March 6, 1990

Warsaw University Preprint *IFD/1/1990*

Genesis of the lognormal multiplicity distribution in the e^+e^- collisions and other stochastic processes

R. Szwed, G. Wrochna and A.K.Wróblewski

Institute of Experimental Physics

Warsaw University

PL-00-681 Warsaw, ul. Hoża 69

Abstract

It has been observed that the e^+e^- multiplicity distributions exhibit the following properties: the dispersions are linear functions of the mean and the distributions obey the KNO-G scaling with the scaling function of the lognormal shape.

In this paper the scale invariant branching is assumed as a mechanism within which all these properties could be derived. It is shown that the lognormal shape of the scaling function can be obtained within proposed mechanism by using the generalization of the Central Limit Theorem. The dependence of average multiplicity on energy is also derived within the postulated framework.

It is also shown that many other phenomena encountered in nature have the similar statistical properties.

1 Introduction

Recent results on the e^+e^- collisions – especially data published by the TASSO collaboration [1] – allow the study of statistical properties of multiplicity distributions with improved precision. It turns out that, in spite of the complicated character of the multiparticle production process, these properties are very simple and have already been encountered in many other phenomena in nature.

In this paper we show that assuming the scale invariant branching as a multiparticle production mechanism one can obtain the scaling of the multiplicity distributions with the scaling function of the lognormal shape.

Section 2 of this paper gives a brief summary of the scaling properties found in the e^+e^- multiplicity distributions [2, 3]. The e^+e^- multiplicity distributions obey the KNO–G scaling [4] with a scaling function of the lognormal shape [3].

In Sect. 3 we describe details of the proposed multiparticle production mechanism. The scale invariant branching is postulated to be responsible for the scaling properties seen in the data. It is also shown that for such process the Central Limit Theorem implies the lognormal shape of the scaling function of the final multiplicity distributions.

The mean multiplicity dependence on the energy is derived within the proposed framework and discussed in Sect. 4.

Section 5 presents some examples of similar scaling properties observed in phenomena not connected with elementary particle physics.

The mathematical formalism and derivation of some general formulae concerning the presented approach are collected in the two appendices. The first one reminds of the ways the lognormal distribution can be derived, and the second one gives the derivation of the linear dependence of dispersions on the mean multiplicity which is equivalent to the derivation of the scaling. The reasoning given in these two appendixes follows that of the works by Kapteyn [5], Wicksell [6] and Kolmogorov [7].

2 Multiplicity distributions in e^+e^- collisions

We start this section by listing the four fundamental properties of multiplicity distributions in e^+e^- collisions which have been described in our previous paper on the subject [3]:

(1) The multiplicity distribution P_n has a continuous density function $f(\tilde{n})$ (we denote the variable by $\tilde{n}$ since it has the meaning of "continuous multiplicity"):

$$P_n = \int_n^{n+1} f(\tilde{n}) d\tilde{n} \ . \tag{2.1}$$

(2) The density $f(\tilde{n})$ has a lognormal distribution (i.e. $\ln \tilde{n}$ is distributed normally):

$$f(\tilde{n}) = \frac{N}{\sqrt{2\pi}\sigma} \cdot \frac{1}{\tilde{n}+c} \exp\left(-\frac{[\ln(\tilde{n}+c)-\mu]^2}{2\,\sigma^2}\right). \tag{2.2}$$

(3) The dispersions of the order k are linear functions of the mean multiplicity:

$$D_k = A_k \left(\langle n \rangle + 0.5 \right) \ , \quad A_k = \mathrm{const}(s) \ , \tag{2.3}$$

where

$$D_k = \left[\sum_{n=0}^{\infty} \left(n - \langle n \rangle \right)^k P_n \right]^{1/k} . \tag{2.4}$$

(4) The density function $f(\tilde{n})$ obeys the scaling:

$$f(\tilde{n}) = \frac{1}{\langle \tilde{n} \rangle} \psi \left(\frac{\tilde{n}}{\langle \tilde{n} \rangle} \right) , \tag{2.5}$$

where $\psi(z)$ is a function independent of energy.

It should be noted that properties (3) and (4) are equivalent to each other [4, 8]. The scaling function $\psi(z)$ can be obtained from the density function by substituting:

$$\frac{\tilde{n}}{\langle \tilde{n} \rangle} = z \ , \qquad c = c' \langle \tilde{n} \rangle \ , \qquad \mu = \mu' + \ln \langle \tilde{n} \rangle \tag{2.6}$$

into the formula (2.2). We then obtain the scaling function of the lognormal shape [3]:

$$\psi(z) = \frac{N}{\sqrt{2\pi}\sigma} \cdot \frac{1}{z + c'} \exp \left(- \frac{\left[\ln(z + c') - \mu' \right]^2}{2 \, \sigma^2} \right) , \tag{2.7}$$

where N is given by the first normalization condition [1]:

$$1 = \int_0^{\infty} \psi(z) dz \quad \Rightarrow \quad N = \frac{2}{\mathrm{erfc} \left((\ln c' - \mu') / \sqrt{2} \, \sigma \right)} . \tag{2.8}$$

The symbol "erfc" stands for the complementary error function [9]:

$$\mathrm{erfc}(\tilde{n}) = \frac{2}{\sqrt{\pi}} \int_{\tilde{n}}^{\infty} e^{-t^2} dt \ . \tag{2.9}$$

The shape of the lognormal scaling function with parameters fitted to the e^+e^- data [3] is presented in Fig. 1.

Needless to say, the combined properties (1) and (4) are equivalent to the KNO-G scaling [4, 8] and can be expressed in the following way [8]:

$$P_n = \int_{n/\langle \tilde{n} \rangle}^{(n+1)/\langle \tilde{n} \rangle} \psi(z) dz \ , \quad z = \frac{n}{\langle \tilde{n} \rangle} \ , \quad \langle \tilde{n} \rangle = \int_0^{\infty} \tilde{n} \psi(z) dz \ . \tag{2.10}$$

Let us introduce the primitive function of $\psi(z)$ denoted by $\phi(z)$. The definition of $\phi(z)$ and its explicit form calculated for the function $\psi(z)$ (2.7) is [2]:

[1] We assume that $c' > 0$. If $c' \leq 0$ then $N = 1$.

[2] If $c' \leq 0$ then z must be greater than c' and $\phi(z) = -0.5 \cdot \mathrm{erfc} \left([\ln(z + c') - \mu'] \ / \ \sqrt{2} \, \sigma \right)$.

$$\phi(z) = -\int_z^\infty \psi(z)dz = -\frac{\mathrm{erfc}\left([\ln(z+c') - \mu'] \;/\; \sqrt{2}\,\sigma\right)}{\mathrm{erfc}\left((\ln c' - \mu') \;/\; \sqrt{2}\,\sigma\right)} \; . \tag{2.11}$$

We advocate the use of the function $\phi(z)$ instead of $\psi(z)$ as the scaling function [8]. It is easy to see that, when the scaling function $\phi(z)$ is used, the formula for P_n assumes a simple non–integral form:

$$P_n = \phi\left(\frac{n+1}{\langle\tilde{n}\rangle}\right) - \phi\left(\frac{n}{\langle\tilde{n}\rangle}\right) . \tag{2.12}$$

Parameters μ', c' and σ are constrained by the second normalization condition:

$$1 = \langle z\rangle = \int_0^\infty z\psi(z)dz = -\int_0^\infty \phi(z)dz \; , \tag{2.13}$$

but it is difficult to derive the analytic formula relating them, so they are treated as independent free parameters when the scaling function is fitted.

It has been found that the scaling function of the lognormal shape (2.7) describes with very good precision the multiplicity distributions in the e^+e^- collisions [3]. Figure 2 presents a graphical test of the KNO-G scaling [8, 3] for 25 experimental e^+e^- multiplicity distributions [1]. We recall that the experimental data, plotted as the quantity S_n versus z, should follow one single curve if the data obey the KNO-G scaling [3, 4, 8]. This can be seen from the following equality:

$$S_n = \sum_{k=n}^\infty P_k = \int_{n/\langle\tilde{n}\rangle}^\infty \psi(z)dz = -\phi\left(\frac{n}{\langle\tilde{n}\rangle}\right) . \tag{2.14}$$

As seen from Fig. 2, the data obey the KNO-G scaling very well.

