Power Law Distribution: Method of Multi-scale Inferential Statistics

Vincenzo Guerriero

Dipartimento di Scienze della Terra, Università degli studi di Napoli ‘Federico II’,80138 Naples, Italy

vincenzo.guerriero@unina.it

***Abstract*-Power law distribution appears in several scientific fields such as physics, earth science, economics, social science and many others. This paper illustrates new practical criteria for inferential statistics involving power law and, namely for estimation of power law distribution exponent and its confidence interval. To calculate this latter interval new expressions, in closed form, are derived. This methodology has been compared with classical least squares method and that of maximum likelihood, showing that it provides a more efficient estimator for power law exponent. In order to describe this criterion by means of a case study, it has been applied to statistical multi-scale analysis of fracture networks in a petroleum reservoir analogue, nevertheless it can be applied in several contexts involving scale free distributions.**

Keywords-Power Law; Poisson Distribution; Parameter Estimation; Confidence Interval; Scaling Analysis

I. INTRODUCTION

In several natural and man made phenomena some quantities show a statistical distribution following (or very well approximated by) a power law, such as e.g. earthquake magnitude [1] or diameter of lunar craters [2]. Namely, considering the example of lunar craters, for a given real number x, the probability $P(x)$ that a lunar crater have diameter greater than x – described by the complementary distribution function [3] – is given by:

$$P(x) = c\, x^{-m}, \tag{1}$$

over a wide range of x values, where c and m are experimental positive constants [2]. In this paper we will refer to complementary distribution function (i.e. *1 – distribution function*) rather than distribution function as use of the former is more appropriated when dealing with power law distributions. Indeed, denoting by $P'(x)$ the distribution function, this latter assumes the form: $P'(x) = 1- P(x) = 1- c\, x^{-m}$, which is not a power law. A review of real cases is provided by Newman [4], in which some quantities are distributed according to a power law, in the field of economics [5], earth sciences [1], physics [6], computer science [7], social science [8] and others more. Furthermore the mentioned paper describes some mechanisms, generating power law distributions, which, in several cases, can explain its occurrence. Let consider again the example of lunar craters [2, 4]. The diameter of such craters shows a probability distribution following a power law. Assume for the moment such craters displaying a random spatial distribution over the lunar surface. Hence we have a set of objects (craters) which are randomly distributed over a surface and exhibit a property (diameter) showing a power law probability distribution.

In this paper we propose a procedure of statistical analysis aimed to study those quantities showing a power law probability distribution and which are associated to objects or events being randomly distributed over the space (within a volume, a surface or a straight line) or along the time. Another example is provided by tensional fractures in rocks (joints). Usually these occur as sets of (roughly) parallel fractures, whose opening displacement (hereafter aperture) can be distributed according to a power law [9-13]. Furthermore, for a given line perpendicular to fracture planes, in several cases, they are randomly distributed over the mentioned line [13]. A further example may be furnished by an event, e.g. a signal, occurring at random and equiprobable instants, whose intensity exhibits a power law probability distribution. For all of these cases, as illustrated in the following sections, statistical sampling can be carried out over more scales of observation in order to improve the estimation of power law distribution parameters. Multi-scale analysis of fracture networks has been carried out by other authors in order to study power law distribution of some fracture features such as aperture, fracture length and spacing, as well as the fractal dimension [11, 12, 14-17]. In these works parameter estimation is achieved by means of least squares methods. Nevertheless, as illustrated in section III, this criterion shows some drawbacks. Main aim of this paper consists in proposing statistical methods of practical use which allow to overcome these problems and significantly improves the precision of power law exponent estimation. In order to illustrate the proposed methodology, a case study from structural geology is described in detail. It deals with statistical structural analysis of fracture networks cropping out within a geologic analogue of a petroleum reservoir. Fracture analysis is a key point of structural characterization of reservoir rock, as fractures strongly affect hydraulic behaviour of reservoir itself, providing the main pathways for ground fluid flow. Analyzed rocks belong to a carbonate cretaceous succession cropping out in Sorrento Peninsula (Naples area, Italy), which has been selected as a surface geological analogue of the buried productive reservoir rocks of Val D’Agri (Basilicata, Italy; [13]).

II. STATISTICAL MULTI-SCALE SAMPLING

A. *Preliminary Concepts About Cumulative Distribution Functions Following a Power Law*

About the complementary distribution function (also known as complementary cumulative distribution function) described by (1), some observation can be drawn. First, the associated probability density $p(x)$ also follows a power law. Indeed, as $P(x)$ denotes the probability $P(X > x) = 1- P(X \le x)$, and taking into account that $p(x)$ is defined as the derivate of $P(X \le x)$, it follows that:

$$p(x) = \frac{d}{dx}\left(1 - P(X > x)\right) = c \cdot m \cdot x^{-(m+1)}, \tag{2}$$

therefore $p(x)$ follows a power law with exponent greater than unity (note that m > 0 as, for definition of $P(x)$, it need to be obviously a decreasing function of x). Second, should be taken into account that, if $P(x)$ follows a power law over the whole positive real axis, it should to assume indefinitely large values when x tends toward zero, whilst cumulative distribution clearly varies within the range 0 - 1 (being a probability). Therefore, usually it is assumed that power law is valid up to a minimum value x_{min} for which it results: $P(x) = 1$ [4]. The condition $P(x_{min}) = c\, x_{min}^{-m} = 1$, implies that:

$$c = 1/\, x_{min}^{-m}, \qquad (3)$$

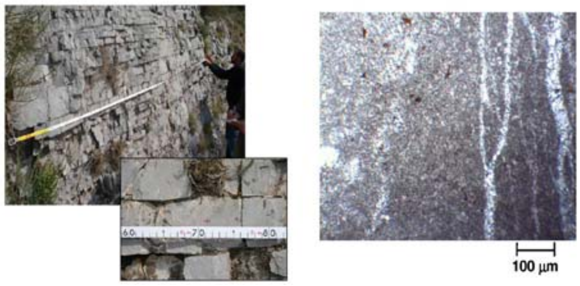
Finally we just remark that the histogram of a power law yields a straight line on a log-log diagram. Indeed, denoting by y' and x' the logarithm of variables in (1), $P(x)$ and x respectively, we obtain: $\ln P(x) = \ln c\, x^{-m} = \ln c + \ln x^{-m} = \ln c - m \ln x$; therefore it results: $y' = \ln c - m\, x'$; i.e. y' is a linear function of x'.

