The normal distribution, an epistemological view

Benjamin Benzon [iD]

Department of Anatomy, University of Split, School of Medicine, Split, Croatia

Correspondence to:
Benjamin Benzon
University of Split School of Medicine,
Šoltanska 2, 21000 Split, Croatia
benjamin.benzon@mefst.hr

The role of the normal distribution in the realm of statistical inference and science is considered from epistemological viewpoint. Quantifiable knowledge is usually embodied in mathematical models. History and emergence of the normal distribution is presented in a close relationship to those models. Furthermore, the role of the normal distribution in estimation of model parameters, starting with Laplace's Central Limit Theorem, through maximum likelihood theory leading to Bronstein von Mises and Convolution Theorems, is discussed. The paper concludes with the claim that our knowledge on the effects of variables in models or laws of nature has a mathematical structure which is identical to the normal distribution. The epistemological consequences of the latter claim are also considered.

"Everybody believes in the exponential law of errors [i.e., the normal distribution]: the experimenters, because they think it can be proved by mathematics; and the mathematicians, because they believe it has been established by observation.", H. Poincare [1]

Introduction

Every student who attended statistics class came in contact with the normal distribution quite rapidly. Many were puzzled with the omnipresence of the normal distribution in statistics and its overrepresentation when compared to other types of distributions. Even a famous statistician such as K. Pearson could not resist the mathematical lore of the normal distribution and hypothesized that even those things that are not distributed normally are actually a mixture of various normal distributions [2]. This paper aims to present a realistic status of the normal distribution within the realms of science and statistics.

Knowledge, measurement, and variability

Natural and some other sciences are based on measurements; a property which universally accompanies any type of measurement is variability. Conceptually speaking, the variability arises due to properties of measuring technique and/or properties of studied system [3]. Purpose of science is to explain i.e. to reduce such variability by making predictions based on the properties of studied system. Such properties are identified by experimentation or observation. Furthermore, they are usually connected in a logically coherent theories, which are then formalized in terms of statistical, i.e. mathematical models. However, due to incomplete understanding of systems, which are typically complex, only occasionally a deterministic ideal of completely diminished variability can be achieved [3]. So it can be said that unexplained variability is a measure of knowledge deficiency i.e. ignorance [3]. This fact leaves us with reality that most models have both deterministic and stochastic (ancient Greek, στόχος, stókhos, to aim, guess) parts [4] (**Figure 1 a**).
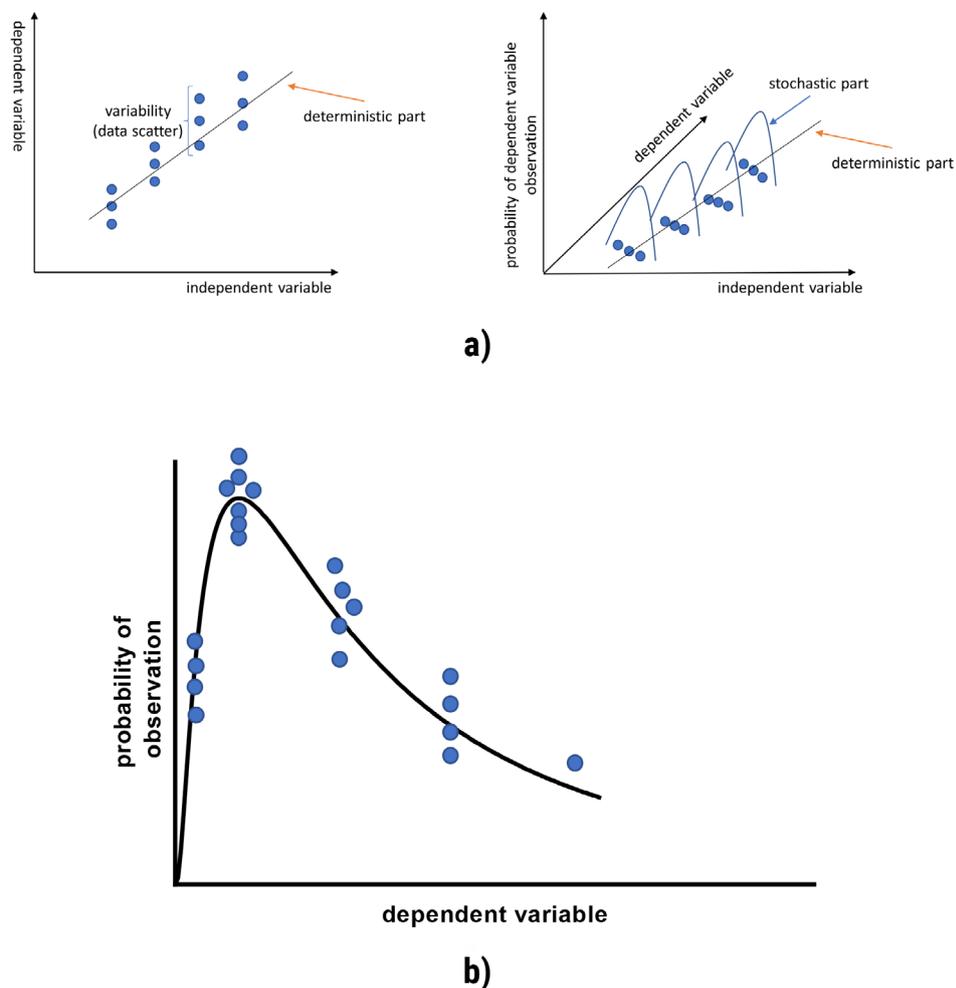

Figure 1. a) Depiction of data variability around deterministic part of model in a left panel. Right panel presents a full model with both deterministic component (line) and a probability distribution (blue) that describes data variability. b) Fully stochastic model, a log normal probability distribution in this case, without any deterministic part.

