JMSP / SPAS

SPAS - SASP

Journal of Mathematical Facts and Short Papers (JMFSP)
Vol. 2, 2019 (1), pages 35 - 41.
DOI: http://dx.doi.org/10.16929/jmfsp/2019.35.005

On the ubiquitous notion of mean in probability and statistics

Aboubakar Maitournam

University Abdou Moumouni of Niamey, faculty of sciences and techniques, department of mathematics and computer sciences, PB 10 662, Niamey, Niger

Received on September 2, 2019; Accepted on September 20, 2019; Published Online on September 27, 2019

Abstract. In probability and statistics, the basic notion of probability of an event can be expressed as a mathematical expectation. The latter is a theoretical mean and is an essential parameter of most probability distributions, in particular of the Gaussian distribution. Last but not least, the notion of mean is at the core of two main theorems of probabilities and statistics, that is : the law of large numbers and the central limit theorem. Whether it is a theoretical or empirical version, the concept of mean is omnipresent in probability and statistics, is consubstantial to these two disciplines and is a bridge between randomness and determinism.

Key words: Mathematical expectation, ϕ-means, Gaussian distribution, law of large numbers, central limit theorem.
AMS 2010 Mathematics Subject Classification Objects : 60-00; 60-01; 60-02; 62-00; 62-01; 62-02

Aboubakar Maitournam, maitourna1@gmail.com

Résumé. (Abstract in French) En théorie des probabilités et statistiques, la notion basique de probabilité d'un événement est une espérance mathématique donc une moyenne théorique. Cette dernière est une caractéristique essentielle des lois de probabilités en particulier de la loi normale. Elle est de plus au cœur des deux théorèmes fondamentaux des statistiques et probabilités : la loi des grands nombres et le théorème central limite. En définitif, le concept de moyenne est omniprésent en statistiques et probabilités, est consubstantiel à ces deux disciplines et est un pont entre le déterminisme et l'indéterminisme.

1. Introduction

The quantitative notion of mean is present tangentially in all scientific fields. It was known implicitly in geometry with the idea of barycentre of a system of points, then in physics with the concept of centre of gravity, thanks to Archimedes. Mőbius (1967) (as cited in Fauvel *et al.* (1993)) gave later a mathematical formulation of barycentre. Indeed, the coordinates of barycentre in a Cartesian system of axes are exactly the arithmetic means of coordinates of points, which are defining that barycentre. In algebra, the abstract idea of linear combination in a linear space Sander (1979) generalizes the concept of barycentre in an affine space, whereas in mathematical analysis, the integral Riemann (1868) of a function is in some sense a mean. The Voiculescu's free probabilities Voiculescu (1986) are also a very theoretical generalization of the mean in connection with the operator theory, probability and quantum mechanics. Like the centre of gravity which concentrates the gravity of a set of material points, in statistics, the mean characteristic (note, income, social stratum, $\cdots$) concentrates intuitively and practically the high number or frequency of individuals. In that era of information Shannon (1948) and data or big data Tukey (1962); Donoho (2000); American Statistical Association (2014), the concept of mean remains the main quantitative and simple parameter of data reduction or synthesis. In particular, it is omnipresent in probability and statistics theories. Here we present first as introduction the origin of the quantitative concept of mean and its occurrence in various mathematical settings followed by the used methodology, our findings, a discussion of its omnipresence in probability and statistics and some concluding remarks.

2. Methodology

We have reviewed the implicit or explicit implications of the concept of mean in key definitions, theorems and characterization of main objects and areas of statistics and probability theory.

3. Results

Due to its omnipresence in key definitions, theorems or characterization of main objects and areas of statistics and probability theory, such as the probability of an event, moments, correlations, characteristic functions, random variables, Brownian motion, Feynman Khac formula, second order stationary processes, ergodicity, estimator, Bayesian rules, fundamental inequalities, Gaussian Distribution, law of large numbers and central limit theorem;

the concept of mean, according to us, is the most important notion in probability and statistics. Furthermore it is a bridge between randomness and determinism.