B. Cumulative Frequency and Multi-Scale Sampling

For a given statistical sample, the empirical cumulative distribution is defined as ratio between the number of elements for which $X \le x$ and the number N of elements constituting the whole sample [3]. As an alternative to distribution function, to describe statistical distribution of a random variable (RV) one can use the cumulative frequency $F(x)$ of a property X, defined as the number of elements per length unit (or surface area or time unit), for which it results $X > x$ [12]. As an example, the cumulative frequency of fracture aperture X is defined as the number of fractures per meter having aperture greater than x. Considering the case of lunar craters, the cumulative frequency can be defined as the number of craters per square kilometer with diameter larger than x and, analogously, for the example of signal (viewed as a time series), $F(x)$ denotes the number of events per time unit whose intensity is greater than x. Note that a RV showing cumulative frequency following a power law with exponent m, exhibits also a distribution function according to a power law with the same exponent value.

Use of cumulative frequency has some advantages if compared with distribution function. It is not constrained to assume values within the range 0 – 1, and therefore does not require to individuate the lower limit of power law validity nor quantify the x_{min} value. Furthermore, use of cumulative frequency allows to put on the same diagram data gathered from sample lines (or surface or time interval) of different lengths at different scales. Let’s consider the following example involving fracture aperture analysis. A common method of fracture statistical sampling consists in detecting joint features along a sample line (scan line) possibly orthogonal to the mean fracture plane and then to record, for each joint, its position - i.e. the abscissa along sample line - and aperture (usually many other characters are recorded, nevertheless here we are interested only in aperture statistics). Once data set has been detected, cumulative frequency can be promptly calculated according to the method, analogue of the rank/frequencies criterion, illustrated in Ortega et al. [12]. In order to achieve a thorough diagram of cumulative frequencies, which reports aperture data over several orders of magnitude, sampling of several hundreds of data may be needed [15]. Detecting a so large sample would require carrying out very long scan lines, that is time consuming and expensive and, in several cases, may be unfeasible a cause of lack of adequate outcrops. One alternative way of achieving aperture data distributed over several orders of magnitude is to carry out statistical sampling at different scales and resolutions [11-13, 18]. A multi-scale sampling can be performed by integrating data from a field scan line on outcrop and a micro scan line carried out over a small rock sample in thin section (usually about 2-3 cm sized), opportunely oriented, by means of microscope (Fig. 1 (a) and (b)).

Cumulative frequencies from these two data sets can be displayed on the same diagram (Fig. 2), providing data over several orders of magnitude, based on data sets including few tens of fractures. Should be noted that use of cumulative distribution rather than cumulative frequency does not allow putting on the same diagram data sets gathered over different scales.



(a) (b)

Fig. 1 Examples of fracture sampling techniques at different scales. (a) Traditional field scan line carried out on outcrop. (b) Micro scan line on thin section of rock (magnification 200x). A 2 cm sized scan line requires using ca. 50 images of this size.

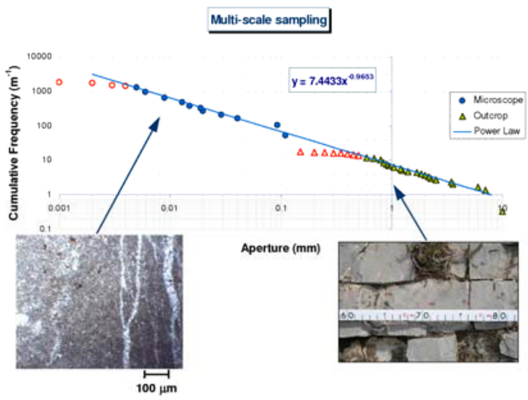


Fig. 2 Cumulative frequencies of fracture apertures, obtained by the integration of outcrop and micro-scale scan line data on: mudstones. The distribution follows a power law over several orders of magnitude. Red dots denote those data which have been discarded because affected by truncation error (see text).

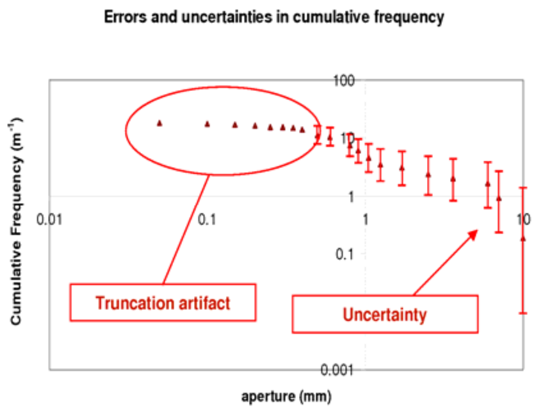


Fig. 3 Main inaccuracies affecting empirical cumulative frequencies at several scales. Below a resolution threshold, a systematic undersampling occurs (truncation artifact) as fractures are scarcely visible. On the right-hand side uncertainty increases considerably as frequency estimations are based on samples including few data.

