CERN TH.5860/90

INTERMITTENCY AND THE CENTRAL LIMIT THEOREM

J. M. Alberty* and A. Bialas**

CERN
Geneva, Switzerland

Abstract

Central limit theorem estimates of anomalous fractal dimensions of self similar random cascades are studied. It is found that, in general, the normal approximation fails badly. A systematic series of approximations which converges to the exact result (both for the fractal dimensions and for the distribution itself) is derived for the α -model. Consequences for the "empty bin effect" are indicated.

(*) Supported by the World Laboratory/HED and the CERN/LAA Projects.

(**) On leave of absence from Institute of Physics, Jagellonian University, Reymonta 4, PL 30059 Cracow, Poland.

1 Introduction

Observation of intermittency in the spectra of particles created in high-energy collisions (for a recent review of experimental evidence, see [1]) focused attention on the cascade nature of the process of particle production. Indeed, as pointed out already in the first papers on this subject [2,3,4], self-similar cascades provide the most natural framework to discuss intermittency. In particular, the power law dependence of factorial moments F_q on the experimental resolution δ

$$F_q(\delta) = F_q(\Delta)(\Delta/\delta)^{(q-1)d_q} \quad (1)$$

where d_q is an anomalous fractal dimension of rank q , is an immediate consequence of the self-similarity in a long cascade [2]¹. Δ is an initial rapidity scale.

It was shown in [3] that, if the underlying self-similar cascade can be well approximated by the central limit theorem, the anomalous fractal dimensions d_q obey the simple linear relation

$$d_q = \frac{1}{2}qd_2 \quad (2)$$

The existing data agree approximately with (2) [1] thus suggesting that indeed (a) a cascade mechanism is responsible for observed effects and (b) that this cascade is not far from its central limit theorem approximation.

At this point it is important to realize that the remark (b) above is not as trivial as it seems. The point is that, although the central limit theorem gives a very good approximation to the distribution not far from its average value [6], it does not (in general) describe correctly the tails of the distribution². It is therefore rather surprising that the normal

¹Distributions which obey (1) with $d_q > 0$ are called fractals [5]. Since they are discontinuous in the limit of very high resolution ($\delta \rightarrow 0$) they are well suited for the description of spectra with large and very narrow peaks and dips, typical for the intermittent behaviour.

²The limits of applicability of the Central limit theorem in the case of random multiplicative processes were recently pointed out in [7], where the important contribution of the rare events to the higher moments of the random products were correctly emphasized. However, the author claimed that the continuum distribution of the product cannot be represented as a single function. In the present work we show that this can be done, and we explicitly construct the distribution.

approximation describes well the data on intermittency, which are known to be particularly sensitive to very large deviations from the average.

In the present paper, using a simple random cascade model which already proved useful in many studies of intermittency, we consider this problem in some detail. Since the model can be solved exactly, it allows one to study unambiguously different approximations involved in application of the central limit theorem. We first discuss the general form of the random cascading model and show explicitly that the application of the central limit theorem for estimates of anomalous dimensions may be very misleading [7]. We then discuss corrections to the central limit theorem which may improve this situation. In the special case of the α -model [2] such correction terms can be explicitly written down, using the well-known generalizations of the central limit theorem which include terms taking into account the possibility of large deviations from the average. We also show that, when all corrections are summed up, one recovers the exact result, thus proving the consistency of the procedure. We then study numerically the approximations obtained to determine how they approach the exact solution for a few moments of low rank. We observe that although the first correction already improves the approximation drastically, in general the series is slowly convergent to the exact value.

To summarize, we find that, although the normal approximation describes well self-similar random cascades in the region close to the average [6], it generally fails to account for fractal properties of the distribution (which are known to be sensitive to large deviations from the average). This is fortunate, because it shows that, by studying the tails of particle distributions, one has a chance to learn about the details of the production process, not masked by general statistical properties of random variables.

In the next section we recall the construction and properties of the random cascade model and compare its anomalous dimensions with those obtained from the central limit theorem. In Section 3 the corrections to the normal approximation to the α -model are discussed and in Section 4 some of our conclusions are spelled out.

