September 15, 1990

Warsaw University Preprint *IFD/8/1990*

HOW TO FIT THE LOGNORMAL DISTRIBUTION

G. Wrochna

Institute of Experimental Physics
Warsaw University
PL-00-681 Warsaw, ul. Hoża 69

Abstract

A set of formulae and some technical remarks concerning the lognormal distribution, KNO-G scaling and its application to the multiplicity distributions in elementary particle collisions are presented.

1 Introduction

Recently it was shown [1,2] that multiplicity distributions in $e^+e^-$and pp collisions obey the scaling and can be described by the lognormal distribution. In this paper I would like to collect formulae concerning this subject which were presented in our various papers [1]–[7]. A more detailed description of some numerical routines is given. I hope that this "handbook" can be useful especially for those who want to study multiplicity distributions themselves.

2 KNO-G scaling

Let us assume that the multiplicity distribution P_n has a continuous density function $f(\tilde{n})$ (we denote the variable by $\tilde{n}$ since it is continuous)

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

Let the density function $f(\tilde{n})$ obey the scaling:

$$f(\tilde{n}) = \frac{1}{\langle\tilde{n}\rangle}\,\psi\left(\frac{\tilde{n}}{\langle\tilde{n}\rangle}\right) , \tag{2.2}$$

where $\psi(z)$ is a function independent of energy. Hereafter z denotes the ratio $z = n/\langle\tilde{n}\rangle$.
The continuous average multiplicity $\langle\tilde{n}\rangle$ is defined as follows:

$$\langle\tilde{n}\rangle = \int_0^\infty \tilde{n} f(\tilde{n})d\tilde{n} = \int_0^\infty \tilde{n}\psi(z)dz \; . \tag{2.3}$$

We can write the formula for P_n in the way called KNO-G scaling [10]

$$P_n = \int_{n/\langle\tilde{n}\rangle}^{(n+1)/\langle\tilde{n}\rangle} \psi(z)dz \; . \tag{2.4}$$

Let us introduce the primitive function of $\psi(z)$ denoted by $\phi(z)$:

$$\phi(z) = -\int_z^\infty \psi(z)dz. \tag{2.5}$$

We advocate the use of $\phi(z)$ instead of $\psi(z)$ as the scaling function. 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:

$$P_n = \phi\left(\frac{n+1}{\langle\tilde{n}\rangle}\right) - \phi\left(\frac{n}{\langle\tilde{n}\rangle}\right) . \tag{2.6}$$

Probability distributions and scaling functions should obey the following normalization conditions:

$$\sum_0^\infty P_n = \int_0^\infty f(\tilde{n})d\tilde{n} = \int_0^\infty \psi(z)dz = -\phi(0) = 1 \; , \tag{2.7}$$

$$\langle z\rangle = \int_0^\infty z\psi(z)dz = -\int_0^\infty \phi(z)dz = 1 \; . \tag{2.8}$$

Experimental data, plotted as the quantity S_n versus z, should follow one single curve if the data obey the KNO-G scaling. This can be seen from the following equality:

$$S_n = \sum_{k=n}^\infty P_k = \int_{n/\langle\tilde{n}\rangle}^\infty \psi(z)dz = -\phi\left(\frac{n}{\langle\tilde{n}\rangle}\right) . \tag{2.9}$$

3 The continuous average multiplicity $\langle\tilde{n}\rangle$

The continuous average multiplicity $\langle\tilde{n}\rangle$ defined by (2.3) is slightly different from the discrete one

$$\langle n\rangle = \sum_{n=0}^\infty n\,P_n \quad (n = 0,1,2,3,\ldots). \tag{3.1}$$

