



# Point process statistics improves particle size analysis

Dietrich Stoyan<sup>1</sup> · Georg Unland<sup>2</sup>

Received: 24 April 2022 / Accepted: 18 August 2022 / Published online: 16 September 2022  
© The Author(s) 2022

## Abstract

This paper re-considers the foundations of particle size statistics. While traditional particle size statistics consider their data as samples of random variables and use methods of classical mathematical statistics, here a particle sample is treated as a point process sample, and a suitable form of statistics is recommended. The whole sequence of ordered particle sizes is considered as a random variable in a suitable sample space. Instead of distribution functions, point process intensity functions are used. The application of point process data analysis is demonstrated for samples of fragments from single-particle crushing of glass balls. Three cases of data handling with point processes are presented: statistics for oversize particles, pooling of independent particle samples and pooling of piecewise particle data. Finally, the problem of goodness-of-fit testing for particle samples is briefly discussed. The point process approach turns out to be an extension of the classical approach, is simpler and more elegant, but retains all valuable traditional ideas. It is particularly strong in the analysis of oversize particles.

**Keywords** Particle size · Point process · Intensity function · Goodness-of-fit test

## 1 Introduction

The statistical analysis of the size (length, surface, cross-section area, volume, or mass) of members of particle collectives is a classical field of engineering statistics. Usually, it starts with a sample of particles, taken by certain rules adapted to the particles' nature. (Often the total number of particles in the sample, which may be very large, is ignored.) These particles are measured by various methods, for example by sieving or by individual particle measurement (e.g. by means of image analysis). The data obtained are transformed into an empirical particle size distribution function, and finally it is tried to fit this distribution to some standard distribution such as Rosin–Rammler–Sperling–Bennett (RRSB), Gates–Gaudin–Schuhmann (GGS) or logarithmic normal (Bernhardt [1], Unland [2]). Of central interest are the cumulative size distribution functions  $Q_0(x)$  and  $Q_3(x)$ ,

where  $Q_0(x)$  denotes the proportion of total number of all particles smaller than size  $x$  and  $Q_3(x)$  the proportion of total volume respectively mass of all particles smaller than  $x$ . The corresponding probability density functions are  $q_0(x)$  and  $q_3(x)$ .

In the long time of using this methodology frequently problems have been observed at the tails of the distributions, both for very small and very large particles. This has led to various modifications of the classical distributions.

For detailed investigations of the tail behavior it makes sense to analyse sub-samples, for example all particles of a size larger than a given  $x_u$ . If for this smaller sample an appropriate distribution function is wanted (which is a conditional distribution), some additional calculations are necessary.

This has led the authors to the decision to change the statistical approach. The new standpoint considers the data as samples of point sequences on an axis of size, where each particle is represented by a point. This leads to the use of methods of point process statistics. By the way, this approach is not fully new. In literature of particle size analysis the idea of point processes has appeared already implicitly in the papers Brown [3], Brown and Wohletz [4] and Bernhardt [5], as explained in the Discussion section.

The basic idea is to consider the whole sequence of all particle size numbers  $x_i$  in the sample as a single random variable,

✉ Dietrich Stoyan  
stoyan@math.tu-freiberg.de

Georg Unland  
Georg.Unland@iam.tu-freiberg.de

<sup>1</sup> Institute for Stochastics, TU Bergakademie Freiberg,  
09596 Freiberg, Germany

<sup>2</sup> Institute for Mineral Processing Machines, TU Bergakademie  
Freiberg, 09596 Freiberg, Germany

as a random vector with a random number of components. This is quite natural in single-particle crushing or in the analysis of rock fragments after a blast. There the whole particle collective appears in one physical step, which may be seen as an analogue of a throw of a dice in elementary probability theory. Perhaps also a sample from a stream of material moving on a conveyor belt can be interpreted similarly. This situation is of course not given if exactly  $n$  particles are taken by some sampling plan in order to study particle properties, as in the sampling model of Gy [6].

In mathematics such random sequences are called ‘point processes’. It should be noted that the word ‘process’ does not necessarily imply a relationship to time. Indeed, early applications studied really events on a time axis and the whole terminology in some books on point processes is time-related. However, many other applications are possible and the present paper demonstrates the use of point process statistics in particle statistics. An early example is [3], where points on a mass axis are considered. The present paper generalizes this by assuming that the points lie on a size axis, where the ‘size’ variable is denoted by the general symbol  $x$ . The particles are ordered with respect to size. ‘Large particles’ are those where the size is larger than a given upper level value  $x_u$  and ‘small’ are those below some lower level  $x_l$ .

The theory of point processes offers various models for such processes, which describe various forms of inner correlations of the point positions. The points may appear in clumps, where a given point tends to have further points in its neighbourhood; or the points may appear regularly, in reduced randomness. The simplest model is the Poisson process, which belongs to the case of complete randomness of point locations. For this model the statistical description is easy: a mean value function, called intensity function, suffices. As it seems, just this simple model is appropriate for particle size statistics.

The present paper first explains why point process statistics may be considered as a natural approach to particle size statistics. Then the necessary fundamentals of the theory of point processes are presented. It follows a thorough analysis of data from single-particle crushing experiments with glass balls. This example is chosen since the samples are given a-priori, without any form of special sampling. However, since the present paper is a methodical one, no new results in the theory of single-particle crushing are aimed. Three typical cases of statistical calculations are discussed in order to show advantages of the point process approach. Finally, a  $\chi^2$ -goodness-of-fit test for particle size statistics is given.

## 2 Particle size analysis in research and industry

Particle size statistics have different aims and methods in research and industry. The diversity of problems is expressed by the various cumulative size distribution

functions used:  $Q_0(x)$ ,  $Q_1(x)$ ,  $Q_2(x)$  and  $Q_3(x)$ , which denote the proportions of number, total length, surface, cross-section area, and volume resp. weight of all particles smaller than size  $x$ . Their application depends on machines, processes and materials which have to be assessed, sized and controlled. In some cases surfaces and volumes are interesting, particles’ surface and volume distribution functions  $Q_2(x)$  and  $Q_3(x)$  and by them the volume-specific surfaces of particles determine drying or combustion processes among other parameters.

In other cases the number and volume distributions  $Q_0(x)$  and  $Q_3(x)$  are important. In mining the number of big rocks determines the type, size and operation of crushers. The amount of very small particles influences the dust development and the necessary dedusting equipment.

In the following problems in mining and comminution applications are briefly sketched.

In research many investigations study the comminution behavior of single particles or limited numbers of particles in order to gain characteristics of comminution processes, like breakage functions or specific power consumption. Figure 1 shows a typical situation of such an experiment. A stone (Greywacke) is crushed by compression with a relative displacement of 15% in order to gain data of the breakage function.

In industrial processes, typically in minerals engineering, representative samples of comminuted material are taken from piles, hoppers, chutes or conveyor belts. The samples consist of a huge number of particles. The aim of statistical size analyses is to get information on the size distribution in order to adjust industrial processes or to assess product qualities. Figure 2 shows blasted rocks with an oversized boulder from the copper mine Los Pelambres in Chile.

In both cases the particles occur in a wide range of sizes. In particular, in mineral processing the biggest boulders come from blasted mining faces and go then into crushers in order to be comminuted. They may show sizes up to more than 2 m. Depending on the kind and size of crusher those boulders can endanger the crusher operation. On the other hand, the smallest particles appear in very fine clays. Clay minerals like illite or montmorillonite may be smaller than 100 nm. These mineral components have to be removed from ore slurries by desliming devices prior to flotation; otherwise flotation will not work properly.

There is no single method of measurement for the whole range of particle sizes [1, 7]. It depends on size and material, which technique is most suitable to gain reliable results. Therefore, there are various measuring systems that apply different physical effects to detect directly or indirectly particle sizes. Electron or atomic force microscopy, dynamic light scattering, centrifugal sedimentation or laser diffraction are used for particles in the nanometer and micrometer range. Sieve and image analysis systems



**Fig. 1** Greywacke particle (equivalent diameter of 50 mm) from Breitenau (Saxony) before and after compression test



**Fig. 2** Pile of blasted copper ore with an oversized boulder (Los Pelambres mine in Chile)

cover the coarser part of the size range. Some techniques can work online, others have to work offline.

The measuring results are structured as well as processed and finally recorded in tables, histograms, graphs and statistically analysed. It is tried to fit theoretical distribution functions to the data, in order to characterize the material by a small number of parameters.

