = \sum_{i=1}^{r} \nu_i \log\left(\frac{\nu_i}{\rho_i}\right) \quad \text{if } \nu \ll \rho$$

$$= \infty \qquad \text{otherwise.} \tag{3.4}$$

As an example consider $\beta_+ = \frac{1}{3}$, $\beta_- = \frac{5}{3}$, $u = \frac{1}{2}$ and $s = 8$. Then we get the results of table 1,where we have divided Z by a normalisation factor so that the average number of particles per bin is 0.21, of the order of the well-known NA22 data. Typically the most probable event (Z_4) is well below the average and the most exceptional one has an average of around 10 particles.

To estimate the likeliness of observing large deviations from the probability measure $\{\rho_r\}$, i.e. exceptional events in finite samples, we have considered the following one-dimensional cuts of the deviation function:

$$I_{c_i} = \nu_i \log\left(\frac{\nu_i}{\rho_i}\right) + (1-\nu_i) \log\left(\frac{1-\nu_i}{\rho_4}\right), \tag{3.5}$$

i.e. the deviation function associated to the occurrence of the empirical measures

$$\nu = (0, \ldots, \nu_i, \ldots, 1-\nu_i, \ldots, 0),$$

where the term $1-\nu_i$ is always placed in the position 4 corresponding to the most probable event. In fig. 1a are plotted I_{c_0} and I_{c_1}, i.e. the functions associated to large deviations to the two events with the largest number of particles. The logarithmic dependence implies a behaviour near the most probable value flatter than in gaussian processes. However the monotonic relatively large growth of the functions implies that exceptional events (with many particles) in finite samples are not very probable.

It would however be misleading not to emphasise that a multiplicative model does have large deviations from the asymptotic measure. What happens however is that the finite-sample effect emphasises large deviations towards values smaller than the asymptotic ones, rather than towards large exceptional events. This is