The dependence of $\langle\tilde{n}\rangle$ on $\langle n\rangle$ is presented in Fig. 1.
Having any value of $\langle\tilde{n}\rangle$ we can calculate probability distribution P_n from (2.4) or (2.6) and then calculate $\langle n\rangle$ from the definition (3.1). However if we want to fit any theoretical distribution P_n^{th} to the probability distribution P_n and we need to know the value of $\langle\tilde{n}\rangle$ to calculate P_n^{th}.
First of all we should convert a usually measured charged multiplicity n_{ch} to n. For example in case of the e^+e^-collisions $n = n_{ch}/2$ and for pp collisions $n = (n_{ch} - 2)/2$. For $\langle n\rangle \gtrsim 1$ it is possible to use an approximation

$$\langle\tilde{n}\rangle \approx \langle n\rangle + 0.5 \; . \tag{3.2}$$

It was proven in Ref. [3] and can be seen in Fig. 1. For lower values of $\langle n\rangle$ we can use the iterative procedure

$$\langle\tilde{n}\rangle_0 = \langle n\rangle + 0.5$$

$$\langle\tilde{n}\rangle_{i+1} = \langle\tilde{n}\rangle_i + \langle n\rangle - \sum_{n=0}^\infty n\,P_n^{th}(\langle\tilde{n}\rangle_i) \; . \tag{3.3}$$

4 The lognormal distribution

The multiplicity distributions e.g. in e^+e^- collisions can be described by the lognormal distribution in the following way:
The density $f(\tilde{n})$ has a lognormal distribution:

$$f(\tilde{n}) = \frac{N}{\sqrt{2\pi}\sigma}\cdot\frac{1}{\tilde{n}+c_f}\,\exp\left(-\frac{[\ln(\tilde{n}+c_f)-\mu_f]^2}{2\,\sigma^2}\right) . \tag{4.1}$$

The scaling function $\psi(z)$ can be obtained from the density function by substituting:

$$\frac{\tilde{n}}{\langle\tilde{n}\rangle} = z \; , \qquad c_f = c\langle\tilde{n}\rangle \; , \qquad \mu_f = \mu + \ln\langle\tilde{n}\rangle \tag{4.2}$$

into formula (4.1). We then obtain the scaling function of the lognormal shape:

$$\psi(z) = \frac{N}{\sqrt{2\pi}\sigma} \cdot \frac{1}{z+c} \exp\left(-\frac{[\ln(z+c)-\mu]^2}{2\sigma^2}\right), \tag{4.3}$$

$$\phi(z) = -\frac{N}{2} \operatorname{erfc}\left(\frac{\ln(z+c)-\mu}{\sqrt{2}\,\sigma}\right) . \tag{4.4}$$

The symbol "erfc" stands for the complementary error function

$$\operatorname{erfc}(x) = \frac{2}{\sqrt{\pi}} \int_x^{\infty} e^{-t^2} dt . \tag{4.5}$$

An effective routine to calculate erfc(x) numerically is given in Ref. [11].

Parameter N is given by the first normalization condition (2.7) [1]

$$N = 2 \ / \ \operatorname{erfc}\left(\frac{\ln c - \mu}{\sqrt{2}\,\sigma}\right) . \tag{4.6}$$

Parameters μ, c and σ are constrained by the second normalization condition (2.8):

$$R \cdot \exp\left(\mu + \frac{\sigma^2}{2}\right) - c = 1 \tag{4.7}$$

where

$$R = \frac{\operatorname{erfc}\left(\nu - \sigma/\sqrt{2}\right)}{\operatorname{erfc}(\nu)} \quad \text{and} \quad \nu = \frac{\ln c - \mu}{\sqrt{2}\,\sigma} . \tag{4.8}$$

The function $\psi(z)$ has a maximum

$$\psi_{max} = \frac{N}{\sqrt{2\pi}\sigma} \exp\left(-\mu + \frac{\sigma^2}{2}\right) \text{ at } \quad z_{max} = \exp(\mu + \sigma^2) - c . \tag{4.9}$$

The dispersion of $\psi(z)$ is equal to

$$D = \sqrt{\langle z^2 \rangle - \langle z \rangle^2} = \exp\left(\mu + \frac{\sigma^2}{2}\right) \cdot \sqrt{\exp(\sigma^2) - 1} . \tag{4.10}$$

5 How to fit the lognormal distribution ?

Because of the normalization conditions (4.6,4.7) the lognormal function $\psi(z)$ has two free parameters. In principle one can fit e.g. μ and σ and calculate N and c from these conditions. However μ and σ are strongly correlated, thus the fitting procedure is unstable and obtained values of μ and σ are more or less accidental. Moreover a correct error analysis is almost impossible. It is seen from Fig. 2 where the χ^2 value is plotted as a function of μ and σ in a form of contour plot for the e^+e^-TASSO multiplicity data [12] as an example.

