Statistical Science
1986, Vol. 1, No. 2, 157–180

Some Statistical Applications of Poisson's Work

I. J. Good

Abstract. Statistical applications and repercussions of Poisson's work are reviewed in historical perspective with special reference to (i) the distinction he made between two kinds of probability; (ii) the law of large numbers; (iii) the Poisson distribution; (iv) the difference between two proportions; (v) legalistic statistics; (vi) Poisson's summation formula; (vii) the Cauchy distribution; and (viii) the Poisson bracket.

Key words and phrases: Bayes factors, Cauchy distribution, judicial decisions, discrimination between treatments, first-digit distribution, kinds of probability, law of large numbers, Poisson distribution, Poisson's summation formula, roulette.

1. INTRODUCTION

Winston Churchill (1951) said that "Everybody has a right to pronounce foreign names as he chooses." I pronounce Poisson's name more or less as a Frenchman would pronounce it, but I have heard it pronounced "Poyson" as if it were Brooklynese for "person."

Poisson (1781–1840), a protégé of Laplace, although outshone by his rivals Cauchy and Fourier (see Grattan-Guinness, 1972), did distinguished work in mechanics, celestial mechanics, the theory of heat, geophysics, wave propagation, electricity and magnetism, potential theory, elasticity, hydrodynamics, the calculus of variations, integral equations, divergent series, complex integration, differential equations, and on the remainder in the Euler-Maclaurin summation formula. (Poisson's work on all these topics is discussed, for example, by Kline, 1972.) But I shall confine my attention to the influences that his work has had on statistics and probability interpreted in a broad sense.

According to Gratton-Guinness (1972, pp. 302, 445, 450, 458, 460–471), Poisson's exposition was often unclear, and his character left a non-negligible amount to be desired, but these matters will not be much discussed in the present essay.

I. J. Good is University Distinguished Professor of Statistics and Adjunct Professor of Philosophy, Department of Statistics, Virginia Polytechnic Institute and State University, Blacksburg, Virginia 24061. This work is based on the text of an invited lecture at the Poisson Bicentennial Commemoration held at George Washington University, March 1982.

Most of what I shall say pertains to (i) the law of large numbers and the distinction between kinds of probability, (ii) the Poisson summation formula, and (iii) the Poisson distribution. There is a good reason for treating the Poisson distribution third instead of first, namely that Poisson was scarcely responsible for introducing this distribution, nor for its applications. But there were other things for which he was responsible that were named after other people, so the eponymy does rough justice.

2. KINDS OF PROBABILITY AND THE LAW OF LARGE NUMBERS

In about 1961 George Barnard told me he was reading Poisson (1837) and that Poisson had repeatedly emphasized the distinction between two kinds of probability. This is historically interesting because philosophers were and still are crediting this distinction to Carnap (1950) who also emphasized it. A few years ago I examined Poisson's book for the first time and found the distinction made on the second page of Chapter 1 (p. 31). He calls a physical probability, to be defined below, a *chance* and reserves the word *probability* for its epistemic or intuitive meaning.

He says on page 30 that the probability of an event is [measures] the reason that we have to believe that it will or has occurred. He goes on to say that probability depends on the knowledge that we have concerning an event and can be different for different people because they can have different knowledge. He does not mention that, in a modern notation, even $P(A \mid B)$ might vary from one person to another, nor from one time to another for a single person. I therefore take it that his concept of probability was not quite that of subjective (= personal) probability, but rather that of logical probability, sometimes called credibility. The logical probability of an event A, when information B is given, is usually assumed, somewhat metaphysically, to have a numerical value (or perhaps an interval of values) that every entirely rational person would accept. It is the unique rational degree (or perhaps interval) of belief in A (or intensity of conviction concerning A) given B, or the *degree* to which B logically implies A. Carnap attempted to define logical probabilities in terms of a given language. The empirical evidence that gives some support for the existence of logical probabilities, or at least multipersonal probabilities, is that, for many pairs (A, B), the judgments of $P(A \mid B)$ by different people do not differ very much. This point is made especially clear when *ratios of related probabilities* are considered. Suppose, for example, that we are making bets about the weight, w kilograms, of the next candidate in a beauty competition. Suppose that Jones was to regard the inequality $65.2 < w < 65.3$ as three times more probable than $65.1 < w < 65.2$, and was prepared to make a conditional bet in accordance with this belief, without having some special information. Then it seems to me that Jones would be *objectively* irrational.