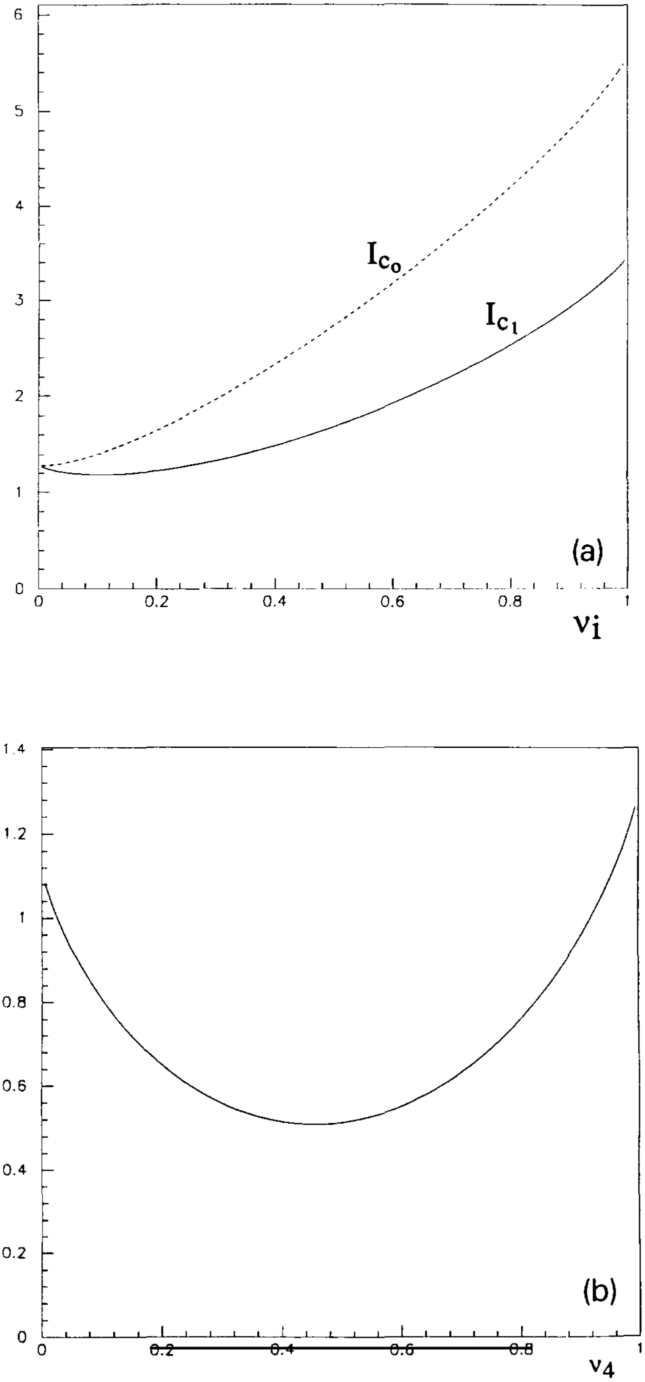


Fig. 1. (a) Deviation functions for large events in the cascading model. (b) Deviation function in the cascading model emphasizing events below the average.

seen for example in fig. 1b where we have plotted the one-dimensional cut of the deviation function corresponding to the empirical measure

$$\nu = \left(0, 0, 0, \tfrac{1}{3}(1-\nu_4), \nu_4, \tfrac{2}{3}(1-\nu_4), 0, 0, 0\right),$$

which emphasises the occurrence of Z_4 events. Recall that the Z_4 events are twelve times smaller than the asymptotic average. One sees that the deviation function, in the region plotted in fig. 1b, has a small value over a large range of ν_4, with a minimum well above the asymptotic Z_4 probability (which is 0.27).

3.2. LARGE DEVIATIONS IN SPIN MODELS

Here we consider a one-dimensional spin model with k spins associated to a bin and the convention that the number of spins that points upwards in each bin represents the number of particles in that bin. We will consider the hamiltonians of the Ising and the Curie–Weiss models,

$$H_{\text{Ising}} = -\tfrac{1}{2}J\sum_{\langle ij\rangle} s_i s_j - h\sum_i s_i, \tag{3.6}$$

$$H_{\text{CW}} = -\frac{J}{2n}\sum_{ij} s_i s_j - h\sum_i s_i, \tag{3.7}$$

J is the ferromagnetic coupling and h an external magnetic field.

We want to study as before the (level-2) large deviations from the probability measure $\rho = \{\rho_i\}$ in state space Y, i.e. deviations from the probability for the number of particles per bin. With k spins per bin the state space Y will have $k+1$ elements.

The level-2 large-deviation probability is determined by the finite-volume Gibbs state summed over all configurations ω for which the n-empirical measure is in the interval $(\nu, \nu + \mathrm{d}\nu)$,

$$Q_n^{(H)}\{\mathrm{d}\nu\} = \int_{\{\omega:\, L_n(\omega,\,\cdot)\in \mathrm{d}\nu\}} \frac{\exp[-\beta H_n(\omega)]}{\int_{\{\omega\}} \exp[-\beta H_n(\omega)] Q_n\{\mathrm{d}\nu\}} Q_n\{\mathrm{d}\nu\}, \tag{3.8}$$

where $Q_n\{\mathrm{d}\nu\}$ is the corresponding probability for independent events. To estimate the deviation function $I^{(2)}(\nu, \rho)$ from eq. (3.8) using theorem 6 of sect. 2 one needs the leading, in n, term of $H_n(\omega)$. However, because in eq. (3.8) one has still to sum over all ω-configurations with the same empirical measure, it will suffice to estimate the average leading n-term of H_n over these configurations. For an empirical measure $\nu = \{\nu_i\}$ and a sample of size-n bins there will be

$$n \sum_{i=0}^{k} \nu_i (k-i) \quad \text{spins pointing upwards and}$$

$$n \sum_{i=0}^{k} \nu_i i \quad \text{spins pointing downwards.}$$

To estimate the leading n-term in H_n averaged over all events with the same empirical measure, we may use a mean-field or a probabilistic method and compute for each spin in the sample the probability that its neighbours point in the same or the opposite direction. It turns out therefore that for the average leading n-term the result in this approximation is the same for the Ising and the Curie–Weiss models. The result is

$$\overline{H}_n = n \left[\frac{J}{2k} \left(\sum_{i=0}^{k} \nu_i (k-2i) \right)^2 - h \sum_{i=0}^{k} \nu_i (k-2i) \right] + o(n). \tag{3.9}$$