Better solution is to take the shift c and the dispersion D as free parameters and calculate μ and σ from (4.7) and (4.10). In case of the e^+e^-data values of c and D are not too large and we can use the approximation

$$N \approx 1 \quad \text{and} \quad R \approx 1 . \tag{5.1}$$

[1]We assume that $c > 0$. If $c \leq 0$ then $N = 1$ and in the next formula $R = 1$.

Hence

$$\sigma \approx \sqrt{\ln\left[\left(\frac{D}{1+c}\right)^2 + 1\right]} , \quad \mu \approx \ln(c+1) - \frac{\sigma^2}{2} . \tag{5.2}$$

The respective contour plot is presented in Fig. 3. In this case the parameters are almost uncorrelated and the minimum is well defined

$$c = 0.81 \pm 0.06 , \quad D = 0.282 \pm 0.002 \quad (\chi^2/NDF = 41.5/68) , \tag{5.3}$$

From (5.2) we have

$$\sigma = 0.155 , \quad \mu = 0.581 . \tag{5.4}$$

However for larger c and D the formulae (5.2) should be used with caution because the approximation (5.1) can be no longer valid. It can be seen from Figs 4 and 5. Figure 4 shows the dependence of N on c and D as a contour plot. The dependence of R on c and D is plotted in Fig. 5.

Neglecting changes of R implies a breaking of the second normalization condition (4.7). It is illustrated in Fig. 6. This contour plot presents the dependence of $\langle z \rangle$ on c and D if R is not calculated from (4.8) but strictly set to 1. For example, for $c = 4.25$ and $D = 0.629$ (fit to the pp data [2]) we have $N = 1.061$ and $R = 1.010$. Now if we neglect deviation of R from the unity we get $\langle z \rangle = 1.055$ instead of 1. It can produce unexpected effects during calculations of P_n.

To avoid such problems we should take into account deviations of N and R from unity. N can be calculated directly using the formula (4.6). In case of R situation is more complicated. First of all we should rewrite the formulae (5.2) including R

$$\sigma = \sqrt{\ln\left[\left(\frac{D \cdot R}{1+c}\right)^2 + 1\right]} , \quad \mu = \ln\left(\frac{c+1}{R}\right) - \frac{\sigma^2}{2} . \tag{5.5}$$

It is impossible to use these formulae directly because the formula (4.8) for R contains σ and μ. However the iterative procedure can be applied

$$\begin{aligned} R_0 &= 1 \\ \sigma_i &= \sigma(R_{i-1}) \qquad \mu_i = \mu(R_{i-1}, \sigma_i) \\ R_i &= R(\sigma_i, \mu_i) \end{aligned} \tag{5.6}$$

where $\sigma(\cdot)$, $\mu(\cdot)$, and $R(\cdot)$ denotes functions defined by (5.5) and (4.8) respectively. The procedure is reasonably fast convergent. It can be seen in Fig. 7 which is a contour plot of $\langle z \rangle$ similar to Fig. 6 but after 5 iterations.

6 The probit diagram

A convenient graphical way to present the lognormal distribution is the so-called "probit diagram" [13]. Let us denote the distribution function of the normal density by $F(x)$

$$F(x) = \frac{1}{\sqrt{\pi}} \int_{-\infty}^{x} \exp(-x^2)dx \; . \tag{6.1}$$

When variable n has a lognormal distribution, we can write

$$\sum_{i=n}^{\infty} P_i = N \cdot \left[1 - F\left(\frac{\ln(n + c_f) - \mu_f}{\sqrt{2}\sigma} \right) \right] \; . \tag{6.2}$$

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

$$F^{-1}\left(F(x)\right) = x \; . \tag{6.3}$$

Hence

$$F^{-1}\left(1 - \sum_{i=n}^{\infty} P_i \,/\, N \right) = \frac{\ln(n + c_f) - \mu_f}{\sqrt{2}\sigma} \; . \tag{6.4}$$

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

$$F^{-1}\left(1 - \sum_{i=n}^{\infty} P_i \,/\, N \right) = \frac{\ln(z + c) - \mu}{\sqrt{2}\sigma} \; . \tag{6.5}$$

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

References

[1] R.Szwed and G.Wrochna: Z. Phys. C – Particles and Fields 47 (1990) 449.