However, there is a principal difficulty in the analysis of particle distributions, which is particularly relevant in the industrial context. Typically, the samples contain a huge number of small particles and only a low number of big ones with a wide size gap in between. Since it is impossible to precisely count the number of small particles, for small particles the distribution is based on the mass of particles within the relevant size classes. On the other side, it is often interesting, which numbers and sizes of big particles can be expected. A typical situation is given by occasionally occurring, but detrimentally very large boulders after blasting a mining face. For plant operation it is important to assess the number and size of those oversized rocks. Here the classical methods of statistics seem to come to their limit.

### 3 Classical statistics and single-particle-crushing statistics

For understanding the basic idea of this paper, it may be useful to remind the reader of the classical idea of a ‘random sample’ (Montgomery and Runger [8], p. 286): There is a random variable  $X$  with cumulative distribution

function  $F(x)$ ,  $F(x) = P(X \leq x)$ . One wants to estimate  $F(x)$  or some corresponding parameters like mean or variance. With this aim  $n$  independent observations are made under fixed conditions, which leads to  $n$  results  $X_i$ , with  $i = 1, 2, \dots, n$ . The  $n$  random variables  $X_i$  have all the same distribution function as  $X$  and they are independent; they form a *random sample*. For the analysis of such data there exist many well-known statistical methods.

A simple example is as follows. In an experiment the strength of (identical) glass balls is studied. A sample of  $n$  balls is taken and the strengths of all these balls are measured. It can be assumed that the data form a random sample if all balls were produced in the same process. Then, the sample can be analysed by the classical methods of statistics as in [8] and at the end, if wanted, the hypothesis can be tested that strength follows a logarithmic normal distribution.

Formally similar is the situation in the classical approach in the case of single-particle crushing if the size of fragments is of interest. There the set of all fragments from one particle is considered as a sample. The fragments are characterized by size characteristics such as weight, volume or equivalent diameter and classical statistics is applied. This leads to cumulative particle size distribution functions  $Q_0(x)$  and  $Q_3(x)$  as mentioned in the introduction. Formally, these distribution functions are empirical distribution functions of the sample of fragment sizes.

However, do the fragment sizes really form a random sample? No, this is not the case!

First, but not that important, the number  $n$  of elements in the sample is also random. It is difficult to predict. Note that if a sub-sample of large fragments is considered, the number may be in the order of some dozens, while the total number of fragments may be in the order of  $10^5$  and of little interest. Second, the fragment size characteristics are not statistically independent. Indeed, consider the fragment volumes. Their total sum is exactly equal to the volume of the crushed particle, i.e. a constant value. However, if the fragment volumes formed a random sample, they would be statistically independent and their sum would be a random variable too and not constant.

In this situation the authors recommend to re-interpret the sampling situation and to assume that a statistician who studies fragments of single-particle crushing does not analyse classical random samples—but point process samples.

Many particle samples contain a lot of very small particles with sizes which cannot be measured precisely. Counting and measuring is impossible. A classical way to circumvent this problem is to assume some theoretical distribution laws for the sizes, for example logarithmic normality. Then, it is possible to apply statistics for the particles larger than the critical size, which is denoted here by  $x_{\text{crit}}$ . The results obtained may be used for extrapolation to values of  $x$  smaller

than  $x_{\text{crit}}$ . This can be justified by super-precise pilot measurement or by the plausibility of the distributional assumption. Otherwise, only the safe data larger than  $x_{\text{crit}}$  should be analysed and presented in histograms with *absolute* frequencies or in closely related continuous functions, to avoid speculations. (*Relative* frequencies are questionable since the total number of particles in the sample is unknown.)

Finally, there is a more formal, mathematical problem in the context of very small particles. Sometimes it is elegant and natural to assume that  $x_{\text{crit}} = 0$ . For example, in statistics with  $Q_3(x)$ , i.e. related to volumes or masses, the equation

$$Q_3(0) = 0 \quad (1)$$

is natural, since the total weight of all particles smaller than some low limit may be negligible. The popular GGS distribution, which is suitable in this context, satisfies equation (1). Unfortunately, for such a distribution the transformation to the number distribution function  $Q_0(x)$  is not possible since it does not exist. However, when working with point processes, this difficulty is circumvented, see Section 4.

## 4 Point processes and their statistics

### 4.1 Fundamentals of the theory of point processes

Point processes are used to describe data sequences that are located in space and/or time. In the case considered here such data are generated in a natural way: We consider the particle sizes of a sample as points on the real line or  $x$ -axis. If arranged with respect to size the points form a sequence, which is interpreted as a one-dimensional point process.

These random points are located in a bounded interval  $I$  with deterministic ends  $x_{\text{min}}$  and  $x_{\text{max}}$ ,  $I = [x_{\text{min}}, x_{\text{max}}]$ . In our setting these limits are usually given in a natural way, often it is  $x_{\text{min}} = 0$ , and when equivalent diameters of fragments of glass balls of constant diameter  $d$  are considered, then  $x_{\text{max}}$  may be just  $d$ .

The random points between these limits are denoted by  $x_i$  with  $i = 1, 2, \dots, n$ . The whole random sequence  $\{x_i\}$  of the  $x_i$  is the point process.

The theory of one-dimensional point processes as used here is presented in detail in the textbook Snyder and Miller [9], for the multi-dimensional case in [10, 11].

For understanding the present paper, the following may be sufficient.

A point process is a random variable which generates point sequences, here in the interval  $I$ . The total number of points of such a sequence is random and finite and denoted by  $N(I) = n$ . For any deterministic subset  $A$  of  $I$  the random number of points of  $\{x_i\}$  in  $A$  is denoted by  $N(A)$ .

The mean behaviour of a point process is determined by the *intensity function*  $\lambda_0(x)$  and the *cumulative intensity function*  $\Lambda_0(x)$ , which are related by

$$\Lambda_0(x) = \int_{x_{\min}}^x \lambda_0(x)dx \text{ for } x_{\min} \leq x \leq x_{\max}. \tag{2}$$

The term  $\Lambda_0(x)$  denotes the mean number (or expectation) of points in the interval  $[x_{\min}, x]$ , in general probabilistic notation written as  $E(N([x_{\min}, x]))$ , where “ $E$ ” is the mathematical expectation operator. The mean total number is  $E(N(I)) = \Lambda_0(x_{\max})$ ; in the following often the simpler symbol  $N_T$  is used. The term  $\lambda_0(x) \cdot dx$  can be interpreted as the mean number of points in the infinitesimal interval  $[x, x + dx]$ . (The use of the subscript “0” will soon be understood, when the subscript “3” appears.)

In the particle size context, both intensity functions are closely related to the cumulative distribution function  $Q_0(x)$  and the corresponding density function  $q_0(x)$ :

$$Q_0(x) = \Lambda_0(x)/\Lambda_0(x_{\max}) \tag{3}$$

and

$$q_0(x) = \lambda_0(x)/\Lambda_0(x_{\max}). \tag{4}$$

Therefore,  $Q_0(x)$  could be called ‘normalized cumulative intensity function’ and  $q_0(x)$  ‘normalized intensity function’.

When working with point processes, often random quantities play a role that are constructed as follows. The members  $x_i$  of the point sequence  $\{x_i\}$  are transformed by a function  $f(x)$  to obtain  $f(x_i)$  and then the sum of all these values is considered. It is written as

$$\sum_{x_i} f(x_i). \tag{5}$$

If, for example,  $f(x) = 1$  for all  $x$ , then the sum is the total number  $N(I)$  of all particles. If  $f(x) = k_v x^3$  with a form factor  $k_v$  [12], then the sum in equation (5) is the total volume of all particles and denoted by  $V(I)$ ; its mean is denoted by  $V_T$ .