III. UNCERTAINTIES ASSOCIATED TO EMPIRICAL DISTRIBUTIONS AND PROBLEMS RELATED TO USE OF LEAST SQUARE METHOD

The empirical cumulative frequency of joint aperture is generally affected by systematic errors occurring at both extremities of the scale of observation. A review of the different types of artifacts in determining such distributions has been provided by Bonnet et al. [15] and Ortega et al. [12]. One of these is the truncation artifact, which is due to an underestimation of fracture number near the lower limit of the scale of observation. Such underestimation occurs because smaller fractures often are not detected since they are scarcely visible. This artifact produces a flattening of empirical distribution near the lower limit of the scale of observation (Fig. 3). This is highlighted in Fig. 2, where the flattening of cumulative frequency, of data associated to field scan line, leads to hypothesize that aperture values below 0.5 mm are practically absent, whilst data gathered by use of microscope, on the same fracture set, point out that on the contrary small fractures are very abundant. This systematic error is not quantifiable a priori and, usually, is corrected by discarding aperture values below a minimum threshold, which often is fixed based on experience. However, a non-systematic, often more dangerous error is that associated with the uncertainty of the obtained sampling estimates, depending on the kind of RVs involved and statistical samples size. Should be noted that each value of cumulative frequency on diagram of Fig. 3 is a sampling estimate of fracture density (defined as number of fractures per meter), which is affected by an intrinsic uncertainty. Under the hypothesis of random spatial distribution of fractures, fracture frequency is a Poisson RV [3] and, therefore, the sampling uncertainty can be quantified by means of the confidence interval of this RV. Note that intervals on the right-hand side of the data distribution are much larger than those on the left-hand side (on a log-log diagram; Fig. 3). This, because on the right-hand side fracture density assumes lower values with respect of those on the left-hand side, where, for the law of large numbers, the sampling mean converges towards the population mean (i.e. the true mean value).

Usually, the power law approximating cumulative distribution of fracture apertures is generally determined by calculating the least squares line interpolating the scan-line data (an alternative criterion is the maximum likelihood method which is addressed in section V). As each confidence interval is characterized by amplitude being a function of the same estimated value (i.e. fracture density) it follows that, as shown in Fig. 3, the standard deviation varies along the abscissa (heteroskedasticity, [3]). Therefore, the application of the least squares method is inappropriate without the application of appropriate weight function for residuals. The main problem of this method consists of a large aleatoric error in estimating the slope of the least squares line (i.e. the exponent of the power law).

A significant reduction of uncertainty on the power-law exponent can be obtained by means of multi-scale analysis, i.e. by the integration of data from micro-scan lines with the traditional analysis performed by means of scan lines at outcrop (Fig. 2). Multi-scale analysis has been applied successfully by many authors for scan-lines and scan-areas [11, 12, 15, and references therein]. However, the problem of the applicability of the least squares method holds also in this case. Bonnet et al. [15] point out that use of cumulative distribution has some drawbacks, suggesting to utilize frequency density distribution (i.e. the derivate of cumulative distribution). Nevertheless use of this latter distribution requires considerably numerous samples which, in many cases, could be not available. Another question related to cumulative distribution consists in establishing how many fractures are needed in order to achieve a meaningful estimate of the power law exponent [15].

IV. THE PROPOSED METHOD

The statistical analysis of joint aperture has been performed by using data sets gathered from field scan line (Fig. 1 (a)) and micro scan lines, carried out by optical microscope (Fig. 1 (b)), both roughly perpendicular to studied fracture sets. Analyzed rocks are fine grained carbonate rocks, such as mudstone and fine dolomite. Namely, three mechanical strata of mudstone (grain size below 5 μm) have been analyzed along scan lines with length of 3-4 m and micro scan lines 2-3 cm sized, carried out on opportunely oriented thin sections by microscope (enlargement 200x); two beds of medium dolomite (grain size in the range 30-80 μm) have been examined either by field scan lines and micro scan lines, performed by microscope (enlargement 50x) on polished slabs of rock, with length of 8-12 cm. The proposed statistical method is articulated according to the following steps:

- Verifying of random spatial distribution of fractures. This step uses some criteria traditionally employed in statistical analysis, including evaluating the variation coefficient of spacing between consecutive fractures (here called spacing), use of probability plots, correlation of contiguous spacing data. Furthermore bootstrap techniques have been developed in order to assess that mean spacing is distributed according to a Chi-squared RV.

- Calculating confidence intervals of fracture frequency estimations associated to opportunely selected aperture values. Here a new formulation of practical use has been derived, in order to evaluate confidence interval for the Poisson distribution.

- Estimation of parameters defining the power law describing aperture distribution. Here a two point fitting method has been suggested rather than least squares method, in order to improve parameter estimation.

- Evaluating the confidence interval for power law exponent. Here a new expression, in closed form, has been derived for upper and lower limits of confidence interval, based on the two point fitting method.

These steps are described in detail in the following sections.

A. *Verifying Uniform Random Spatial Distribution of Fractures*

A preliminary investigation uses the variation coefficient of fracture spacing c_v (defined as ratio between standard deviation and mean; Fig. 4 (a)). Convergence of c_v towards the unity (i.e. of standard deviation and mean towards the same value) is typically associated with a uniform random spatial distribution of fractures [19]. Based on probabilistic theory [e.g. 3, 20], a uniform random spatial distribution of fractures is expected to produce a fracture spacing distribution following an exponential law. Moreover, under this hypothesis, fracture density follows a Poisson distribution. A main property of exponential distribution consists in showing a unitary coefficient of variation [20]. Nevertheless, this latter is a necessary but not sufficient condition for inferring a uniform fracture spatial distribution, since several distributions can exhibit a unitary coefficient of variation.