The following parameters [3] have been obtained as a result of a fit to the TASSO data [3]:

$$\mu' = 0.581 \pm 0.001 \; , \qquad c' = 0.811 \pm 0.002 \; , \qquad \sigma = 0.155 \pm 0.001 \; . \tag{2.15}$$

A convenient graphical way to present the lognormal distribution is the so–called "probit diagram" [10]. Let us denote the distribution function of the normal density by $F(x)$:

$$F(x) = \frac{1}{\sqrt{\pi}} \int_{-\infty}^x \exp(-x^2)dx \; . \tag{2.16}$$

When variable n has a lognormal distribution, we can write the distribution function of n in the form:

$$\sum_{i=0}^n P_i = F\left(\frac{\ln(n+c) - \mu}{\sqrt{2}\sigma}\right) \; . \tag{2.17}$$

Let $F^{-1}(x)$ denote the inverse function of $F(x)$:

[3] In our previous publication [3] we have used a slightly different parametrization with parameters a and b, which however could be easily converted to μ' and c': $\mu' = \ln(1/a)$ and $c' = b/a$.

$$F^{-1}\left(F(x)\right) = x \ . \tag{2.18}$$

Hence

$$F^{-1}\left(\sum_{i=0}^{n} P_i\right) = \frac{\ln(n+c) - \mu}{\sqrt{2}\sigma} \ . \tag{2.19}$$

Thus when we plot $F^{-1}\left(\sum_{i=0}^{n} P_i\right)$ against $\ln(n+c)$ for lognormal distribution, we obtain a straight line. Using the scaled variable $z = n/\langle n \rangle$, instead of n, we can check the scaling:

$$F^{-1}\left(\sum_{i=0}^{n} P_i\right) = \frac{\ln(z+c') - \mu'}{\sqrt{2}\sigma} \ . \tag{2.20}$$

If the parameters μ', c' and σ are constants, all the data should follow a single curve.

Figure 3 presents the same 25 experimental e^+e^- multiplicity distributions [1] expressed as a probit diagram.

We recall that the experimental data should follow one single straight line if they are distributed according to the lognormal scaling function. As seen from Fig. 3 this is exactly the case.

An equivalent way to check the scaling properties of the multiplicity distribution is by plotting the dependence of the k–order dispersions on the mean multiplicity [8, 3]. Such a plot is presented in Fig. 4. The dispersions are indeed linear functions of the mean multiplicity, which again confirms the fact that the data obey the KNO–G scaling.

3 Genesis of the lognormal distribution observed in the e^+e^- collisions

In the present section we describe a possible mechanism of the multiparticle production, which in a natural way explains the scaling properties and the lognormal shape of the scaling function established for the data.

In Appendix A we summarize the derivation of the lognormal distribution as a consequence of the Central Limit Theorem. In the following it will be shown how to apply this mathematical concept to the e^+e^- collisions.

Let us consider, after Polyakov [11], that multiparticle production is a scale invariant branching process. The process starts with two primary quarks coming from a reaction $e^+e^- \rightarrow q\bar{q}$. The kinetic energy of colliding electrons, $\sqrt{s}$, is converted into the potential energy of a string between the quarks. The string can break forming a second pair of quarks. The sum of the potential energy of both pairs is smaller than the energy of the primary string because a fraction of the latter converted into the kinetic energy of the two pairs. Each of the strings can break again and the process is continued until the potential energy of the strings is of the order of the pion mass.

The number of particles created in the i-th generation is proportional to the number of particles in the $(i-1)$-th generation. Thus, denoting the number of particles in the i-th generation by $\tilde{n}_i$ we have:

$$\tilde{n}_i = \tilde{n}_{i-1} + \varepsilon_i \, \tilde{n}_{i-1} \ , \tag{3.1}$$

where ε_i is a random variable.

The potential energy per particle, E_i, can change for two reasons. Firstly, the energy should be divided into a larger number of particles (decrement by the factor $1+\varepsilon_i$). Secondly, a fraction δ_i of the potential energy is converted into kinetic energy.

$$E_i = E_{i-1} \cdot (1-\delta_i)/(1+\varepsilon_i) \; . \tag{3.2}$$

Thus after n steps:

$$\tilde{n} = \tilde{n}_0 \cdot (1+\varepsilon_1) \cdot \ldots \cdot (1+\varepsilon_n) \; , \tag{3.3}$$

$$E = E_0 \cdot \frac{1-\delta_1}{1+\varepsilon_1} \cdot \ldots \cdot \frac{1-\delta_n}{1+\varepsilon_n} \; . \tag{3.4}$$

Let us introduce new variables:

$$\eta = \prod_i (1+\varepsilon_i) = \frac{\tilde{n}}{\tilde{n}_0} \; , \qquad\qquad \xi = \prod_i (1-\delta_i) = \eta \cdot \frac{E}{E_0} \; . \tag{3.5}$$

Variables η and ξ are independent because of the independence of ε_i and δ_i. Both are distributed lognormally as it is shown in Appendix A. The common distribution of η and ξ can be written in the form (see Appendix B):

$$f(\eta,\xi) = \frac{1}{2\pi\sigma_\eta\sigma_\xi\,\eta\xi} \cdot \exp\left\{-\frac{1}{2}\left[\frac{(\ln\eta-\mu_\eta)^2}{\sigma_\eta^2} + \frac{(-\ln\xi-\mu_\xi)^2}{\sigma_\xi^2}\right]\right\} \; . \tag{3.6}$$

After substitution of η and ξ by $\tilde{n}$ and E one obtains:

$$f\left(\frac{\tilde{n}}{\tilde{n}_0},\frac{E}{E_0}\right) = \frac{1}{2\pi\sigma_1\sigma_2\,\frac{\tilde{n}}{\tilde{n}_0}\frac{E}{E_0}\sqrt{1-\rho^2}} \cdot \exp \tag{3.7}$$

$$\left\{-\frac{1}{2(1-\rho^2)}\left[\frac{\left(\ln\frac{\tilde{n}}{\tilde{n}_0}-\mu_1\right)^2}{\sigma_1^2} - 2\rho\frac{\left(\ln\frac{\tilde{n}}{\tilde{n}_0}-\mu_1\right)\left(\ln\frac{E_0}{E}-\mu_2\right)}{\sigma_1\sigma_2} + \frac{\left(\ln\frac{E_0}{E}-\mu_2\right)^2}{\sigma_2^2}\right]\right\},$$

where parameters $\mu_1,\mu_2,\sigma_1,\sigma_2$ and ρ can be expressed by $\mu_\eta,\mu_\xi,\sigma_\eta$ and σ_ξ:

$$\begin{aligned} &\mu_1 = \mu_\eta \; , \quad \mu_2 = \mu_\xi + \mu_\eta \\ &\sigma_1 = \sigma_\eta \; , \quad \sigma_2 = \sqrt{\sigma_\xi^2+\sigma_\eta^2} \\ &\rho = \sigma_\eta \;/\; \sqrt{\sigma_\xi^2+\sigma_\eta^2} \end{aligned} \tag{3.8}$$

The parameters μ_1 and μ_2 have interpretation of the mean logarithms of $\tilde{n}$ and E respectively, and σ_1 and σ_2 are dispersions of $\ln\tilde{n}$ and $\ln E$, and ρ is a correlation coefficient. Thus, $f(\frac{\tilde{n}}{\tilde{n}_0},\frac{E}{E_0})$ is a bivariate lognormal distribution.

If we fix the primary energy E_0 at $\sqrt{s}$ and the final energy E at the pion mass m_π, then from rel.(B.2) we obtain a formula for conditional probability density:

$$f(\tilde{n}|\sqrt{s}) = \frac{1}{\sqrt{2\pi}\sigma_1\sqrt{1-\rho^2}\,\tilde{n}} \cdot \tag{3.9}$$

$$\exp\left\{-\frac{1}{2(1-\rho^2)\,\sigma_1^2}\left[(\ln\tilde{n}-\mu_1) - \rho\frac{\sigma_1}{\sigma_2}\left(\ln\frac{\sqrt{s}}{m_\pi}-\mu_2\right)\right]^2\right\} .$$

This is again a lognormal distribution with parameters:

$$\mu = \mu_1 + \rho\frac{\sigma_1}{\sigma_2}\left(\ln\frac{\sqrt{s}}{m_\pi} - \mu_2\right), \qquad \sigma^2 = \sigma_1^2\,(1-\rho^2)\,. \tag{3.10}$$

Now we can use formulae (B.4–B.6) from Appendix B to derive the scaling:

$$\langle \tilde{n}^k \rangle = \exp\left[k\,\mu(\sqrt{s}) + \frac{1}{2}\,k\,\sigma^2\right]. \tag{3.11}$$

For the dispersion of the second order we have:

$$\tilde{D} = A\,\langle \tilde{n} \rangle \quad , \quad \text{where} \quad A = \sqrt{\exp(\sigma^2) - 1} \tag{3.12}$$

and similarly for the higher orders.