Using eq. (3.9), theorem 6 of sect. 2 and the probability $\rho = \{\rho_i\}$ for independent events (with k spins per bin),

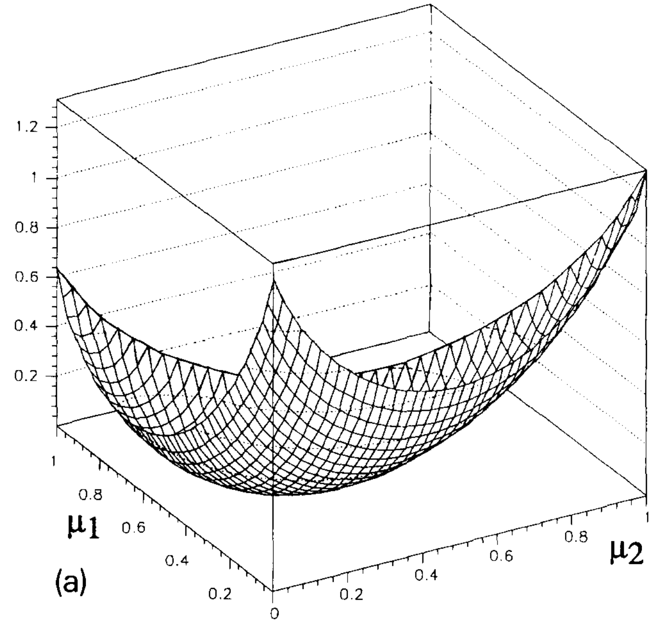
$$\rho_i = \left(\frac{1}{2} \right)^k \binom{k}{i}, \tag{3.10}$$

one obtains

$$I^{(2)}(\nu, \rho) = \sum_{i=0}^{k} \nu_i \log \frac{2^k \nu_i}{\binom{k}{i}} - \beta \frac{J}{2k} \left(\sum_{i=0}^{k} \nu_i (k-2i) \right)^2$$
$$- \beta h \sum_{i=0}^{k} \nu_i (k-2i) - \inf\{\dots\}, \tag{3.11}$$

the expression $\inf\{\dots\}$ meaning the infimum over all ν of the first three terms. This simply adjusts to zero the minimum of $I^{(2)}(\nu, \rho)$.

To understand the meaning of this deviation function we have plotted its variation for several values of the coupling constant and the external magnetic field. Let $k=2$, i.e. two spins per bin. In fig. 2a we plot $I^{(2)}(\nu, \rho)$ for $\beta J = 0$ and $\beta h = 0$. The notation is $\mu_i = \nu_{k-i}$ and, because ν_i means the probability of i spins pointing down in a bin, in our particle interpretation μ_2 means the probability of having two particles in a bin and μ_1 the probability of having one particle. One sees that without coupling nor magnetic field there is a large probability for one and two particles in a bin but that the deviation function grows uniformly from the