2 The random cascade model and the central limit theorem

Let us begin by defining the random cascade model we are going to study [2,8]. It has a tree structure with λ branches at each splitting. The average density in any of the bins of size $\delta_{\min}$ at the bottom of the cascade is a product of N positive random factors (denoted by w) ascribed to the N links forming a path leading from the origin of the cascade to a given bin:

$$\rho(\delta) = w(1) \dots w(N) \rho(\Delta) \quad (3)$$

where $\rho(\delta)$ is the density in a bin of size δ and $\rho(\Delta)$ is the density in the bin of size Δ , with $\delta/\Delta = 1/\lambda^N$.

The distribution $\tau(w)$ of these random factors is arbitrary but it is the same for all links (self-similarity). Different factors are independent of each other. For simplicity we normalize w by requiring

$$\int \tau(w) w dw = 1 \quad (4)$$

All these relations are illustrated in Fig.1³.

The anomalous fractal dimensions of the random cascading model can be easily evaluated [2,3] with the result⁴

$$(q-1)d_q = \frac{\ln \langle w^q \rangle}{\ln \lambda} \quad (5)$$

To apply the central limit theorem to the distribution of $\rho(\delta)$ we define

$$\xi = \ln \rho(\delta) \quad (6)$$

³At this stage only particle densities in each bin are determined. In the present paper we need not go beyond that. Let us only mention that the particle distributions can be constructed by requiring that in any bin of size δ in which the particle density equals ρ the distribution of particles is given by the Poisson formula with the average $\rho\delta$.

⁴The eq. (5) describes the anomalous fractal dimensions as long as $d_q \leq 1$. For $d_q > 1$ the meaning of d_q is different. See [8] for a detailed discussion of this point.

It follows from (3) that ξ is a sum of N independent random variables and thus its distribution can, according to the central limit theorem, be approximated by a normal distribution:

$$P(\xi)d\xi = \frac{d\xi}{\sqrt{2\pi N\sigma^2}} \exp \left[-\frac{(\xi - N\xi_0)^2}{2N\sigma^2} \right] \quad (7)$$

with

$$\xi_0 = \langle \ln w \rangle \quad (8)$$

and

$$\sigma^2 = \langle (\ln w)^2 \rangle - \langle \ln w \rangle^2 \quad (9)$$

The corresponding distribution for ρ (obtained from (7) using (6)) is called the log-normal distribution.

Let us now discuss the anomalous dimensions of the distribution of density following from (7) [3]. To this end we first evaluate the moments $\langle \rho^q \rangle$. A straightforward integration gives

$$\langle \rho^q \rangle = \langle \exp(q\xi) \rangle = \exp \left\{ \frac{N\sigma^2}{2} q \left(q + \frac{2\xi_0}{\sigma^2} \right) \right\} \quad (10)$$

This implies

$$d_q = \frac{\sigma^2}{2 \ln \lambda} \frac{q}{(q-1)} \left(q + \frac{2\xi_0}{\sigma^2} \right) \quad (11)$$

Comparing (11) with the exact formula (2) one sees immediately that they are in general rather different. Nevertheless, it is not difficult to show [9,10] that they give similar results provided the fluctuations of the variable w are small. To be more precise, if we write

$$w = 1 + \tau \quad (12)$$

then the condition (4) implies $\langle \tau \rangle = 0$, and up to second order in τ one obtains

$$\ln(\langle w^q \rangle) \simeq \ln \left(1 + \frac{1}{2} q(q-1) \langle \tau^2 \rangle \right) \simeq \frac{q(q-1)}{2} \langle \tau^2 \rangle \quad (13)$$

From this equation, using (5) one recovers easily the normal approximation (2). Thus, as long as we can neglect higher order terms in τ the normal approximation describes correctly the fractal structure of the distribution.

If the fluctuations are not small, one observes that the eq. (11) gives a finite d_1 only if

$$\xi_0 = -\sigma^2/2 \quad (14)$$

This condition is of course not valid for a general distribution $\tau(w)$. The conclusion is that the approximation (11) can only be maintained if ξ_0 is treated as a free parameter and not if it is calculated according to the central limit theorem from eq. (8). This approach was taken in Ref. [2]. Then, using (14) one obtains (11) with

$$d_q = \frac{q \sigma^2}{2 \ln \lambda} \quad (15)$$

If σ^2 is treated as a free parameter, one obtains the eq. (2) where d_2 is arbitrary. This obviously improves the approximation.