And if the summation is restricted to all  $x$  smaller than some limit  $x_0$ , then the sum is written as

$$\sum_{x_i < x_0} f(x_i).$$

For the means of such sums there are formulas known, yielded by the so-called Campbell’s theorem [13]:

$$E\left(\sum_{x_i} f(x_i)\right) = \int_{x_{\min}}^{\infty} f(x)\lambda_0(x)dx \tag{6}$$

and

$$E\left(\sum_{x_i < x_0} f(x_i)\right) = \int_{x_{\min}}^{x_0} f(x)\lambda_0(x)dx. \tag{7}$$

For the particular case of  $f(x) = 1$  this yields  $\Lambda_0(\infty)$  and  $\Lambda_0(x_0)$ , respectively. The result for  $f(x) = k_v x^3$  is denoted by  $\Lambda_3(x)$  with

$$\Lambda_3(x) = E\left(\sum_{x_i \leq x} k_v x_i^3\right). \tag{8}$$

This leads for particle statistics to

$$Q_3(x) = \Lambda_3(x)/\Lambda_3(x_{\max}) = \int_{x_{\min}}^x t^3 \lambda_0(t)dt / V_T \text{ for } x_{\min} \leq x \leq x_{\max} \tag{9}$$

and

$$q_3(x) = x^3 \lambda_0(x) / \Lambda_3(x_{\max}) = x^3 \lambda_0(x) / V_T \text{ for } x_{\min} \leq x \leq x_{\max}, \tag{10}$$

where the form factor  $k_v$  cancels out.

Finally, equation (9) yields together with (4) the following relationship between  $q_0(x)$  and  $Q_3(x)$ ,

$$Q_3(x) = \int_{x_{\min}}^x t^3 q_0(t)dt / \int_{x_{\min}}^{x_{\max}} t^3 q_0(t)dt, \tag{11}$$

see e.g. [1], p. 256, [12], and [14]. The latter paper contains particle data for this relationship.

### 4.2 Basics of point process statistics

The two intensity functions  $\lambda_0(x)$  and  $\Lambda_0(x)$  are statistically estimated as follows. One takes  $n$  samples of particles and determines the sample sizes  $n_i$  and their particle sizes or points  $\{x_{ij}\}$  with  $i = 1, 2, \dots, n$  and  $j = 1, 2, \dots, n_i$ . Then the estimator  $\hat{\Lambda}_0(x)$  of  $\Lambda_0(x)$  is

$$\hat{\Lambda}_0(x) = \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^{n_i} \mathbf{1}_{[x_{\min}, x]}(x_{ij}) \text{ for } x_{\min} \leq x \leq x_{\max}, \tag{12}$$

where  $\mathbf{1}_{[x_{\min}, x]}(z) = 1$  if  $x_{\min} \leq z \leq x$  and  $= 0$  otherwise. The inner sum in (12) is just the number of points respectively particle sizes of the  $i$ th sample in the interval  $[x_{\min}, x]$ .

For the estimation of  $\lambda_0(x)$  histogram-based methods may be used: Dividing the  $x$ -axis into class intervals or bins and counting of numbers of  $x_{ij}$  in the class intervals or bins to obtain a frequency distribution, absolute frequencies, not relative. An alternative is the use of kernel estimators, i.e. estimators of the form

$$\hat{\lambda}_0(x) = \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^{n_i} k(x - x_{ij}), \tag{13}$$

where  $k(z)$  is a kernel function, a suitable probability density. These estimators yield automatically smooth curves, their results are just continuous counterparts of histograms. This paper uses the Gaussian kernel, where  $k(x)$  is the probability density function of a Gaussian distribution with mean 0 and a suitable variance  $\sigma$ . The  $\sigma$  controls the smoothness of the curve for  $\hat{\lambda}_0(x)$ : Large values of  $\sigma$  yield smooth curves, while for small the curves may be spiky. The choice of  $\sigma$  is perhaps a weak point of the method, since it introduces a subjective factor into the estimation, which is, however, also given by the choice of bin size when histogram-based methods are used.

Note here that the value of  $\hat{\lambda}_0(x)$  is given by the points close to  $x$ . In contrast, any estimator of the probability density  $q_0(x)$  at  $x$  also needs information from the whole sample, in particular of the total number of points, since relative frequencies are considered.

The estimators of  $\Lambda_3(x)$  and  $\lambda_3(x)$  are analogous,

$$\hat{\Lambda}_3(x) = \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^{n_i} k_v x_{ij}^3 \mathbf{1}_{[x_{\min}, x]}(x_{ij}), \tag{14}$$

and

$$\hat{\lambda}_3(x) = \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^{n_i} k_v x_{ij}^3 k(x - x_{ij}) \tag{15}$$

or

$$\hat{\Lambda}_3(x) = \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^{n_i} v_{ij} \mathbf{1}_{[x_{\min}, x]}(x_{ij}), \tag{16}$$

and

$$\hat{\lambda}_3(x) = \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^{n_i} v_{ij} k(x - x_{ij}), \tag{17}$$

where  $v_{ij}$  is the volume of the  $j$ th particle of the  $i$ th sample. (The kernel functions  $k(\cdot)$  in the equations above are of course different.)

If there is a lower limit  $x_{\text{crit}}$  below which size data  $x$  are questionable, then  $\hat{\lambda}_0(x)$  and  $\hat{\lambda}_3(x)$  can be determined for all  $x \geq x_{\text{crit}}$  by the estimators in the equations (13) and (15). In contrast, the probability density  $q_0(x)$  (defined for the whole interval  $[x_{\min}, x_{\max}]$ ) cannot be estimated in this situation since the number of particles between  $x_{\min}$  and  $x_{\text{crit}}$  is questionable and consequently also the total number of particles between  $x_{\min}$  and  $x_{\max}$ . In contrast, the intensity estimator  $\hat{\lambda}_0(x)$  uses only the local data around  $x$ , while any probability density estimator needs also global information, namely the

total sample size, which is questionable. The same also holds for the distribution function  $Q_0(x)$ .

### 4.3 The inhomogeneous Poisson process

The simplest non-trivial example of a point process on  $I$  is the Poisson process, more precisely the inhomogeneous Poisson process.

#### 4.3.1 Definition of the process

The inhomogeneous Poisson process has two characteristic properties:

- the numbers of points in disjoint subintervals of  $[x_{\min}, x_{\max}]$  are independent random variables,
- the probability distributions of these numbers are Poisson distributions.

The latter means, in particular, that for an interval  $[a, b]$  (with  $x_{\min} \leq a < b \leq x_{\max}$ ) the probability to have  $i$  points in the interval is

$$P(N([a, b]) = i) = \frac{\mu^i}{i!} \exp(-\mu) \text{ for } i = 0, 1, \dots, \tag{18}$$

where

$$\mu = \int_a^b \lambda_0(x) dx. \tag{19}$$

The function  $\lambda_0(x)$  is the intensity function of the process. If  $\lambda_0(x)$  is constant and the process is defined on the full real line, the point process is called *homogeneous* Poisson process.

The definition shows that the distribution of the inhomogeneous Poisson process is completely given by the intensity function  $\lambda_0(x)$ . Therefore, it is a model without additional features. It plays, also in our context, the role of a ‘null model’: if the Poisson process hypothesis is rejected, then a model more complicated has to be employed—if a model is wanted. Only then there is interaction between the points, which are in the case of a Poisson process completely random, following only the trend given by the intensity function.

#### 4.3.2 Simulation

A possible way to understand the inhomogeneous Poisson process is to learn how it is simulated, i.e., how to simulate samples of the process, irregular sequences of points between  $x_{\min}$  and  $x_{\max}$  that follow suitable laws. The standard method for this purpose has two steps: At first a Poisson-distributed random number is generated, a sample of the

total number of points  $N(I)$  in the final point sequence. The corresponding parameter  $\mu_0$  of the Poisson distribution is

$$\mu_0 = \int_{x_{\min}}^{x_{\max}} \lambda_0(x) dx = \Lambda_0(x_{\max}). \tag{20}$$

The literature offers various methods for ‘generating Poisson variates’. The obtained realization of  $N(I)$  is denoted by  $n_p$ , it is the number of points to simulate.

In a second step  $n_p$  random points are generated, independently, with probability density function  $q_0(x) = \lambda_0(x)/\Lambda_0(x_{\max})$ . The literature offers again various methods for ‘generating random variates’.

Note that the second step is just the method of simulating a random sample of deterministic size  $n_p$  for a distribution function  $Q_0(x)$ . This shows that in the case of a Poisson process the point process statistics is close to classical statistics. The point process standpoint considers just a higher degree of variability, since the sample size is random too and has to be statistically analysed.

### 4.3.3 Testing the Poisson process hypothesis

A test of the hypothesis that a sample of  $n$  point sequences belongs to an inhomogeneous Poisson process consists usually of two sub-tests. At first it is tested whether the total numbers of points  $n_i$  per sequence follow a Poisson distribution. Second it is tested whether the point positions behave as expected for an inhomogeneous Poisson process.