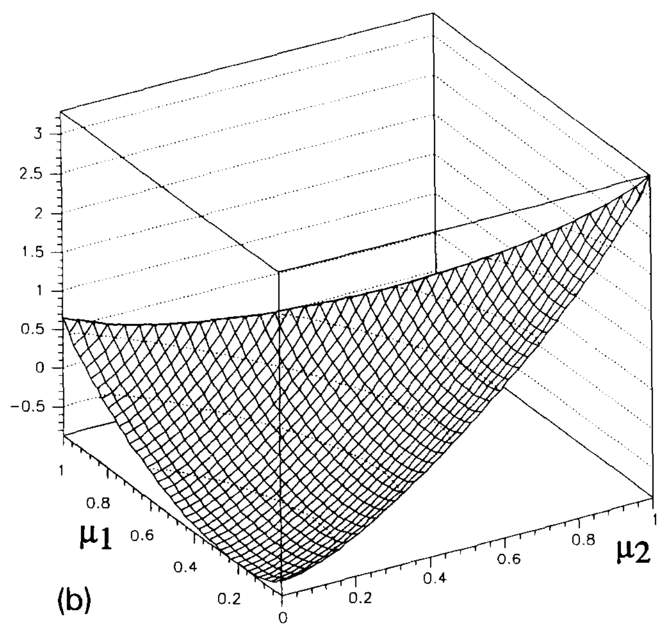


Fig. 2. Level-2 deviation function for two spins per bin, (a) $\beta J = 0$ and $\beta h = 0$; (b) $\beta J = 0$ and $\beta h = -1$; (c) $\beta J = 0.4$ and $\beta h = -1$; (d) $\beta J = 4$ and $\beta h = -1$; (e) $\beta J = 12$ and $\beta h = -1$.

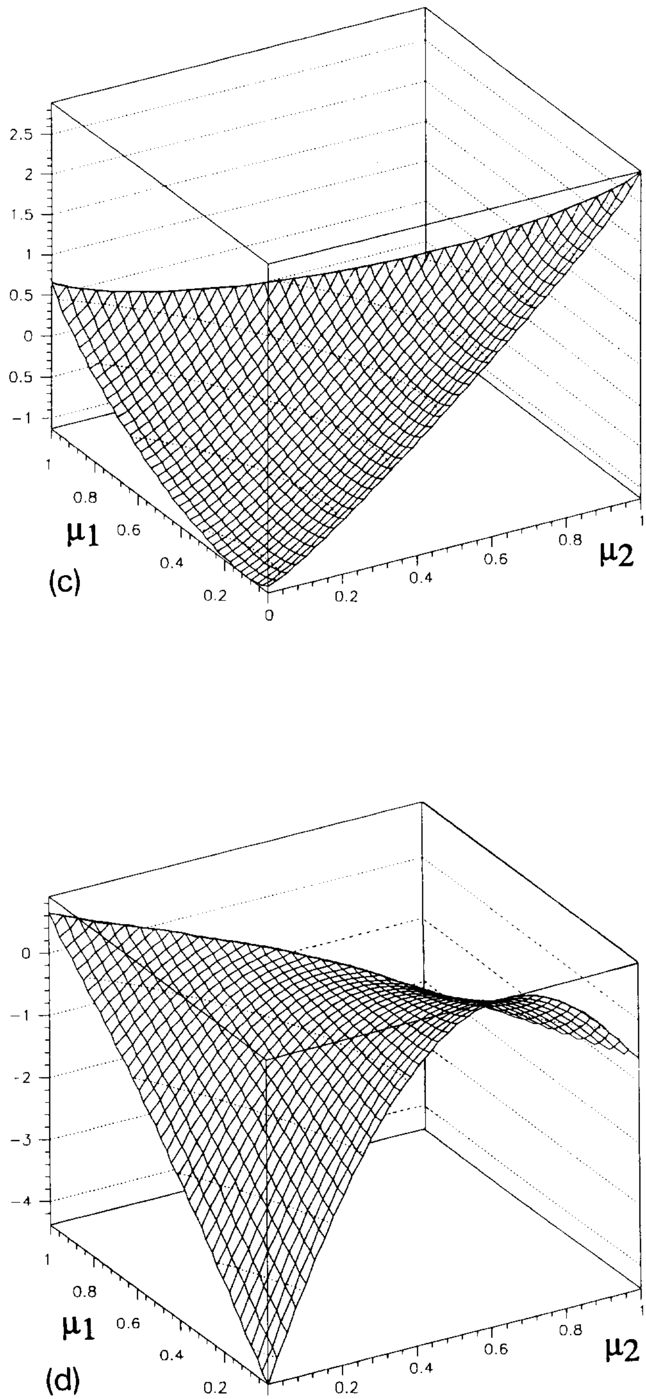


Fig. 2. (continued)