It is customary, when talking about subjective probability, to assume that a set of (rational) subjective probabilities that a person has over a short time period, or *expresses in a single document*, ought to obey the usual axioms, at least approximately; in other words, that the subjective probabilities should be in some sense at least approximately self-consistent. But even precise self-consistency is not enough to guarantee rationality.

The expression "epistemic probability," used before, means a probability that is either logical, subjective, or multisubjective. Strictly speaking, logical probabilities need to be entirely self-consistent and in many circumstances can be regarded as an unattainable ideal. The existence of logical probabilities is controversial, but presumably it was this concept that Poisson had in mind.

On the other hand, a physical probability, or intrinsic probability, or chance, or propensity, is supposed to be a probability that exists in the "outside world," and would exist even if there were no conscious beings considering it. The notation $P(A \mid B)$ can still be used, but now B is intended to be a complete description of the true or hypothetical state of the world. Poisson does not describe the matter in such terms, but the following free translation of some of his page 31 seems to agree with this description:

> For example, in the game of *heads* and *tails*, the chance of obtaining *heads*, and that of obtaining *tails*, depend on the constitution of the coin; one can regard it as physically [almost] impossible that these two chances are equal [to one-half]. On the other hand, if the constitution of the coin is not known to us, and if we have not already made trials, the probability of obtaining *heads* is, for us, exactly the same as that of obtaining *tails*. We have in fact no reason to believe in the one event rather than the other. It is no longer the same if the coin has been tossed several times. The "chance" of each face does not change during these trials, but for someone who knows the results of the trials, the probability varies with the numbers of times that the two faces are presented.

A loaded die would have been a somewhat better example because loaded dice are easier to construct than biased coins.

If Poisson had known about quantum mechanics he might have used the probabilities derived from Schrödinger's equation as examples of physical probabilities because these probabilities are calculable in terms of the "state" of a physical system and of the experimental set-up. Most, although by no means all, physicists regard these quantum mechanical probabilities as physically intrinsic and not dependent upon the beliefs or existence of the experimenter or of any one else. Of course if a rational person somehow *knows* that a physical probability is equal to some value p, then that person's subjective probability must also equal p.

The existence of physical probabilities is controversial. Einstein, for example, said that "God does not play dice" and that quantum mechanics is an incomplete theory that does not describe "the whole of reality" (see, for example, Pais, 1982, p. 456). Just as in classical (prequantum) statistical mechanics, it is natural to think of the so-called physical probabilities as merely a way of describing our own (inevitably) incomplete knowledge.

Most card-carrying Bayesians regard epistemic probabilties as primary in the sense that physical probabilities, if they exist, can be estimated only by making use of epistemic probabilities. de Finetti (1974, p. x) goes further and claims in capital letters that [physical] probabilities do not exist, a statement that can only be properly understood in terms of his representation theorem for permutable (exchangeable) events. For a statement and proof of the theorem for binomial and multinomial sampling see, for example, Good (1965b, pp. 13, 21–23) and for a historical account and a vast generalization see Hewitt and Savage (1955). In a nutshell these representation theorems show that the Bayesian formalism used by Poisson, depending on two kinds of probability, can be reinterpreted and justified in terms of epistemic probability

Poisson

alone. One way of expressing the matter is that a Bayesian behaves *as if* he believes in both epistemic and physical probabilities. de Finetti's theorem really proved only that it is not *essential* to assume that physical probabilities exist, not that they *cannot* exist.

