F.S.A.I., Vol. 10, No. 3, pp. 14-17, 2004

Some Mathematical Statistics for the Socio-technical Systems

Ennio Cortellini
University of Chieti - Pescara, Italy
e-mail: enniocortellini@inwind.it

Abstract. In this paper some new ways for the developing of mathematical statistics are proposed in the social field.

Key words: information, random variables, descriptive statistics

The first records of statistic methods appear in the seventeenth century, when John Graunt (1620 - 1674) together with W. Petty have invented the "political arithmetic" as method of study of the social phenomena using numbers and measures. In the same time, in Germany, the descriptive statistics develops as the analysis of the economic resources, of the commercial development, interconnected with population growth.

In the literature, many authors attribute the term of "statistics" to the German scientist G. Achenwall, even if the other investigators deny this, without mentioning an alternative origin.

The first epistemological step is represented by the work of Laplace, Legendre and Gauss in the Theory of Error. In fact, these represent the beginning of the Mathematical Statistics. In the mid 1800's A. Quételet (1796 - 1874) has introduced the Theory of Probabilities in the analysis of some social phenomena.

A considerable contribution to the mathematical statistics is due to F. Galton (1822 – 1907), which has developed the Theory of Correlations. Among the other many studies in the mathematical statistics we mention the work of K. Pearson (the Theory of selections), R. A. Fisher (the Dispersional Analysis, the Theory of the maximal likelihood, and so on).

Usually, in the dictionaries, or more generally in the technical literature, the definitions of the mathematical statistics, are not exhaustive. In our days, the mathematical statistics means "the development of the methods which permit to obtain, from experimental data, scientific conclusions of the casual phenomena formed by a great number of individual phenomena produced like the consequences of the practical accomplishment of some determined conditions and which are repeated every time when these conditions are estimated".

Nowadays in mathematical statistics we distinguish there directions of research:

- the descriptive – documentable statistics;
- the inferential statistics (classical and predictive);
- the correlational statistics.

Omitting the descriptive statistics (because it has no mathematical characterization), one can say that the main problem of statistics can be formalized in this way.

One considers a population (which is an arbitrary set of elements) as the object of the analysis, and one of its characteristic or more than one when one intends to examine the correlations between them.

Choosing a sample, adequately selected, we estimate its elements from the point of view of the considered characteristic, obtaining a set of corresponding values (the experimental data).

Through the inference processes, these data reveal the intrinsec informational messages, also called the conclusions on the population.

From the mathematical point of view, we can reformulate: is it possible to extend the function determined by the sample to a function defined on the set of all possible values of the considered characteristic into the set of natural numbers (in the classical frequential case)?

In the inference process three aspects are distinguished: frequential (called "Misses – Wald" by M. Frechet); logical, introduced by J. M. Keynes; and subjective which derives from the subjective concept of probability introduced by De Finetti.

The last expression suggests a new way of research (the subjective mathematical statistics), obtainable by modifying the static concept of the element of the sample in a dynamic one; in more mathematical terms, for the dynamic element the sequence to which the element belongs or the context in which the element appears is taken into account.

Moreover, it is necessary to extend the meaning of the statistical value of the characteristic. An examination points out the interconnections with other disciplines, namely the information theory and the set theory.

From the point of view of information theory we know that information is obtained either by measuring or by perceptions. Through the measuring we get the metric or quasimetric information – which are expressible with fuzzy numbers – while when the perceptions are used the information is represented by linguistic attributes. It is remarkable that in some cases the perceptive information has quasimetric corresponding. Also a set can be determined in two ways: the first one by enumeration (just for the finite sets) and the second one specifying a characteristic property. The properties can be qualitative or quantitative; a correspondence between qualitative properties and perceptions respectively quantitative properties and measuring emerges. When taking a sample of *n* elements and the characteristic *c*, evaluating the elements from the point of view of *c* one obtain the values $x_1, x_2, ..., x_n$. Using the remarks measures we have the following cases:

1) the characteristic *c* can take the discrete values; consequently $x_1, x_2, ..., x_n$ are numbers;

2) the values $x_1, x_2, ..., x_n$ are numbers i.e. discrete values, but the characteristic *c* can have continues values in a real interval;

3) the values $x_1, x_2, ..., x_n$ can be real intervals;

4) the values $x_1, x_2, ..., x_n$ can be linguistic attributes;

5) the values $x_1, x_2, ..., x_n$ can be fuzzy numbers.

