Comment

Persi Diaconis and Eduardo Engel

I. J. Good has given us a marvelous historical perspective, blending the subjective/objective dichotomy with the mathematics of Poisson's summation formula. These subjects are intimately related to recent work on probability in classical physics. We first develop this connection, and then compare Poisson's techniques with Poincaré's method of arbitrary functions. We conclude by outlining generalizations connected to Selberg's trace formula.

1. A SUBJECTIVE GUIDE TO OBJECTIVE CHANCE

Physics and mathematics offer a useful way of relating subjective probability to objective chance devices. Consider throwing a real dart at a real wall. If the left half of the wall is painted black, and the right half painted white, there is nothing very random about the outcome: by aiming a bit to the left, the dart winds up in the black section.

Now suppose the paint is rearranged to form stripes which are alternately painted black and white. If the distance between the stripes is large, things still aren't random, but as the distance gets smaller, black and white will be judged nearly equally likely by almost anyone.

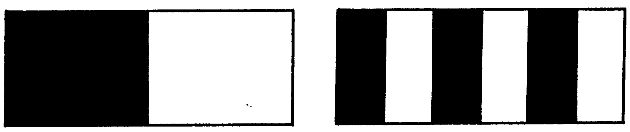

FIG. 1

It is not difficult to give quite sharp quantitative estimates: Suppose the thrower stands a distance l from the wall, and the stripes have width d. Clearly only the ratio l/d matters, so without loss of generality, take $l = 1$. The situation is pictured in Figure 2.

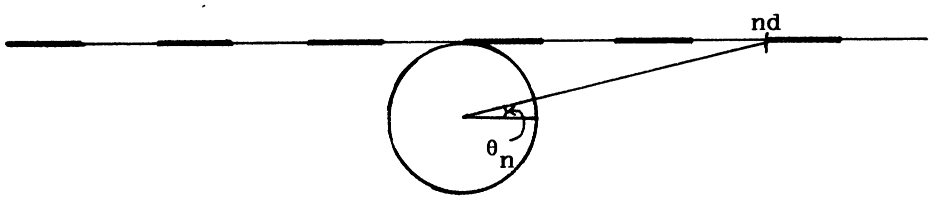

FIG. 2

Let θ be the angle of release of the dart and suppose $f(\theta)$ is a probability density on $(-\pi/2, \pi/2)$. If θ is chosen from f,

$$P\{\text{Black}\} = \sum_{n=-\infty}^{+\infty} \int_{\theta_{2n}}^{\theta_{2n+1}} f(\theta)\, d\theta \tag{1.1}$$

with

$$\theta_n = \tan^{-1}\left(\frac{1}{nd}\right).$$

A result due to Kemperman (1975) (discussed in Section 2) and straightforward calculus lead to the bound

$$|P\{\text{Black}\} - \tfrac{1}{2}| \le cd$$

with

$$c = \frac{V(f) + 1}{4} \quad \text{and} \quad V(f) = \int_{-\pi/2}^{\pi/2} |f'(\theta)|\, d\theta.$$

Now let us consider the philosophical implications of the mathematics. If $f(\theta)$ is Smith's subjective distribution of the angle of release, and if $f(\theta)$ is not too sharply peaked, then Smith is forced to assign probability about ½ to the dart landing in the black region.

This gives an objective chance device in the following sense: Most people will assign the same probability to the outcome black even though they may have very different prior beliefs about θ.

This is quite different from the usual argument for multisubjective agreement ("the data swamp the prior"). It applies to a single event: Agreement is reached without the need of any data.

A similar analysis can be given for many other objective chance devices. Consider flipping a coin in the air and catching it as it lands. When the coin leaves the hand, its upward velocity and angular momentum completely determine which side will land uppermost. It is possible to carry out the physics and

Persi Diaconis is Professor of Statistics and Eduardo Engel is a graduate student in the Department of Statistics, Stanford University, Sequoia Hall, Stanford, California 94305.

show that the partitioning of the velocity-momentum space into regions corresponding to heads and tails become very close together when the initial velocity is sufficiently large. Thus small changes in initial conditions make for the difference between heads and tails.

Quantitative versions of the coin tossing argument have been carried out by Keller (1985), Yue and Zhang (1985), and Vulovic and Prange (1985). We have carried out experiments to measure how fast a coin spins in real tosses. In a toss of about 1 foot, a coin turns over about 15 times. Experiments combined with theory lead to the conclusion that coin tossing is fair to two decimal places but not three.

The analyses above are versions of Poincaré's "method of arbitrary functions." The phenomena considered are such that the conclusions we draw about their outcome are valid for an *arbitrary* density f describing the initial conditions.