[2] R.Szwed, G.Wrochna and A.K.Wróblewski: Warsaw University Preprint IFD/7/1990.

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

[4] R.Szwed: "Novel Approach to Scaling in Multiplicity", Warsaw University Preprint IFD/3/1985, Proceedings of the VIII Warsaw Symposium on Elementary Particle Physics, Kazimierz, Poland, 1985, p.301.

[5] R.Szwed: "Some conceptional and technical aspects of the multiplicity distribution study", Talk given at the Perugia Workshop on Multiparticle Dynamics, Perugia, Italy 1988

[6] R.Szwed: "Multiplicity Distributions in Proton-Proton and Nucleus-Nucleus Collisions", Warsaw University Preprint IFD/6/1989.

[7] R.Szwed, G.Wrochna and A.K.Wróblewski: "Genesis of the Lognormal Multiplicity Distribution in the e^+e^-Collisions and Other Stochastic Processes", Warsaw University Preprint IFD/1/1990, Modern Physics Letters A 23 (1990) 1851.

[8] G.Wrochna: "Multiparticle Production as a Bivariate Branching Process", Warsaw University Preprint IFD/5/1990, to be published in Proceedings of the XIII Warsaw Symposium on Elementary Particle Physics, Kazimierz, Poland, May 1990.

[9] R.Szwed, G.Wrochna and A.K.Wróblewski: "New AMY and DELPHI Multiplicity Data and the Lognormal Distribution", Warsaw University Preprint IFD/6/1990, to be published in Modern Physics Letters A.

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

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

[12] *TASSO* (14.0, 22.0, 34.8, 43.6 GeV) W.Braunschweig et al.: Z. Phys. C – Particles and Fields 45 (1989) 1939