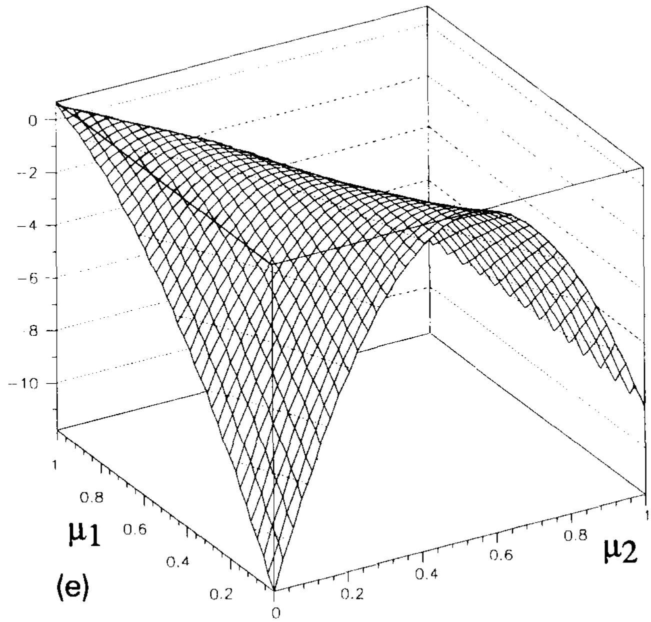


Fig. 2. (continued)

minimum value and large deviations from the most probable values in finite samples are strongly suppressed.

To simulate the particle production situation where typically the average number of particles per bin is small we have in figs. 2b–2e considered a strong magnetic field ($\beta h = -1$) pointing in the downwards direction. In fig. 2b there is only the magnetic field and no coupling. The probability of one particle per bin becomes small and for two particles per bin it is almost zero. Also the deviation function grows steeply and monotonously from the minimum point and large deviations will be strongly suppressed. In fig. 2c we have the same magnetic field and a small ferromagnetic coupling $\beta J = 0.4$. One sees that the situation is similar to the case of no coupling. In fact the probability of having one particle per bin is even smaller than with no coupling. This is easy to understand because the coupling introduces a pairing interaction and, because most spins are pointing downwards because of the magnetic field, it does not pay energetically to have an isolated spin pointing upwards. However, when the coupling is increased, the situation is completely different as shown in figs. 2d, 2e for $\beta J = 4$ and $\beta J = 12$. The most probable values are essentially at zero and the deviation function is even steeper near the origin. However, along the two-particles direction, the deviation function becomes small again and large deviations for finite samples become likely.

In fig. 3 we have considered the case of five spins per bin ($k = 5$) and plotted for $\beta h = 1$ and $\beta J = 12$ the one-dimensional cuts of $I^{(2)}(\nu, \rho)$ for the empirical

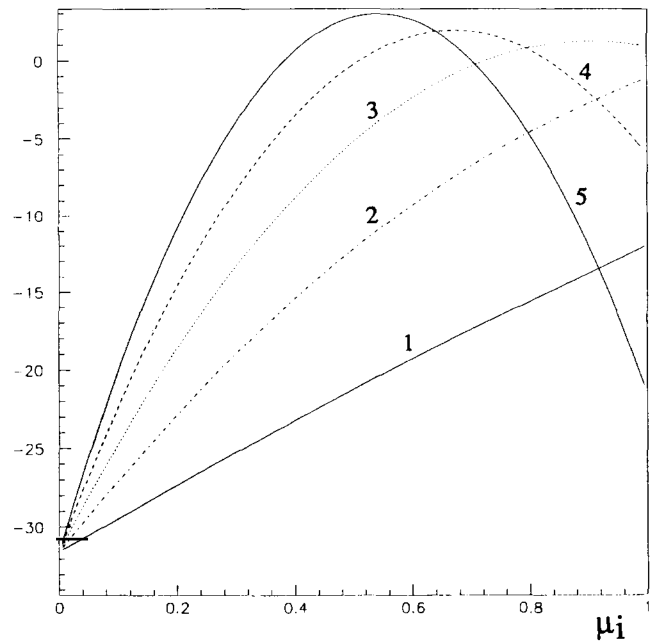


Fig. 3. One-dimensional cuts of the deviation function for five spins per bin, $\beta J = 12$ and $\beta h = -1$.