Poincaré's method was brilliantly developed by E. Hopf in a series of papers published in the 1930's. Hopf treated problems such as Buffon's needle and the stopping place of a ball rolled on a floor with general frictional force. Hopf began a classification of low order differential equations damped by friction to determine to what extent uncertainty about initial conditions affects the distribution of the resting state. A useful review, containing references, is given by Von Plato (1983). Wolfram (1985) contains further discussion.

Hopf and earlier workers interpreted $f(\theta)$ as a smooth approximation to the empirical distribution of initial conditions. The subjective interpretation is clearly discussed in the charming article by Savage (1973).

We are currently actively involved in working on both special cases and a general theory which we hope to report on soon.

Thinking hard about randomness reveals that many of the usual examples of randomizers are not very good when looked at critically. For a discussion of problems in drawing from an urn, see Fienberg (1971). For a physical analysis of roulette, see Bass (1985).

Finally we note that many simple methods of randomizing cannot be analyzed by "doing the physics." Rolling dice or coins flipped on the floor require physics beyond our power. A proper analysis of even the simplest methods of shuffling cards (e.g., put n cards in a row and randomly switch pairs of cards) requires some insight into the psychology of the shuffler. We are a long way from being able to give a physical analysis. Aldous and Diaconis (1986) report on some methods of analyzing card shuffling.

Poisson and Good both emphasize the distinction between physical and subjective probabilities. The argument given above identifies features of classical chance devices that push most of us to make similar predictions. It helps bridge the gap between the two positions, identifying a variety of situations where both lead to essentially the same results.

2. POISSON'S FORMULA AND THE METHOD OF ARBITRARY FUNCTIONS

A general ingredient of the arguments in the last section may be called the method of arbitrary functions. If a space is divided into small pieces and alternately colored white and black, any density on the space which is sufficiently spread out will assign mass of about ½ to white.

In the dart and wall example, the space was the unit circle. In the coin tossing example, it was the velocity-momentum space of initial conditions.

Such useful tools are worthy of careful analysis and extension. In Section 5 of his paper, Good uses Poisson's summation formula to give bounds when the space is the line with equal distance between the pieces. Good's bounds are most effective for scale mixtures of normals.

In unpublished work, Joop Kemperman has developed sharp bounds for general densities. We would like to describe a simple case of these bounds and compare bounds on the normal distribution.

Let X be a random variable with distribution function F. Let F_σ denote the distribution function of $\sigma X \pmod 1$. Thus, for $x < y \le x + 1$, $F_\sigma(y) - F_\sigma(x) = P\{x < \sigma X \le y \pmod 1\}$. Kemperman uses the discrepancy $D(\sigma) = \sup_{x,y} | F_\sigma(y) - F_\sigma(x) - (y - x)|$ to measure the distance between F_σ and the uniform distribution.

To begin with, $D(\sigma)$ tends to zero as σ tends to infinity if and only if the characteristic function of F tends to zero at infinity. Hence a density for F is not required.

To get bounds, suppose X has density $f(x)$ having k derivatives ($k \ge 0$) such that the kth derivative $f^{(k)}$ is of bounded total variation V_k.

THEOREM (Kemperman). *There is a universal constant c_k such that for all $\sigma > 0$*

$$D(\sigma) \le \frac{c_k V_k}{\sigma^{k+1}}.$$

In particular, $c_0 = 1/8$, $c_1 = 1/64$, $c_2 = 1/384$.

As an example, suppose $k = 0$ and f' is integrable. Then:

$$D(\sigma) \le \frac{1}{8\sigma} \int_{-\infty}^{+\infty} |f'(x)|\, dx. \tag{2.1}$$

Thus, the rate of convergence depends on how much the density of f "wiggles."

Kemperman shows that the theorem is sharp in the sense that there are densities where asymptotic equality is achieved in the upper bound. On the other hand, convergence is often more rapid than linear in σ.

Using Poisson's summation formula as Good does, one can show that for the standard normal density $D(\sigma)$ converges to zero at a rate of $\exp[-(2\pi\sigma)^2]$. For the standard Cauchy the rate is $\exp[-(2\pi\sigma)]$.

David Aldous has suggested that the normal density has the fastest rate under all possible centerings of the initial lattice.

The method of arbitrary functions can easily be varied to take account of variations in the underlying assumptions. For example, consider the dart and the wall again, but this time suppose that the distance l is not precisely specified.