4. Discussion of the omnipresence of the notion of mean in probability and statistics

In probability and statistics, the theoretical mean that is that of a random variable is defined by the term of mathematical expectation if the latter exists. For instance, the Cauchy random variable has no mathematical expectation. We can approximate the latter based on realizations $x_1 = X_1(\omega), \dots x_n = X_n(\omega)$ of a random sample $X_1, \dots X_n$ of independent identically distributed random variables with ω belonging to Ω the sample space, that is the set of results of a random experiment. The most usual approximation of mathematical expectation is what is called an empirical mean, for instance the arithmetic mean. The latter is available in two versions : as an estimator which is a random variable and has the following expression

$$\overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i \tag{1}$$

or as an estimation

$$\overline{X}_n(\omega) = \overline{x}_n = \frac{1}{n}\sum_{i=1}^{n} x_i = \frac{1}{n}\sum_{i=1}^{n} X_i(\omega) \tag{2}$$

which is the realization of the estimator $\overline{X}_n$. Beyond the single case of estimator $\overline{X}_n$, the mathematical expectation is used to evaluate the bias of any estimator of a given parameter.

In addition to the arithmetic mean, other types of means exist such as the quadratic, geometric or harmonic means grouped under the name of ϕ means where ϕ is a monotonic continuous function Kolmogorov (1930); Hardy *et al.* (1952); Lovric (2011). The Kolmogorov's axiomatic view of the mean Kolmogorov (1930); Carvalho (2016), unifies all these concepts of mean.

The concept of mean is also present in regression techniques Stanton (2001) and in the theory of stochastic processes in general Doob (1934) and in that of stationary processes Brockwell and Davis (1987) in particular. Indeed, in the regression approach, explanatory or independent variables are means and in the theory of stochastic processes, the Feynman Khac formula Karatzas and Shreve (1998) establishes a bridge between deterministic PDEs and their stochastic solutions through Brownian motion Bachelier (1900, 1901); Wiener (1923) and conditional expectation Kolmogorov (1933); Neveu (1970). The Feynman Khac formula and stochastic representations of deterministic PDEs show in some way that a smooth function is an average of irregular functions. The ergodic theorem of Von Neumann and Birkhoff Von Neumann (1932); Birkhoff (1932) a key result of probabilities and dynamical systems is based also on the notion of mean. In the theory of stationary processes, the second order stationarity signifies that means and covariances are invariant by temporal or spatial translation. Furthermore, seasonal and trend components are means. Concerning the notion of conditional expectation, basis of Bayesian Statistics Lindley

(1972) and stochastic processes, it is a generalization of mathematical expectation as it is an average depending on random or deterministic information.

In fact, the notion of mean is ubiquitous in probability and statistics because the fundamental brick of probability theory Kallenberg (1997) that is the probability $P(A)$ of an event A is the mathematical expectation $E(1_A)$ of the indicator function of A. The latter is a random variable defined by $1_A(\omega) = 1$ if $\omega \in A$ and $= 0$ if $\omega \notin A$. Moreover, we characterize in general a random variable by its moments (mean, variance, covariance, characteristic function,..) and we characterize the linear dependence between two random variables by their covariance and thus by their linear correlation coefficient. But behind the characteristic function, moments of various orders, covariance and linear correlation coefficient, lies the notion of mean that is the mathematical expectation. Thus the characteristic function of the random variable X is defined by $E(e^{itX})$ for all real number t. The variance of X is expressed as $E\big(X - E(X)\big)^2$. The covariance between X and Y noted $COV(X, Y)$ is a mean of scalar product of centered values of X and Y, that is $E\Big((X - E(X))(Y - E(Y))\Big)$. Last but not least, the linear correlation coefficient between X and Y, is a mean of centered and standardized values of X and Y, that is $E\Big(\frac{(X-E(X))(Y-E(Y))}{\sqrt{Var(X)}\sqrt{Var(Y)}}\Big)$. In addition, all major inequalities Hardy *et al.* (1952) in probability theory involve the mathematical expectation or its variants.