For the experimentally measured charged multiplicity, formula (3.12) assumes the form [8]:

$$D_{ch} \approx A\,(\langle n_{ch} \rangle + 1)\,. \tag{3.13}$$

Formulae (B.4–B.6) are still valid, i.e. the dispersions are linear functions of the mean, which is equivalent to the scaling. Needless to say, the linear dependence of dispersion on the mean multiplicity was firstly observed for the *pp* data [12].

It is obvious, that the obtained scaling function cannot be directly applyed to the *pp* collisions, because protons are not point-like objects. We have checked that indeed the shape of the scaling function for the *pp* multiplicity distribution is slightly different than the lognormal one. The results of investigation of this problem will be published elsewhere [13].

In conclusion, it has thus been shown that the lognormal multiplicity distribution, the KNO-G scaling and the linear relation between dispersions and the mean are consequences of the Central Limit Theorem within the postulated framework.

4 Average multiplicity

Formulae (B.4) and (3.10) can be used to evaluate the dependence of the average multiplicity on the energy. If we substitute (3.10) into (3.11) we obtain:

$$\langle \tilde{n} \rangle = \exp\left[\mu_1 + \rho\frac{\sigma_1}{\sigma_2}\left(\ln\frac{\sqrt{s}}{m_\pi} - \mu_2\right) + \frac{1}{2}\,\sigma_1^2\,(1-\rho^2)\right]. \tag{4.1}$$

Simplifying, we have:

$$\langle \tilde{n} \rangle = \exp\left(\alpha \ln s + \gamma\right) \tag{4.2}$$

or

$$\langle \tilde{n} \rangle = \tilde{\beta}\, s^{\alpha}\,. \tag{4.3}$$

The relation between $\langle \tilde{n} \rangle$ and $\langle n_{ch} \rangle$ [8]:

$$\langle n_{ch} \rangle \approx 2\,(\langle \tilde{n} \rangle - 0.5) \tag{4.4}$$

can be used to obtain the formula for the average charge multiplicity:

$$\langle n_{ch} \rangle = \beta \, s^{\alpha} - 1 \, . \tag{4.5}$$

The validity of this formula is shown in Fig. 5 where the data from Ref. [1, 14, 15] are presented. The fit gives $\alpha = 0.442$ and $\beta = 2.96$. As seen in Fig. 5 there are some systematic differences between data obtained in various experiments. This is mainly due to the two reasons: (1) in the majority of the data the strange particle decay products are either not subtracted from the sample or they are subtracted in an ambiguous way, (2) the lowest multiplicity channels are not measured experimentally but are taken from various Monte Carlo programs. Therefore, the values of the fitted parameters should be taken with caution.

5 Universal character of the statistical properties observed in the e^+e^- multiplicity data

We would like to point out that the properties (1) – (3) discussed in the Sect. 2 have the universal character and are often encountered in nature. This was already observed in the forties [16]. There exists a vast literature on the subject with numerous examples illustrating the occurrence of properties (1) - (3) in many different phenomena (for the bibliography see [17]). We present some evidence for the universal character of these properties:

Property (1): The way the discrete multiplicity distribution is calculated from the continuous density distribution (2.1) Usually for limit theorems, when the number of objects tends to infinity, the limiting distribution is continuous. But for a finite number of objects a good approximation is given simply by an integral over the limiting density ([18], p.92 of Ref. [17]). The best known example is the approximation of the binomial distribution by the normal distribution (Moivre-Laplace Theorem).

Property (2): The lognormal shape of the density distribution. The lognormal distribution is very often encountered in nature. Many examples of this can be found in the literature (see Refs. [16, 17] and references therein). In the following we quote some examples – from various branches of science – of distributions which were found to be of the lognormal shape [16, 17]: distribution of size of particles coming from grinding rocks (geology), distribution of incomes and family budgets (economy), distribution of size of organisms, abundance of species, number of lesions in plants and animals (biology), distribution of age of marriage (sociology), distribution of the number of words in sample sentences (philology).

Property (3): Linear dependence of dispersion on the mean. Linear dependence of dispersion on the mean is also very often encountered in nature. First of all it is a basic feature of a wide class of branching processes [19, 20, 21]. Let us also list examples such as: the hourly lemon sole catches [22], the number of pines in forests [22], and again in the processes mentioned in previous paragraph: incomes [23], family budgets [24], number of non-resident species of birds [22], lesions in plants [25], the tick counts on sheep [22], and the age of marriage [26].

Thus each property characterizing the e^+e^- multiplicity distribution is also found in many natural phenomena. But the crucial point is that all these properties are very often

characteristic for the same phenomenon. The correlation is so strong that some authors treat, e.g., the observation of property (3) as a signature of (2) [25]. There is a standard way of analysing skew distributions, namely by applying the so–called lognormal transformation [5, 22, 27]. The frequency distribution is investigated as a function of $\ln x$ rather than x itself. Obviously, linearity of dispersion of x implies constancy of dispersion of $\ln x$ for the lognormal distribution. In a number of cases, when the distribution of $\ln x$ was plotted instead of x, the distribution became normal and dispersions of all distributions became equal [5, 22, 28]. Thus the properties (2) and (3) appear together.

In summary, it seems that many various phenomena in nature have similar statistical features as the multiplicity distribution in e^+e^- collisions. For illustration we present the distribution of age of marriage [26]. We have chosen this set of distributions because of a clear interpretation of bivariate distribution [26]. It happens that for a given age of the bridegroom, a distribution of the age of a bride is lognormal with a certain mean and dispersion. These parameters vary with the age of the bridegroom in such a way that dispersions are linear functions of the mean which is illustrated in Fig. 6. Obviously, this means that the data obey the KNO–G scaling. We illustrate this fact also by presenting a standard graphical KNO–G plot (S_n vs $\tilde{n}/\langle\tilde{n}\rangle$) in Fig. 7 and a probit diagram in Fig. 8. Figs. 6, 7 and 8 can be compared with the equivalent plots for the e^+e^- data (Figs. 2, 3 and 4). Small deviations from linearity in Fig. 8 for large values of age mean that distributions are limited due to the final human life time.

6 Conclusions

It has been shown that the e^+e^- multiplicity distributions and many other phenomena of completely different "dynamics" have the same simple statistical properties. This fact can be understood in terms of the Central Limit Theorem. When a final distribution is created by many contributions, its shape is not determined by the properties of the particular contributions. Especially, if a random variable is varied by a large number of random impulses in such a way that the change is proportional to the actual value of the variable, the distribution is lognormal. In the case of a distribution of two variables varied by proportional impulses, if one of the variables is fixed, the dispersions of the second variable are linear functions of the mean and the distributions obey a scaling law.

In the case of the e^+e^- collisions the role of the second variable is played by the available energy. It seems that the scale invariant branching process provides an explanation of multiparticle production. Such a picture has a deep theoretical foundation in the scale invariant field theories [11]. This approach can be successfully applied to the e^+e^- multiplicities [3]. The lognormal distribution with its scaling properties describes the experimental data very well.

We claim that due to the predictive power of the Central Limit Theorem it is not necessary to create any complicated model with many details to predict the general shape of the inclusive distributions. Such details, unless they change basic statistical assumptions, cannot change the final result. It is just the statistics, which determines the scaling properties and the general shape of distributions. However, the Central Limit Theorem does not give values of the parameters, which have to be fitted. Should the values of parameters be obtained within a theory, the theory has to be enriched by a dynamics. It seems that the modern QCD branching models, such as [29], which have intrinsically

the statistical properties proposed in this paper supply in the best way the necessary dynamics.