If l is chosen from the density $h(l)$ and, for a given distance l, θ is chosen according to the density $f_l(\theta)$, the same argument gives

$$|P\{\text{Black}\} - 1/2| \leq \frac{d}{4} E\left(\frac{V_l(f) + 1}{l}\right),$$

where

$$V_l(f) = \int |f_l'(\theta)|\, d\theta.$$

If $l \geq l_0$ is assumed and $V_l(f) \leq V_0$, the bound can be replaced by $((V_0 + 1)/4l_0)\, d$. This gets smaller as l_0 gets larger, and is still linear in d.

Poisson's formula exploits the symmetry arising from considering random variables mod 1. Kemperman's argument is basically integration by parts. It gives bounds for two colorings of the line which are not equally spaced. Both arguments are important in generalizing to more complicated spaces that arise naturally in physical problems. This is discussed further in Section 3.

3. GENERALIZATIONS AND OTHER APPLICATIONS OF POISSON'S FORMULA

In recent years, the mathematics community has evolved and applied extensions of the Poisson summation formula to more general groups. To take a special case, let G be a commutative group and Γ a closed subgroup. Let $\Gamma^{\perp}$ be the set of characters in the dual group which equal 1 for every γ in Γ. Then, for all reasonable functions f,

$$\int_{\Gamma} f(\gamma)\, d\gamma = \int_{\Gamma^{\perp}} \hat{f}(\sigma)\, d\sigma.$$

For example, if G is the real line and Γ the subgroup of integers, then $\Gamma^{\perp} = \{t: e^{itj} = 1 \text{ for all integers } j\}$. Thus $\Gamma^{\perp} = \{2\pi k: k \text{ integer}\}$ and the formula reduces to the usual result. When G is taken as a finite commutative group we get Good's (5.2.4).

There are also noncommutative versions, collectively known as Selberg's trace formula. The interested reader can find details in Mackey (1978, pp. 325, 347–353) or Gelbart (1975, Chapter 9).

One interesting result connected to Poisson's formula but not emphasized in Good's review is the construction of two characteristic functions which agree in an interval. One first constructs a characteristic function of finite support and then considers its periodic extension. The latter is shown to be a characteristic function using Poisson's formula. To appreciate this, try thinking of a direct probabilistic interpretation of the periodic continuation.

Feller (1971, Chapter 19) gives a motivated version of this construction which apparently goes back to Khintchine (see Lévy, 1961). Diaconis and Ylvisaker (1985) use the idea to construct two different priors with the same posterior mean for a location problem.

We have used Good's version of the Poisson summation formula to construct versions of the curiosities in Feller (1971, Chapter 15.2) for finite Abelian groups.

Closely related to Poisson's formula is the idea of considering an integer random variable Z mod m. If Z is sufficiently spread out we expect its remainder will be close to uniformly distributed on $\{0, 1, \ldots, m-1\}$. In (6.2) Good computes explicit bounds for a Poisson Z. Diaconis and Stein (1978) and Bingham (1981) give explicit bounds related to the central limit theorem.

Finally, Diaconis (1982) uses the discrete version of Poincaré's argument to get bounds in random walk problems on finite Abelian groups.

4. LAST WORDS

One of the wounderful features of Good's writing is its pointers to goodies near and far. We have benefitted from following several of his leads. A card trick has been invented starting from Good's amazing formula for the number of permutations when repeats are allowed. We thank him and Morrie DeGroot for the chance to have a closer look at this material.

ACKNOWLEDGMENTS

We thank Joseph Keller and Bradley Efron for their help in formulating our basic example. We thank Joop Kemperman for communicating the bounds of Section 2.

ADDITIONAL REFERENCES

ALDOUS, D. and DIACONIS, P. (1986). Shuffling cards and stopping times. Technical Report No. 231, Dept. of Statistics, Stanford Univ. *Amer. Math. Monthly*, to appear.

BASS, T. A. (1985). *The Eudaemonic Pie: Or Why Would Anyone Play Roulette Without a Computer in His Shoe*? Houghton Mifflin, Boston.

BINGHAM, N. H. (1981). Tauberian theorems and the central limit theorem. *Ann. Probability* **9** 221–231.

DIACONIS, P. (1982). *Group Theory in Statistics*. Lecture notes, Harvard Univ.

DIACONIS, P. and STEIN, C. (1978). Some Tauberian theorems related to coin tossing. *Ann. Probability* **6** 483–490.

DIACONIS, P. and YLVISAKER, D. (1985). Quantifying prior opinions. In *Bayesian Statistics* **2** (J. M. Bernardo, M. H. DeGroot, D. V. Lindley, and A. F. M. Smith, eds.) 133–156. North Holland, Amsterdam.