But the decisive importance of the concept of mean in statistics and probability the "geometry of randomness" by quoting Blaise Pascal, is due to the law of large numbers of Bernoulli and Kolmogorov Bernoulli (1713); Bienayme (1853); Tchebychev (1867); Markov (1899); Kolmogorov (1928, 1930) and to the ubiquity of Gaussian distribution in nature resumed by the central limit theorem of De Moivre and Laplace De Moivre (1967); Laplace (1774, 1812). Indeed, the law of large numbers postulates that the limit of the arithmetic mean $\overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$ of independent identically distributed random variables when n tends towards infinite, is deterministic, precisely equals to the common mathematical expectation of those random variables. A consequence of this theorem is that any regular integral in a deterministic sense defined on a compact set can be approximated by using samples of uniform distributions on the interval $[0, 1]$. This idea has given rise to the birth of the powerful and popular Monte Carlo method Eckhardt (1987); Devroye (1986) dedicated to the simulation of complex random phenomena. As for the central limit theorem, it postulates that the random variable $\overline{X}_n$ is asymptotically normally distributed regardless of the common distribution of underlying samples of identically distributed random variables from which it was derived. But the Gaussian distribution discovered by Abraham de Moivre, Laplace and Gauss De Moivre (1967); Laplace (1774, 1812); Gauss (1809), is perfectly common because it models many natural phenomena or characteristics as the size, the weight, the height, physical measures and errors. It lies even at the core of the most emblematic stochastic process namely the Brownian motion, which is in particular a Gaussian process.

But one of the essential properties of the Gaussian distribution is that its density function the famous "bell curve" is symmetrical in relation to its mathematical expectation. The latter coincides with the median and the mode in the Gaussian distribution setting. Thus for the Gaussian distribution, the maximal frequency is associated with the mean which represents the central tendency. This fact can be expressed by the equation

$$P\big[-z_{\frac{\alpha}{2}}\sigma < X - m < z_{\frac{\alpha}{2}}\sigma\big] = 100(1-\alpha)\% \tag{3}$$

where X is a Gaussian random variable with mean m and variance σ^2, $z_{\frac{\alpha}{2}}$ and $-z_{\frac{\alpha}{2}}$ are respectively the $1-\frac{\alpha}{2}$ and $\frac{\alpha}{2}$ percentiles of the standard Gaussian distribution, α varying between 0 and 1. The above equation shows that for the Gaussian distribution, normal points are concentrated around the mean value on one hand. On the other hand, abnormal points, called outliers in statistics, are uncommon with low frequency and are far from the mean which can be expressed by the following equation

$$P\big[|X-m| > z_{\frac{\alpha}{2}}\sigma\big] = \alpha = 100\alpha\%. \tag{4}$$

5. Conclusion

Whether in its theoretical version as a mathematical expectation or in its empirical variation as an arithmetic average, the mean is the parameter, which in general is close to every value in a sample. This explains on one hand its ubiquity in probability and statistics. On the other hand there is a fundamental equivalence between the probability of an event and the mathematical expectation of this event's indicative function. Furthermore numerous key parameters of random variables and major inequalities and properties in probabilities and statistics are mean-based. Last but not least, the Gaussian distribution catches the fact that many physical quantities, are in terms of frequency concentrated around their mean value, which acts therefore as a magnet, a focal point thanks to the law of large numbers and the central limit theorem. But the importance of mean goes beyond physics, geometry, algebra, analysis and the commutative setting of traditional probabilities. Indeed, the very theoretical non-commutative probabilities or free probabilities are also a high level formalization of the conventional mean. Finally the mean is a bridge between randomness and determinism.

Acknowledgments. I wish to thank Dr. Carmen Buchrieser Senior Researcher at Pasteur Institute for her valuable comments.

References

(Edited) Fauvel, J., Flood, R. & Wilson, R. (1993). *Mobius and his Band: Mathematics and Astronomy in Nineteenth-century Germany*. Oxford; New York: Oxford University Press.

Sander.D.F. (1979). Hermann Grassmann and the Creation of Linear Algebra. *American Mathematical Monthly* 86, 809-817.

Riemann. R. (1868). On the representability of a function by a trigonometric series. *Proceedings of the Royal Philosophical Society at Gőttingen*, vol. 13, pages 87-132.

Voiculescu. D. (1986). Addition of certain non-commuting random variables. *J. Funct. Anal.* 66, 323–346.

Shannon. C.E. (1948). A Mathematical Theory of Communication. *The Bell System Technical Journal*, Vol. 27, pp. 379–423, 623–656.

Tukey. J.W. (1962). The future of data analysis. *Ann. Math. Statist*; 33:1-67.

Donoho. D. (2000) *High-dimensional data analysis: The curses and blessings of dimensionality.* Aide-Mémoire, Stanford University.

ASA (American Statistical Association). (2014). *Discovery with data. Leveraging statistics with computer science to transform science and society.* A working group of the American statistical association; 27 pages. www.amstat.org/newsroom/pressreleases/2014-ASAWhitePaper.pdf

Kolmogoroff. A. (1930). Sur la notion de la moyenne. *Rend. Accad dei lincei* (6), 12, 388-391.