In order to ascertain that joints are randomly distributed, in the present study the analysis is carried out following a series of steps. First, the bootstrap method [3, 21, 22] is applied to spacing data in order to determine whether mean spacing distribution conforms to a Chi-squared RV. For each scan-line data set, the mean of ten randomly chosen spacing values has been calculated. This process has been repeated 10000 times, obtaining 10000 mean values. Subsequently, the cumulative distributions of these values have been calculated. Let us now consider the following function: $y = 2 \cdot n / \mu \cdot x$, where x is the mean of n determinations of an exponential RV, and μ is the mean of the whole population of such a RV (i.e. the 'true' mean). Since y shows a Chi-squared distribution [20], the inverse function of this distribution has been applied to the observed cumulative distributions. It has then been verified whether the calculated values fall along a line of equation: $y = 2 \cdot N_f / S \cdot x$, where N_f is the number of spacing values and S is the mean spacing (note that S is calculated from the whole data set, whilst x is calculated on ten values). This has been carried out using a diagram where the mean (x) is plotted along the abscissa axis and the inverse function along the ordinate axis (Fig. 4 (b)).

Second, probability plots are employed to verify that spacing values display an exponential distribution. These diagrams are usually used in statistics to verify whether an analyzed RV is characterized by a certain probability distribution and, in this latter instance, to obtain the equation of the theoretical distribution best fitting the data. If the analyzed RV is exponential with a mean value μ, then the complementary cumulative distribution approaches to an exponential curve of equation $y = e^{-x/\mu}$, (i.e. a straight line in a diagram with logarithmic ordinates; Fig. 4 (c)).

Finally, the stochastic independence of consecutive spacing values (required for a random spatial distribution of fractures) is verified by means of autocorrelation analysis. A useful criterion used in geostatistics for evaluating the dependence between consecutive values attained by an RV consists in the analysis of the variance, covariance or correlation of elements put at increasing distance [23]. In this paper, the correlation between contiguous values of spacing has been analyzed. The first correlation value is calculated between the n-th and n + 1 spacing values, for each n. Successive values of correlation are calculated between n-th and n + 2, … n + k, … values and so on. Then the calculated correlation values are plotted versus the variable k (correlogram). If neighboring spacing values are dependent correlogram shows a correlation value closely approaching unity for $k = 1$ (and for small k values), gradually decreasing as k increases. Otherwise the correlation attains very low values, with respect to unity, for any value of k and the related correlogram shows a very irregular trend (for details see [18]). Fig. 4 shows the results of the illustrated analysis carried out on fracture data gathered by means of field scan line and micro scan line on mudstone. For the illustrated data, and for all those detected on the study area, we have achieved the following results: (i) standard deviation and mean of spacing values converge toward the same value; (ii) bootstrap method points out that sampling mean of spacing values follows a Chi-squared distribution; (iii) for most of the data set analyzed in this study, probability plots show that an exponential distribution provides good fit of observed data points, at both outcrop- and micro-scale; (iv) autocorrelation diagrams show that the correlation attains low values (even for small values of k) and a rather irregular trend, pointing out a lack of statistical correlation between each spacing value and the contiguous ones (Fig. 4 (d)).

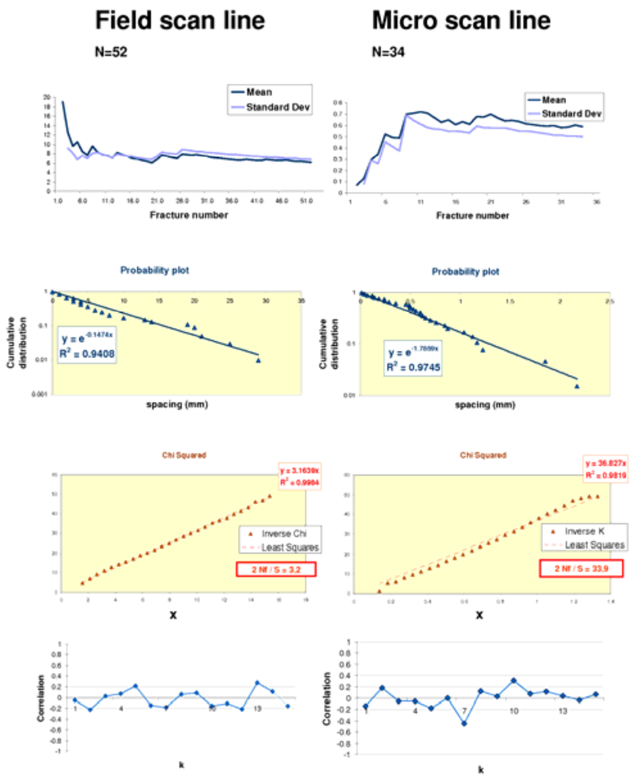


Fig. 4 Integrated analysis of spatial fracture distribution in mudstone, at both field and micro scales, including: diagram showing mean and standard deviation of spacing values as a function of the sample size, probability plot of spacing, probability plot of bootstrap mean spacing (note that the least squares line slope is consistent with the theoretical slope given by: $2 \cdot N_f / S$; see text) and spacing autocorrelation diagram. These results are consistent with a random spatial distribution of fractures.