Acknowledgements

We would like to thank drs. R.Nowak, M.Gaździcki and G.Wilk for valuable remarks during preparation of this paper and prof. J.A.Zakrzewski for useful comments on the manuscript.

A Derivation of the lognormal distribution

In this appendix we shall remind the derivation of the lognormal distribution after Kapteyn [5]. Let us consider a process controlled by a random variable x. The variable x can be changed by random impulses ε_i, whose effects can in general depend on the actual value of x. Thus we can write:

$$x_i = x_{i-1} + \varepsilon_i \, g\left(x_{i-1}\right) . \tag{A.1}$$

Now we take the sum of ε_i:

$$\sum_{i=1}^{n} \frac{x_i - x_{i-1}}{g\left(x_{i-1}\right)} = \sum_{i=1}^{n} \varepsilon_i \ . \tag{A.2}$$

If every individual impulse is small, we can replace the sum on the left hand side by the integral and denote it by $G(x)$:

$$\sum_{i=1}^{n} \frac{x_i - x_{i-1}}{g\left(x_{i-1}\right)} \approx \int_{x_0}^{x} \frac{dx}{g(x)} = G(x) \ . \tag{A.3}$$

If the ε_i are independent of each other, the sum on the right-hand side of (A.2) should be distributed normally due to the Central Limit Theorem. Hence, the sum on the left-hand side ($G(x)$) is also distributed normally and we obtain the following formula for the distribution of x:

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma \, g(x)} \ \exp\left\{-\frac{\left[G(x) - \mu\right]^2}{2\,\sigma^2}\right\} . \tag{A.4}$$

Functions of the form (A.4) are called "Kapteyn distributions" [5]. The case most interesting for us is when the effect of an impulse is proportional to the actual value of x:

$$x_i = x_{i-1} + \varepsilon_i \, x_{i-1} \ . \tag{A.5}$$

Thus:

$$g(x) = x \ . \tag{A.6}$$

In this case we obtain:

$$G(x) = \ln x \tag{A.7}$$

and the lognormal distribution:

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma \, x} \ \exp\left\{-\frac{\left(\ln x - \mu\right)^2}{2\,\sigma^2}\right\} . \tag{A.8}$$

This result one can obtain in a less rigorous but a more intuitive way. If the effect of an impulse is proportional to the actual value of x, the formula (A.1) assumes the form:

$$x_i = x_{i-1} + \epsilon_i \, x_i \ . \tag{A.9}$$

Thus the final value of x is a product of a large number of individual factors:

$$x = x_0 \prod_i (1 + \epsilon_i) \,. \tag{A.10}$$

Taking logarithm of the both sides and expanding $\ln(1 + \epsilon_i)$ near 1 one obtains:

$$\ln(x/x_0) = \sum_{i=1} \epsilon_i \,. \tag{A.11}$$

The sum of large number of such contributions ϵ_i is distributed normally due to the Central Limit Theorem. Hence $\ln(x)$ is distributed normally and x lognormally.

In the more general case, when the effect is a linear function of x, we obtain a normal distribution in the variable $\ln(x + c)$.

Thus we see that the lognormal distribution is obtained when the effect of random impulse is proportional to the actual value of x.

B Linearity of dispersions

Now let us consider the generalization of the process for two variables x and y. In principle the random impulses can affect x and y in different ways. Thus, under arguments similar to those given in the previous section, the final distribution is a bivariate normal distribution of $G_1(x)$ and $G_2(y)$ [6]:

$$f(x,y) = \frac{1}{2\pi\sigma_1\sigma_2\, g_1(x) g_2(x)\sqrt{1-\rho^2}} \cdot \exp \tag{B.1}$$
$$\left\{ -\frac{1}{2(1-\rho^2)} \left[\frac{(G_1(x)-\mu_1)^2}{\sigma_1^2} - 2\rho \frac{(G_1(x)-\mu_1)\,(G_2(y)-\mu_2)}{\sigma_1\sigma_2} + \frac{(G_2(y)-\mu_2)^2}{\sigma_2^2} \right] \right\} .$$

Let us again consider the case when x is distributed lognormally, i.e. $G_1(x) = \ln x$. For a fixed value of $y = \zeta$ we have the conditional probability:

$$f(x \mid y = \zeta) \;=\; f(x,\zeta) \,/ \int_{-\infty}^{\infty} f(x,\zeta)dx = \tag{B.2}$$
$$= \frac{1}{\sqrt{2\pi}\sigma_1 \sqrt{1-\rho^2}\, x} \exp\left\{ -\frac{1}{2(1-\rho^2)\,\sigma_1^2} \left[(\ln(x) - \mu_1) - \rho\frac{\sigma_1}{\sigma_2}\,(G_2(\zeta) - \mu_2) \right]^2 \right\} .$$

This is again a lognormal distribution (A.8) with parameters:

$$\mu = \mu_1 + \rho\frac{\sigma_1}{\sigma_2}\,(G_2(\zeta) - \mu_2)\,, \qquad \sigma^2 = \sigma_1^2\,(1-\rho^2)\,. \tag{B.3}$$

Thus, if we measure the distribution of x for various values of ζ, the parameter μ ($\mu = \langle \ln x \rangle$) will change, while the dispersion of $\ln x$ (denoted by σ) will be constant.

Now we can calculate moments of x as a function of parameter ζ:

$$\langle x^k \rangle(\zeta) = \exp\left[k\,\mu(\zeta) + \frac{1}{2}\,k^2\,\sigma^2 \right], \tag{B.4}$$

$$D_x^2(\zeta) = \langle x^2 \rangle(\zeta) - \langle x \rangle^2(\zeta) = \left[\exp(\sigma^2) - 1 \right] \langle x \rangle^2(\zeta)\,. \tag{B.5}$$

Thus, since $\sigma = const(\zeta)$, the dispersion is a linear function of the mean:

$$D_x = \sqrt{(\exp(\sigma^2) - 1)} \cdot \langle x \rangle\,. \tag{B.6}$$

One can easily find similar formulae for higher moments.

References

[1] *MARK I* (3.0, 4.0, 4.8, 7.4 GeV) J.L.Siegrist et al.: Phys. Rev. D26, (1982) 969
LENA (7.4, 8.9, 9.3 GeV) B.Niczyporuk et al.: Z. Phys. C – Particles and Fields 9 (1981) 1
JADE (12, 30, 35 GeV) W.Bartel et al.: Z. Phys. C – Particles and Fields 20 (1983) 187
HRS (29 GeV) M.Derrick et al.: Phys. Rev. D34 (1986) 3304
PLUTO (3.6, 5.0, 7.7, 9.4, 12.0, 13.0, 17.0, 22.0, 27.5, 30.6 GeV) private communication
TASSO (14.0, 22.0, 34.8, 43.6 GeV) W.Braunschweig et al.: Z. Phys. C – Particles and Fields 45 (1989) 1939

[2] A.I.Golokhvastov: Dubna preprint, JINR E2-87-484 (1987)

[3] R.Szwed and G.Wrochna: Warsaw University Preprint IFD/3/1989, to appear in Z. Phys. C

[4] A.I.Golokhvastov: Sov. J. Nucl. Phys. 27 (1978) 430; 30 (1979) 128

[5] J.C.Kapteyn: "Skew Frequency Curves in Biology and Statistics.", P.Noordhoff, Groningen 1903

[6] S.D.Wicksell: Arkiv for Matematik, Astronomi och Fysik band 12, No 20 (1917) 1

[7] A.N.Kolmogorov: Comptes Rendus de l'Academie des Sciences de l'URSS (Doklady Akademii Nauk SSSR) 31 (1941) 99

[8] R.Szwed and G.Wrochna: Z. Phys. C – Particles and Fields 29 (1985) 255

[9] W.H.Press, B.P.Flannery, S.A.Teukolsky and W.T.Vetterling: "Numerical Recipes", Cambridge University Press 1986, p. 164

[10] N.Arley and K.R.Buch: Introduction to the Theory of Probability and Statistics, New York 1950, Chapter 10, (John Wiley and Sons)

[11] A.M.Polyakov: Sov. Phys. JETP 32 (1971) 296; 33 (1971) 850 (in Russian: Zh. Eksp. Teor. Fiz. 59 (1970) 542; 60 (1971) 296)