The first step can be carried out by classical goodness-of-fit tests for the Poisson distribution, applied to the particle numbers  $n_i$ . For very small  $n$  comparison of sample mean  $\bar{x}$  and variance  $s^2$  of the  $n_i$  we recommend the following ‘better-than-nothing’ solution: Since for a Poisson random variable mean and variance are equal, the Poisson hypothesis may be rejected if the difference between  $\bar{x}$  and  $s^2$  is large.

The second step is carried out separately for the  $n$  samples, conditional to the numbers  $n_i$ . It is helpful to eliminate the special form of the intensity function by a trick, called here ‘size transformation’. (In the time-related literature it is called ‘time transformation’.) It uses the cumulative intensity function  $\Lambda_0(x)$ , which is assumed to be known.

By the transformation the original sizes  $x_{ij}$  are transformed into values  $z_{ij}$  by the relation

$$z_{ij} = \Lambda_0(x_{ij})/\Lambda_0(x_{\max}). \tag{21}$$

If the  $x_{ij}$  for fixed  $j$  belong to an inhomogeneous Poisson process with cumulative intensity function  $\Lambda_0(x)$ , then the  $z_{ij}$  form a piece of a homogeneous Poisson process in the interval  $[0, 1]$ , i.e., the points  $z_{ij}$  are distributed as independent uniformly distributed points in  $[0, 1]$ .

For testing the hypothesis of independence and uniformity various methods exist. A simple  $\chi^2$ -test is described in Subsection 5.3.

## 5 Probability density functions and intensity functions

Classical particle size statistics uses the distribution functions  $Q_0(x)$  and  $Q_3(x)$  and the corresponding density functions. In point process terms these functions are given by the equations (3) and (9), they are simply multiples of the cumulative intensity functions  $\Lambda_0(x)$  and  $\Lambda_3(x)$ . Therefore, there is no reason not to work with them also in the point process approach.

In particle statistics various forms of  $Q_3(x)$  are used [1] [2]. Two important cases are:

- the RRSB (Rosin–Rammler–Sperling–Bennett) distribution function,

$$Q_{3R}(x) = 1 - \exp\left(-\left(\frac{x}{x'}\right)^r\right) \text{ for } x \geq 0, \tag{22}$$

where  $r$  and  $x'$  are positive parameters

- the GGS (Gates-Gaudin-Schuhmann) distribution function,

$$Q_{3G}(x) = \left(\frac{x}{x'}\right)^k \text{ for } 0 \leq x \leq x', \tag{23}$$

where  $k$  and  $x'$  are positive parameters.

Note that (23) is an approximation of (22) for very small  $x/x'$  with  $k = r$ . In both cases it is  $x_{\min} = 0$ .

Sometimes so-called three-parameter or shifted distributions are used. For RRSB it is

$$Q_{3R}(x) = 1 - \exp\left(-\left(\frac{x - x_{\min}}{x'}\right)^r\right) \text{ for } x \geq x_{\min}, \tag{24}$$

for use for particle collectives of size larger than  $x_{\min} > 0$ .

Since the present paper is particularly interested in  $q_0(x)$  and  $Q_0(x)$ , the transformation of  $Q_3(x)$  into  $Q_0(x)$  is considered now. The transformation formula is for particle sizes in the range  $0 < x_{\min} \leq x \leq x_{\max}$

$$Q_0(x) = \int_{x_{\min}}^x t^{-3} q_3(t) dt / \int_{x_{\min}}^{x_{\max}} t^{-3} q_3(t) dt, \tag{25}$$

which is a consequence of equation (11).

In case of  $x_{\min} = 0$  known difficulties arise: Assume that  $Q_3(x)$  is the GGS distribution function with a parameter  $k$  smaller than 3. Then the corresponding probability density function is

$$q_3(t) = k\left(\frac{t}{x'}\right)^{k-1}/x' \tag{26}$$

and the integrand in (25) has the form  $ct^{-m}$  with  $m$  larger than 1 and some constant  $c$ . Thus the integral takes, caused by  $x_{\min} = 0$ , the value  $\infty$  and no  $Q_0(x)$  and  $q_0(x)$  do exist. Rumpf and Ebert [12] and Alderliesten [15] demonstrated the same effect for the RRSB distribution. The lognormal distribution is free of this disadvantage [12].

We remark that Bernhardt [1] gives the transformation formulas with  $x_{\min} = 0$ , while Stieß [16] correctly has a positive lower limit of integration.

The authors consider the problem with the transformation just discussed not a weak point of the classical theory: The RRSB and GGS distributions were originally introduced for  $Q_3(x)$ , where they have worked well, while  $Q_0(x)$  was out of consideration. In contrast, Alderliesten [15] points out that the RRSB (and GGS) distribution should not be expected to describe correctly the small particles outside the range of measurement.

However, if one works with intensity functions, all goes smoothly. Intensity functions do not need to satisfy an equation of the form

$$\int_0^\infty f(x)dx = 1$$

as a probability density function  $f(x)$  has to fulfil. An intensity function

$$\lambda_0(x) = q_3(x)x^{-3}V_T \text{ for } 0 \leq x \leq \infty \tag{27}$$

with the  $q_3(x)$  in equation (26) (with dummy variable  $t$  replaced by  $x$ ) is finite for all  $x > 0$ , i.e. a regular object. Integrals of the form  $\int_u^z \lambda_0(x)dx$  with this  $\lambda_0(x)$  and  $z > u$  are finite for  $u > 0$ , giving the mean number of particle sizes between  $u$  and  $z$ . And  $\int_0^z \lambda_0(x)dx = \infty$  means simply that there are many small particles.

The smooth operation with poles at  $x = 0$  is a mathematical argument for the use of point process methods in particle statistics as already sketched at the end of Section 3.

## 6 Statistics for fragments of glass balls

### 6.1 General

Now an example demonstrates the application of methods of point process statistics for real particles. For this purpose experimental results from the Diploma thesis Lehmann [17] are used, which studied the fragmentation of glass balls. The aim is to show how point process statistics works, not to find new results in the field

of single-particle crushing. Therefore, only a small set of data is considered and only one method of particle size measurement is used.

Lehmann made systematic crushing experiments with 50 glass balls of diameter 50 mm. The balls were destroyed by the blow of a hammer, with different offered energies between 70 and 100 J. Not all glass balls were completely destroyed, but only these are of interest here. The following reports about the six balls completely destroyed by offered energy 75 J. (The results for the other energies are similar.) Here the fragment-size parameter is the equivalent diameter, therefore  $x_{\max} = 50$  while  $x_{\min}$  is assumed to be 0. The equivalent diameters are measured by means of image analysis.

Imaging Particle Analysis is a suitable approach for the measurement of collective samples of broken glass particles, as it enables each particle to be assessed individually. Lehmann used a HAVER CPA 4-2 for this task. This device uses dynamic imaging particle analysis to measure the size and various shape properties of bulk material in the size range from approx. 34  $\mu\text{m}$  to 90 mm. The CPA takes photos of the single particles that fall from a vibrating chute and uses these particle projections for evaluation. The results can then be further processed with the corresponding HAVER CpaServ V2.0.2 software. It is possible to assess statistical diameters of irregular particles, such as the Feret's, Martin's or the maximum linear diameters as well as diameters of reference circles with equal perimeters or projected areas. In this paper the diameters of circles with equal projected areas are used.

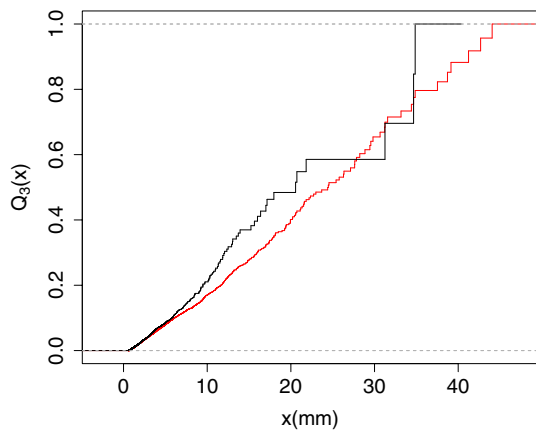
### 6.2 Classical statistical analysis

Figure 3 shows the empirical distribution function  $\hat{Q}_3(x)$  for one of the six samples. For small  $x$  it has a nearly continuous behavior, and it makes sense to write  $\hat{Q}_3(0) = 0$ . In contrast, for large  $x$  the graph shows steps, already for  $x$  larger than 10 mm. The corresponding fragments may be considered as 'large' particles. The distribution function curve is smoothed when calculated for six particles, as the red curve shows.