Generally speaking the observations on a sample are represented as a random variable of the form:

$$X = \begin{pmatrix} x_1 & x_2 & ... & x_k \\ n_1 & n_2 & ... & n_k \end{pmatrix} \qquad (1)$$

where n_i is the frequency of x_i.

Since the values $x_1, x_2, ..., x_n$ are not necessarily pairwise distinct, we can assume (relabelling if necessary) that the random variable *X* takes only *k* distinct increasing values $x_1, x_2, ..., x_k$.

Now, the effective mathematical tools are just those used in the discrete field or rather in the continues case, but only starting with the repartition function H. It seems necessary to examine all the possible passages between all the previous mentioned cases.

- The passage from 2 to 3.

In this case it is possible an algorithmic migration of the type:

- let $a = \min\{x_1, x_2, ..., x_n\}, b = \max\{x_1, x_2, ..., x_n\}$, $w = b - a$

- we construct the number $p = \dfrac{w}{1 + 3,222\ln(n)}$ (the discretization step of Sturges);

- we determine $N = \min\{k \mid w < k \cdot p\}$;

- construct the covering interval $[A, B]$, where $A = a - \dfrac{N \cdot p - w}{2}$, $B = b + \dfrac{N \cdot p - w}{2}$;

- construct the intervals $[d_0, d_1), [d_2, d_3), ..., [d_{n-1}, d_n)$, where $d_0 = A, d_1 = A + p, ..., d_n = B$.

In this case the frequencies (n_i) for the constructed intervals are summated.

- The passage from 4 to 1.

It is possible to pass from linguistic attributes to discrete values of the variable. This passage takes place just in the case when the linguistic attributes are ordered (for example from small to big). The relative algorithm is: if *n* is the number $x_1 = 1, x_2 = a + h, ..., x_n = a + (n-1)h$ of pairwise distinct attributes, then compute *a* and *h*:

$$a=-\sqrt{\frac{3(n-1)}{n+1}}, \quad h=\sqrt{\frac{12}{n^2-1}};$$

(knowing that *a* and *h* are obtained setting the mean value to zero and the square difference equal to one for the starting variables).

Substitute the values $x_1=1, x_2=a+h, ..., x_n=a+(n-1)h$.

It is clear that values of x_i are such that - $\sqrt{3}<x_i<\sqrt{3}$.

- The passage from 3 to 1.

Given $x_1=[1_0,1_1), ..., x_n=[1_{n-1},1_n)$, it is possible to replace every interval with the arithmetic mean of its limits.

We omit the fifth case of passage from 5 to1, because this will be the subject of another work for which it is necessary a specific presentation and a close examination of the fuzzy mathematical instruments.

Moreover, the algorithms of the transformation permit to present in the same time (in many other cases it is suitable to present them separately) the various types of the research results.

For the study of the variable $_iX=\begin{pmatrix} x_1 & x_2 & ... & x_k \\ n_1 & n_2 & ... & n_k \end{pmatrix}$, generally one studies its arithmetic mean and consequently other statistic functions and parameters. It is clear that if a priori we replace in the above the arithmetic mean with other type of mean all the other statistic parameters to be modified.

A criterion for the appropriate choice of the mean is required. The only known result in this direction is the Chisini - Boiarski test.

Another open problem appears when passing from the absolute to the relative frequencies, which implies passing to the probability defined by the classical theory. Thus the necessity to investigate the behaviour of the variables and of the statistic functions in the case when the probability becomes the subjective one proposed by Finetti.

Finally, but not less important, it seems inevitable to pay attention to a further epistemological necessity, namely to research perturbation phenomena (in the sense of Schrödinger) in the processes of extrapolation of the additional information; the process will have to be considered a systemic process and thus one enters the wider problem of complexity.

References

[1] DARIO ANTISERI, *Trattato di metodologia delle Scienze Sociali*, UTET Libreria, 1996

[2] J.P. BENZECRI, *L'analise des donnes*, I,II, Dunod Paris, 1980

[3] C.CALOT, *Cours de Statistique Descriptive*, Dunod, Paris, 1973

[4] B. De FINETTI, *Teoria delle Probabilita*, I,II, Einaudi, Torino, 1970

[5] G. GIRONE, T. Salvemini, *Lezioni di statistica*,I,II, Cacucci, Bari,1983

[6] G. LETI, *Statistica Descriptiva*, Il Mulino, Bologna, 1983

[7] ALFREDO RIZZI, *Analisi dei dati*, La nuova Italia Scientifica, Roma, 1985