Savage (1954) showed that a perfectly rational being would behave as if she had subjective probabilities and utilities, and de Finetti's theorem in effect adds another "as if": see also Vaihinger (1911) for a general "Als Ob" philosophy, also known as fictionalism. (Vaihinger was influenced by Kant who said "the world that is the object of our knowledge is a world of appearances, existing only insofar as it is constructed"; Walsh, 1967, p. 316ii. Earlier influences were William of Occam and Thomas Hobbes, while Jeremy Bentham forestalled many of Vaihinger's conclusions according to Ogden, 1932.) Vaihinger (1935, p. viii) says "An idea whose theoretical untruth ... is admitted ... may have great practical importance." The use of pseudorandom numbers as if they were random is a good example of fictionalism, better than the use of imaginary numbers which constituted one of Vaihinger's examples. The assumption that physical probabilities exist is, in some contexts, not even theoretically self-contradictory in my opinion, and in such contexts their use can be regarded as an exemplification of pragmatism rather than fictionalism. Not knowing any description of reality that clearly goes beyond the "as if," it seems to me that the easiest procedure is to talk about epistemic and physical probabilities as if they are both real, just as Poisson did. When someone succeeds in defining reality clearly sans Als Ob we might be able to resolve some of the controversies.

The notion that physical probabilities are only a theoretical construction of our minds somewhat resembles but is less extreme than philosophical solipsism interpreted in the sense that the self is the only reality. I don't know any card-carrying solipsists but I like to think of de Finetti's theorem as showing that solipsism cannot be disproved although he didn't put it that way. I believe that most statisticians today would agree with Poisson that physical probability and epistemic probability are both useful concepts. This distinction is still in need of emphasis, for it is often not taught to students of statistics. Although the distinction is controversial, it will be adopted in this article. Those who prefer de Finetti's approach will be able mentally to "translate" what is said into the terminology of permutability (exchangeability).

In modern times distinctions have been made between at least five kinds of probability, for example by Kemble (1942), Good (1959, 1966), and Fine (1973), but the main distinction is still between physical probability on the one hand and epistemic probability on the other. In connection with physical probability, Poisson was much concerned with the law of large numbers which he extended from the Bernoulli case to sequences of trials in which the probability $p_i (i = 1, 2, \cdots, n)$ varies from trial to trial. Putting it roughly, Poisson's Law of Large Numbers states that in a long sequence of trials the fraction of successes will very probably be close to the average of the "chances" for the individual trials even when this average does not tend to a limit (see Uspensky, 1937, pp. 208 and 294). This theorem is perhaps Poisson's main direct contribution to the mathematical theory of probability and statistics.

Feller (1968, p. 218) comments that Bernoulli trials with variable probabilities are "known under the confusing name of 'Poisson trials'." But this name is historically much more justifiable than the name "Poisson distribution" in its familiar sense. To avoid confusion with the Poisson distribution, the number of "successes" in a sequence of Poisson trials is sometimes said to have a "generalized binomial distribution" although this expression can be misinterpreted as a multinomial distribution.

Poisson (1837, Chapter 4) bases his discussion on the generating function

$$\Pi_i(up_i + vq_i) \qquad (q_i = 1 - p_i).$$

He puts $u = e^{ix}$ and $v = e^{-ix}$, so his approach is more or less equivalent to the use of characteristic functions (which had been introduced by Laplace). In arriving at an asymptotic formula he does not use the assumption that the p_i's do not vary very much, so his argument cannot be rigorous. According to Heyde and Seneta (1977, p. 49), the formula for the variance of the number of successes, namely

$$n\bar{p}(1 - \bar{p}) - \sum (p_i - \bar{p})^2 \qquad (\bar{p} = \sum p_i/n),$$

might have been given first by Czuber in 1899. Heyde and Seneta also point out how the notion of Poisson trials led to the study of homogeneity and stability in repeated trials, or dispersion theory, by Bienaymé, Lexis, and other well known statisticians. "Stability" here refers to the degree of constancy of the relative frequencies over different parts of a series or population, as in Keynes (1921, pp. 392–393). The work of Lexis anticipated Fisher's F statistic, the distribution theory being obtained by Helmert in 1876 (see Heiss, 1978). Thus there seems to be a causal chain from Poisson trials to the analysis of variance.

The mathematical theory of probability originated largely from games of chance where probabilities can often be calculated based on judgments of equal physical probability and of physical statistical independence. Poisson may have been the first to emphasize that, outside of games of chance, it is usually necessary to estimate physical probabilities by repeated sampling, and this was one reason why he was interested in the Law of Large Numbers. He wrote:

> Things of all kinds are subject to a universal law that may be called the law of great numbers ... From cases of all kinds, it follows that the universal law of great numbers is a general and incontestable fact, resulting from experiments that never contradict it. (Quoted by Keynes, 1921, p. 333, in the French.)