[12] A.K.Wróblewski: Acta Phys. Pol. B4 (1973) 857

[13] R.Szwed, G.Wrochna and A.K.Wróblewski: paper in preparation

[14] *ADONE-MEA* (1.46–1.79 GeV) B.Esposito et al.: Lett. Nuovo Cim. 30 (1981) 65
ADONE $\gamma\gamma - 2$ (1.43–2.87 GeV) C.Bacci et al.: Phys. Lett. 86B (1979) 234
MARK I (2.6–6.3 GeV) J.L.Siegrist et al.: Phys. Rev. D26, (1982) 969
CLEO (10.5 GeV) M.S.Alam et al.: Phys. Rev. Lett.49 (1982) 357

[15] *ALEPH* (91 GeV) D.Decamp et al.: Preprint CERN–EP/89–139
TOPAZ (52.5, 55.5 GeV) Proc. of the XXIV Int.Conf.on HEP (Springer–Verlag,1988)
MARK II (91 GeV) Proc. of Madrid Conf. 1989

[16] J.H.Gaddum: Nature 156 (1945) 463

[17] J.Aitchison and J.A.C.Brown: "The Lognormal Distribution", Cambridge University Press 1957

[18] H.R.Thompson: Biometrika 38 (1951) 414

[19] T.E.Harris: "The Theory of Branching Processes", Springer-Verlag 1963

[20] K.B.Althreya: P.E.Ney, "Branching Processes", Springer-Verlag 1972

[21] A.T.Bharucha-Reid: "Elements of the Theory of Markov Processes and Their Application", Mc Graw-Hill Book Co.Inc. 1960

[22] M.H.Quenouille: "Introductory Statistics", London Pergamon Press Ltd 1950, p.162-183.

[23] S.J.Prais and J.Aitchison: Revue de l'Institut int. de Statist. 22 (1954) 1

[24] D.S.Brady: Journal of the American Statistical Association 33 (1938) 385

[25] A.Kleczkowski: The Annals of Applied Biology 36 (1949) 139

[26] S.Nydell: Skand. Aktuarieforeningens Tidskrift 7 (1924) 36

[27] W.G.Cochran: Emp. J. Exp. Agric. 6 (1938) 157

[28] J.H.Curtiss: Ann. Math. Statist. 14 (1943) 107

[29] G.Marchesini and B.R.Webber: Nucl. Phys. B310 (1988) 461

Figure captions

Fig. 1
Shape of the lognormal scaling function $\psi(z)$ (eq. (2.7) for the e^+e^- data [1]).
Fig. 2
Graphical test of the KNO–G scaling. $S_n = \sum_{k=n}^{\infty} P_k$ plotted as a function of reduced multiplicity ($z = n/\langle\tilde{n}\rangle$) for the whole collection of the e^+e^- data [1].
Fig. 3
Probit diagram for the whole collection of the e^+e^- data [1].
Fig. 4
D_2, D_4 and D_6 plotted as a function of mean multiplicity for collection of the e^+e^- data [1].
Fig. 5
The average multiplicity $\langle n_{ch}\rangle$ plotted as a function of centre of mass energy $\sqrt{s}$ for the multiplicity data measured by the indicated collaboration teams. The solid line presents the fit to the data with formula (4.5).
Fig. 6
Dispersions of the distributions of ages of bride (D_2, D_4 and D_6) plotted as a function of mean age of bride for a given age of bridegroom [26].
Fig. 7
Graphical test of the scaling properties of the age of marriage distributions [26]. $S_n = \sum_{k=n}^{\infty} P_k$ plotted as a function of reduced age ($age/\langle age\rangle$) for the distribution of age of marriage [26].
Fig. 8
Probit diagram for the age of marriage [26].

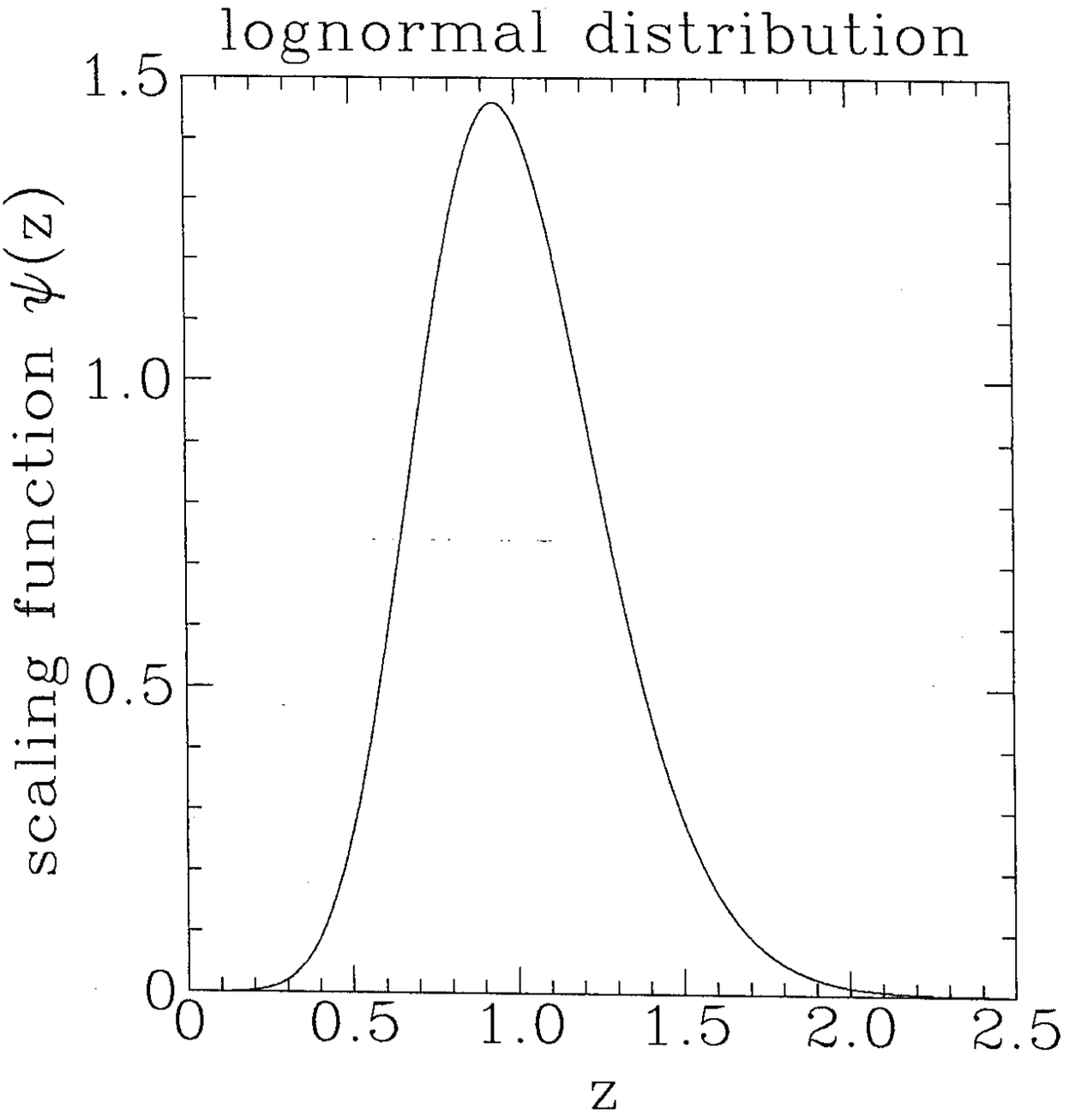
lognormal distribution
scaling function $\psi(z)$
1.5
1.0
0.5
0
0
0.5
1.0
1.5
2.0
2.5
z