Deterministic part is some kind of mathematical function which contains variables that are considered to govern the system, whereas stochastic part of a model is also a mathematical function which describes data scatter around deterministic part i.e. the variability which was left unaccounted for by deterministic part of model [4]. The latter type of mathematical functions is called probability distribution [4]. It describes probability of observation occurrence around some kind of central point which might or might not be defined by deterministic part of model (**Figure 1 a**). If a deterministic part of a model is replaced by probability distribution, then the model can be considered purely stochastic which means that properties of system cannot predict the outcome i.e. outcome is treated as random variable (**Figure 1 b**) [4]. Models that are most common in science are models with finite number of variables and they are called parametric models. Since they lend themselves to extrapolation and intrapolation very easily, their application in decision-making process is most common, therefore this paper will consider only parametric models [3, 5, 6].

Estimating the effects of variables

Another problem that is ubiquitous in sciences that use measurement is so called estimation or inverse problem. System properties which are used to predict outcomes of interest are called independent, input, explanatory or predictor variables [4]. The estimation procedures seek to infer values of model parameters ("estimands"), i.e. effects of each variable, from the collected data [4]. Estimation procedure is a mathematical algorithm which is used to select the "the best" value of a model parameter by some kind of criterion [7]. This criterion is calculated by mathematical function derived from model equation which takes data as an input variable. This function is called an estimator. The model parameter value selected in such manner is called point estimate. Furthermore, in addition to point estimate, estimation procedure usually gives a range of plausible values which are called interval estimate and a value of estimator (i.e. criterion), which can be used to judge model validity in relation to other competing models [7, 8]. If those plausible values are plotted with their respective value of the criterion, a graph of a so-called sampling distribution of a model parameter estimate will arise (**Figure 2**) [7]. The latter often has the same mathematical form as some common probability distributions. However, whether it represents actual probability or something else is a matter of centuries long debate in statistics [9]. To conclude, in estimation procedure a scientist inductively reasons from a sample (i.e. collected data) to a population [2].

The discovery of the normal distribution and its place within statistical inference is intimately connected to the above-mentioned problems of estimation and variability.

How did the normal distribution emerge?

Problems of variability

Sciences that were first to introduce measurement to their method were astronomy and geodesy both of which were essential for the European society at the time of Renascence and Modern age [10]. Typical problems of that age were calculating an orbit of a celes-
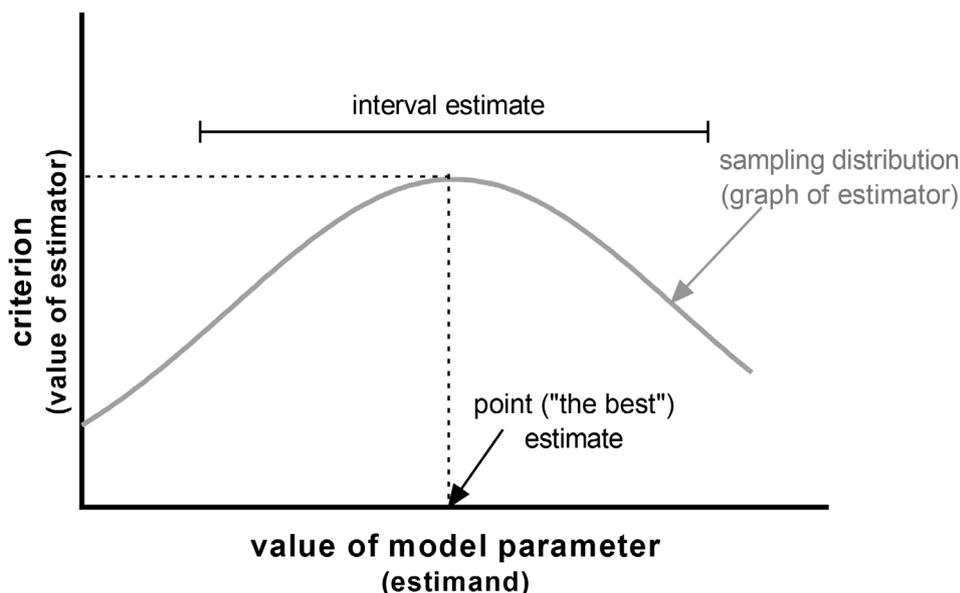
Figure 2. An example of hypothetical sampling distribution of model parameter (estimand) along with point estimate (a value of model parameter associated with the greatest estimator value) and an interval estimate. Estimator is defined as function of data which is used to calculate the criterion based on which the optimal estimate of parameter is chosen.

tial body based on series of observations or calculating an arc of meridian also based on instrumental observations. If those problems are looked at in more general manner two questions arise. The first one is how to deal with variability in multiple measurements taken under same conditions [10]. The second one is how to deal with variability in measurement taken under different conditions, for example, a measurement of same variable under two different experimental conditions or on different occasions of an observation [10]. Answering the first question would eventually result in development of different probability distributions or error functions as they were called at the time [11]. Likewise answering the second question would result in development of least absolute difference and least squares estimation techniques and eventually in maximum likelihood estimation procedure [12]. From both historical and mathematical viewpoints answers to these two questions turned out to be directly connected.

