

basis the data can then be used to select d so as to obtain a desirable balance between these two aspects. Criteria for selecting d (and making analogous choices within a family of proposed models in other problems) have been developed by Akaike, Mallows, Schwarz, and others (for more details, see, for example, Linhart and Zucchini 1986):

Bayesian solutions to this problem differ in a number of ways, reflecting whether we assume that the true model belongs to the hypothesized class of models (e.g., is really a polynomial) or can merely be approximated arbitrarily closely by such models. For more on this topic see Shao (1997).

See also: Estimation: Point and Interval; Robustness in Statistics.

Bibliography

- Bickel P, Doksum K 2000 *Mathematical Statistics*, 2nd edn. Prentice Hall, New York, Vol. 1
- Blackwell D, Dubins L 1962 Merging of opinions with increasing information. *Annals of Mathematical Statistics* 33: 882–6
- Ferguson T S 1996 *A Course in Large Sample Theory*. Chapman & Hall, London
- Fisher R A 1922 On the mathematical foundations of theoretical statistics. *Philosophical Transactions of the Royal Society of London, Series A* 222: 309–68
- Fisher R A 1973 *Statistical Methods and Scientific Inference*, 3rd edn. Hafner Press, New York
- Gauss C F 1809 *Theoria Motus Corporum Celestium*. Perthes, Hamburg, Germany
- Girshick M A, Savage L J 1951 *Bayes and Minimax Estimates for Quadratic Loss Functions. Proceedings of the 2nd Berkeley Symposium of Mathematics, Statistics, and Probability*. University of California Press, Berkeley, CA
- Hampel F, Ronchetti E, Rousseeuw P, Stahel W 1986 *Robust Statistics*. Wiley, New York
- Jeffreys H 1939 *Theory of Probability*. Clarendon Press, Oxford, UK
- Keynes J M 1921 *A Treatise on Probability*. Macmillan, London
- LeCam L 1990 Maximum likelihood: An introduction. *International Statistical Review* 58: 153–71
- Lehmann E L 1986 *Testing Statistical Hypotheses*, 2nd edn. Springer, New York
- Lehmann E L, Casella G 1998 *Theory of Point Estimation*, 2nd edn. Springer, New York
- Linhart H, Zucchini W 1986 *Model Selection*. Wiley, New York
- Neyman J 1937 Outline of a theory of statistical estimation based on the classical theory of probability. *Philosophical Transactions of the Royal Society Series A* 236: 333–80
- Neyman J 1961 Silver Jubilee of my dispute with Fisher. *Journal of the Operations Research Society of Japan* 3: 145–54
- Neyman J, Pearson E S 1933 On the problem of the most efficient tests of statistical hypothesis. *Philosophical Transactions of the Royal Society Series A* 231: 289–337
- Shao J 1997 An asymptotic theory for linear model selection (with discussion). *Statistica Sinica* 7: 221–66
- Staudte R G, Sheather S J 1990 *Robust Estimation and Testing*. Wiley, New York
- Stein C 1956 Inadmissibility of the usual estimator for the mean of a multivariate distribution. *Proceedings of the 3rd Berkeley*

- Symposium of Mathematics, Statistics, and Probability*. University of California Press, Vol. 1, pp. 187–95
- Tukey J W 1960 Conclusions versus decisions. *Technology* 2: 423–33
- von Mises R 1928 *Wahrscheinlichkeit, Statistik and Wahrheit*. Springer, Wien, Austria
- Wald A 1950 *Statistical Decision Functions*. Wiley, New York
- Wilson E B 1952 *An Introduction to Scientific Research*. McGraw Hill, New York

P. J. Bickel and E. L. Lehmann

Frequentist Interpretation of Probability

If the outcome of an event is observed in a large number of independent repetitions of the event under roughly the same conditions, then it is a fact of the real world that the frequency of the outcome stabilizes as the number of repetitions increases. If another long sequence of such events is observed, the frequency of the outcome will typically be approximately the same as it was in the first sequence.

Unfortunately, the real world is not very tidy. For this reason it was necessary in the above statement to insert several weasel words. The use of ‘roughly the same,’ ‘typically,’ ‘approximately,’ and ‘long sequence’ make it clear that the stability phenomenon being described cannot be stated very precisely. A much clearer statement is possible within a mathematical model of this phenomenon. This discovery is due to Jacob Bernoulli who raised the following question.

It is well known, Bernoulli says in Part IV of his book *Ars Conjectandi* (published posthumously in 1713), that the degree to which the frequency of an observed event varies about the probability of the event decreases as the number of events increases. He goes on to say that an important question that has never been asked before concerns the behavior of this variability as the number n of events increases indefinitely. He envisages two possibilities.