B. *Practical Formulation for the Confidence Interval of the Mean of Poisson Random Variable*

A simple and rapid method to calculate the confidence interval for mean fracture density, for any aperture value, utilizes the determination of the confidence interval of mean spacing. Although we know the probabilistic distribution of the mean fracture density (Poisson distribution), we suggest an easier way to calculate its confidence interval, based on the application of the central limit theorem [e.g. 3]. This allows us considering the estimated mean spacing S as a RV with: (i) a normal distribution, (ii) a mean value equal to the mean spacing of the whole population μ, and (iii) a standard deviation of $\mu/n^{1/2}$, where n is the number of sampled spacing values. Therefore, we impose that the probability – that is the standardized variable $u = (S-\mu) / (\mu/\sqrt{n})$ being comprised within the interval $[-u_{\alpha/2} ; u_{\alpha/2}]$ – is equal to 0.95:

$$Pr\ \{-u_{\alpha/2} < (S-\mu) / (\mu/\sqrt{n}) < u_{\alpha/2}\} = 0.95, \tag{4}$$

Where: $u_{\alpha/2} = 1.96$ is the value of the normal standard RV with probability = 0.025. The lower limit μ_{low} of the interval is obtained as:

$$u_{\alpha/2} = (S-\mu_{low}) / (\mu_{low} /\sqrt{n}), \tag{5}$$

Therefore:

$$\mu_{low} = S / (1 + u_{\alpha/2} / \sqrt{n}). \tag{6}$$

Similarly, the upper limit μ_{upp} of the interval is obtained as:

$$\mu_{upp} = S / (1 - u_{\alpha/2} / \sqrt{n}). \tag{7}$$

Taking into account that the mean fracture density is given by: $F = N / L \approx 1 / S$ [12], where $N = n + 1$ is the number of fractures detected along the scan line, therefore we put: $F_{low} \approx 1 / \mu_{upp}$; $F_{upp} \approx 1 / \mu_{low}$. Substituting the values in the equations above, we obtain two simple equations for the lower and upper limits of the confidence intervals for mean fracture density, respectively:

$$F_{low} = \left(1 - \frac{1.96}{\sqrt{N-1}}\right) \cdot \frac{N}{L} \tag{8}$$

$$F_{upp} = \left(1 + \frac{1.96}{\sqrt{N-1}}\right) \cdot \frac{N}{L} \tag{9}$$

Equations (8) and (9) permit a rapid calculation of the 95% confidence interval for each value of fracture density. Based on our simulations, a sample including 20 measurements is adequate to furnish significant results. For a first approximation estimate, a value of $N \geq 10$ is sufficient. In case the number of detected fractures is very low, it is necessary to use an exact method for the evaluation of the confidence interval. Terming F the estimated mean fracture density and f the fracture density described by the Poisson RV (dependent on μ and L), the method consists on imposing $\Pr\{F \leq f\} = 0.975$ and solving it numerically in μ, obtaining the lower limit of the confidence interval. Similarly, solving the equation $\Pr\{F \leq f\} = 0.025$, the upper limit of the confidence interval is obtained.

C. Two Point Fitting Method

The problems related to application of the least squares method for estimating the power law best fitting empirical cumulative distributions (illustrated in section III) can be overcome by an analysis based on the definition of a straight line passing for two points only (on a log-log diagram), i.e. one for each – outcrop and micro-scale – data set (Fig. 5). These points are chosen in such a way as to obtain the maximum number of sampled fractures for each data set, these being affected by the minimum truncation error [12, 15]. Based on our experience and practice we suggest a fracture aperture threshold of at least 0.5 mm for outcrop-scale fracture analysis, whereas for thin section images under the optical microscope we propose a threshold of 0.005 mm.

Denoting by x_1 the aperture value chosen for field data (e.g. 0.5 mm) and by x_2 that for micro scale data (e.g. 0.005 mm), by F_1 and F_2 the associated fracture frequency estimations, the exponent m can be calculated as ratio $\Delta \ln F / \Delta \ln x$, by changing its sign, as m is positive (see section I). Therefore it results:

$$m = -\frac{\ln\left(\frac{F_1}{F_2}\right)}{\ln\left(\frac{x_1}{x_2}\right)} \tag{10}$$

The coefficient can be calculated by using the expression for a straight line for two points (on a log-log diagram):

$$\ln c = - m \ln x_2 + \ln F_2 . \tag{11}$$

In order to compare this estimation criterion with least squares method and that of maximum likelihood, Monte Carlo experiments have been performed, simulating field and micro scale scan line data. The results are illustrated in section V and point out that the two point fitting method improves significantly the estimation of power law exponent, if compared with least squares and maximum likelihood methods.

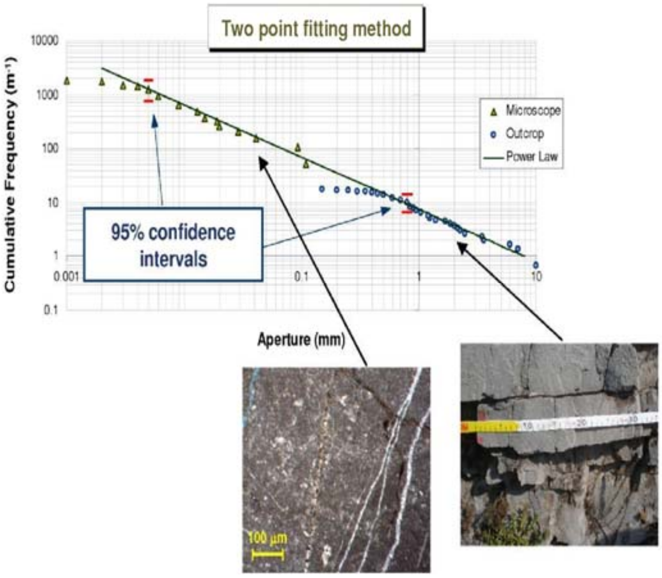


Fig. 5 A power-law distribution, interpolating empirical data, can be individuated by using a line for two opportunely chosen points, rather than a least-squares line. For each data set, the points are chosen as much as possible on the left side of the curve (i.e. where there is a lesser uncertainty), but before this attains a sub-horizontal trend (associated to truncation error).