measure of the form $\nu = \{0, \dots, \nu_i, \dots, 1 - \nu_i\}$. We see once again that large deviations are likely and that are most likely for the most exceptional events, i.e. five particles per bin (ν_0 or μ_5). This is easy to understand physically because, due to the ferromagnetic coupling, if there is a deviation, what pays, in energetic terms, is to have a really large deviation.

The inflection points in the deviation function reflect the fact that the long-range interaction of the Curie–Weiss model (or the mean-field Ising) is compatible with the existence of several phases. Then the free energy has points of non-differentiability and the deviation function, not being a Legendre transform (theorem 3 in sect. 2) need not be convex. For a short-range one-dimensional ferromagnetic model, one would have the inflection replaced by a flat region in the deviation function.

In conclusion: the random cascading model and the ferromagnetic spin model (for an appropriate set of parameters) both have events far from the average and clustering effects. However, our results show that they can be distinguished by their large-deviation behaviour, models of the spin type seeming to be more prone to large deviation to large exceptional events in finite samples.

Of course for particle physics the challenge is to decide, from QCD or from QCD inspired models, what is the structure, both geometrical and probabilistic, of the multiparticle process. The models that we studied had the main merit of displaying diversified behaviour in their multiparticle distribution behaviour and statistical properties. Rather than analysing the proposed particle models which are still at a phenomenological level, we will take the point of view that, in addition to the factorial moments which characterise in a certain sense the geometrical or dimensional structure of the data, useful insight on the production process might be obtained if one were able to extract the deviation function from the data itself. A method to carry out large-deviation analysis from the data is described in sect. 4.

4. Large-deviation tools for data analysis

There are several ways to analyse a set of data points from the point of view of large deviations. One may for example look in a small domain $(\nu, \nu+\Delta\nu)$ in probability space and find for each sample size n the number of times that the empirical measure falls in that interval. This should of course be repeated for other small domains in probability space. The rate of variation of the logarithm of that number with n gives the deviation function at ν. This should of course be repeated for other small domains in probability space to obtain the deviation function throughout the space of probability measures. This method has however the shortcoming that, each time the analysis is performed, only a small amount of the available data is used and the statistics is poor.

A better method is to construct the free energy

$$c(t) = \lim_{n\to\infty} \frac{1}{n} \log \mathrm{E}^{P_n}\{\exp(t, W_n)\}, \qquad t \in \mathrm{Y}, \tag{4.1}$$

and obtain the deviation function from the Legendre transform,

$$I(x) = \sup_{t\in \mathrm{Y}} \{tx - c(t)\}. \tag{4.2}$$

Of course in eq. (4.1) the $\lim_{n\to\infty}$ is estimated from the approach to the largest possible available sample.

When a deviation function is either obtained from the data or conjectured in some other way, the data provides a consistency check from the fact that the deviation function is the deviation function of its own distribution

$$Q_n(\mathrm{d}I) \asymp \exp(-a_n I)\ \mathrm{d}I. \tag{4.3}$$

This follows from the contraction principle (theorem 5) using I as the function ψ. This property is useful, for example, to check whether the relative entropy

$$I_r = \sum_{i=1}^{r} \nu_i \log\left(\frac{\nu_i}{\mu_i}\right) \tag{4.4}$$

(with μ being the empirical measure computed for the largest possible sample) is a good guess for the deviation function. Recall that the relative entropy coincides with the deviation function if the events are independent (see sect. 2).