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

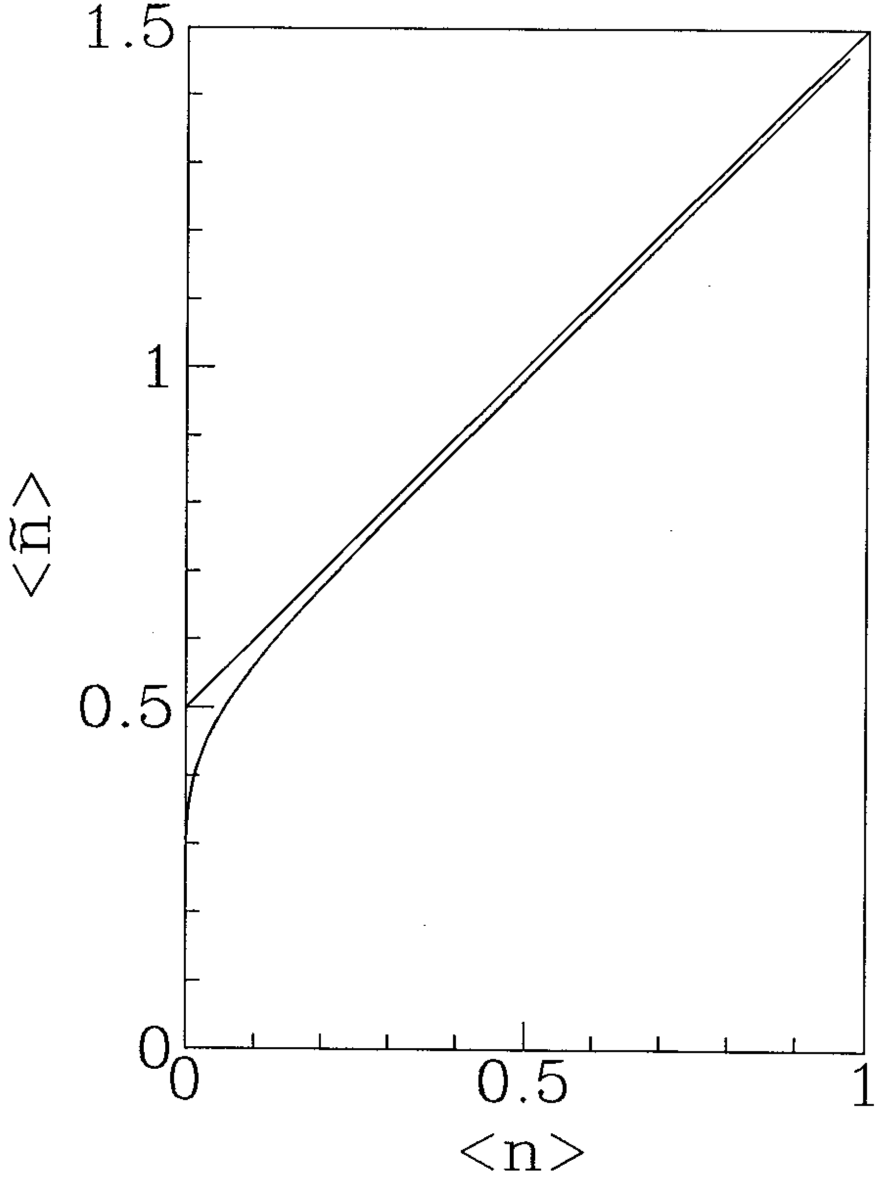

Fig.1.