D. Expressions for the Confidence Interval of Power Law Exponent

The proposed two point fitting method furnishes a rapid way to calculate the confidence interval of the power-law exponent. Indeed, the real values of fracture density, for the two points, fall in the respective 95% confidence intervals (Fig. 6). Furthermore, also the real power law (represented by a straight line in a bi-logarithmic diagram) falls within one of these confidence intervals, with the same probability. The joint probability that the line falls within the two intervals simultaneously is equal to the product of the probabilities that the line passes through each interval, being the two data set (quasi) stochastically independent (assuming that outcrop- and microscale scan lines have no common fractures). Therefore, the final probability is equal to 0.95^2, i.e. ca. 0.9.

By indicating with F_{low1} and F_{upp1} the confidence interval limits of the outcrop-scale scan line, F_{low2} and F_{upp2} those relative to the micro scan line, x_1 and x_2 the abscissas of the two selected points, the following equations for the 90% confidence interval of the exponent m have been obtained:

$$m_{upp} = -\frac{\ln\left(\frac{F_{low1}}{F_{upp2}}\right)}{\ln\left(\frac{x_1}{x_2}\right)}, \tag{12}$$

$$m_{low} = -\frac{\ln\left(\frac{F_{upp1}}{F_{low2}}\right)}{\ln\left(\frac{x_1}{x_2}\right)}, \tag{13}$$

where the interval limits (F_{low1} , F_{upp1} etc.) can be calculated by applying Eqs. (8) and (9).

This estimation method has general validity and can be usefully applied in all those cases in which a studied RV is distributed in accordance to a power law [4], requiring the only hypothesis that each cumulative frequency value follows a Poisson distribution in the sampling domain. The method may also be modified in order to include all those cases in which spatial/temporal frequency follows different statistical distributions, in so doing becoming of more general application to statistical as well as fractal analysis (e.g. to box counting techniques). This goes well beyond the scope of this paper and will be object of future research.

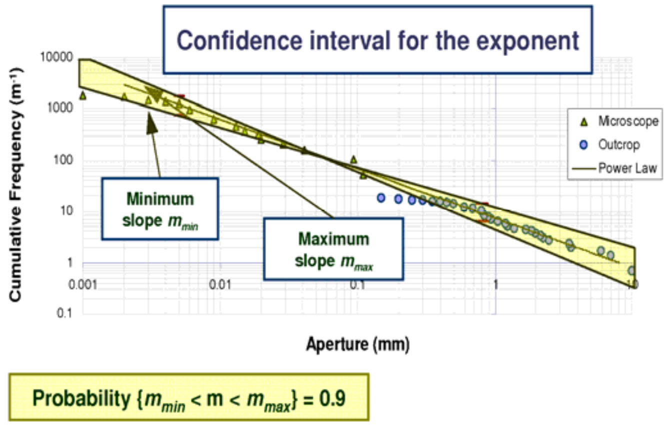


Fig. 6 The line is contained within each one of the two confidence intervals with a probability of 95%. The joint probability that the line is contained simultaneously within the two intervals, is equal to $0.95^2 \approx 0.9$.

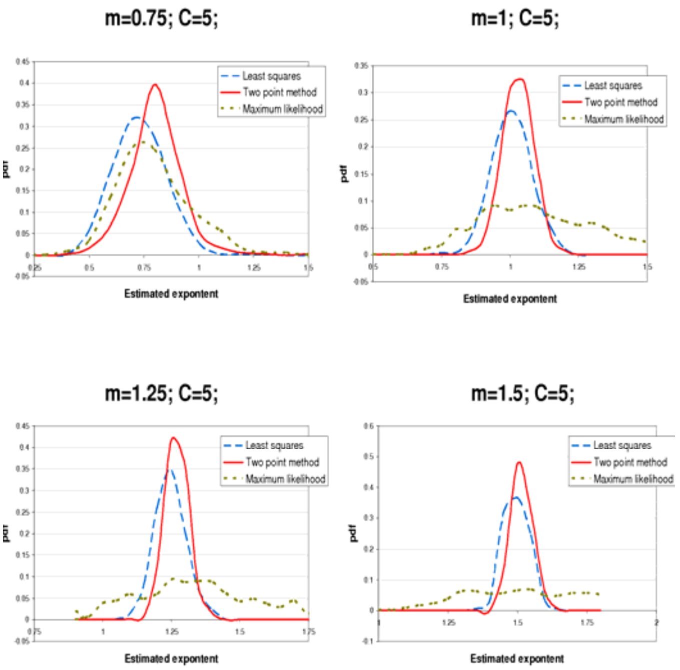


Fig. 7 Figure. 1 Probability density function (pdf) of exponent estimates from Monte Carlo simulations. The estimation has been performed, for four values of the exponent, according to three methods: (i) maximum likelihood applied on field scan line data (green line), (ii) least squares (blue line) and two point fitting method (red line) applied on multi scale sampling data. For all cases the two point interpolation criterion provides a smaller dispersion of exponent estimates. Maximum likelihood method provides acceptable results, for the simulated case of a 4 m sized scan line, only for a low value of the exponent.