In the mid-18th century, a problem of exact shape of Earth was at the center of scientific interest. One of the methods employed to study the problem was measurement of length of meridian arc. A scientific expedition would travel along the meridian and measure lengths of the arc at different latitudes [10]. Model which was used to translate length and latitude to information about Earth shape was a simple linear model derived from geometric considerations [10]. However, due to variability in measurement the data points collected at different latitudes (i.e. occasions) could not be aligned with a single straight line when plotted. First to solve this problem of fitting a straight line through set of scattered points was Roger J. Boscovich (*Ruđer Josip Bošković*) a Croatian scientist, philosopher and Jesuit priest [13]. His solution was based on idea that line should be drawn in such manner that sum of absolute deviations, on y axis, (i.e. differences) between the line and individual data points should be minimized [2, 10, 12] (**Figure 3**). Boscovich did not model the data

Boscovich drew a line between data points in a such way that the sum of *absolute values* of distances parallel to y axis between data points and line was minimized, whereas Gauss minimized the sum of *squares* of those distances.
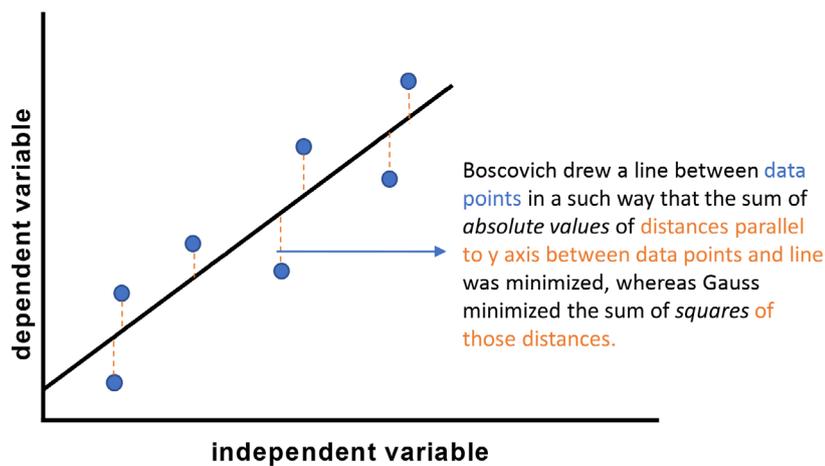
Figure 3. Illustration of least absolute difference and least squares estimation techniques.

scatter around the deterministic part of model (i.e. the straight line) explicitly since the concept of modelling data scatter with probability distributions was most probably unknown to him in 1757, when he developed his solution [11, 12].

The notion that variability in measurement can be treated as probability density (distribution) or error curve was first put forward by Thomas Simpson in 1756 [11, 12]. Simpson modeled the dispersion of measurements around a center by what we would call today uniform and triangular probability distributions [11]. Laplace followed by proposing two error curves, both with the property of exponential decay in probability of measurements as the latter deviates more from central, i.e. most common, value (**Figure 4**) [11]. Daniel Bernoulli went few steps further and suggested that parameter estimation for fitting his semicircular error curve should be done by choosing those parameter values, among the whole group of potential parameter values, that will maximize the likelihood, as we would call it today, of generating the observed data [12]. This was the birth of what is today known as maximum likelihood estimation procedure [14].

Between 1794 and 1798 Johann Carl Friedrich Gauss made derivation of what we call today normal probability distribution. In solving the problems of orbits of celestial bodies Gauss faced the difficulties that were essentially identical to those of Boscovich. However, since Gauss was primarily an excellent mathematician, he discovered that minimizing the sum of squared differences, on y axis, between the deterministic part of model (e.g. straight line) and individual data points is far more advantageous in terms of algebra and arithmetic then minimizing the sum of absolute differences [12]. Despite this elegant idea of squaring the differences, Gauss was not satisfied with its justification solely on computational ground [12]. In his attempt to find conceptual grounding for minimizing the sum of squared differences, he concluded that such thing is impossible without explicitly modelling data scatter with an error curve [12]. After a theoretical consideration on
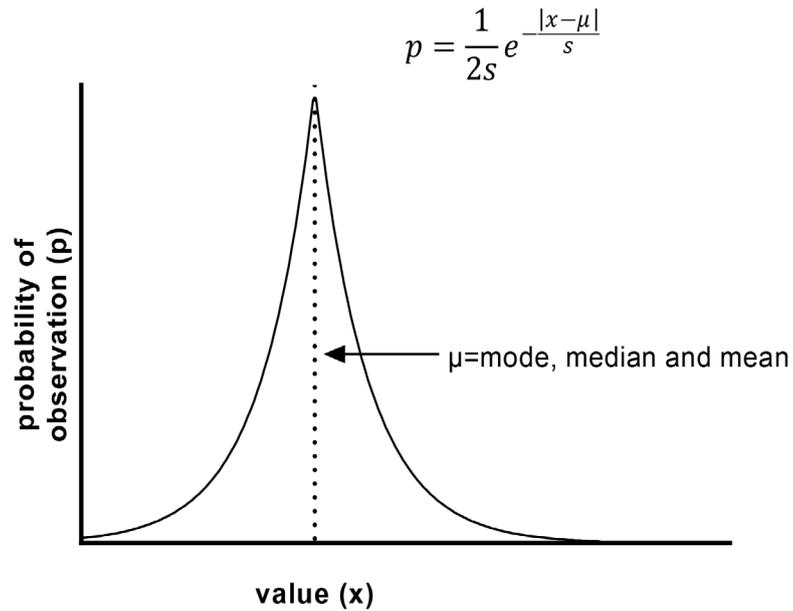
$$p = \frac{1}{2s} e^{-\frac{|x-\mu|}{s}}$$

Figure 4. Laplace's error curve. s,mean absolute deviation(MAD).

properties of observational errors Gauss came to three assumptions: 1) small errors are more likely than larger errors, 2) error curve must be symmetrical i.e. equal deviations in negative and positive directions from most common measurement have equal probability of occurrence and 3) the most likely quantity being measured i.e. the most common measurement in a set of measurements is an arithmetic average of all measurements [11]. When these three assumptions are translated in mathematical language the equation of normal probability distribution arises (**Figure 5**) [11]. Gauss went even further and by use