(a) As n gets larger and larger, the variability eventually shrinks to zero, so that for sufficiently large n the frequency will essentially pinpoint the probability p of the outcome.

(b) Alternatively, it is conceivable that there is a positive lower bound below which the variability can never fall so that p will always be surrounded by a cloud of uncertainty, no matter how large a number of events we observe.

Bernoulli then proceeds to prove the law of large numbers, which shows that it is (a) rather than (b) that pertains. More precisely, he proves that for any $a > 0$

$$\Pr \left(\left| \frac{X}{n} - p \right| < a \right) \rightarrow 1 \quad \text{as } n \rightarrow \infty \quad (1)$$

where X/n is the frequency under consideration.

It is easy to be misled into the belief that this theorem proves something about the behavior of frequencies in the real world. It does not. The result is only concerned with properties of the mathematical model. What it does show is that the behavior of frequencies in the model is mirrored in a way that is much neater and more precise, the very imprecise stability phenomenon stated in the first paragraph of this article.

In fact, a result for the model much more precise than (1) was obtained by De Moivre (1733). It gives the normal approximation

$$\Pr \left(\left| \frac{X}{n} - p \right| < c/\sqrt{n} \right) \rightarrow \int_{-c/\sqrt{pq}}^{c/\sqrt{pq}} \varphi(x) dx \quad (2)$$

where φ denotes the standard normal density. This is again a theorem in the model. The approximate corresponding real-world phenomenon can be seen, for example, by observing a quincunx, a mechanical device using balls falling through 'random' paths to generate a histogram.

De Moivre's result was given a far reaching generalization by Laplace (1810) in the central limit theorem (CLT) concerning the behavior of the average \bar{X} of n identically, independently distributed random variable X_1, \dots, X_n with mean ξ and finite variance σ^2 . It shows that

$$\Pr \left(|\bar{X} - \xi| < c/\sqrt{n} \right) \rightarrow \int_{-c/\sigma}^{c/\sigma} \varphi(x) dx. \quad (3)$$

This reduces to (2) when X takes on the values of 1 and 0 with probabilities p and q , respectively. The CLT formed the basis of most frequentist inference throughout the nineteenth century.

The first systematic discussion of the frequentist approach was given by Venn (1866), and an axiomatization based on frequencies in infinite random sequences (*Kollektives*) was attempted by von Mises (1928). Because of technical difficulties his concept of a random sequence was modified by Solomonoff (1964), Martin-Löf (1966), and Kolmogorov (1968), with the introduction of computational complexity. (An entirely different axiomatization based on events and their probabilities rather than random sequences was put forward by Kolmogorov in 1933, and has successfully served as a basis for both frequentist and subjective interpretations of probability). For more details on these different approaches see Barnett (1982).

The frequentist concept of probability described so far has met considerable criticism. One of the main objections is that it is not applicable to many situations to which one might want to apply probability assessments. To see this, consider the following three possibilities.

(a) An actual sequence of repetitions may be available; for example, a sequence of coin tosses or a sequence of independent measurements of the same quantity.

(b) A sequence of repetitions may be available in principle but not likely to be carried out in practice; for example, the polio experiment of 1954 involving a sample of over a million children.

(c) A unique event which by its very nature can never be replicated, such as the outcome of a particular historical event; for example, whether a particular president will survive an impeachment trial. The conditions of this experiment cannot be duplicated.

The frequentist concept of probability can be applied in cases (a) and (b) but not in the third situation. An alternative approach to probability which is applicable in all cases is the notion of probability as degree of belief; i.e., of a state of mind (for a discussion of this approach, see Robert (1994)). The inference methods based on these two interpretations of the meaning of probability are called frequentist and Bayesian, respectively.

Although frequentist probability is considered objective, it has the following subjective feature. Its impact on a particular person will differ from one person to another. One patient facing a surgical procedure with a 1 percent mortality rate will consider this a dire prospect and emphasize the possibility of a fatal outcome. Another will shrug it off as so rare as not to be worth worrying about.

There exists a class of situations in which both approaches will lead to the same probability assessment. Suppose there is complete symmetry between the various outcomes; for example, in random sampling which is performed so that the drawing favors no sample over another. Then we expect the frequencies of the various outcomes to be roughly the same and will also, in our beliefs, assign the same probability to each of them.

Let us now turn to a second criticism of frequentist probability. This concerns the difficulty of specifying what is meant by a repetition in the first sentence of this section. Consider once more the surgical procedure with 1 percent fatalities. This figure may represent the experience of thousands of cases, with the operation performed by different surgeons in different hospitals and—of course—on different patients. The rate of fatalities may vary from one hospital or surgeon to another and may, in particular, vary drastically with the condition, for example, the age and general health, of the patient.