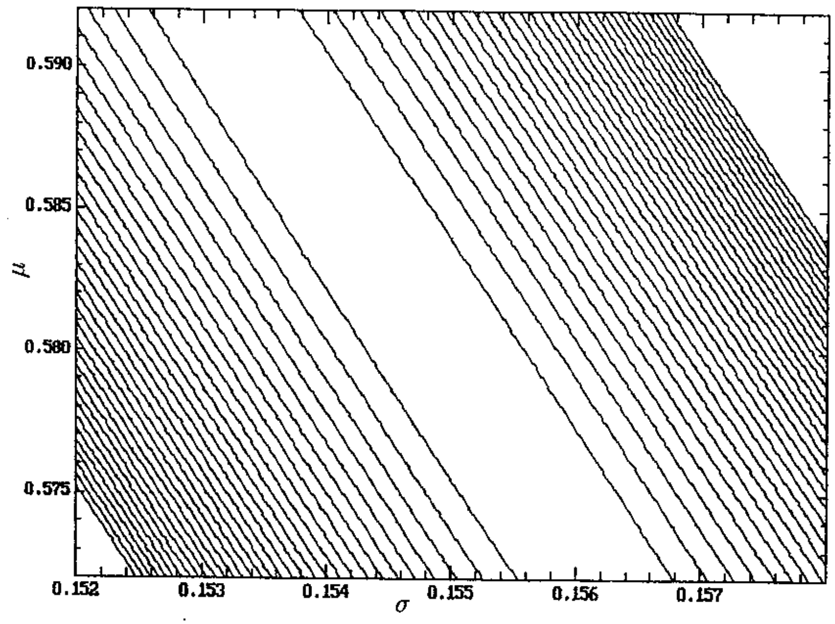

Fig.2. Dependence of χ^2 on μ and σ for fit to TASSO data

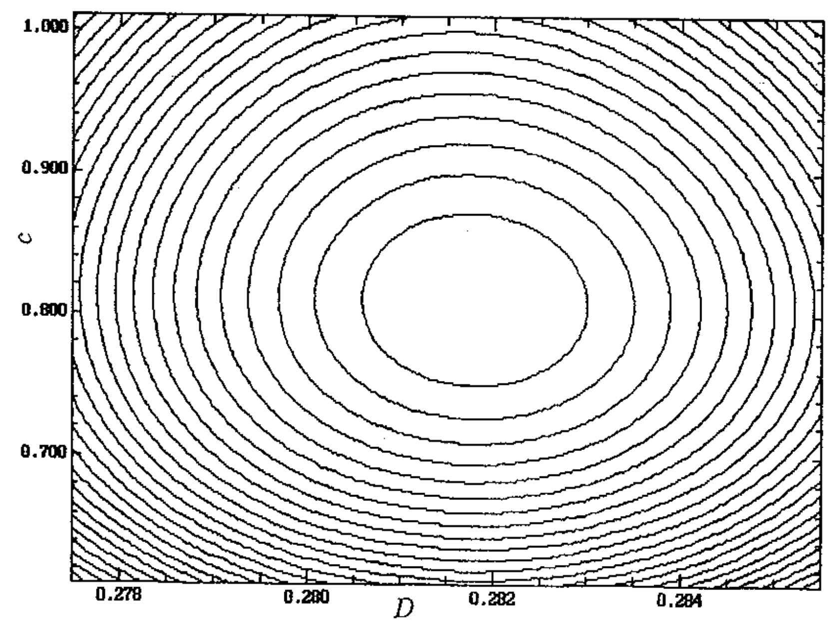

Fig.3. Dependence of χ^2 on c and D for fit to TASSO data

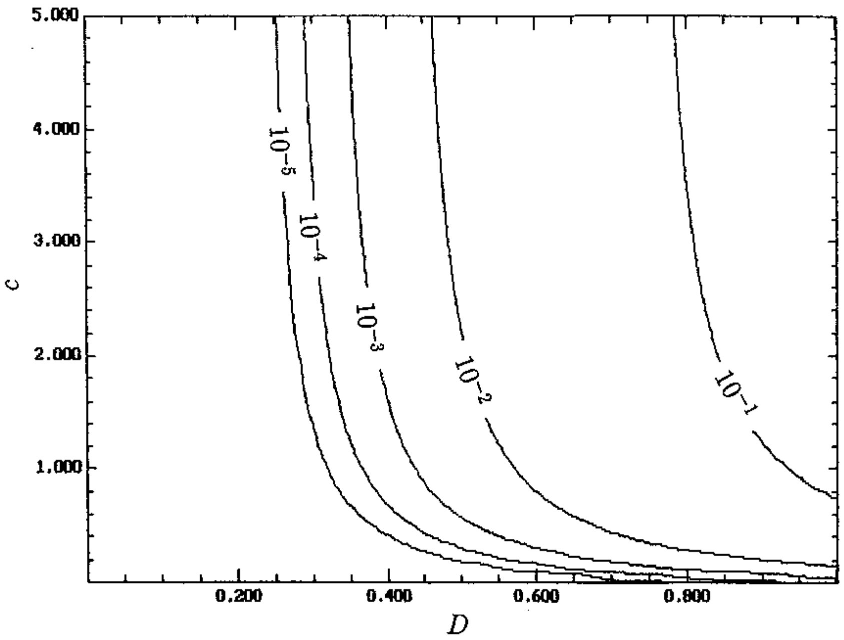

Fig.4. Dependence of $N - 1$ on c and D.

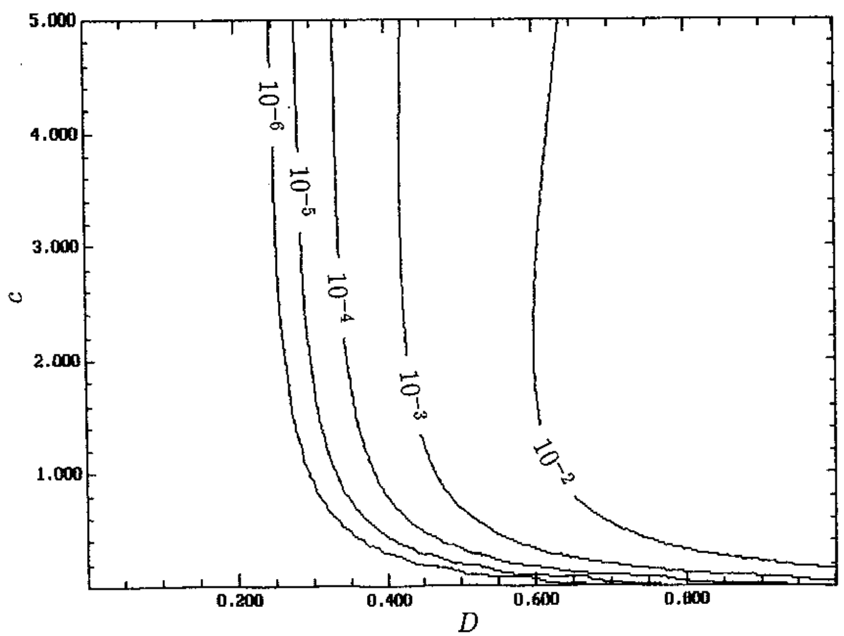

Fig.5. Dependence of $R - 1$ on c and D.