Fig.1

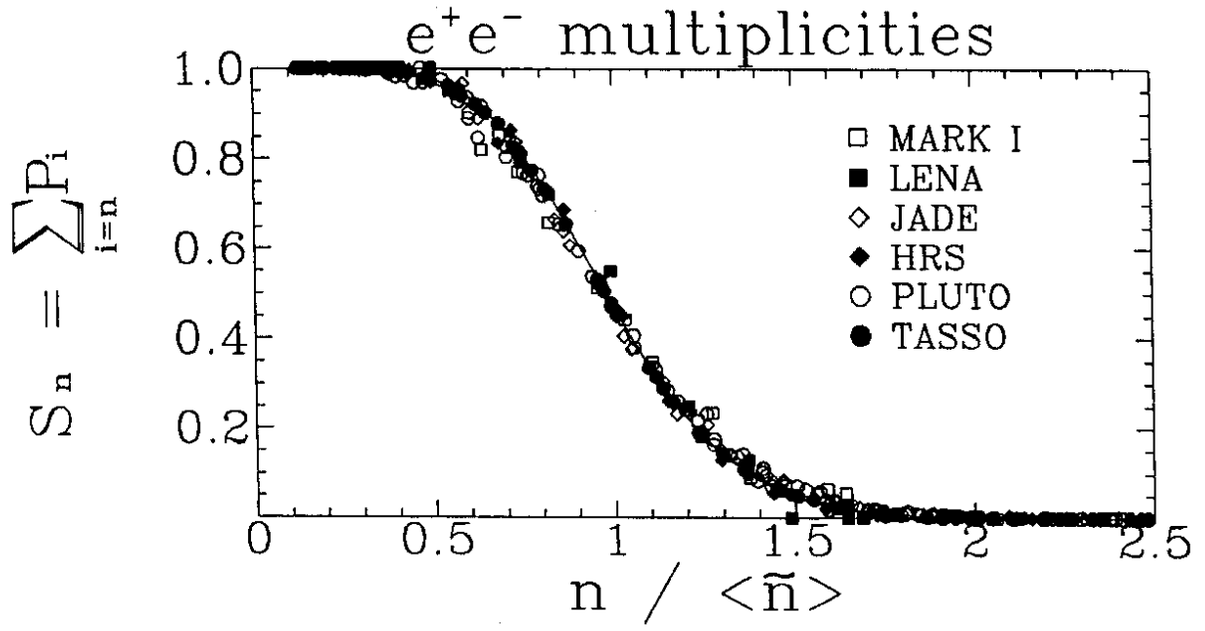
e+e- multiplicities
MARK I
LENA
JADE
HRS
PLUTO
TASSO
1.0
0.8
0.6
0.4
0.2
0
0.5
1
1.5
2
2.5
n / <ñ>
Sn = ∑ Pi
i=n

Fig. 2

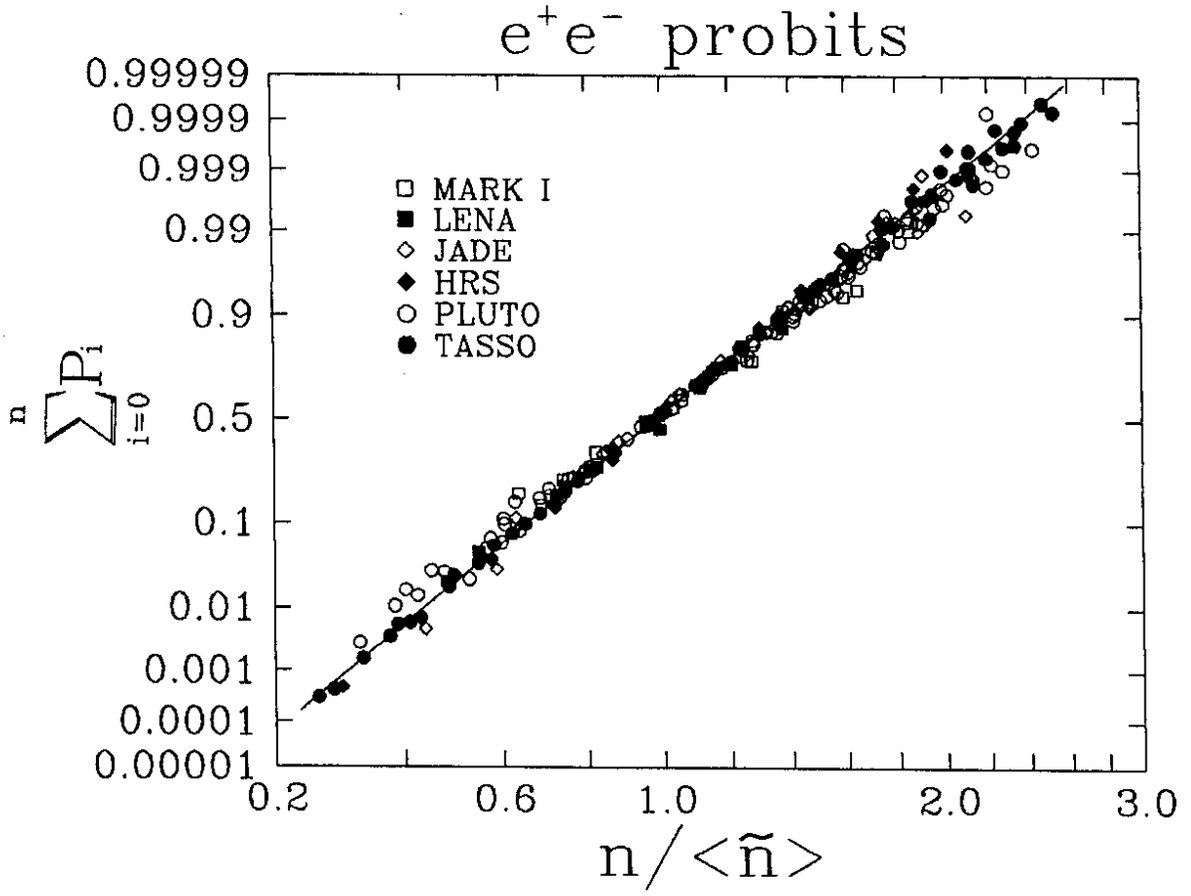
e+e- probits
MARK I
LENA
JADE
HRS
PLUTO
TASSO
0.99999
0.9999
0.999
0.99
0.9
0.5
0.1
0.01
0.001
0.0001
0.00001
0.2
0.6
1.0
2.0
3.0
n/<ñ>
n ∑ Pi
i=0

Fig. 3

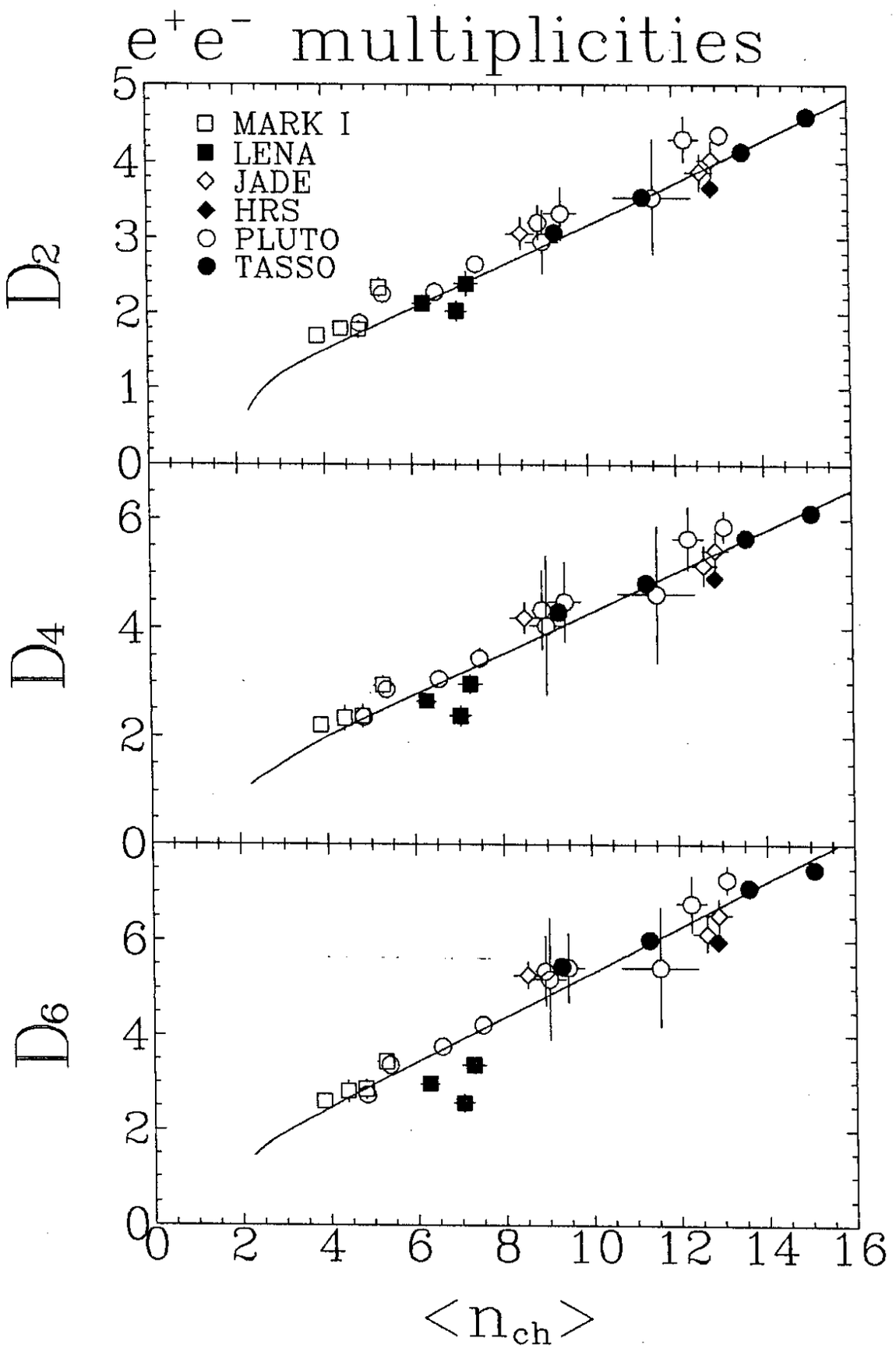
e+e- multiplicities
MARK I
LENA
JADE
HRS
PLUTO
TASSO
D2
D4
D6
<nch>
0
2
4
6
8
10
12
14
16