Hardy. G.H., Littewood. J.E., Polya. G. (1952). *Inequalities.* Cambridge university press.

Lovric M. (ed.). (2011). *International Encyclopedia of Statistical Science.* Springer-Verlag Berlin Heidelberg.

De Carvalho. M. (2016). "Mean, What do you Mean?". *The American Statistician.* 70 (3), 270-274.

Stanton. J.M. (2001). Galton, Pearson, and the Peas: A Brief History of Linear Regression for Statistics Instructors. *Journal of Statistics Education* Volume 9, Number 3. Online : jse.amstat.org/v9n3/stanton.html

Doob. J. (1934). Stochastic Processes and Statistics". *Proceedings of the National Academy of Sciences of the United States of America.* 20 (6): 376–379.

Brockwell. P.J., Davis. R.A. (1987). *Time Series: Theory and Methods.* Springer.

I. Karatzas. I. and Shreve. S.E. (1998). *Brownian Motion and Stochastic Calculus.* Second edition. Springer.

Bachelier. L. (1900). Théorie de la spéculation. *Ann. Sci. École Norm. Sup.* 17, 21–86.

Bachelier. L. (1901). Théorie mathématique du jeu. *Ann. Sci. École Norm. Sup.* 18, 143–210.

Wiener. N. (1923). Differential space. *J. Math. Phys.* 2, 131–174.

Kolmogorov. A.N. (1933-56). *Foundations of the Theory of Probability.* Engl. trans., Chelsea, New York. 1933-56.

Neveu. J. (1970). *Bases mathématiques du calcul des probabilités.* Masson.

Von Neumann. J. (1932). Proof of the quasi-ergodic hypothesis. *Proc. Natl. Acad. Sci.* USA 18, 70–82.

Birkhoff. G.D. (1932). Proof of the ergodic theorem. *Proc. Natl. Acad. Sci.* USA 17, 656–660.

Lindley. D.V. (1972). Bayesian Statistics, A review. *SIAM*, 31 janv. 83 pages.

Kallenberg. O. (1997). *Foundations of modern Probability.* Springer.

Bernoulli. J. (1713). *Ars Conjectandi.* Thurnisiorum, Basel.

Bienaymé.I. (1853). Considérations à l'appui de la découverte de Laplace sur la loi de probabilité dans la méthode des moindres carrés. *C.R. Acad. Sci.* Paris 37, 309–324.

Chebyshev. P.L. (1867). Des valeurs moyennes. *J. Math. Pures Appl.* 12, 177–184.

Markov. A.A. (1899). The law of large numbers and the method of least squares (in Russian). *Izv. Fiz.-Mat. Obshch.* Kazan Univ. (2) 8, 110–128.

A.N. Kolmogorov. A.N. (1928). Ű̈ber die Summen durch den Zufall bestimmter unabhg̋angiger Grw̋ossen. *Math. Ann.* 99, 309–319; 102, 484–488.

Kolmogorov. A.N. (1930). Sur la loi forte des grandes nombres. *C.R. Acad. Sci.* Paris 191, 910–912.

De Moivre. A. (1967). *Approximatio ad Summam Terminorum Binomii in Seriem Expansi.* 1733-56. Translated and edited in The Doctrine of Chances, 2nd and 3rd eds. Reprint ed., F. Case and Chelsea, London and New York.

De Laplace. P.S. (1774). Mémoire sur la probabilité des causes par les évènements. *Mémoires de l'Académie royale des sciences de Paris (Savants étrangers)*, t. VI. p. 621; p.27-65.

De Laplace. P.S. (1812-20). *Théorie Analytique des Probabilités*, 3rd ed. Repr. in œuvres Complètes de Laplace 7. Gauthier-Villars, Paris 1886–1912. 1812-20.

Roger. E. (1987). "Stan Ulam, John von Neumann, and the Monte Carlo method". *Los Alamos Science*, Special Issue (15): 131–137.

Devroye. L. (1986). *Non Uniform Random Variate Generation.* Springer Verlag, New York.

Gauss. C.F. (1963). *Theory of Motion of the Heavenly Bodies.* 1809. Engl. trans., Dover, New York.

Mőbius A. (1967). *Gesammelte Werke*, 4 volumes (edited by R. Baltzer, F Klein, W Scheibner), Leipzig, 1885-7; reprinted Dr Maring Sặndig, oHG, Wiesbaden