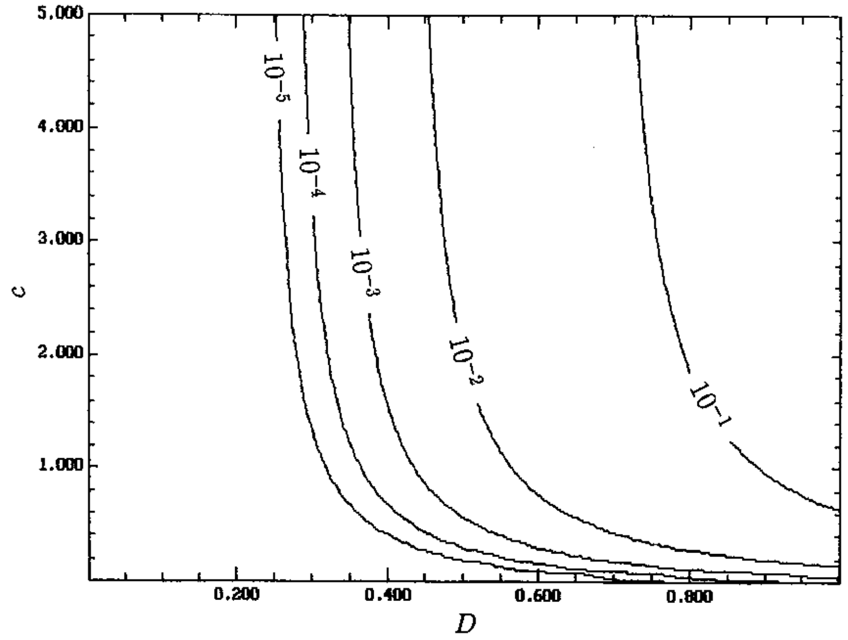

Fig.6. Dependence of $\langle z \rangle - 1$ on c and D for $R = 1$.

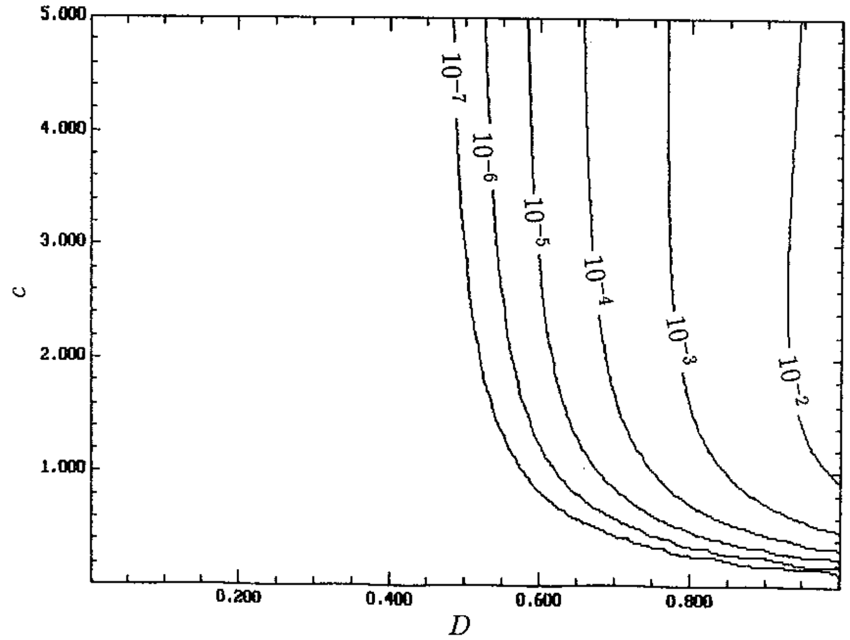

Fig.7. Dependence of $\langle z \rangle - 1$ on c and D after 5 iterations.