To investigate closer these approximations we considered as an example a random cascade model with

$$\tau(w)dw = \frac{\Gamma(k)}{\Gamma(b)\Gamma(k-b)} \left(\frac{w}{\lambda}\right)^{b-1} \left(1 - \frac{w}{\lambda}\right)^{k-b-1} \frac{dw}{\lambda} \quad (16)$$

where k is a free parameter and $b = k/\lambda$. The variable w is restricted to the interval $0 < w < \lambda$. In this case the moments of the distribution can be explicitly calculated [11] and one obtains

$$(q-1)d_q = \frac{1}{\ln \lambda} \ln \left\{ \frac{[\lambda+k] \dots [(q-1)\lambda+k]}{(1+k) \dots (q-1+k)} \right\} \quad (17)$$

This exact result can be directly compared with the normal approximations (2), (11) and (15).

Such a comparison is shown in Fig.2 for $\lambda = 2$ and $k = 5$. One sees a serious discrepancy between the normal approximation and the exact result even in the case of eq.(2), whereas (11) and (15) are very far from the actual values of anomalous dimensions. We have found that a reasonable approximation is obtained only for $k \geq 50$.

This result shows clearly that, in general, the anomalous dimensions of a cascade process have nothing to do with those calculated from the central limit theorem: it just emphasizes the well-known fact that the moments of a distribution are, in general, not correctly approximated by the central limit theorem. Consequently, the validity of the formula (2) can only be justified if the fluctuations are weak [9,10] or if the fluctuations of the variable w in the cascade are themselves well approximated by the log-normal distribution. It will thus be of great interest to analyse the data on intermittency in two and in three dimensions, where there are indications that the fluctuations are much stronger and thus a real test of log-normal distribution should be possible.

In the next section we discuss some improvements of the central limit theorem which allow a better estimate of the anomalous dimensions of a cascade.

3 Central limit theorem and the α -model

We have shown in the previous section that the central limit theorem does not, in general, describe correctly the anomalous dimensions of a random cascade. The reason is easy to recognize: the central limit theorem describes only the distribution close to the average value and fails badly at the tails. To improve the approximation it is therefore necessary to treat more carefully large deviations from the average. In the present section we discuss this problem for the special case of the random cascades, the so-called α -model [2]. In this model, the random factors w take only two values

$$w = \beta_{\epsilon} \quad (18)$$

where ϵ is a random variable (subscript) taking the value $+$ with the probability u and the value $-$ with the probability $v = 1 - u$. In the following we assume $\beta_+ > \beta_-$. The normalization condition (4) implies

$$u = 1 - v = \frac{(1 - \beta_-)}{(\beta_+ - \beta_-)} \quad (19)$$

Using (5) we obtain for the anomalous dimensions

$$d_\gamma = \frac{\ln \{u(\beta_+)^q + v(\beta_-)^q\}}{(q-1) \ln \lambda} \quad (20)$$

This is the exact formula. We are now going to study how it can be approximated by the central limit theorem and its generalizations.

To this end we observe that the α -model is isomorphic to a chain of N Bernoulli trials: indeed, the probability distribution of k successes out of N Bernoulli trials is the same as the probability that the density ρ at the end of the cascade takes the value

$$\rho(k; N) = (\beta_+)^k (\beta_-)^{N-k} \quad (21)$$

This allows the application of the well-known theory of the Bernoulli trials to discuss the α -model⁵.

For a long chain of Bernoulli trials, the probability of k successes can be approximated by the formula [13]

$$P(k; N) \sim \exp[-\Phi(z)] \quad (22)$$

where

$$\Phi(z) = Nuv \sum_{\nu=2}^{\infty} \frac{u^{\nu-1} - (-v)^{\nu-1}}{\nu(\nu-1)} z^\nu \equiv Nuv \sum_{\nu=2}^{\infty} a_\nu z^\nu \quad (23)$$

with

$$z = (k - Nu)/Nuv \quad (24)$$

This formula gives a good approximation as long as z is a small number (the formula is exact in the limit $z \rightarrow 0$). If only the first term is kept, one obtains the normal approximation. One sees that its validity requires a much stronger condition

$$Nuvz^3 \rightarrow 0 \quad (25)$$

and thus it is not valid in the tail of the distribution. This is precisely the reason for the problems encountered in the evaluation of the moments.