The method based on eqs. (4.1) and (4.2) when applied to the numerically simulated data of the spin model [5] yields results compatible with the analytical estimates of subsect. 3.2. Here we present another simple example to show how the large-deviation analysis may provide additional insight on the statistical structure of the data. We have constructed a multiplicative process $\{X_n\}$ with expectation value 1, i.e.

$$\lim_{n\to\infty} \frac{1}{n} W_n = 1,$$

where

$$W_n = \sum_{i=1}^{n} X_i$$

(by the law of large numbers) and used this numerical process to generate sets of data which are then analysed for large deviations. This is done by constructing a numerical approximation to the free energy and its Legendre transform. Fig. 4a shows a typical set of 10^4 events (bin width = 10^{-6}). The sample average in this set is 0.285, very far from the asymptotic value. From a simple examination of the distribution of the data it would be hard to guess that one is so far away from the real average value of the process.

For the numerical computation of the free energy the data is divided into k slices of length n, such that $k \times n$ equals the total number of data points and one considers

$$c_{k,n}(t) = \frac{1}{n} \log\left(\frac{1}{k} \sum_{j=0}^{k-1} \exp\left[t W_j^{(n)} \right] \right), \tag{4.5}$$

where

$$W_j^{(n)} = \sum_{s=jn}^{jn+n-1} X_s .$$

A numerical estimate of the limit (4.1) is obtained by increasing n and decreasing k maintaining the product fixed. In practice this should stop at a level where there are yet a sufficient number of data slices to have reasonable statistics in the estimation of $\mathrm{E}\{\exp(t, W_n)\}$. In our example we have stopped at $k = n = 100$. Once the free energy is computed the Legendre transform is easily computed from eq. (4.2).