V. VALIDATION OF THE TWO POINT FITTING METHOD AND COMPARISON WITH LEAST SQUARES AND MAXIMUM LIKELIHOOD METHODS

A. *Monte Carlo Simulation*

In order to validate the method described in the previous sections, Monte Carlo simulations [3, 22] have been performed. The experiment simulates a set of sampling estimates of the power-law exponent and coefficient, according to the following three methods: (i) field scale scan line and exponent estimated by maximum likelihood method, (ii) multi-scale scan line and estimation by least squares method and (iii) multi-scale sampling with estimation by two point method. The applied theoretical distribution is a power law characterized by reliable exponent and coefficient values. Namely the coefficient has been fixed to 5 with the following values for the exponent: 0.7, 1, 1.25 and 1.5. Furthermore, random spatial fracture distributions over a 4 m long scan line and a 2 cm long micro-scan line have been assumed. Each simulation produces an empirical aperture cumulative distribution according to the following criterion. In a scan line of length L, the number of detected fractures with aperture greater than a fixed minimum value x_{min} is a Poisson RV, whose mean is provided by the power law. For example, consider the case of exponent equal to 1. The mean fracture number per meter whose aperture is greater than 0.8 mm is 6.25 (provided by: $F(x_{min})=5\cdot 0.8^{-1}$). Thus, in a 4 m long scan line the expected fracture number is 25 (given by: $6.25 \cdot 4$). The detected fracture number is a Poisson RV with a mean value equal to 25. Applying the Monte Carlo method [3, 22], a determination of this latter RV (i.e. the detected fracture number) is simulated. Then an aperture value is associated to each fracture, according to a power law distribution. The final result is an artificial probability distribution of aperture values. The simulation has been repeated 1000 times for each one of the mentioned exponent values. In the simulation of a traditional scan-line data set (gathered on outcrop), only fractures with apertures larger than 0.8 mm have been considered in the calculation of the best-fit power law by means of the maximum likelihood criterion. In the case of multi-scale analysis, aperture values in the range 0.005-0.8 mm, for micro-scan lines, and larger than 0.8, for outcrop-scale scan lines, have been considered. Finally, the least squares line and the line passing for two points (with abscissa values of 0.005 and 0.8) have been calculated. In this way we produced 1000 simulated sampling estimates of the power-law exponent for each experiment. Results of the Monte Carlo simulations are shown in Fig. 7, as probability density function of exponent estimation according the three mentioned methods.

B. *Efficiency of the Compared Estimation Methods*

Often the parameters of power law best fitting the empirical cumulative frequency are evaluated by means of the least squares method. A more effective criterion than that of least squares, aimed at estimating the power-law exponent, is the maximum likelihood method. The basic theory of this method is thoroughly illustrated by Newman [4], where practical equations are also provided to obtain the power-law exponent estimate. In statistical inference, the maximum likelihood method generally provides very robust and highly converging estimators. Although this method provides a more efficient estimator than that of least squares also in analyzing power-law distributions, its formulation for estimating the exponent and confidence interval cannot be applied to a multi-scale data set like that utilized in this paper, as it can be employed only on cumulative distribution functions rather than cumulative frequency. Fig. 7 points out that estimation achieved by multi-scale sampling (either by least squares or two point method) provides a notably smaller dispersion of estimates with respect to maximum likelihood method. For all exponent values chosen for the performed simulations, the two point fitting method provides a more convergent estimator (i.e. exhibiting the smaller standard deviation) than lest squares method. In order to emphasize the significant advantage provided by use of this criterion we address to an example. For the data set associated to the exponent value of 1 (4 m sized scan line and 2 cm micro-scan line, coefficient c=5), the exponent estimation showed a standard deviation equal to 0.057. Taking into account that a micro scan line carried out over a thin section cannot be long more than 2-3 cm, in order to achieve the same standard deviation by use of least squares criterion, we had to carry out a 18 m sized scan line (Table 1).

TABLE I COMPARISON AMONG CRITERIA USED TO ESTIMATE THE EXPONENT OF THE POWER-LAW DISTRIBUTION. NOTE THAT THE TWO-POINT FITTING METHOD REQUIRES THE MINIMUM SCAN LINE AND MICRO SCAN LINE LENGTH TO ACHIEVE THE SAME VALUE OF STANDARD DEVIATION FOR THE EXPONENT

Method	Standard deviation of exponent estimate	Scan line length (m)	Micro scan line length (cm)	Number of sampled fractures
Least squares	0.057	18	2	132
Two points	0.0569	4	2	45

VI. CONCLUSIONS

A procedure for multi-scale analysis of power law probability distributions is suggested. This includes:

- new practical formulations, in closed form, for the approximate confidence interval of the mean of Poisson random variable;

- a novel criterion for estimation of power law parameters – two point fitting method - based on the definition of a straight line passing for two opportunely selected points only, rather than the least squares method. These two points belong to data sets gathered at different scales of observation;

- new expressions, in closed form, for estimating the confidence interval of the power law exponent;

This procedure can be applied to scaling analysis of those quantities showing probability distribution according to a power law and which are associated to objects, or events, randomly distributed on the space (e.g. over a line or a surface) or along the time. It is articulated according to the following steps: (i) verifying of random spatial/temporal distribution of studied objects/events, (ii) calculating confidence intervals for opportunely selected cumulative frequency values, (iii) estimation of power law parameters and (iv) evaluating the confidence interval for the estimated power law exponent.

Monte Carlo simulations have proved that the two point fitting method increases significantly the precision of power law exponent estimation, providing a more efficient estimator than least squares and maximum likelihood methods.

ACKNOWLEDGMENT

The author would like to thank Stefano Mazzoli and Stefano Vitale for introducing him to the studied reservoir analogue and for the collaborative work on these outcrops. Financial support by the SHELL ITALIA E&P for carrying out research on fractured carbonate reservoirs is gratefully acknowledged.

REFERENCES

[1] B. Gutenberg and R. F. Richter, “Frequency of Earthquakes in California,” Bull. Seismol. Soc. Am., 34, 185, 1944.

[2] G. Neukum and B.A. Ivanov, in Hazards Due to Comets and Asteroids. edited by T. Gehrels, University of Arizona Press, Tucson, AZ, 1994, 359 – 416.