⁵Another generalization of the normal approximation to intermittent spectra (based on the Levy-stable laws) was proposed recently in [12].

One sees that eq. (22) represents a scheme of systematic approximations to $P(k; N)$. By retaining more terms in the sum (23), we obtain formulae which are valid for larger deviations of k from its average value (Nu). To be more specific, if ν_0 terms are kept, the formula is correct in the region where

$$Nuvz^{\nu_0+1} \rightarrow 0 \quad (26)$$

One may thus hope that such improved formulae will also provide better estimates of the moments of the distribution than the normal approximation. We are now going to study this problem.

Let us first rewrite the formula (22) in a form which is more suitable for the discussion of the α -model. Introducing the variable

$$\xi = \ln \rho \quad (27)$$

we obtain

$$P(\xi) \sim \exp[-\Phi(y)] \quad (28)$$

where

$$y = \frac{\xi - N\xi_0}{Nuv \ln(\beta_+/\beta_-)} \quad (29)$$

and

$$\xi_0 = \langle \ln w \rangle = u \ln \beta_+ + v \ln \beta_- \quad (30)$$

To calculate the moments $\langle \exp(q\xi) \rangle$ of this distribution we use the saddle-point method (which is exact in the limit of large N). Let us first see what happens when the series (23) is truncated at some value of $\nu = \nu_0$. In this case the equation for the saddle-point reads

$$\sum_{\nu=2}^{\nu_0} \nu a_\nu y^{\nu-1} = q \ln(\beta_+/\beta_-) \quad (31)$$

and in the large N limit we obtain for the moments

$$\langle \exp(q\xi) \rangle = \exp \left[Nuv \left(- \sum_{\nu=2}^{\nu_0} a_\nu y_\nu^* + qy, \ln(\beta_+/\beta_-) + \frac{q\xi_0}{uv} \right) \right] \quad (32)$$

where $y_s = y_s(q)$ is the solution of the saddle-point equation (31). Consequently, using (5), the anomalous dimensions are given by

$$d_q = \frac{uv}{(q-1)\ln\lambda} \left[\frac{q\xi_0}{uv} + qy_s \ln(\beta_+/ \beta_-) - \sum_{v=2}^{\nu_0} a_v y_v^p \right] \quad (33)$$

The saddle-point equation is an algebraic equation of order $\nu_0 - 1$. For $\nu_0 = 2$ one recovers the result from normal approximation. For $\nu_0 = 3$ the result is

$$d_q = \frac{1}{(q-1)\ln\lambda} \left\{ \frac{q}{2} \ln(\beta_+ \beta_-) + \frac{uv}{3(u-v)^2} \{ (q-1) + (1+2q(u-v)\ln(\beta_+/\beta_-))^{3/2} - (1+2(u-v)\ln(\beta_+/\beta_-))^{3/2} q \} \right\} \quad (34)$$

For larger values of ν_0 (31) must be solved numerically. Without an explicit solution, however, one can still derive the asymptotic behaviour of anomalous dimensions in the limit of large q . Indeed, in this limit (31) gives

$$y_s \sim q^{1/(\nu_0-1)} \quad (35)$$

Substituting this relation into (33) we obtain ⁶

$$d_q \sim q^{1/(\nu_0-1)} \quad (36)$$

Thus we see that, in the limit $\nu_0 \rightarrow \infty$, d_q tends to a constant value at large q , as required by the exact formula (20).