Fig. 4

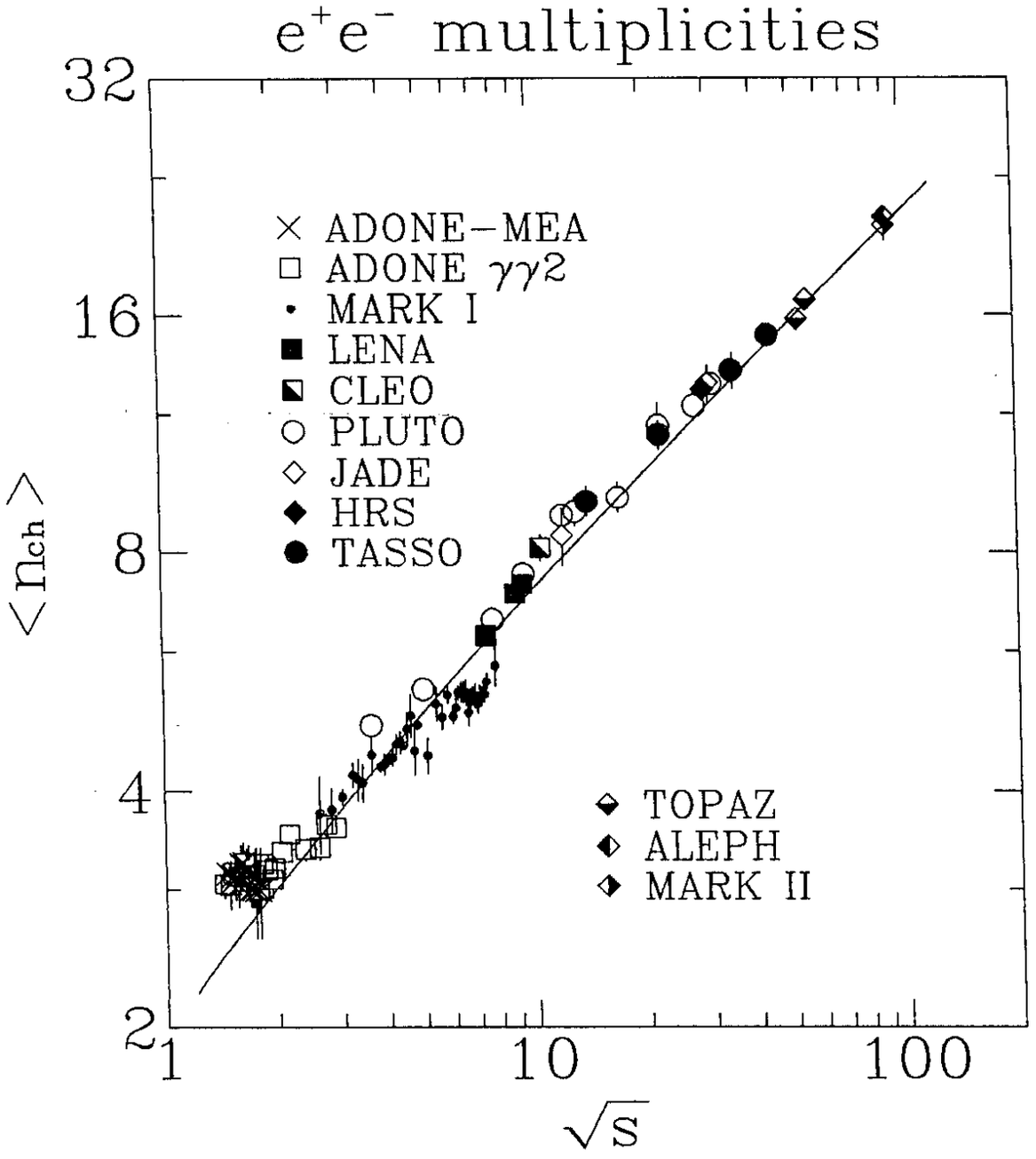
e+e- multiplicities
32
16
8
4
2
1
10
100
√s
<nch>
× ADONE−MEA
□ ADONE γγ2
• MARK I
■ LENA
◪ CLEO
○ PLUTO
◇ JADE
◆ HRS
● TASSO
◆ TOPAZ
◆ ALEPH
◆ MARK II