Keynes goes on to say:

> It is not clear how far Poisson's result is due to *a priori* reasoning, and how far it is a natural law based on experience; but it is represented as displaying a certain harmony between natural law and the *a priori* reasoning of probabilities. Poisson's conception was mainly popularized through the writings of Quetelet ... and he [Quetelet] has a very fair claim ... to be regarded as the parent of modern statistical method.

That would perhaps make Poisson and Gauss grandfathers, and Laplace a great-grandfather. But we now need to add a couple of generations, and Karl Pearson has been described as the grandfather of 20th century statistics.

3. THE DIFFERENCE BETWEEN TWO PROPORTIONS

Suppose that we have two binomial samples, with unknown parameters, or physical probabilities, p and

p'. These samples might correspond to two medical treatments. Let the numbers of "successes" be r and r', respectively, let the sample sizes be n and n', and let $\delta = (r/n) - (r'/n')$ which is assumed to be positive. Poisson (1837, p. 225) asks what is the asymptotic probability that $p - p' > \epsilon$ when r, n, r', and n' are large? He uses a Bayesian approach (naturally: the *question* is itself Bayesian) and arrives at the asymptotic formula

$$(3.1) \quad P(p - p' > \epsilon) = (2\pi)^{-1/2} \int_v^\infty e^{-t^2/2}\, dt$$

where

$$(3.2) \quad v = (\epsilon - \delta)\left[\frac{r(n-r)}{n^3} + \frac{r'(n'-r')}{n'^3}\right]^{-1/2}.$$

According to Westergaard (1932, pp. 149–150), Poisson's work on the difference between two proportions was popularized, for medical applications, by Gavarret (1840), "an enthusiastic pupil of Poisson." It is unclear that "popularized" is the right word because Westergaard goes on to say that unfortunately Gavarret's "contemporaries took very little notice of his book."

4. JUDICIAL DECISIONS

Laplace (1820, first supplement, p. 33) had given a formula for the probability that an accused person is innocent if he is acquitted by r members of a jury of n people, under certain Bayesian assumptions. Independently, Good and Tullock (1984, 1985) dealt in a similar manner with the probability that the Supreme Court reaches a "correct" decision when the voting is say 5 to 4, where "correctness" now has to be given a definition. We decided that the probability in this case is *no more* than 0.63. Poisson (1837, p. 364) discusses Laplace's result somewhat critically and states:

> . . . it is fair to say that Condorcet should be credited with the ingenious idea of regarding the guilt or innocence of the accused as the cause of the verdict reached, the latter being the observed event from which one can infer the probability of the cause by using the principle of Blayes [sic]. (Poisson, 1837, p. 2; free translation)

It is safe to assume that Poisson means Bayes, not Blayes (although he spells it this way twice), otherwise we have a new candidate for Stigler's question "Who discovered Bayes's theorem?" (Stigler, 1983).

But Poisson's main interest in the application of statistics to judicial decisions is again connected with the law of large numbers. A simple example of this work of Poisson, based on the records of the Cours d'Assises, is given by Féron (1978). The required majority for conviction used to be 7 to 5, but became 8 to 4 in 1831. Poisson noted that 0.07 was the proportion of the time that the vote was 7 to 5 and the accused was convicted, and that therefore this could have been used as an estimate of the additional proportion of acquittals once the law was changed. Moreover the prediction would have been correct for 1831. As Féron says "at the time it [the reasoning] was received with skepticism."

Poisson's idea has been extended by Gelfand and Solomon (1973, 1974, 1975) in considerable detail, and their papers contain further references on the topic.