Suppose a young woman requires this operation although her general health is very good. The frequency of a fatal outcome with patients sharing these characteristics may be much lower, and the 1 percent figure in that sense would be quite misleading for her. And yet she might be considered to have been obtained under 'roughly the same conditions,' namely to be drawn at random from the total population of persons

requiring this surgery. To obtain the most useful figures one should identify the most important variables, classify the cases accordingly (for example, young, middle-aged, old; male, female; etc.) and then provide the frequency for each class. They will, of course, be meaningful only for the classes which contain a reasonable number of cases.

1. A Terminology Note

A source of much confusion in the discussion of probability is the fact that 'probability' is used both as a mathematical term, i.e., as a concept in the mathematical model and also in everyday language when talking about real events. When reading about probability, it is important to be aware of these two meanings and to keep them distinct.

See also: Frequentist Inference; Probability: Formal; Probability: Interpretations; Statistical Methods, History of: Post-1900; Statistical Methods, History of: Pre-1900

Bibliography

- Barnett V 1982 *Comparative Statistical Inference*, 2nd edn. Wiley, New York
- De Moivre A 1733 1738, 1756. In: *Doctrine of Chance*, Millar, London, Reprinted by Chelsea Publishers, New York (1967), pp. 235–43 (1738), 243–54 (1754) 1878–1912
- Kolmogorov A 1933 *Grundbegriffe der Wahrscheinlichkeitsrechnung*. Ergebnisse der Mathematik, Vol. 2. Springer, Berlin
- Kolmogorov A 1968 Logical basis for information theory and probability theory. *IEEE Transactions on Information Theory* 14: 662–4
- Laplace P S *Oeuvres complètes de Laplace*. Gauthier-Villars, Paris
- Martin-Löf P 1966 The definition of random sequences. *Information and Control* 9: 602–19
- Robert C P 1994 *The Bayesian Choice*. Springer, New York
- Solomonoff R J 1964 The formal theory of inductive inference. *Information and Control* 7: 224–54
- Venn J 1866 *The Logic of Change*. Reprinted by Chelsea Publishers, New York, (1967)
- von Mises R 1957 *Probability, Statistics and Truth*. Macmillan, New York

P. J. Bickel and E. L. Lehmann

Freud, Sigmund (1856–1939)

1. Family Background and Early Life

Sigmund Freud, neuroscientist turned founder of psychoanalysis, was born on May 6, 1856, in Freiberg, Moravia, now part of the Czech Republic. His father,

Jakob (1815–96), a wool merchant, came from Galicia, and married Amalie (1835–1930, neé Nathansohn), 20 years younger, when he was 40. Jakob had two sons by a previous marriage, Emanuel and Philipp. Jakob was warm and affectionate, with a strong sense of humor, and Freud evidently admired and respected him. Amalie was a lively and attractive woman who was proud of Sigmund, her first-born.

The family multiplied quickly: five daughters and two more sons were born within 10 years. The first of these, Julius, died when 8 months old and Sigmund was aged 19 months. Both parents were Jews, with no religious affiliations, though a devout Catholic Nanny took Sigmund to Church. She left when he was two and a half. A nephew, John, a year older than Freud was very close to him: the two fought and played together. That childhood ambivalence, Freud reflected years later, must have had a profound effect on his character and on his ability to defend himself (Freud 1900).

Financially, times were hard. In 1859 the family moved to Leipzig, and in the following year to Vienna. Freud missed the beauty of the countryside round Freiberg. Although the children were well looked after, poverty was not assuaged by rapid family growth. By 1875 family fortunes improved: Amelie's family were helping, making possible a move to a larger house. Freud had a room of his own that served as a study.

Freud's memories of his childhood as revealed in his writings are without early detail, but one unforgettable recollection was of urinating in his parents' bedroom while they were present. Reprimanding him, his father said bluntly: 'The boy will come to nothing!' Freud's wounded ambitions were reflected recurrently in dreams in which he listed his achievements as if to say: 'You see! I have come to something!' (Freud 1900; see also Shengold 1993).

2. Education and Early Interests

Freud's schooling began with his mother, until his father took over before sending him to a private school. Freud learned rapidly, and was reading Shakespeare (whom he loved all his life) from the age of eight years. When nine years of age, he won a place at the Sperl Gymnasium, where he was a distinguished pupil, passing out, at 17, with the distinction '*summa cum laude*.' He became an accomplished linguist, with a firm foundation in Latin and Greek, a sound knowledge of Hebrew, fluency in English and French, and good Italian. He corresponded with an adolescent friend in Spanish. He became one of the greatest stylists in the German language (Kaufmann 1980). He studied at every opportunity. He loved the Arts, but took an early dislike to music, which he never felt able to appreciate. He was fond of walking, swimming, and skating. He liked travel, and made his first visit to England at the age of 19.