Fig. 5

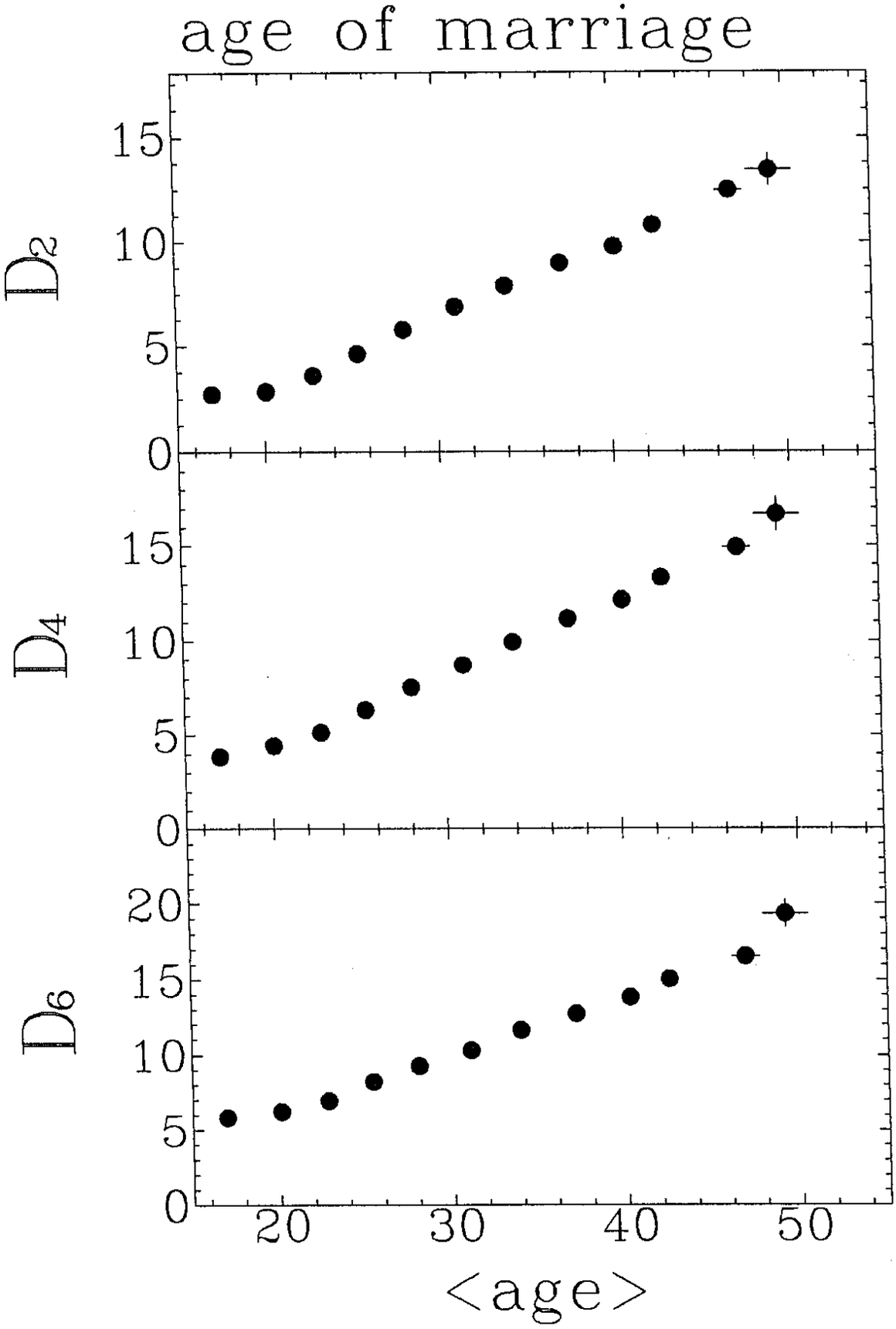

Fig. 6

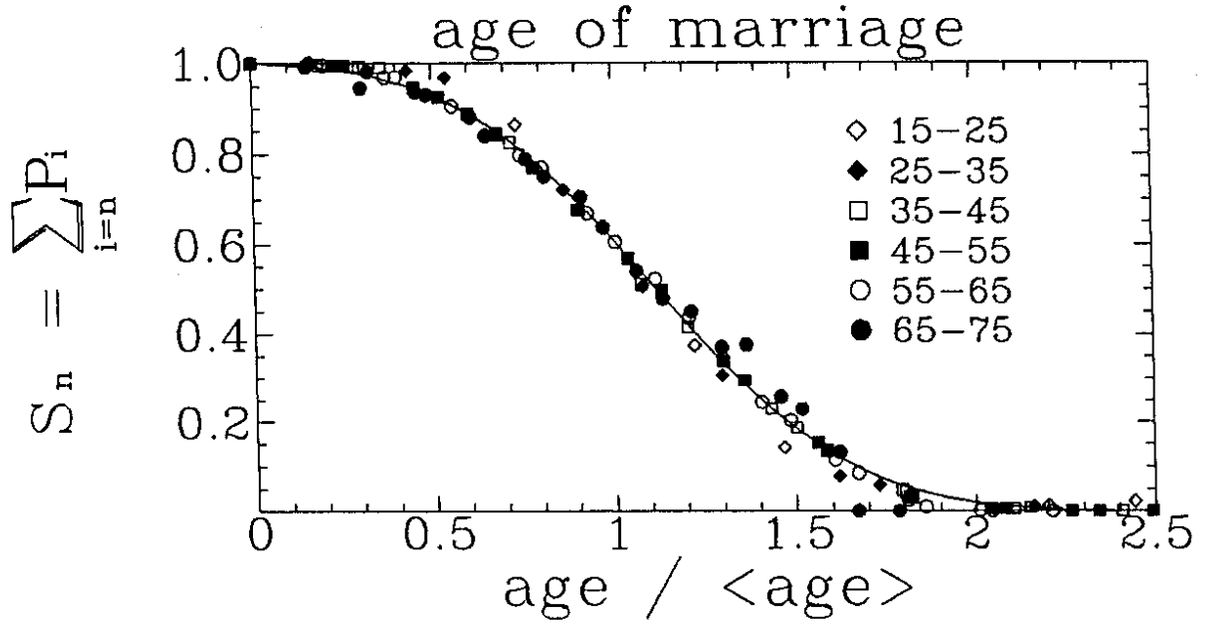
age of marriage
$S_n = \sum_{i=n} P_i$
1.0
0.8
0.6
0.4
0.2
0
0.5
1
1.5
2
2.5
15–25
25–35
35–45
45–55
55–65
65–75
age / <age>

Fig. 7

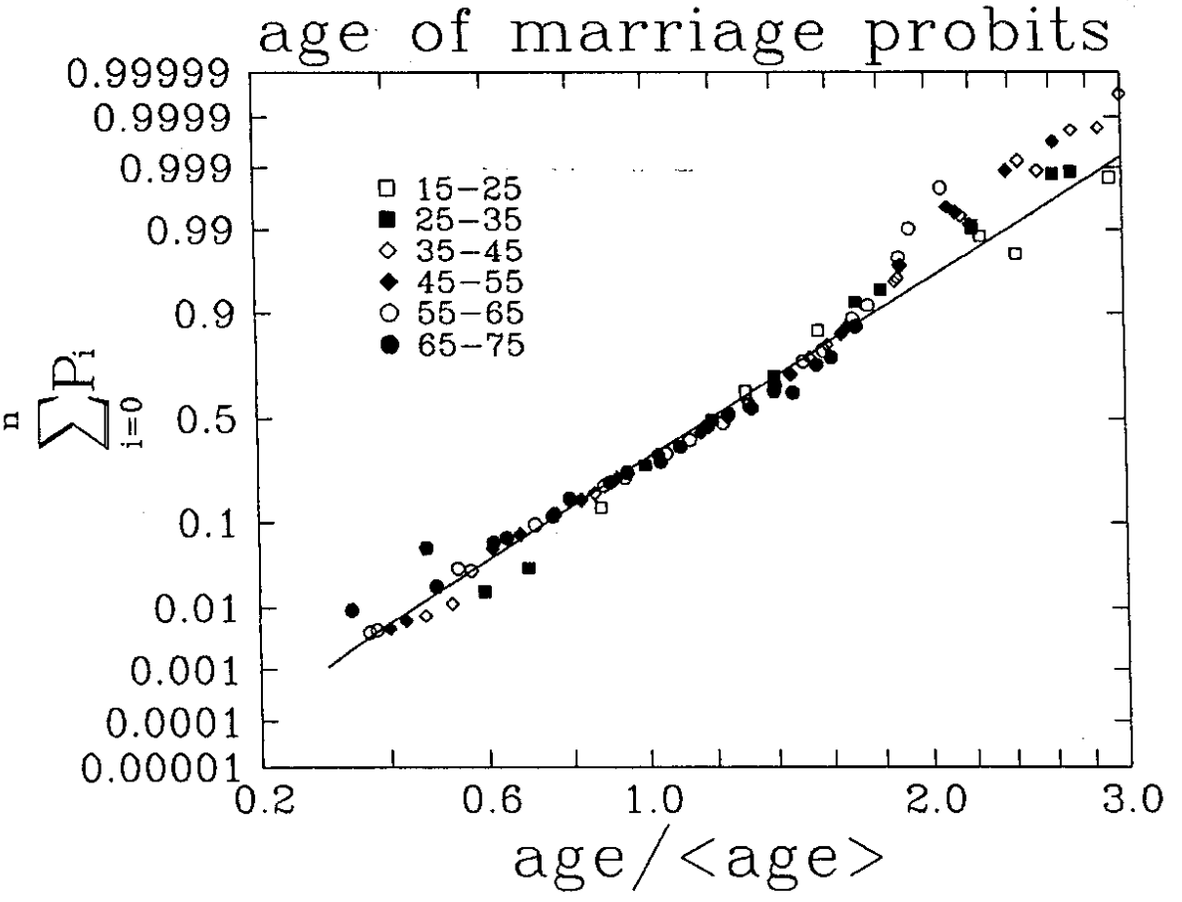
age of marriage probits
$\sum_{i=0}^{n} P_i$
0.99999
0.9999
0.999
0.99
0.9
0.5
0.1
0.01
0.001
0.0001
0.00001
0.2
0.6
1.0
2.0
3.0
15–25
25–35
35–45
45–55
55–65
65–75
age/<age>

Fig. 8