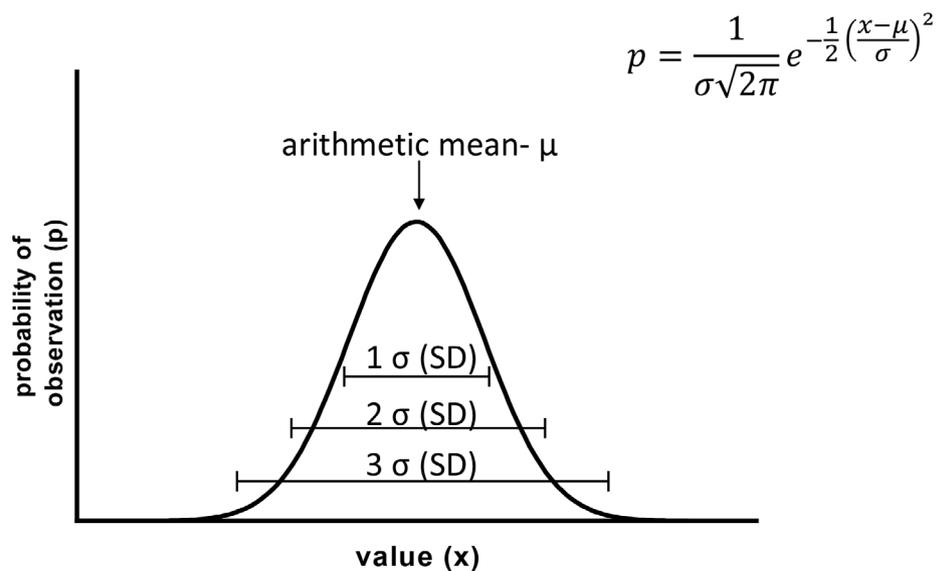$$p = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

Figure 5. A graph and equation of the normal distribution. One of the nice properties of the normal distribution is that 68% of events that occur are captured by an interval of 1 standard deviation (SD, σ) around arithmetic mean (μ), 95% by an interval of 2 SD and 99.7 by an interval of 3 SD.

of Bayes theorem and Bernoulli's maximum likelihood principle he showed that fitting the line with a normal distribution around it, is actually done by minimizing the sum of squared differences between the line or a curve and data points [12, 15]. Thus, if one assumes that data is scattered equally around the deterministic part of a model in shape that is best described by the normal distribution, the most likely values of parameters in such model are reached by minimizing the sum of squared differences between deterministic part of a model and individual data points. Today, the procedure of fitting a line or a curve by minimizing sum of squared differences is called the least squares estimation and is a specific case of maximum likelihood estimation procedure [4].

Problems of estimation

The very first discovery of what we call today the normal curve was done by Abraham de Moivre in 1733 [16]. De Moivre was studying different problems from, then popular games of chances. One of such problems is best illustrated by the question: what is the probability of observing 10 or more heads in 30 coin tosses. Or to put it in a more general or modern language, what is the probability of certain numbers of outcomes in fixed number of trials, given that outcomes have fixed probability of appearing in each trial. De Moivre elegantly addressed the problem by applying the binomial distribution, however, calculating those probabilities for large number of trials based on binomial distribution proved to be very tedious [17]. In search of simplification, he discovered that as number of trials increases the binomial distribution approximates and becomes almost indistinguishable from a bell shaped curve that we know today as the normal curve (**Figure 6**) [17].
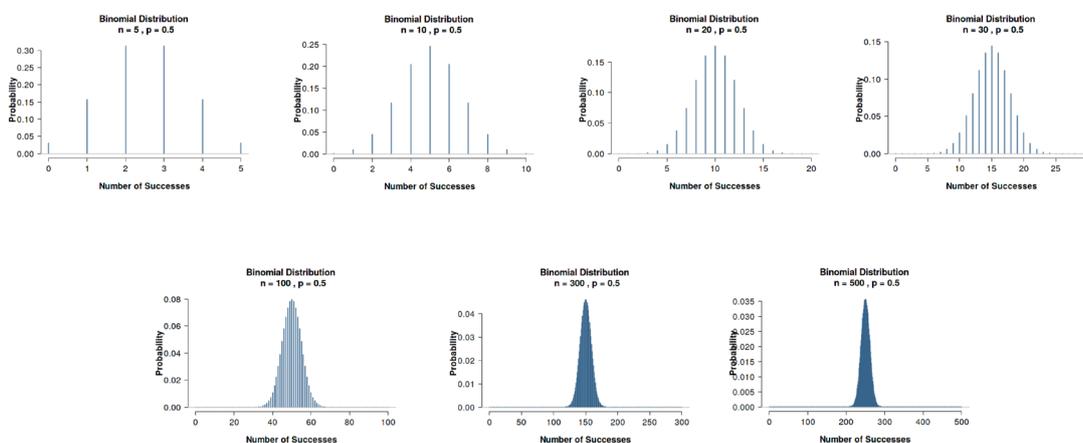

Figure 6. An illustration of convergence of binomial distribution to normal distribution as number of trials (n) increases.

Shortly after Gauss, Laplace published his derivation of the normal curve, which will have far more reaching consequences on quantitative science as a whole [2]. Laplace was interested in studying probability distributions of sums and averages of great number of random variables. Most often these variables were astronomical data; however he was also concerned with daily barometric pressures and actuary data [18]. One of specific problems on which he worked was studying the distribution of deviations or differences between arithmetic mean of data and true value that he was trying to estimate from mul-