[3] Dekking, F. M., Kraaikamp, C., Lopuhaa, H. P., Meester, L. E., A Modern Introduction to Probability and Statistics: Understanding Why and How. Springer-Verlag, London (UK), 2005.

[4] Newman, M., “Power laws, Pareto distributions and Zipf's law,” Contemporary Physics 46, 323–351, 2005, doi: 10.1080/00107510500052444.

[5] V. Pareto, Cours d’Economie Politique. Droz, Geneva, 1896.

[6] D. Sornette, Critical Phenomena in Natural Sciences: Chaos, Fractals, Selforganization and Disorder: Concepts and Tools. Springer, Berlin, 2000.

[7] L. A. Adamic and B. A. Huberman, “The Nature of Markets on the World Wide Web,” Q. J. Electron. Commerce, 1, 512, 2000.

[8] G.K. Zipf, Human Behaviour and the Principle of Least Effort. Addison-Wesley, Reading, MA, 1949.

[9] Belfield, W. C., “Multifractal characteristic of natural fracture apertures,” Geophysical Research Letters 21, No. 24, 2641–2644, 1994.

[10] Belfield, W. C., Sovich, J., “Fracture statistics from horizontal wellbores,” Journal of Canadian Petroleum Technology 34, 47–50, 1995.

[11] Gale, J. F. W., S. E. Laubach, R. A. Marrett, J. E. Olson, J. Holder, R. M. Reed, “Predicting and characterizing fractures in dolostone reservoirs: using the link between diagenesis and fracturing,” Geological Society, London, Special Publications, 235, 177-192, 2004, doi:10.1144/GSL.SP.2004.235.01.08.

[12] Ortega, O., Marrett, R., Laubach, E., “Scale-independent approach to fracture intensity and average spacing measurement,” AAPG Bulletin 90, 193–208, 2006.

[13] Guerriero, V., Iannace, A., Mazzoli, S., Parente, M., Vitale, S., Giorgioni M., “Quantifying uncertainties in multi-scale studies of fractured reservoir analogues: Implemented statistical analysis of scan line data,” Journal of Structural Geology, 2010, doi:10.1016/j.jsg.2009.04.016.

[14] Odling, N., “Scaling and connectivity of joint systems in sandstones from western Norway,” Journal of Structural Geology 19, 10, 1257-1271, 1997.

[15] Bonnet, E., Bour, O., Odling, N., Davy, P., Main, I., Cowie, P. A., Berkowitz, B., “Scaling of Fracture Systems in Geological Media,” Reviews of Geophysics 39, 3, 347–383, 2001.

[16] Bour, O., Davy, P., Darcel, C., Odling, N., “A statistical scaling model for fracture network geometry, with validation on a multiscale mapping of a joint network (Hornelen Basin, Norway),” Journal of Geophysical Research 107, B6, 2002, doi:10.129/2001JB000176.

[17] Davy, P., Bour, O., De Dreuzy, J. R., Darcel, C., “Flow in multiscale fractal fracture networks,” Geological Society, London, Special Publications 261, 31-45, 2006, doi:10.1144/GSL.SP.2006.261.01.03.

[18] Guerriero, V., Vitale, S., Ciracia, S., Mazzoli, S., “Improved statistical multi-scale analysis of fractured reservoir analogues,” Tectonophysics, 2011, doi:10.1016/j.tecto.2011.01.003.

[19] Gillespie, P.A., Johnston, J.D., Loriga, M.A., McCaffrey, K.L.W, Walsh, J.J., Watterson, J., “Influence of layering on vein systematics in line samples,” In: McCaffrey, K.L.W., Lonergan, L., Wilkinson, J.J. (Eds.), Fractures, Fluid Flow and Mineralisation. Geological Society, London, Special Publication 155, pp. 35–56, 1999.

[20] Balakrishnan, N., Basu, A. P., The Exponential distribution: theory, methods and applications. Gordon and Breach, Newark, NJ, USA, 1996.

[21] Wasserman, L., All of Nonparametric Statistics. Springer Science, New York (USA), 2006.

[22] Good, P. I., Resampling Methods. Birkhauser, Boston, 2006.

[23] Diggle, P. J., Ribeiro, P. J., Model-based geostatistics. Springer Science+Business, New York, 2007.

Vincenzo Guerriero - Dipartimento di Scienze della Terra, Università degli studi di Napoli 'Federico II', Largo San Marcellino 10, 80138 Napoli, Italy. E-mail: vincenzo.guerriero@unina.it.
He received his Ph.D. from the University of Naples 'Federico II', Italy. Statistical analysis/characterization of complex fracture networks in petroleum reservoir analogues is a major area of his research. Some of his main publications are:

Guerriero et al., "Fracture Network Properties in Limestone and Dolostone: Implication for Fluid Flow and Reservoir Behavior," in Horizons in Earth Science Research. Volume 7, Nova Publishers, 2012, ISBN: 978-1-62100-622-0.

Guerriero et al., "Improved statistical multi-scale analysis of fractured reservoir analogues," Tectonophysics, 2011, doi:10.1016/j.tecto.2011.01.003.

Guerriero et al., "Quantifying uncertainties in multi-scale studies of fractured reservoir analogues: Implemented statistical analysis of scan line data," Journal of Structural Geology, 2010, doi:10.1016/j.jsg.2009.04.016

Dr. Guerriero is a member of the Geo Africa Sciences Society and of the editorial board for the following Journals:

Advances in Petroleum Exploration and Development, Academic Journals CSCanada,

American Journal of Geoscience, Science Publications,

Journal of Petroleum Science Research (JPSR), American V-King Scientific Publishing.

Furthermore he is reviewer for:

Advances in Water Resources, Elsevier,

Mathematical Geosciences, Springer.