5. POISSON'S SUMMATION FORMULA

Poisson is usually credited with an extremely elegant formula in infinite series which can be usefully applied in probability and statistics although Poisson did not make such applications as far as I know. Let us first consider a special case. For positive values of t, let

$$(5.1) \quad \psi(t) = 1 + 2\sum_{n=1}^{\infty} e^{-\pi n^2 t}.$$

Then Poisson (1823, p. 420) proved that

$$(5.2) \quad \psi(t)\sqrt{t} = \psi(1/t).$$

An entertaining way of writing (5.2) is

$$(5.3) \quad (-\log x)^{1/4}\left(\frac{1}{2} + \sum_1^\infty x^{n^2}\right) = (-\log y)^{1/4}\left(\frac{1}{2} + \sum_1^\infty y^{n^2}\right)$$

where

$$0 < x < 1 \quad \text{and} \quad \log x \log y = \pi^2.$$

Formula (5.2) is about θ functions (Whittaker and Watson, 1927, pp. 124 and 475). The θ function appears in the theory of heat but Poisson's writings on this theory appeared later. (Series of the form $\sum a_n e^{-kn^2 t} \sin nx$ appear in Fourier, 1812, in relation to the propagation of heat in a solid sphere.) Poisson (1827, p. 592) generalized this identity to give

$$(5.4) \quad \pi^{1/2} \sum_{n=-\infty}^{\infty} e^{-(a+n\pi/\omega)^2} = \omega \sum_{n=-\infty}^{\infty} e^{-n^2\omega^2}\cos(2na\omega).$$

As Whittaker and Watson state, Jacobi's "imaginary transformations" for all four θ functions, published in 1828, can then be readily derived by elementary algebra. Courant (1962, pp. 199–200) shows how (5.4) can be obtained by solving, in two different ways, the heat equation for a wire bent into a circle (compare Lévy, 1939, p. 37).

A further generalization is the following result known as Poisson's Summation Formula (Poisson, 1827, p. 591):

$$\text{(5.5)}\qquad \sqrt{a}\sum_{n=-\infty}^{\infty} f(an) = \sqrt{b}\sum_{n=-\infty}^{\infty} g(bn) \quad \text{if} \quad ab = 2\pi \qquad (a > 0)$$

where g and f are Fourier transforms, that is

$$g(t) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} f(x)e^{-itx}\, dx,$$

and where certain regularity conditions are satisfied. A pair of simple sufficient conditions for (5.5) are (Mordell, 1928):

(a) $f(x)$ and $f'(x)$ are continuous for all real x and tend to zero as $|x| \to \infty$; and

(b) the integrals $\int_{-\infty}^{\infty} f(x)\, dx$ and $\int_{-\infty}^{\infty} |f''(x)|\, dx$ converge, and f' is an integral of f''.

Other sets of sufficient conditions, at least for cosine transforms, are given by Linfoot (1928), Titchmarsh (1932, pp. 443–444; 1937, p. 61), and Courant (1962, p. 77). Feller (1971, p. 630) shows that it is sufficient to have f non-negative and g absolutely integrable over $(-\infty, \infty)$.

The formula can be written in the negligibly more general but convenient form

$$\text{(5.6)}\qquad \sqrt{a}\sum_{n=-\infty}^{\infty} f(na)e^{inay} = \sqrt{b}\sum_{m=-\infty}^{\infty} g(mb + y),$$

which we obtain by applying (5.5) to the function $g(\cdot + y)$ instead of to $g(\cdot)$. We now give a well known formal proof of (5.6) without attending to points of rigor.

Let $\psi(y)$ denote the right side of (5.6). It has period b and so (formally at least) has a Fourier series

$$\psi(y) = \sum_{n=-\infty}^{\infty} c_n e^{inay} \quad (\text{where } a = 2\pi/b)$$

where

$$c_n = \frac{1}{b}\int_{-b/2}^{b/2} \psi(y)e^{-inay}\, dy$$

$$= \frac{1}{\sqrt{b}}\int_{-b/2}^{b/2} e^{-inay}\sum_{m=-\infty}^{\infty} g(mb + y)\, dy$$

$$= \frac{1}{\sqrt{b}}\sum_{m=-\infty}^{\infty}\int_{-b/2}^{b/2} e^{-inay} g(mb + y)\, dy$$

$$= \frac{1}{\sqrt{b}}\sum_{m=-\infty}^{\infty}\int_{(m-1/2)b}^{(m+1/2)b} e^{-ina(y-mb)} g(y)\, dy$$

$$= \frac{1}{\sqrt{b}}\sum_{m=-\infty}^{\infty}\int_{(m-1/2)b}^{(m+1/2)b} e^{-inay} g(y)\, dy$$

$$= \frac{\sqrt{a}}{\sqrt{(2\pi)}}\int_{-\infty}^{\infty} e^{-inay} g(y)\, dy$$

$$= \sqrt{a} f(na).$$

Therefore the left side of (5.6) is the Fourier series for the right side with period b.