tiple series of measurement [18]. In today's language we would say that he was studying behavior of averages and sums of variables by repeated sampling from population. Again, when the problem is translated in mathematical framework, Laplace was able to derive the normal distribution equation [18]. In other words, arithmetic means of multiple series of measurements will be approximately normally distributed around population arithmetic mean as a number of the series of observations tends to infinity [18]. The latter fact is independent of actual distribution of estimated variable, and it was rigorously proven in first half of 20th century and named central limit theorem (CLT) [18]. More intuitively and with less technicalities the central limit theorem might be explained in the fallowing way (**Figure 7**). If a certain variable or a property (e.g., height) in population is considered, it can be described with some kind of probability distribution curve (**Figure 7 a-c**). Furthermore, if the goal of scientific inquiry is to estimate population arithmetic mean by random sampling, then the fallowing can be said. By random sampling or just by virtue of representative sampling the arithmetic mean of drawn samples is likely to be the same as or near to population mean (**Figure 7 d-f**). Samples which have means characterized by large deviations from population mean are more unlikely to be drawn from population then the samples with smaller deviations. Finally, again by the virtue of random (representative) sampling the arithmetic means of samples are equally likely to deviate in both directions from the population arithmetic mean. These latter three facts, that are derived from an assumption of random (representative) sampling, are actually identical to those from which Gauss derived the normal curve. Therefore, distribution of arithmetic means
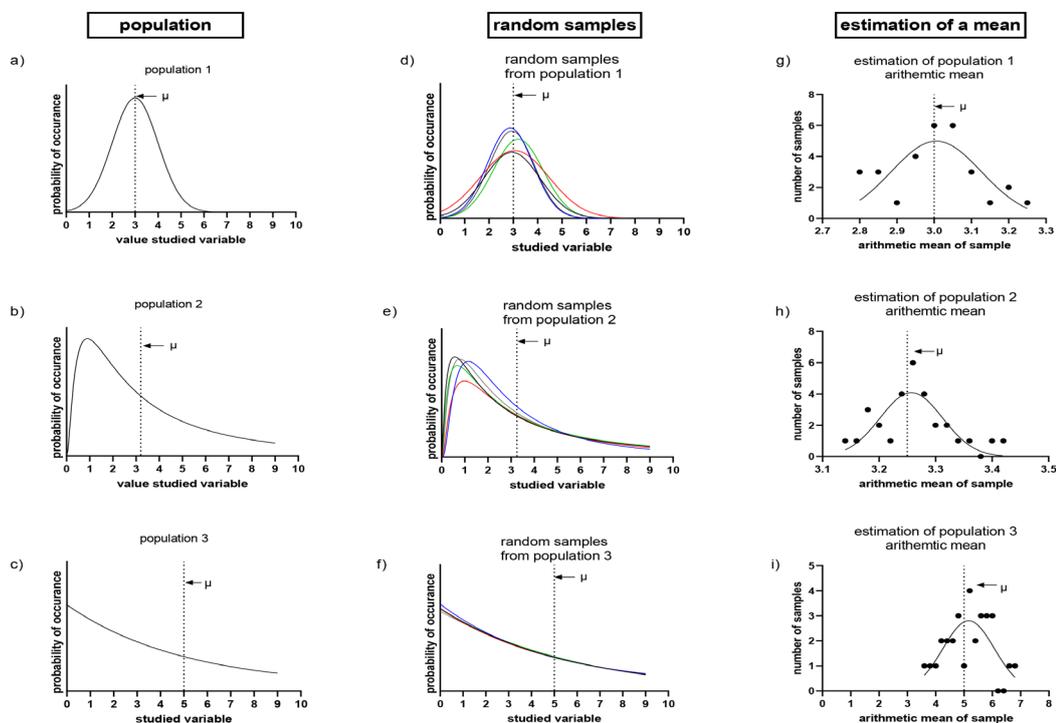

Figure 7. Intuitive explanation of the central limit theorem. Three different populations were simulated, 30 random samples (only 5 are displayed on d, e and f) with size of 10 measurements per sample were drawn from each population. If unbiased and random sampling is done from populations (a-c) the drawn samples (d-f) are likely going to have arithmetic means that are equal or near to arithmetic mean of population (μ). Arithmetic means of samples, also due to random sampling, are more likely to deviate less then more from population mean. Furthermore, again due to random sampling, direction of such deviation from population mean in both directions is equally likely. All of these facts that are derived from properties of random (unbiased or representative) sampling determine that the arithmetic means of samples are normally distributed around population arithmetic mean (μ) (g-i).

of multiple random samples drawn from population is a normal distribution centered around population mean (**Figure 7 g-i**).

The consequence of the discovery was immediately clear to Laplace. He concluded that from the distribution of arithmetic means of samples, a population arithmetic mean can be estimated along with the estimation error around it, without need of knowing the real distribution of measured variable; this formed the first alternative to Bayesian inference [2]. Today, the distribution of arithmetic means of multiple samples (i.e. series of observations), which is normal by CLT, is called sampling distribution, center of such distribution which is again arithmetic mean calculated from multiple arithmetic means of the samples is called point estimate of population arithmetic mean; dispersion around the arithmetic mean of sampling distribution is called interval estimate or confidence interval [2]. Later, Laplace was also able to derive a version of central limit theorem which applies to medians [19].

Asymptotics

Laplace's discovery is foundation of two large and significant disciplines in statistics, one of them is called frequentists estimation and another one asymptotics [2]. The latter studies the behavior of different estimating techniques or sampling distributions of their estimates as sample size tends to infinity i.e. when the observation is repeated under same conditions ad infinitum without ever accounting for whole population. Sampling to infinity appeared as a consequence of mathematical technique used in Laplace's derivation of CLT [18]. Its interpretation in terms of logic was given at the beginning of 20th century by C.D. Broad and H. Jeffreys. They concluded that for scientific law to be true it must withstand the test of infinite validation or replication [20]. Since infinity due to its definition is not reachable, the scientists can never be completely confident in their laws (models). This is basically, a mathematical restatement of denying the consequent (modus tollens) rule of inference, or to put in more simple view, phenomena can be caused by multiple system properties some of which might be unknown to investigator [21]. Therefore, scientist can never claim that the causes of certain phenomena are completely known since there always exists a possibility that some unknown causes (variables) are not included in model [21].

Normal curve as a stochastic model

First one to apply normal distribution as a stochastic model was Adolphe Quetelet in mid-19th century. He modeled a chest girth of 5000 Scottish soldiers with the normal distribution [11]. Other social and biological scientist soon fallowed and many traits such as IQ, height, body mass were modeled with the normal curve, at least to some degree of accuracy [11]. In biomedical field and especially in laboratory medicines a lot of variables from usual blood tests can be modeled with the normal curve [22]. Furthermore, since most of linear and nonlinear regression techniques use normal distribution as stochastic part of model, we can argue that the normal distribution is definitely most used one in science. However, there are some other natural phenomena that are not normally distributed. For instance, lifespans of human beings or machines are often exponentially, or Weibull distributed [23].