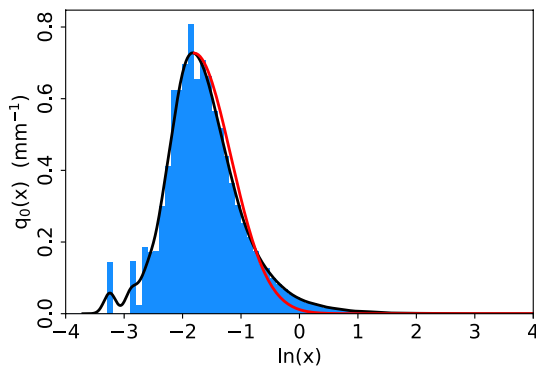
Now turning to number statistics, Fig. 4 shows in blue the number histogram of logarithmized sizes (equivalent diameters)  $\ln x_i$  for the same sample as in Fig. 3. The total number of fragments in the sample is 109.128. Note that the sizes were logarithmized first and then the obtained numbers are processed. The histogram is interpreted as an estimate of the probability density  $q_0(x)$  for  $x > 0.08$  mm.

The problems with ultra-fine fragments become visible. Obviously, for values smaller than  $\ln x = -2.5$ , i.e.  $x < 0.082$  mm, the histogram is questionable. The last column at  $\ln x = -3.2$  represents 1576 fragments, the next at  $\ln x = -2.88$  has 1608. It is unknown how many fragments are so small that they were not registered.





**Fig. 3** Empirical distribution function  $\hat{Q}_3(x)$  for equivalent fragment diameters of one glass ball with offered energy 75 J in black. The steps of the curve for  $x > 10$  mm correspond to large fragments. Note that the number of fragments larger than 10 mm is 40, while the total fragment number is 109.128. The empirical  $\hat{Q}_3(x)$  for six particles shown in red is already smooth (colour figure online)



**Fig. 4** Histogram of logarithmized equivalent fragment diameters of a glass ball with offered energy 75 J, kernel estimate of the probability density function  $q_0(x)$  (in black), and fitted normal probability density with variance 0.625 (in red) (colour figure online)

Nevertheless, on first view the distribution is close to logarithmic normal, some of the readers may consider this as a result expected.

The black curve in Fig. 4 is the corresponding empirical intensity function or probability density function obtained with a kernel estimator with  $\sigma = 0.1$ . Its shape is indeed similar to a probability density function of a normal distribution, i.e., in classical thinking the equivalent diameters may be considered approximately logarithmic normally distributed. Under this assumption the parameters of the normal distribution were estimated. Because of the missing data at the left tail, the standard methods of statistics cannot be applied. Therefore, the right side of the histogram was the starting point of parameter estimation.

It is plausible to choose for the mean  $\mu$  the value  $-1.8$ . Then, the standard deviation  $\sigma$  was chosen with  $\sigma = 0.625$ , which guaranties a value of the probability density function at  $\ln x = -1.8$  equal to the value of the black curve at  $\ln x = -1.8$ . As Fig. 4 shows this leads to a realistic fit for the distribution for  $\ln x > -1.8$ .

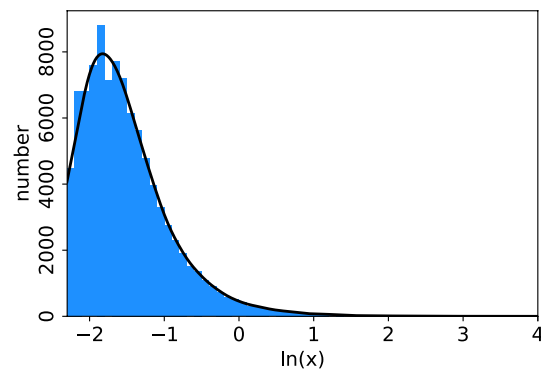
Finally two critical remarks against the normal distribution are given. (1) Even if the values smaller than  $\ln x = -2.5$  are ignored, some asymmetry of the histogram with respect to  $\ln x = -1.8$  is obvious: On the right-hand side the decrease is slower than left. (2) There are too many  $x_i$  in the right tail of the empirical distribution, which are not visible in Fig. 4. In order to demonstrate this discrepancy, the expected number of values  $\ln x_i$  larger than  $\mu + 3\sigma = 0.075$  or of  $x_i$  larger than 1.08 mm is approximately determined assuming normal distribution: The number of values  $\ln x_i$  larger than  $-1.8$  is 62.154, thus the total number of relevant values is assumed to be  $2 \times 62.154 = 124.308$ . The probability that a normally distributed random variable with parameters  $\mu$  and  $\sigma$  is larger than  $\mu + 3\sigma$  is 0.00135. Consequently, the expected number is  $124.308 \times 0.00135 = 167.8$ , while the empirical number is 718. This shows that the log-normal distribution fails to represent the large fragments.

### 6.3 Point process statistics

Figure 5 shows the empirical intensity function  $\hat{\lambda}_0(x)$ . The main difference to Fig. 4 is the scale at the ordinate, which shows now numbers and not  $\text{mm}^{-1}$ .

Now the tails of the intensity function are considered in detail.

The number distribution for the small fragments cannot be determined for very small fragments, because the data obtained are doubtful. Maybe the sizes larger than 0.1 mm, i.e.  $\ln x > 2.30$  merit analysis. Their total number is 96.641. A distributional description in the spirit of point process statistics is the black curve in Fig. 5 for  $\ln x > -2.30$ . This



**Fig. 5** Empirical intensity function  $\hat{\lambda}_0(x)$  of logarithmized equivalent fragment diameters of a glass ball with offered energy 75 J, in form of a histogram and as kernel estimate (black curve)

curve is shown in a way that the integral over  $\hat{\lambda}_0(x)$  from  $-2.30$  to  $4$  is  $96.641$ , as it is expected for an estimate of the intensity function  $\lambda_0(x)$ .

More interesting is the statistical analysis of large fragments. Figure 6 shows the empirical intensity function  $\hat{\lambda}_0(x)$  for the fragments with  $x \geq 9$  mm of the same glass ball as above, which results from the 1000 largest glass fragments.

When turning to large fragments, the long right-hand side tail of the distribution in Fig. 5 becomes visible, which theoretically can go until  $x_{\max} = 50$  mm. And just this tail is the point of interest, when large fragments are studied. Figure 6b shows the corresponding intensity function, a zoomed part of the curve in Fig. 6a, which is based on only the 40 fragments with equivalent diameter larger than 10 mm.

For all offered energies the value  $x = 10$  mm turned out to be a natural limit for defining ‘large fragments’, i.e., it is assumed  $x_u = 10$  mm.

Figure 7 shows the empirical intensity functions  $\hat{\lambda}_0(x)$  for the large fragments for all six completely destroyed glass balls for offered energy 75 J. There is a general similarity tendency but also some natural variability of the fragment distributions of large fragments. Additionally the corresponding averaged intensity function is shown in blue.

In order to demonstrate a possible influence of offered energy, Fig. 8 shows the empirical intensity functions for the fragments of three glass balls for offered energy 95 J. These few data do not support a hypothesis of some dependence of the distribution of large fragments on offered energy. For all cases of offered energy, [17] comes to the same result, which is in agreement with [18], who also did not observe an effect for squeezed glass balls when the strain rate was varied. Therefore, in the following only the 75 J samples are considered.