Poisson's summation formula appears in Poisson (1827, pp. 591–592) where he calls it a "nouvelle formule." In accordance with a literal reading of Stigler's Law that eponymy is always wrong (Stigler, 1980), it is impossible that Poisson could have been the original discoverer of Poisson's summation formula. Indeed it appears in Cauchy (1817) who says, with French understatement, "Cette proposition nouvelle nous paraît digne d'être remarquée" (translated into British understatement: this new theorem seems to us to deserve notice), and he mentions (5.2) as a special case. But perhaps Poisson discovered the formula independently of Cauchy and seems to have made more applications of it, so the eponymy is not totally unreasonable and is well entrenched.

One statistical application of Poisson's summation formula is for the theory of roulette. Without using Poisson's summation formula, Poincaré (1912, pp. 148–150) made the following contribution to this theory:

> Suppose a wheel, divided into a very large number of equal sectors alternately colored red and black, is given a rapid rotation. When it stops, one of its sectors is opposite a certain fixed point: what is the probability P that this sector is red?

Poincaré gives the following simple solution:

Let $f(x)$ denote the probability density for the total angle x rotated by the wheel, starting from some given position, where $f(x) = 0$ when x exceeds some value A. Suppose $|f'(x)| < M$ for all x. Then it is easily proved that $|2P - 1| < MA\epsilon$ where ϵ is the angular width of a sector. Hence $P \to ½$ when $\epsilon \to 0$. Feller (1971, pp. 62–63) gives another easy proof of a similar result on the assumption that $f(x)$ is unimodal and is small at its maximum value.

Much stronger results can be obtained when more is assumed about $f(x)$. Suppose that we measure angles in units of 2π radians ("circumferences"). Then, after "wrapping $f(x)$ round the circle," the density function is

$$\text{(5.7)}\qquad p(x) = \sum_{n=-\infty}^{\infty} f(x + n),$$

and we are interested in whether $p(x)$ is close to 1 for all x where $0 \le x \le 1$. Let us take, as a measure of nonuniformity,

$$\kappa = \int_0^1 [p(x) - 1]^2\, dx = \int_0^1 [p(x)]^2\, dx - 1 \tag{5.8}$$

which is of course non-negative. We would like to show that κ is small under some assumptions about f.

It would hardly be reasonable to assume that f is a "normal" density because the croupier's mood might vary from one occasion to another. Instead, let us assume that $f(x)$ is obtained as a mixture of a parameterized set of densities $f(x, \theta)$, the mixing function being $g(\theta)$, where θ might be a scalar or a vector parameter. In other symbols, we assume that

$$f(x) = \int f(x, \theta) g(\theta)\, d\theta$$

where, for each θ, $f(x, \theta)$ has a simple form, say normal. Then

$$p(x) = \int p(x, \theta) g(\theta)\, d\theta$$

where

$$p(x, \theta) = \sum_{n=-\infty}^{\infty} f(x + n, \theta).$$

Denote the nonuniformity corresponding to $p(x, \theta)$ by $\kappa(\theta)$. To obtain an upper bound to $\kappa(\theta)$, note first that

$$[p(x)]^2 = \int\int p(x, \theta) p(x, \psi) g(\theta) g(\psi)\, d\theta\, d\psi.$$

Therefore,

$$\begin{aligned}
\int_0^1 &[p(x)]^2\, dx \\
&= \int\int g(\theta)g(\psi) \int_0^1 p(x, \theta)p(x, \psi)\, dx\, d\theta\, d\psi \\
&\le \int\int g(\theta)g(\psi) \left\{\int_0^1 [p(x, \theta)]^2\, dx \right. \\
&\qquad \left. \times \int_0^1 [p(x, \psi)]^2\, dx\right\}^{1/2} d\theta\, d\psi \\
&= \left[\int g(\theta)\left\{\int_0^1 [p(x, \theta)]^2\, dx\right\}^{1/2} d\theta\right]^2 \\
&= \left\{\int g(\theta)[1 + \kappa(\theta)]^{1/2}\, d\theta\right\}^2 \\
&\le \left[1 + \frac{1}{2}\int g(\theta)\kappa(\theta)\, d\theta\right]^2 \\
&= (1 + \beta/2)^2
\end{aligned}$$

where

$$\beta = \int g(\theta)\kappa(\theta)\, d\theta \le \left\{\int [g(\theta)]^2\, d\theta \int [\kappa(\theta)]^2\, d\theta\right\}^{1/2}. \tag{5.9}$$