Normal distribution and model parameters estimation

Maximum likelihood estimation

Some 150 years after D. Bernoulli stated a principle that estimates of model parameters should be chosen in a such manner that selected values of model parameters maximize the likelihood of data generation by the model, R. A. Fisher published his three influential papers in which he named the principle as maximum likelihood estimation method [14]. In his 1922 seminal paper, which is considered by many as establishment of mathematical statistics as we know it today, he defined four key properties of estimators [24] (**Table 1**). Consistent estimator, as Fisher defined it, is the one which has its sampling distribution centered on true value of model parameter, as sample size tends to infinity [25]. Efficient estimator was defined by Fisher as the estimator which asymptotically has a normal probability distribution with the smallest possible standard deviation [25]. Modern definition of efficiency is somewhat different, and it states that the efficient estimator has the least possible variance of a sampling distribution [26]. Sufficient estimator is the one which gives as much information about the estimated parameter as possible for a given sample [25]. Fisher was able to prove, although with deficiencies, that maximal likelihood estimates are efficient, asymptotically normally distributed (i.e. their sampling distribution approximates the normal distribution as sample size tends to infinity), consistent, and sufficient [14].

Table 1. Properties of estimators

Property of an estimator	Definition
Consistency	Sampling distribution of estimator is centered on true value of parameter that is being estimated
Efficiency	The sampling distribution of parameter has the smallest possible variance
Sufficiency	Sampling distribution contains as much as possible information about the parameter being estimated. That is there is no other estimating technique that would provide more information on parameter.
Regularity	Regular estimator is the one which has a stable variance of sampling distribution i.e. the variance does not behave erratically in the neighborhood of the center of sampling distribution.

When considering likelihood as a technical term it should be noted that although closely related to probability it is not the probability since it does not satisfy properties of probability (e.g. unlike probability, total sum of likelihood does not sum to 1) [27]. Therefore, when speaking of probability distribution of maximum likelihood function, it merely means that shape of maximum likelihood function can be described by formula which is identical to that of certain probability distribution. However, by using the Bayes theorem, likelihood can easily be converted to proper probability [23]. If the problem of asymptotic distribution of parameter estimate is put in a Bayesian framework, a very influential result know as Bernstein-von Mises theorem can be reached. The theorem states that under certain regularity conditions such as smooth, well specified parametric model with well-behaved prior distribution of parameters, the posterior distribution of parameters estimates asymptotically converges to normal distribution centered on maximum likelihood estimate of parameters [28].

Convolution theorems

By the middle 20th century all of Fishers claims were more rigorously proven by Wald, Cramer, Wolfowitz and others [14]. In their proofs they had to resort to assumptions which are now known as *usual regularity conditions* [14]. Even in the 1960s and 1970s some gaps in evidence for such optimal properties of maximum likelihood existed. Work by Kaufman finally resolved the problem of asymptotic normality of maximum likelihood estimate and its optimality [29]. In doing so Kaufman, with a contribution of Inagaki, discovered a very general and far reaching concept in theory of estimation known as Convolution Theorem [29]. Recently, Geyer was able to conceptually simplify conditions for asymptotic normality of maximum likelihood estimates and show that asymptotic normality stems from mathematical properties of likelihood function itself. In doing this he did not have to use asymptotics (i.e. CLT) [30].

Maximum likelihood is only one of many estimation methods. Theoretically an estimator can be almost any mathematical function that maps from a sample space (i.e. set of all possible outcomes of experiment) to a set of parameter estimates. Given this vast diversity of possible estimators, their properties are what qualifies them as desirable or not [4]. As mentioned above, the introduction of estimator properties are in large part due to Fisher's work. Building on Kaufman's work Hajek and later Le Cam were able to prove that asymptotic or limiting sampling distribution of any regular estimator in parametric model is a sum of two independent probability distributions [29]. One of those distributions is normal with the least possible variance and the other one is of arbitrary form. Today this result is known as *Hajek-Le Cam Convolution Theorem* [29]. The consequence of such theorem is that the efficient estimator is the one for which arbitrary component equals zero i.e. the one which is asymptotically normally distributed.

Epistemological considerations and the normal distribution

Truth, knowledge and decision making

The purpose of all statistical analysis ultimately is to inform process of decision making [31]. For example, in science of medicine therapeutic decision are informed by statistical models of drug effectiveness, in business, stock market models help to inform subjects which shares to buy or sell. If models are describing the truth about the system that is being studied, then by necessity all of the decisions are going to be well informed. The truth itself is defined by some philosophers as everything that an entity, or a system in our case, is [32]. In natural sciences prior to a study, experimentation, or observation the truth about any system is unknown. The main reason for this is that man is not a creator of those systems, moreover, even in a system created by the men (e.g an airplane or a building) the elements of system (e.g atoms) and laws governing interaction of those elements are not created by men. Therefore, in order to gain insight into a truth (i.e to obtain knowledge) about any system, in natural science, investigators have to study previous knowledge, observe and experiment with a system. Statements derived from the latter three are usually united in a theory or mathematical model that pertains to be the truth about the sys-

tem. However, the validity of such claims must be gauged by the set of standards or rules known in epistemology as the tests or criteria of truth.

The most adequate criterion of truth is coherence [33]. It states that theory or a model that gives consistent and overarching explanation for all observations is most likely to be true. Furthermore, another criterion of truth that is built in definition of coherence is consistency. Consistency can be understood in two ways: as mere consistency, which is the same as the principle of non-contradiction, or as a strict consistency which states that all claims in theory should logically proceed one from another [33]. Mathematics would be good example of the latter, on the other hand most natural sciences are not consistent in a strict sense.