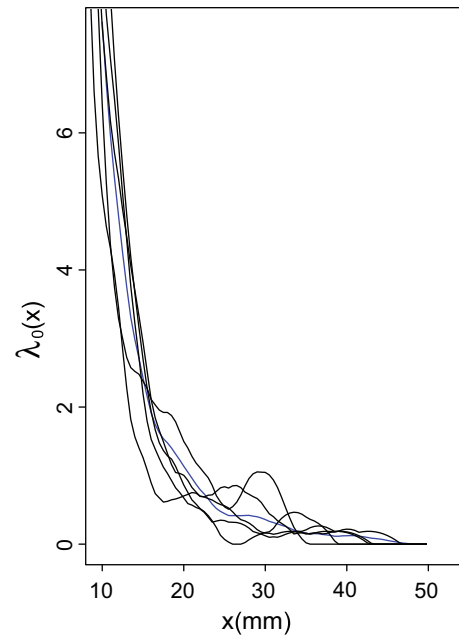
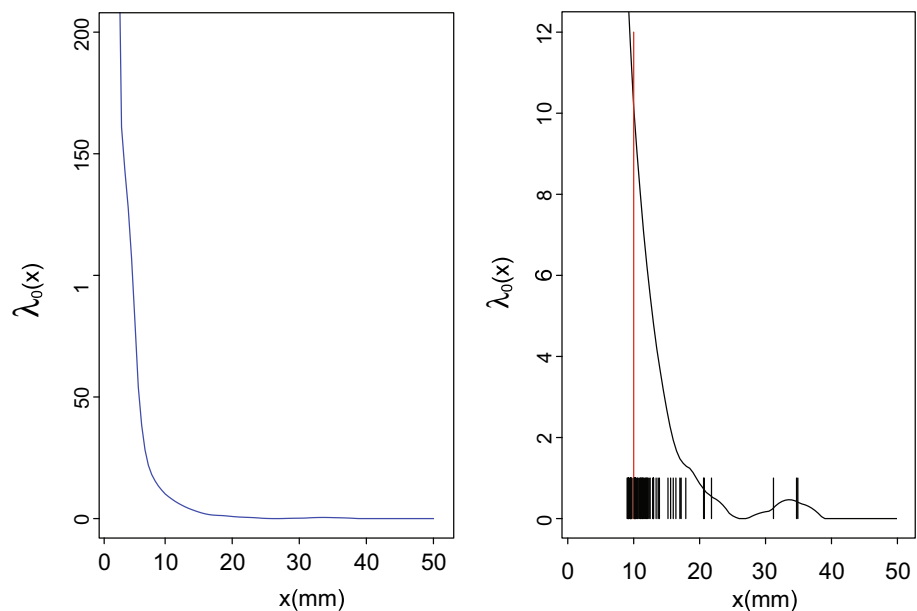


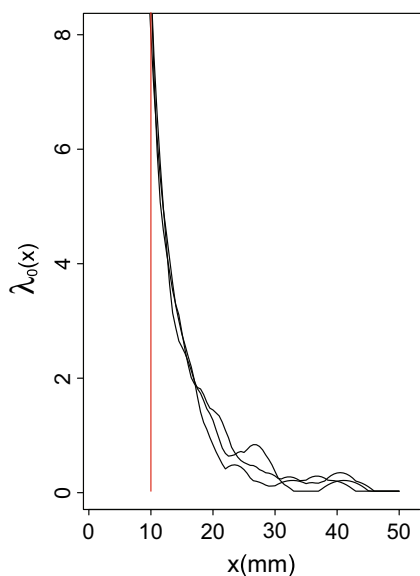
Fig. 7 Plot of the empirical intensity functions  $\hat{\lambda}_0(x)$  for the large fragments for all six glass balls with offered energy 75 J. The blue curve is the averaged intensity function (colour figure online)

Finally, the hypothesis is tested that the six black curves in Fig. 7 can be understood as they belong to independent samples of an inhomogeneous Poisson process with a fixed intensity function. Only fragments larger than  $x_u = 10$  mm are considered.

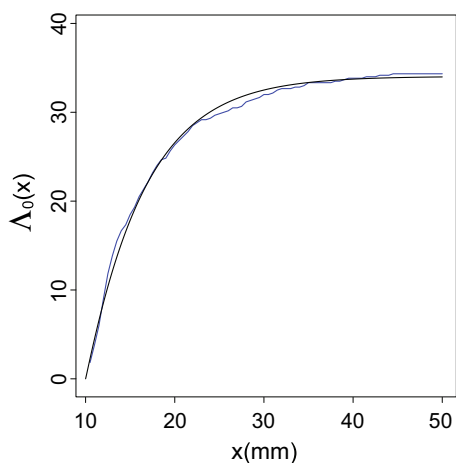
For this purpose the size transform is used, i.e. equation (21), which generates six samples of numbers between 0 and 1, which should be uniform in the interval [0, 1]. As cumulative

Fig. 6 a Plot of the empirical intensity function  $\hat{\lambda}_0(x)$  for all fragments larger than 1.5 mm for a single glass ball with offered energy 75 J. b Subplot for the fragments larger than approximately 9 mm. The small black segments stand for the fragments of sizes larger than 9 mm. Note the different scales at the ordinates





**Fig. 8** Plot of the empirical intensity functions  $\hat{\lambda}_0(x)$  for the large fragments for three glass balls for offered energy 95 J



**Fig. 9** Plot of the mean empirical cumulative intensity function  $\hat{\Lambda}_0(x)$  for the large fragments of all six glass balls with offered energy 75 J, in blue. Furthermore, the approximating function  $\tilde{\Lambda}_0(x)$  is shown in black (colour figure online)

intensity function the empirical mean cumulative intensity function  $\hat{\Lambda}_0(x)$  was used, approximated by a smooth function. For this purpose the shifted RRSB distribution was used as a formal ansatz, for a number distribution, i.e.

$$\Lambda_0(x) = c \left[ 1 - \exp \left( - \left( \frac{x - 10}{x'} \right)^r \right) \right] \text{ for } x \geq 10. \tag{28}$$

With the parameters  $r = 1.03$ ,  $x' = 6.67$  and  $c = 34.04$  estimated by a least squares method a good result was found. Figure 9 shows the empirical cumulative intensity function  $\hat{\Lambda}_0(x)$  and the graph of

$$\tilde{\Lambda}_0(x) = 34.04 \left[ 1 - \exp \left( - \left( \frac{x - 10}{6.67} \right)^{1.03} \right) \right] \text{ for } x \geq 10. \tag{29}$$

By the way, since the exponent  $r$  is so close to 1, one could also speak about a shifted exponential distribution. (One is close to a ‘heavy tail’, as said by statisticians if  $r$  were smaller than 1.)

The size transformation led to six sets of points between 0 and 1, which were tested for uniform distribution. For this the index-of-dispersion test was used, see [19], p. 87, which is suitable for the small numbers of points in the samples. In the test, the interval  $[0, 1]$  was divided into  $k = 10$  subintervals of length 0.1, the numbers of points within these intervals were counted and for each ball the dispersion index

$$c = \frac{(k - 1)s^2}{\bar{x}}, \tag{30}$$

was calculated, where  $\bar{x}$  is the mean of the ten numbers and  $s^2$  the sample variance. If  $c$  is between  $\chi^2_{k-1, \alpha/2}$  and  $\chi^2_{k-1, 1-\alpha/2}$ , the uniformity hypothesis is not rejected. For the error probability  $\alpha = 0.05$  the critical  $\chi^2$ -values are 2.9 and 19.0. For all balls values of  $c$  between 4.1 and 17.4 were observed.

Since for all six balls the Poisson hypothesis is not rejected, it is assumed that the inhomogeneous Poisson process is an acceptable model for the large fragments of the glass balls.

### 7 Goodness-of-fit testing

In many branches of engineering statistics, goodness-of-fit tests are frequently used. An example is the test of the Weibull distribution, which is, by the way, mathematically the same as the RRSB distribution.

In contrast, the authors do not know any paper on formal goodness-of-fit testing in particle statistics. Figures of a nature as Fig. 8 above are a popular equivalent. (The book [1] considers some tests in the particle context, but only parameter tests such as  $t$ - and  $F$ -tests and not goodness-of-fit tests.) Three facts may explain this: (a) particle statistics has no random samples in the classical sense, (b) often the data are of a minor quality, contain outliers and long tails, and (c) the application of tests for  $Q_3(x)$  is not obvious.

In the following an approximate goodness-of-fit  $\chi^2$ -test for  $Q_0(x)$  in the Poisson process case is proposed. This follows the example of the test described in [11] for the homogeneous Poisson process.

Assume that there is a sample of  $N$  particles. The test is made conditionally, for the number  $N$ . The null hypothesis is that the data follow some theoretical distribution function  $Q_0(x)$ . The form of the distribution is determined by  $p$  parameters, which are estimated from the sample data.

For the test, the interval  $[x_{\min}, x_{\max}]$  is divided in  $k$  classes or bins  $I_1, \dots, I_k, I_i = [a_{i-1}, a_i]$ . (These intervals may belong to classes defined by sieving.) For each class a theoretical frequency  $m_i$  is determined and compared with the empirical class frequency  $N_i$ . The  $m_i$  are calculated by

$$m_i = N \cdot p_i$$

with

$$p_i = Q_0(a_i) - Q_0(a_{i-1}).$$

Then

$$X_0^2 = \sum_{i=1}^k \frac{(N_i - m_i)^2}{m_i^2} \tag{31}$$

is the test statistic.