FIENBERG, S. E. (1971). Randomization and social affairs: the 1970 draft lottery. *Science* **171** 255–261.
GELBART, S. (1975). *Automorphic Forms on Adele Groups.* Princeton Univ. Press, Princeton, N. J.
HOPF, E. (1934). On causality, statistics and probability. *J. Mathematical Phys.* **17** 51–102.
HOPF, E. (1937). Ein Verteilungsproblem bei dissipativen dynamischen Systemen. *Math. Ann.* **114** 161–186.
KELLER, J. (1985). The probability of heads. *Amer. Math. Monthly,* in press.
KEMPERMAN, J. (1975). Sharp bounds for discrepancies (mod 1) with application to the first digit problem. Unpublished manuscript.
LÉVY, P. (1961). Quelques problèmes non résolus de la theorie des functions caractéristiques. *Ann. Mat. Pura Appl.* **53** 315–332.
MACKEY, G. W. (1978). *Unitary Group Representations in Physics, Probability, and Number Theory.* Benjamin/Cummings, Reading, Mass.
SAVAGE, L. J. (1973). Probability in science: A personalistic account. In *Logic, Methodology and Philosophy of Science* (P. Suppes et al., eds.) 417–428. North Holland, Amsterdam.
VON PLATO, J. (1983). The method of arbitrary functions. *Br. J. Philos. Sci.* **34** 37–47.
VULOVIC, V. Z. and PRANGE, R. E. (1985). Is the toss of a coin really random? Univ. of Maryland. Preprint.
WOLFRAM, S. (1985). Origins of randomness in physical systems. *Phys. Rev. Lett.* **55** 449–452.
YUE, Z. and ZHANG, B. (1985). On the sensitive dynamical system and the transition from the apparently deterministic process to the completely random process. *Appl. Math. Mech.* (English ed.) **6** 193–211.

Comment

Herbert Solomon

I. J. Good has done a service in highlighting some of Simeon Denis Poisson's (1781–1840) contributions to statistics and probability. While his name is a commonplace to us, the breadth and variety of Poisson's work is neglected in formal courses undertaken by even the most advanced students. It is good to view Poisson as a member of the French school of probabilists who thrived from the late 18th century to the middle of the 19th century. Laplace overwhelms this group but the contributions of Condorcet, Cournot, and Bienaymé, as well as Poisson, among others, must receive attention. This is especially so in model building and estimation in the behavioral sciences. Applying the calculus of probabilities to important societal problems such as jury behavior was not beneath them. All lived in dramatically changing times in France where individual rights had assumed an importance that did not exist before the late 18th century.

Unfortunately, this kind of endeavor was frowned upon in the latter half of the 19th century by European mathematicians and it was not until the middle of this century that we found this activity again receiving the attention it deserves. For some reason or other, R. A. Fisher also refers in a rather negative manner to that earlier era when probabilists concerned themselves with the veracity of witnesses and group decisions. Yet one of Poisson's most important works was his 1837 volume on the *Calculus of Probability Applied to Civil and Criminal Proceedings*—the book in which what we now call the Poisson distribution first appears; albeit as a mathematical approximation artifice.

Herbert Solomon is Professor of Statistics, Department of Statistics, Stanford University, Sequoia Hall, Stanford, California 94305.

Much as we wish that Good would have elaborated more on the themes in his paper, editorial constraints and his own tastes no doubt limited the size of his effort. Let us now look into the Poisson jury model in some detail to catch the flavor of the statistical thinking and the concern with moral and societal values demonstrated by Poisson.

It is important to note that Poisson in developing his model paid heed to the data available in his day. For the period 1825–1830, jury decisions were based on 7 or more out of 12 jurors favoring either conviction or acquittal. Cases based on verdicts of exactly 7 out of 12 went to a higher court which could change the verdict. For each year, the number of trials and number of convictions were registered and listed for crimes against persons and crimes against property. Note here that this distinction in crimes is definitely drawn over 150 years ago. In the period 1831–1833, listings were also available, except the majority required was 8 or more out of 12. In 1832 and 1833, the jury could find extenuating circumstances in a conviction that would then lead to a lighter penalty.

What impressed Poisson was the stability of the conviction ratios over each of the years 1825–1830 and 1832–1833. He felt this was a basis for developing a model that in some parsimonious way could reproduce the data, and if so, lead to the computation of the probabilities of the two kinds of errors important in judging the effects of size and decision-making rules of a jury, namely, the probability of acquitting a guilty defendant, and the probability of convicting an innocent defendant.