Finally, let us remark that, if all terms in the sum (23) are kept, the formula gives exact values of all moments (in the limit of large N). To see this, let us observe that the series (23) can be summed up with the result [14]

$$\Phi(y) = Nuv \left[\left(\frac{1}{u} - y \right) \ln(1 - uy) + \left(\frac{1}{v} + y \right) \ln(1 + vy) \right] \quad (37)$$

⁶This is a special case of the general asymptotic behaviour discussed in [12].

and thus the probability $P(\xi)$ can be written in the closed form:

$$\ln[P(\xi)] = \frac{\xi - N \ln \beta_+}{\ln(\beta_+/\beta_-)} \ln \left[\frac{N \ln \beta_+ - \xi}{Nv \ln(\beta_+/\beta_-)} \right] + \frac{N \ln \beta_- - \xi}{\ln(\beta_+/\beta_-)} \ln \left[\frac{\xi - N \ln \beta_-}{Nu \ln(\beta_+/\beta_-)} \right] \quad (38)$$

The saddle-point equation (31) gives then

$$-\ln \left[\frac{u N \ln \beta_+ - \xi_s}{v \xi_s - N \ln \beta_-} \right] = q \ln(\beta_+/\beta_-) \quad (39)$$

from which one easily finds the solution ξ_s . The moments are obtained by substituting this solution into the formula

$$\langle \exp(q\xi) \rangle = P(\xi_s) \exp(q\xi_s) \quad (40)$$

It is not difficult to verify that they coincide with the ones given by the exact formula (10).

In Fig.3 the anomalous dimensions of the α -model calculated from the exact formula and from the truncated series (31) are plotted versus the rank, using two different sets of model parameters. One sees again that the normal approximation fails badly. One sees also that already taking $\nu_4 = 0$ improves the approximation a good deal. Unfortunately, further terms are not so effective and one may expect that many of them are needed to recover the exact result.

In Fig.4 we compare directly the exact distribution (38) and the central limit theorem results for $\nu_0 = 2$ and $\nu_0 = 4$ with the distribution obtained from 5000 Monte-Carlo runs. Three points are to be noted:

(a) One sees that for small deviations, the normal distribution is indeed close to the actual one, whereas deviations occur at the tails.
(b) The deviations are substantial when $u \neq v$, whereas they are barely visible for $u = v$.

(c) Although in the case when $u = v$ the deviations can hardly be seen, they nevertheless lead to large discrepancies when the moments $\langle \exp(q\xi) \rangle$ are calculated (as seen in Fig.2).

4 Discussion and conclusions

We have studied the application of the central limit theorem to random cascade models.

It is known that the central limit theorem describes well the distribution of a sum of independent random variables close to its average value but fails, in general, at the tails. It was thus interesting to discuss the fractal properties of random cascades which are particularly sensitive to large deviations from the average.

Our conclusions can be summarized as follows:

(i) For a general random cascade model the log-normal approximation fails badly in describing the anomalous fractal dimensions unless (a) the fluctuations are very small or (b) the cascade itself is of very special type (log-normal cascade).

(ii) For the α -model, it is possible to construct the series of approximations which converge to the exact values of the anomalous dimensions. Unfortunately the series does not seem to converge very fast.

(iii) An explicit closed expression for the distribution in a general α -model (in the limit of a very long cascade) was found and compared successfully with the Monte-Carlo simulation.

(iv) The deviations of the log-normal approximation from this exact expression were studied. It was found that even seemingly unimportant deviations at the tails of the distribution lead to large discrepancies in the calculated fractal dimensions.

(v) This last point may be of practical importance: It confirms that the evaluation of moments of the distribution from experimental data may be very sensitive to the statistics reached in an experiment ("empty bin" effect [15]). Indeed, with increasing statistics one can reach larger deviations from the average value and thus one is more sensitive to the tails of the distribution [7]. Our calculation shows that, at least in the case of the α -model, this may lead to rather large effects. A more detailed study of this problem will be given elsewhere.

5 Acknowledgements

AB thanks A. Szerba for enlightening discussions on applications of the central limit theorem, R. Buschbeck and A. Kittel for informing us of their results prior to publication, and E. De Wolf and A. Wroblewski for pointing out the ref. [7]. JMA thanks R. Peschanski for very stimulating discussions on the α -model, and R. Ribeiro for his help in the use of a graphics package.