Statistical measures of explanation are logical extensions of coherence criterion into quantitative realm. These, so called, goodness of fit measures quantify which of many alternative models (theories) preforms the best in explaining the observations, that is in reducing data scatter (variability) around deterministic part of model [34, 35]. The most widely used such measures are deviance, likelihood ratios and various informational criterions (e.g AIC, BIC). All of the latter measures are calculated from estimators used in model parameters estimation procedures [34, 35]. Thus, it is tempting to speculate that goodness of fit measures also fallow the normal distribution in asymptotic case, since as it has been shown in previous section, the estimators, which are the building blocks of their calculation, do. However, this is not the case because in those calculations estimators are mathematically manipulated (e.g., logged or divided) and therefore their distributions are changed from the normal distribution to some other more or less know distributions [34]. Based of goodness of fit criteria and some other graphical tools that basically also deal with goodness of fit, a scientist can decide which model explains the data best and therefore which are likely to be true [4, 34]. Furthermore, if more than one model has similar explanatory power, a scientist can resort to model averaging techniques given that those models are not mutually exclusive [36]. In a decision-making process once the question of most likely true model is settled, the next question to consider is the question of effects of different variables governing the system (i.e. model parameters) [31]. These are of peculiar interest when it comes to application of science since manipulating these variables changes the states of studied system. Sizes of these effects are normally distributed, as it was stated in a previous section.

After all that has been said in this section, it can be stated that both measures of goodness of fit and measures of effects of model's variables are quantities that have their distributions i.e. they can be defined by mathematical functions knowns as probability distributions. Thus, it can be concluded that epistemological insight (i.e. acquired knowledge) has a mathematical structure. That structure is the normal distribution for measures of effects of variables in model (i.e., model parameters), on the other hand when it comes to measures of explanatory power of model (i.e. goodness of fit measures) the mathematical structure is not normal distribution, but its calculation is based on estimators that are normally distributed.

The structure of acquired knowledge and its implications

Since acquired knowledge has mathematical structure of a certain probability distribution that implies certain things can be said about it from mathematical perspective.

Estimates of model parameters have normal distribution, this means that qualitative traits that describe the normal distribution can be applied in description of our knowledge of model parameters (i.e. the effects of variables in model). So, it can be said that such estimates tend to aggregate around average estimate, their deviation form average estimate is symmetrical and of equal magnitude. Furthermore, from quantitative sides it can be stated which estimates of magnitudes or sizes of effect are more likely than others (**Figure 5**]. Similar descriptions can be done based on properties of distributions of goodness of fit measures. However, the most striking property of probability, by definition, is its randomness i.e., uncertainty of actual outcome. This points to the fact that we cannot be deterministically certain in our predictions of outcomes, that is our knowledge is in the realm of uncertainty or belief. The latter term would be preferred by Bayesians [37]. Thus, increasing the variances (i.e. scale measures) of those distributions translates in to more uncertainty in one's estimates that is in one's gained knowledge.

On the other hand, one can argue that if the system is described with a model that has a negligible data scatter around deterministic part and interval estimates (e.g., 95% CIs) of effects of variables that would be infinitesimally close to the point estimates, then a deterministic prediction could be done. Indeed, such models exist, manly in sciences that deal with systems that are composed of few entities and few interactions [38].

Even if such deterministic ideal is reached and mathematical structure of knowledge that describes uncertainty (i.e., probability distributions) is reduced to negligibility, the notion of uncertainty can not be eliminated. This mainly has to do with philosophical definition of truth and logic. Since, man is not a creator of any system in natural sciences, one can never know with certainty if his knowledge of system is equal to the truth of the system (i.e., everything that systems is). Thus, if knowledge is defined as a subset of truth, one can never be certain if that subset is equal to a set (i.e., truth) or if there is more left to be observed and studied. In a more applied language this means that some unknown variables governing the system, or their magnitudes have not been present in an instance or a period of time when data samples, which were used to infer about validity of models, were collected. Consequentially their effects are not captured either by deterministic or stochastic part of model. This ultimately leads to a problem of a definition of a studied system and experimental design and control. Namely, a model that had a very good properties in one instance can fail in another instance because the two instances only appear to be identical, whereas they actually differ due to presence of unknown variables that had different or negligible effect on system in a previous instance. The latter, again points to the requirement of infinite replication (asymptotics) if the model is to be considered true with certainty, which is just a another manifestation of a fact that scientist is not a creator of any natural system and thus he cannot know if the truth in its fullness has been reached by ones experimentation, observation and analysis.

Conclusion

In conclusion it can be summarized that scientific knowledge in natural sciences must always be considered within a realm of uncertainty and consequentially when it comes to decision-making within a realm of belief. This knowledge and associated uncertainty or a belief, as Bayesians would say, can be quantified as probability distributions of effects of model variables (i.e., model parameters) and goodness of fit (i.e. explanatory power) measures. The former fallow the normal distribution, whereas the latter are derived from the normal distribution. However, this mathematical structure does not describe episte-mological uncertainty in its fullness. To accomplish this, a problem of ignorance towards unknown variables in definition of system and consequentially, experimental design has to be added to the mathematical structure.

ORCID

Benjamin Benzon ⓘ https://orcid.org/0000-0003-3681-8499

References

1. Poincaré H. Calcul des probabilités Paris: Gauthier-Villars; 1912. 333 p.

2. Wilcox RR. Fundamentals of Modern Statistical Methods. 2nd ed. Heidelberg: Springer; 2010. 249 p.

3. National Research Council (US). Introduction. Assessing the Reliability of Complex Models: Mathematical and Statistical Foundations of Verification, Validation, and Uncertainty Quantification. Washington, D.C.: The National Academies Press; 2012. p. 7-18.