Therefore

$$\kappa \le \beta(1 + \tfrac{1}{4}\beta). \tag{5.10}$$

Corollary. From (5.9) and (5.10) we see that

$$\kappa \le \kappa_0(1 + \kappa_0/4) \tag{5.11}$$

where

$$\kappa_0 = \sup_\theta \kappa(\theta),$$

and

$$\kappa \le \beta_0(1 + \tfrac{1}{4}\beta_0) \tag{5.12}$$

where

$$\beta_0^2 = \int g^2\, d\theta \int \kappa^2\, d\theta.$$

From (5.11) we see that the nonuniformity of a mixture is bounded above by a number slightly larger than the "largest" nonuniformity of the distributions that are mixed.

As an example, suppose that the distributions that are mixed are all normal distributions, a typical one being

$$f(x; \mu, \sigma) = \sigma^{-1}(2\pi)^{-1/2}\exp[-\tfrac{1}{2}(x - \mu)^2\sigma^{-2}].$$

Of course θ is now the pair (μ, σ). Then, for $0 < x < 1$, we have (compare Lévy, 1939, p. 37)

$$\begin{aligned}
p(x, \theta) &= \sum_{n=-\infty}^{\infty} f(x + n, \theta) \\
&= \sigma^{-1}(2\pi)^{-1/2} \sum_{n=-\infty}^{\infty} e^{-(x-\mu+n)^2/(2\sigma^2)} \\
&= 1 + 2\sum_{n=1}^{\infty} e^{-2\pi^2 n^2 \sigma^2} \cos[2\pi n(x - \mu)]
\end{aligned} \tag{5.13}$$

by Poisson's summation formula or its special case (5.2). Hence

$$|p(x, \theta) - 1| \le 2\sum_{n=1}^{\infty} e^{-2\pi^2 n^2 \sigma^2} \quad \text{(for all } x \text{ and } \mu) \tag{5.14}$$

and

$$\kappa(\theta) \le 4\left(\sum_{n=1}^{\infty} e^{-2\pi^2 n^2 \sigma^2}\right)^2 \approx 4e^{-4\pi^2\sigma^2} \tag{5.15}$$

if σ exceeds say one-tenth of a "circumference," as it always would in practice. It now follows from (5.11) that κ does not exceed a number approximately equal to $4\exp(-4\pi^2\sigma_0^2)$ where σ_0 is the greatest lower bound of all the values of σ that occur. For example, $\kappa < 10^{-16}$ if $\sigma_0 = 1$.

We have oversimplified the real roulette problem in a few ways: (i) We need to assume also that μ/σ is not too small to justify using a normal distribution when x is necessarily positive; (ii) strictly in roulette, in addition to rotating the wheel, the croupier sends an ivory ball round in the opposite direction, and it comes to rest in a numbered pocket near the circumference after a certain amount of bouncing around. The angle traveled by the ball relative to the wheel is the sum of two random variables, namely the angle traveled by the wheel and the angle traveled by the ball; (iii) the wheel might not be exactly horizontal and the pockets might not be exactly equal; (iv) we have analyzed a measure of nonuniformity for this relative angle, reduced modulo 1, whereas it would be of more direct interest to the gambler and to the casino to use a discrete measure of nonuniformity for the probabilities $p_1, p_2, \cdots, p_t$ corresponding to the t slots (where t is 37 in Monte Carlo and is 38 at Las Vegas where the casinoes are greedier). The discrete measure analogous to κ is $t \sum_r (p_r - t^{-1})^2 = t\rho - 1$ where ρ is the "repeat rate" $\sum p_i^2$. A relationship between κ and $t\rho - 1$ can be worked out by using properties of the midpoint method of integration, or by using the Euler-Maclaurin summation formula, but the details are omitted; and (v) perhaps more realistic than a mixture of normal distributions would be a mixture of gamma distributions, with

$$(5.16)\qquad f(x, \theta) = \beta^{-\alpha}x^{\alpha-1}e^{-x/\beta}/\Gamma(\alpha) \qquad (\alpha > 0, \beta > 0, x > 0).$$