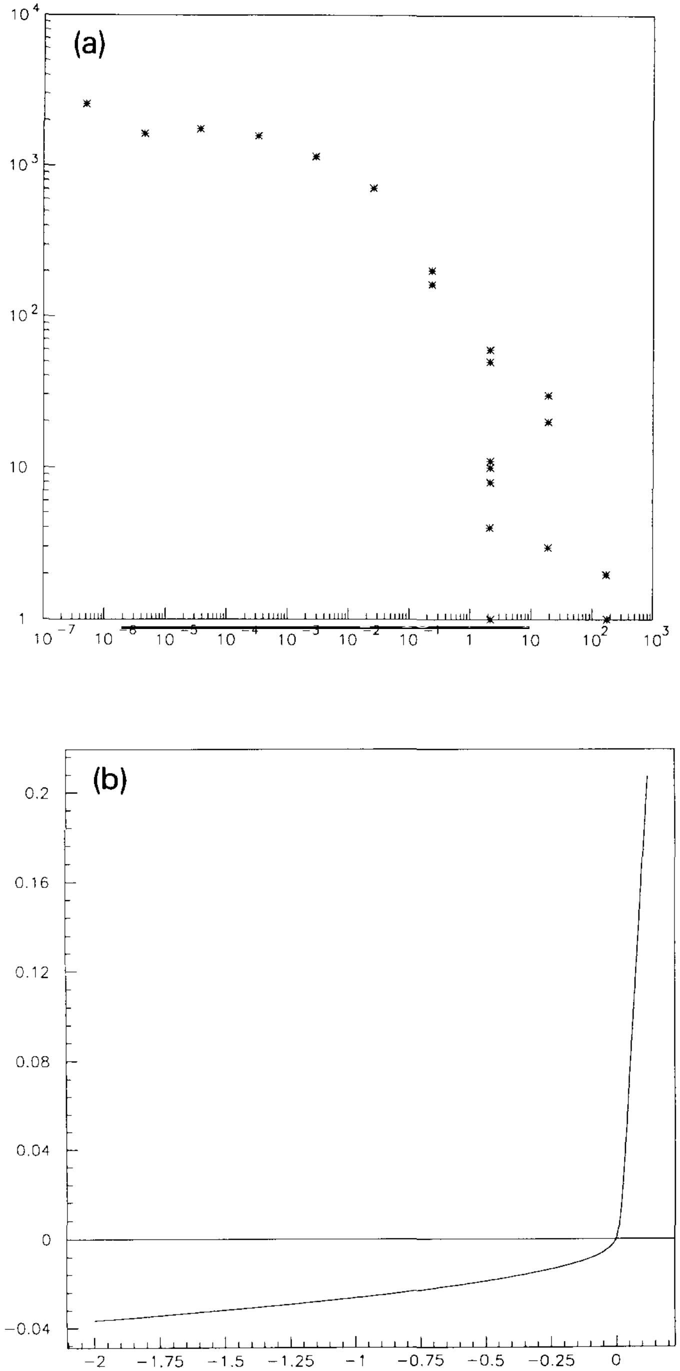


Fig. 4. (a) Set of 10^4 events of a multiplicative process with $\langle X \rangle = 1$. Sample average = 0.285. Bin width = 10^{-6}. (b) Free energy estimated from the data in (a). (c) Deviation function estimated from the data in (a).

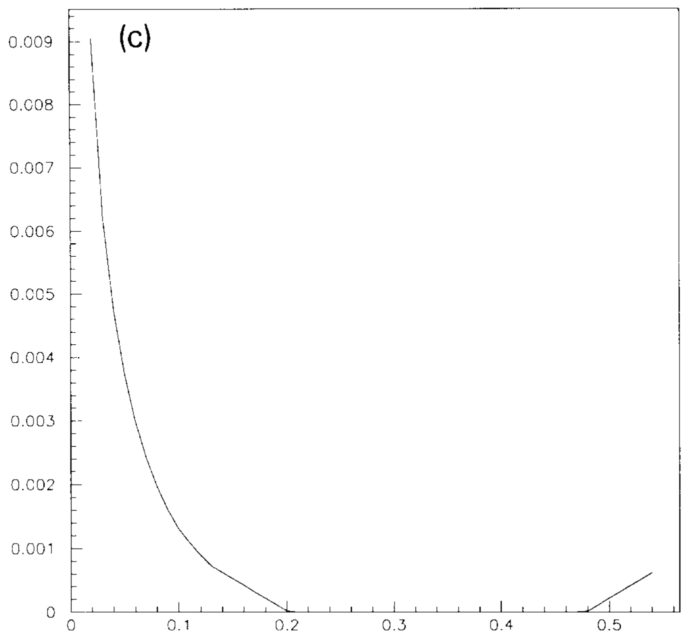


Fig. 4. (continued)

Figs. 4b, 4c show the numerical computed estimates of the free energy and the deviation function obtained from the data in fig. 4a. Recall (from sect. 2) that the exact deviation function should have a unique minimum at the average value of the process. However, the function that is obtained has instead a large flat minimum region. This shows that the sample average cannot be guaranteed to be close to the actual average value. This one could be anywhere in the flat region or above. The solid conclusion therefore is that there is insufficient data to make any statement about the average value of the process. This conclusion is very clear from the examination of the estimate of the deviation function whereas it might be debatable, at first sight, from the data distribution. There is however nothing magic in this result. It is simply the effect of the exponential sum in eq. (4.5) which as soon as t becomes positive emphasizes large events, whenever they exist, and makes $c(t)$ grow very fast for $t > 0$.

The example we have presented deals with the numerical computation, from the data, of level-1 large-deviation properties. For level-2 and level-3 the method is similar, choosing for W_n the appropriate random vector associated to the slicing of the data

References

[1] A. Bialas and R. Peschanski, Nucl. Phys. B273 (1986) 703; B308 (1988) 857
[2] S.R.S. Varadhan, Large deviations and applications (SIAM, Philadelphia, 1984)
[3] D.W. Stroock, An introduction to the theory of large deviations (Springer, Berlin, 1984)
[4] R.S. Ellis, Entropy, large deviations and statistical mechanics (Springer, Berlin, 1985)
[5] J. Seixas, Mod. Phys. Lett. A6 (1991) 1237