If  $X_0^2$  is larger than the  $\alpha$ -percentage point of the Chi square distribution with  $k - p - 1$  degrees of freedom, denoted by  $\chi_{\alpha, k-p-1}^2$ , then the null hypothesis is rejected with error probability  $\alpha$ .

In the case of an inhomogeneous Poisson process exactly the situation in the classical  $\chi^2$ -test is given:  $N_i$  follow a so-called multinomial distribution. The standard condition that the parameters are estimated with the maximum likelihood method is often not satisfied in particle statistics; remember the graphical methods in RRSB-statistics. However, with ‘good’ estimates the test may work nevertheless.

Probably, for  $Q_3(x)$  a similar test can be constructed. However, this problem has not been solved yet. An alternative could be a simulation test, but this should be the theme of another paper.

### 8 Three cases of statistical calculations

In order to compare the classical distribution-function approach (DF approach) with the point-process approach (PP approach), now three typical cases of statistical calculations are considered: pooling some samples to a larger sample, pooling two sub-samples to a full sample and statistics for very large particles.

#### 8.1 Case 1: pooling of samples

Assume that there are  $n$  statistically independent samples of particles and it is wanted to describe the union of all these as one unit by distribution and intensity function. (In point process terminology one speaks about ‘superposition’.)

The solution of this problem is simple, presented here for  $Q_3(x)$  and  $\lambda_3(x)$ . Let  $V_i$  the volume of sample  $i$ ,  $V = \sum_{i=1}^n V_i$ , and let  $Q_{3,i}(x)$  be the distribution functions and  $\lambda_{3,i}(x)$  be the intensity functions. Then

$$Q_3(x) = \sum_{i=1}^n V_i Q_{3,i}(x) / V \tag{32}$$

and

$$\lambda_3(x) = \sum_{i=1}^n \lambda_{3,i}(x). \tag{33}$$

Note that the superposition of independent Poisson processes is again a Poisson process with intensity function equal to the sum of the intensity functions of the components. Furthermore, in the theory of point processes it is shown that the superposition of many independent point processes is approximately a Poisson process [20], section 11.2.

Mixtures of standard distributions, i.e. multimodal size distributions, may be seen as results of superpositions of point process samples. However, point process statistics does not simplify the difficult problem of decomposition of distributions [21].

#### 8.2 Case 2: piecewise analysis

Sometimes a sample of particles is divided into two parts determined by a size limit  $x_{\text{div}}$ , into  $\{x \leq x_{\text{div}}\}$  and  $\{x > x_{\text{div}}\}$ , which are measured with different methods. At the end the sub-results have to be pooled to a final result for the whole sample. This case is studied in [1], page 273, in the distribution-function approach. It is re-considered here for  $Q_3(x)$  and  $\Lambda_3(x)$ . (It is analogous for  $Q_0(x)$  and  $\Lambda_0(x)$ .)

*DF approach* Analyzing only the particles with sizes larger than  $x_{\text{div}}$  the true distribution function  $Q_3(x)$  is measured for  $\{x > x_{\text{div}}\}$ . It is positive at  $x = x_{\text{div}}$  (equal to the volume fraction of all particles smaller than  $x_{\text{div}}$ ) and satisfies  $Q_3(x_{\text{max}}) = 1$ . For the particles smaller than  $x_{\text{div}}$  the distribution function  $Q_3^*(x)$  is estimated with a distribution-function estimator, with  $Q_3^*(x_{\text{min}}) = 0$  and  $Q_3^*(x_{\text{div}}) = 1$ . Then

$$Q_3(x) = Q_3(x_{\text{div}})Q_3^*(x) \text{ for } x_{\text{min}} \leq x \leq x_{\text{div}}. \tag{34}$$

*PP approach* The intensity function  $\lambda_3(x)$  can be directly estimated for all  $x$ . The cumulative intensity function  $\Lambda_3(x)$  can be directly estimated for the  $x$  smaller than  $x_{\text{div}}$ . For the larger  $x$  (ignoring the particles smaller than  $x_{\text{div}}$ ) a cumulative intensity function  $\Lambda_3^*(x)$  is estimated, which satisfies  $\Lambda_3^*(x_{\text{div}}) = 0$ . Then

$$\Lambda_3(x) = \Lambda_3(x_{\text{div}}) + \Lambda_3^*(x) \text{ for } x \geq x_{\text{div}}. \tag{35}$$

#### 8.3 Case 3: statistics for very large particles

An important problem of particle statistics is that of statistics for very large particles, statistics of the tail of extremely

big particles. In particular, the extent of appearance of particles larger than a limit size  $x_L$  is of considerable interest.

For this purpose the complements of distribution functions and intensity functions are used,

$$\bar{Q}_0(x) = 1 - Q_0(x), \bar{Q}_3(x) = 1 - Q_3(x)$$

and

$$\begin{aligned} \bar{\Lambda}_0(x) &= \Lambda_0(x_{\max}) - \Lambda_0(x), \bar{\Lambda}_3(x) \\ &= \Lambda_0(x_{\max}) - \Lambda_3(x) = \int_{x_L}^{x_{\max}} \lambda_3(t) dt. \end{aligned}$$

*DF approach* The quantity  $\bar{Q}_3(x_L)$  is equal to the mean volume fraction of all particles larger than  $x_L$ . This gives valuable information on the occurrence of large particles, which is, however, not sufficient. Also the number of large particles is of interest! Therefore, somewhat like  $Q_0(x)$  must be taken into account. However, this usually cannot be the cumulative distribution function for all particles, since often small particles cannot be measured precisely. Instead, a conditional distribution function should be used that corresponds to all particles larger than some limit  $x_u$ . This function could be denoted as  $Q_0(x|x_u)$ .

The corresponding quantity  $\bar{Q}_0(x_L|x_u)$  denotes the mean number proportion of particles larger than  $x_L$  from all particles larger than  $x_u$ . Given the total number  $N_u$  of particles larger than  $x_u$ ,  $N_u \cdot \bar{Q}_0(x_L|x_u)$  is the mean number of particles larger than  $x_L$ .

*PP approach* The quantity  $\bar{\Lambda}_3(x_L)$  is equal to the mean volume of all particles larger than  $x_L$  in the sample considered and analogously  $\bar{\Lambda}_0(x_L)$  denotes the mean number of particles larger than  $x_L$ . If the point process can be considered as a Poisson process, the number of these large particles follows a Poisson distribution with mean  $\bar{\Lambda}_0(x_L)$ .

We remark that the problem of statistics for very small particles in a wide size distribution is not addressed here. This issue could be subject to future research.

## 9 Discussion

The Introduction already mentioned that the use of point process ideas in particle size statistics, in particular the use of the intensity function, is not entirely new in the context of particle statistics. Indeed, it appears in hidden form and with another notation in the papers [3, 4], known for a physically-based derivation of the Weibull and RRSB distribution in the context of fragmentation.

Brown and Wohletz [4] consider fragment or particle masses  $m$  (instead of  $x$  in the present paper) and use a function  $n(m)$  which “is the number distribution in units of

particles per unit mass of mass  $m$  between  $m$  and  $m + dm$ ” and derive the equation

$$n(m) = \frac{N_T}{m_1} \left( \frac{m}{m_1} \right)^\gamma \exp \left[ -\frac{(m/m_1)^{\gamma+1}}{\gamma + 1} \right].$$

Then it is written as

$$N_T = \int_0^\infty n(m) dm,$$

“where  $N_T$  is the total number of fragments in the distribution.”

The latter seems to be not correct: Since  $n(m)$  is by definition a deterministic quantity, also  $N_T$  should be deterministic, while the total number of fragments is, of course, a random variable. Therefore,  $N_T$  should be correctly the *mean* total number of fragments, in the notation of the present paper

$$N_T = E(N([0, \infty))) = \Lambda_0(\infty).$$

And  $n(m)$  is not a probability density function, but an intensity function.

Nevertheless, the main result of the paper [4] is true.

In the introduction of his text [5] Bernhardt, the author of [1], writes on page 118: “Das Ergebnis einer Partikel... größenmessung liegt primär immer als ein System von diskreten Werten vor. Gleichwohl hat es sich eingebürgert, ihre Darstellung und Auswertung analog zur mathematischen Statistik als Wahrscheinlichkeitsverteilung einer stetigen Zufallsgröße  $X$  zu behandeln.” In English: “The result of a particle size analysis is always given as a system of discrete values. Nevertheless it is standard to treat its presentation and analysis analogously to mathematical statistics as for a probability distribution of a continuous random variable  $X$ .”