4. Guthrie W, Filliben J, Hackert A. Process Modeling. 2012 [cited 3/11/2020]. In: NIST/SEMATECH e-Handbook of Statistical Methods [Internet]. National Institute of Standards and Tehchnology (US), [cited 3/11/2020]. Available from: https://itl.nist.gov/div898/handbook/pmd/section1/pmd11.htm.

5. Reid N. A Conversation with Sir David Cox. Statistical Science. 1994;9(3):439-55.

6. Breiman L. Statistical Modeling: The Two Cultures (with comments and a rejoinder by the author). Statist Sci. 2001;16(3):199-231.

7. Bol'shev LN. Statistical estimator. [cited 17.12.2020]. In: Encyclopeadia of Mathematics [Internet]. EMS press, [cited 17.12.2020]. Available from: https://encyclopediaofmath.org/index.php?title=Statistical_estimator&oldid=13861.

8. Anderson DR. Appendix A: Likelihood theory. Model Based Inference in the Life Sciences. 1 ed. New York: Springer-Verlag 2008. p. 147-54.

9. Kass RE. Statistical Inference: The Big Picture. Stat Sci. 2011;26(1):1-9.

10. Stigler MS. Least Squares and Combinations of Observations. The history of statistics : the measurement of uncertainty before 1900. Cambridge (MA, US): Belknap Press of Harvard University Press; 1986. p. 11-55.

11. Stahl S. The evolution of the normal distribution. Mathematics Magazine. 2006;79(2):96-113.

12. Eisenhart C. The Meaning of "Least" In Least Squares. Journal of the Washington Academy of Sciences. 1964;54(2):24-33.

13. Frankovic B. Ruđer Bošković (1711-1787) Scientist and Humanist. Strojarstvo. 2011;53(3):151.

14. Stigler SM. The Epic Story of Maximum Likelihood. Statist Sci. 2007;22(4):598-620.

15. Sprott DA. Gauss's contributions to statistics. Historia Mathematica. 1978;5(2):183-203.

16. Daw RH. Studies in the History of Probability and Statistics. XXX. Abraham De Moivre's 1733 derivation of the normal curve: A bibliographical note. Biometrika. 1972;59(3):677-80.

17. Bellhouse DR, Genest C. Maty's Biography of Abraham De Moivre, Translated, Annotated and Augmented. Statistical Science. 2007;22(1):109-36.

18. Fischer H. A History of the Central Limit Theorem From Classical to Modern Probability Theory. Heidelberg: Springer; 2010. 402 p.

19. Stigler SM. Studies in the History of Probability and Statistics. XXXII: Laplace, Fisher and the Discovery of the Concept of Sufficiency. Biometrika. 1973;60(3):439-45.

20. Senn S. Of Dice and Men. Dicing with Death Chance, Risk and Health. London: Cambridge University Press; 2003. p. 69-90.

21. Gruner R. Plurality of Causes. Philosophy. 1967;42(162):367-74.

22. Whyte MB, Kelly P. The normal range: it is not normal and it is not a range. Postgrad Med J. 2018;94(1117):613-6.

23. Tobias P. Assessing Product Reliability. 2012. In: NIST/SEMATECH e-Handbook of Statistical Methods [Internet]. National Institute of Standards and Tehchnology (US). Available from: https://www.itl.nist.gov/div898/handbook/apr/section1/apr1a.htm.

24. Hand DJ. From evidence to understanding: a commentary on Fisher (1922) 'On the mathematical foundations of theoretical statistics'. Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences. 2015;373(2039):20140252.

25. Fisher RA, Russell EJ. On the mathematical foundations of theoretical statistics. Philosophical Transactions of the Royal Society of London Series A, Containing Papers of a Mathematical or Physical Character. 1922;222(594-604):309-68.

26. Dekking FM, Kraaikamp C, Lopuhaä HP, Meester LE. Efficiency and mean squared error. A Modern Introduction to Probability and Statistics. London: Springer-Verlag; 2005.

27. Aldrich J. R.A. Fisher and the making of maximum likelihood 1912-1922. Statist Sci. 1997;12(3):162-76.

28. Prakasa-Rao BLS. Bernstein-von Mises theorem. [cited 3/11/2020]. In: Encyclopedia of Mathematics [Internet]. EMS Press, [cited 3/11/2020]. Available from: https://encyclopediaofmath.org/wiki/Bernstein-von_Mises_theorem#References.

29. Pfanzagl J. Mathematical Statistics Essays on History and Methodology. 1st ed. Berlin Heidelberg: Springer-Verlag; 2017. 316 p.

30. Geyer CJ. Asymptotics of Maximum Likelihood without the LLN or CLT or Sample Size Going to Infinity. Advances in Modern Statistical Theory and Applications: A Festschrift in honor of Morris L Eaton. Collections. Volume 10. Beachwood, Ohio, USA: Institute of Mathematical Statistics; 2013. p. 1-24.

31. Chernoff H, Moses LE. Elementary Decision Theory. New York: Dover publications; 1986.

32. Walker L. Truth. The Catholic Encyclopedia. New York: Robert Appleton Company; 1912.

33. Sahakian WS, Sahakian ML. Criteria of truth. Ideas of the great philosophers. New York: Barnes & Noble; 1993. p. 10-1.

34. Anderson DR. Model Based Inference in the Life Sciences. 1 ed. New York: Springer-Verlag 2008. 184 p.

35. Rice JR, White JS. Norms for Smoothing and Estimation. SIAM Review. 1964;6(3):243-56.

36. Anderson DR. Multimodel Inference. Model Based Inference in the Life Sciences. 1 ed. New York: Springer-Verlag 2008. p. 105-24.

37. Higham DJ. Bayesian Inference in Applied Mathematics. In: Higham NJ, editor. The Princeton Companinon to Applied Mathematics: Princeton University Press; 2016. p. 685-61.

38. Glendinning P. Complex systems. In: Higham NJ, editor. The Princeton Companinon to Applied Mathematics: Princeton University Press; 2016. p. 83-4.

RESEARCH ARTICLE

ST open