References

- [1] For a recent review of experimental and theoretical work on intermittency, see A. Bialas, "Intermittency '90", *Preprint CERN-TH.5791/90*, to be published in *Nucl. Phys. B* (Proc. Suppl.).
- [2] A. Bialas and R. Peschanski, *Nucl. Phys. B* **273** (1986) 703.
- [3] A. Bialas and R. Peschanski, *Nucl. Phys. B* **308** (1988) 857.
- [4] W. Ochs and J. Wosiek, *Phys. Lett. B* **214** (1988) 617; **B232** (1989) 271.
- [5] B. Mandelbrot, *The Fractal Geometry of Nature*, Freeman, N. Y. 1982.
- [6] For a recent application of this property, see R. Szved, G. Wrochna and A. K. Wroblewski, "Genesis of the lognormal multiplicity distribution in the e^+e^- collisions and other stochastic processes", Warsaw Univ. preprint IFD/1/1990 (March 1990).
- [7] S. Redner, *Am. J. Phys.* **58** (1990) 267.
- [8] A. Bialas and K. Zalewski, *Phys. Lett. B* **228** (1989) 155.
- [9] A. Capella, K. Fialkowski and A. Krzywicki, *Phys. Lett. B* **230** (1989) 271.
- [10] D. Seibert, *Phys. Lett. B* **240** (1990) 215.
- [11] A. Bialas and R. Peschanski, *Phys. Lett. B* **207** (1988) 59.
- [12] Ph. Brax and R. Peschanski, "Levy Stable Law Description of Intermittent Behaviour and Quark-Gluon Plasma Phase Transitions", *Preprint Saclay SPHT/90-102* (1990).
- [13] W. J. Feller, *An Introduction to Probability Theory and its Applications*, 3rd ed., Vol. 1, Wiley, New York, 1968.
- [14] J. M. Albery, "Intermediate Asymptotics of Random Walks", *Preprint CERN 1990*, to appear.
- [15] B. Buschbeck and W. Kittel, private communication.

Figure Captions

Fig.1

Illustration of the cascading mechanism through a collection of branched and non-overlapping paths (tree), depicted as vertical arrows, from the initial rapidity interval Δ to the final rapidity bins $\delta \equiv \delta_N$. In the simplest case represented here, each rapidity interval is split into $\lambda = 2$ smaller ones at the next step. Along each path, the density is changed by a random factor $w(j) = \beta \delta(j)$ (with $\delta(j) = \pm 1$) at each step j ($j = 1, 2, \dots, N$).

Fig.2

The anomalous dimensions of the random cascade model with $r(w)$ given by (16) ($\lambda = 2, k = 5$), calculated from the exact formula and from the normal approximation using Eqs. (2), (11) and (15), are plotted versus the rank q . The failure of the normal approximation to describe the exact values of d_q is clear.

Fig.3

The anomalous dimensions in the α -model calculated from the exact formula and from the truncated series (31), are plotted versus the rank q . One sees a serious discrepancy between the exact result and the normal approximation ($\nu_0 = 2$). One can also observe that the inclusion of higher terms improves the agreement:

- (a) With $\beta_+ = 1.5, \beta_- = 0.5, \lambda = 2$;
- (b) With $\beta_+ = 1.1, \beta_- = 0.1, \lambda = 2$.

Fig.4

Histogram of the number of random walks ending ξ steps away from the origin, after a total number of steps equal to N . We also plot the exact

distribution (38) (solid line) and the normal distribution (7) (dashed line) of $P(\xi)$ as a function of ξ . The dashed-dotted line is $P(\xi)$ in the approximation used in (33) with $\nu_0 = 4$. The vertical scale is the logarithm of the number of walks. The numerical simulations and the curve normalizations were done with $N = 300$ and 5000 walks. The parameters were chosen with the following values:

- a) $u = 0.50$, $\beta_+ = 1.5$, $\beta_- = 0.5$,
- b) $u = 0.90$, $\beta_+ = 1.1$, $\beta_- = 0.1$.

We see that there is a noticeable difference between the exact curve (38) and the normal one (7) when u is close to 1 (or to 0) and that the exact curve seems to describe the actual numerical data rather well, as opposed to the normal curve. For $u \sim 0.5$, on the other hand, the two curves are almost indistinguishable and they both describe the data very well.

Fig.1

Fig.2

Fig.3 (a)

Fig.3 (b)

Fig.4 (a)

Fig.4 (b) $\xi - N \ln(\beta\beta)/2$