These words can be interpreted as a statement that Bernhardt had the same basic idea as that of this paper, but he did not apply it. Indeed, the ‘discrete values’ are the ‘points’ of the present paper.

The assumption of an inhomogeneous Poisson process is very important. If it is really true, the whole statistics is greatly simplified. This means first that there is no inner structure or order in the sequence of size-ordered fragments; nobody has mentioned until now such phenomenon. Second, for the number statistics of particle sizes the whole point-process statistics is the classical distribution-function-based statistics with  $Q_0(x)$  plus some statistics for the number of large particles. The latter is quite simple, it is only statistics for a Poisson variate, where the mean completely determines the distribution.

In order to prove the Poisson assumption, the authors recommend systematic statistical analyses of particles of

different origin with testing the Poisson process hypothesis. The superposition fact in Sect. 8 strengthens the belief that the Poisson process is indeed a universal model for particles collectives.

Since number statistics makes sense only for large particles, for particle sizes larger than a limit  $x_u$ , the forms of  $q_0(x)$  have to be adapted to this case. This means in the case of the RRSB distribution the shifted form of the distribution as in (28) should be used.

Our data are in the order of typical lab sieve analysis—and already there it makes sense to study sub-samples of very large particles. However, the main problem we have in mind are statistics for rocks in quarries after blasting. We are still looking for suitable datasets and hope to get such, which may result from image analysis.

## 10 Conclusions

The present paper has shown that particle size statistics can be refined by application of ideas of point process statistics. Nevertheless, if only one sample is considered ( $n = 1$ ) and only  $Q_3(x)$  and  $q_3(x)$  are wanted to be estimated, the classical distribution-function approach is often sufficient. Problems appear when the large jumps in the empirical distribution function  $\hat{Q}_3(x)$  for large  $x$  cannot be ignored, see Fig. 3. Then, for precise statistics more than one sample should be analysed and important additional variability may become visible if not fixed as in the experiment with the glass balls): the total volume  $V(I)$  turns out to be variable and has to be analysed statistically, too.

The cumulative intensity function  $\Lambda_3(x)$  has then the form

$$\Lambda_3(x) = V_T \cdot Q_3(x) \text{ for } x_{\min} \leq x \leq x_{\max}, \quad (36)$$

where  $Q_3(x)$  can be determined as usually and the determination of the mean volume  $V_T$  requires extra statistics. The result of the statistical analysis may be given as  $Q_3(x)$  plus  $V_T$ .

It is similar with the number of particles. The cumulative intensity function  $\Lambda_0(x)$  has the form

$$\Lambda_0(x) = N_T \cdot Q_0(x) \text{ for } x_{\min} \leq x \leq x_{\max}. \quad (37)$$

However, for small  $x$  the determination of  $Q_0(x)$  can be questionable or impossible. The number intensity  $\lambda_0(x)$  can be used to determine the mean number of particles larger than a limit  $x_L$  as

$$E(N([x_L, x_{\max}])) = \int_{x_L}^{x_{\max}} \lambda_0(x) dx \quad (38)$$

and the latter integral is  $\bar{\Lambda}_0(x_L)$ .

In the Poisson process case, the number of particles larger than  $x_L$  has a Poisson distribution with mean  $\bar{\Lambda}_0(x_L)$ .

**Acknowledgements** The authors thank Dr. Felix Ballani, Dr. Claus Bernhardt, Dr. Michael Klichowicz, and Dr. Bernd Kubier for constructive discussions about particle statistics. They are grateful to the three referees for constructive and helpful remarks.

**Funding** Open Access funding enabled and organized by Projekt DEAL.

## Declarations

**Conflict of interest** The authors declare that they have no conflict of interests.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

- Bernhardt, C.: Particle Size Analysis. Classification and Sedimentation Methods. Particle Technology Series. Chapman & Hall, London (1994)
- Unland, G.: The principles of single-particle crushing. In: Salman, A.D., Ghadiri, M., Hounslow, M. (eds.) Particle Breakage, vol. 12, pp. 117–225. Elsevier, Amsterdam (2007)
- Brown, W.K.: The theory of sequential fragmentation and its astronomical application. *J. Astrophys. Astron.* **10**, 89–112 (1989)
- Brown, W.K., Wohletz, K.H.: Derivation of the Weibull distribution based on physical principles and its connection to the Rosin–Rammler and lognormal distributions. *J. Appl. Phys.* **78**, 2758–2763 (1995). <https://doi.org/10.1063/1.360073>
- Bernhardt, C.: Characterization of the size of particles and pieces (in German). In: Rasemann, W. (ed.) Handbook Sampling and Quality Assurance in the Investigation and Evaluation of Materials Systems, Vol. 1 (Handbuch Probenahme und Qualitätssicherung bei der Untersuchung und Bewertung von Stoffsystemen), pp. 116–165. Institut für Qualitätssicherung Freiberg, Freiberg, Germany (2005)
- Gy, P.M.: Sampling of Particulate Materials. Theory and Practice, 2nd edn. Developments in Geomathematics 4. Elsevier, Amsterdam (1982)
- Klichowicz, M., Lieberwirth, H.: In: Daya Sagar, B.S., Cheng, Q., McKinley, J., Agterberg, F. (eds.) Grain Size Analysis, pp. 1–6. Springer, Cham (2020). [https://doi.org/10.1007/978-3-030-26050-7\\_147-1](https://doi.org/10.1007/978-3-030-26050-7_147-1)
- Montgomery, D.C., Runger, G.C.: Applied Statistics and Probability for Engineers. Wiley, New York (1994)
- Snyder, D.L., Miller, M.I.: Random Point Processes in Time and Space. Springer, New York (1991)
- Reiss, R.-D.: A Course on Point Processes. Springer Series in Statistics. Springer, New York (1993)

11. Karr, A.F.: Point Processes and Their Statistical Inference, 2nd edn. Routledge, Boca Raton (2017). <https://doi.org/10.1201/9780203743027>
12. Rumpf, H., Ebert, K.F.: Darstellung von Kornverteilungen und Berechnung der spezifischen Oberfläche. *Chemie Ingenieur Technik* **36**, 523–537 (1964)
13. Chiu, S.N., Stoyan, D., Kendall, W.S., Mecke, J.: Stochastic Geometry and Its Applications, 3rd edn. Wiley, Chichester (2013). <https://doi.org/10.1002/9781118658222>
14. Dang, H.K., Meguid, M.A.: Algorithm to generate a discrete element specimen with predefined properties. *Int. J. Geomech.* **10**, 85–91 (2010). [https://doi.org/10.1061/\(ASCE\)GM.1943-5622.0000028](https://doi.org/10.1061/(ASCE)GM.1943-5622.0000028)
15. Alderliesten, M.: Mean particle diameters. Part VII: the Rosin–Rammler size distribution: physical and mathematical properties and relationships to moment-ratio defined mean particle diameters. *Part. Part. Syst. Charact.* **30**, 244–257 (2013)
16. Stieß, M.: Mechanical Process Engineering and Particle Technology (in German). Springer, Berlin (2009). <https://doi.org/10.1007/9783-540-32552-9>
17. Lehmann, J.: Investigations of Impact on Balls (in German). Diplomarbeit, TU Bergakademie Freiberg, Freiberg, Germany (2020)
18. Bergstrom, B.H.: Energy and size distribution aspects of single particle crushing. In: Fairhurst, C. (ed.) Proceedings of 5th U.S. Symposium on Rock Mechanics, pp. 155–172. Pergamon, New York (1962)
19. Illian, J., Penttinen, A., Stoyan, H., Stoyan, D.: Statistical Analysis and Modelling of Spatial Point Patterns. Wiley, Chichester (2008). <https://doi.org/10.1002/9780470725160>
20. Daley, D.J., Vere-Jones, D.: An Introduction to the Theory of Point Processes, 2nd edn. Springer, New York (2008). <https://doi.org/10.1007/978-0-387-49835-5>
21. Alderliesten, M.: Mean particle diameters. Part VIII: computer program to decompose mixtures of (truncated) lognormal particle size distributions using differential evolution to generate starting values for nonlinear least squares. Part. Part. Syst. Charact. **33**, 675–697 (2016)

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.