It then turns out, by means of Poisson's summation formula, that

$$(5.17)\qquad p(x, \theta) = 1 + \sum_{n=1}^{\infty} \frac{2 \cos \gamma_n}{[1 + 4\pi^2\beta^2(n + x)^2]^{\alpha/2}}$$

where $\tan \gamma_n = 2\pi(n + x)\beta$. Hence

$$(5.18)\qquad \kappa(\theta) < \left[2 \sum_{n=1}^{\infty} (1 + 4\pi^2\beta^2n^2)^{-\alpha/2}\right]^2$$

and

$$(5.19)\qquad \kappa(\theta) < 4e^{-4\pi^2\beta^2\alpha} = 4e^{-4\pi^2 \mathrm{Var}(x)}$$

if this is small, and we can then apply (5.11) as before. Thus the result is essentially the same as if we had assumed a mixture of normal distributions. Perhaps this result is true under much less restrictive assumptions than have been made here.

The roulette theory leads to a fairly adequate explanation of why the distributions of the initial digits of numbers in large data sets are *not* uniform, in fact the digit 1 often occurs about 30% of the time and digit 9 only 5% (Benford, 1938; Feller, 1971, p. 63; Raimi, 1976). Feller states that "A distinguished applied mathematician was extremely successful in bets that a number chosen at random in the *Farmer's Almanac*, or the *Census Report* or a similar compendium, would have the first digit less than 5." Wallis and Roberts (1957, pp. 331–332) refer to the use of the phenomenon as "a parlor game or swindle." Even with third digits one would be unlucky not to be able to discriminate a sample of 100,000 digits from a flat random population (Good, 1965a). A random number x, that is obtained by multiplying together several other numbers, has a logarithm obtained by adding logarithms. The mantissa of $\log x$ can be expected to be roughly uniformly distributed on a circular slide rule. It may be noted too that Jeffreys (1939, p. 100) proposed a uniform "improper" prior for the logarithm of a positive random variable x, in other words the prior density $1/x$ for x. He pointed out that this prior is invariant for powers. We seem to have a case here where a Bayesian prior roughly mirrors actual frequencies of occurrence in the real world (over some range of values of x) (compare Raimi, 1985). Note that the "first digit" phenomenon is intended to apply to small numbers like 2.31×10^{-17} as well as to large numbers like 3.7×10^6.

Poisson's summation formula can be used to prove the important sampling theorem of communication theory concerning band-limited time series (see, for example, Feller, 1971, p. 631). It states that a function f, the Fourier transform of which vanishes outside $(-\tfrac{1}{2}b, \tfrac{1}{2}b)$ is uniquely determined by the values of f at the points $2n\pi/b (n = 0, \pm1, \pm2, \ldots)$; in other words, by two values per shortest wavelength. The theorem is often attributed to H. Nyquist or to C. E. Shannon, but it occurs in Whittaker (1915) and there are even earlier anticipations as reviewed by Higgins (1985). If f satisfies the band-limited condition, then the right side of (5.6) when $|y| < \tfrac{1}{2}b$ reduces to the single term $\sqrt{b}g(y)$. On then taking the inverse Fourier transform of (5.6) we obtain $f(x)$ in terms of the values of f at the points na, that is, the points $2\pi n/b$.

Poisson's summation formula has also been applied to a problem in Brownian motion (Feller, 1971, pp. 341–342), to Kolmogorov-Smirnov tests (for example, Pelz and Good, 1976), to a problem in inventory control (Good, 1962a), and to a statistical theory of "remnants" related to a manufacturing process (Aitchison, 1959). The spectral density functions corresponding to the autocorrelation functions $\rho^{|\tau|}$ and ρ^{τ^2} can both be neatly expressed by using the summation formula (for example, Good, 1981b).

An entertaining combinatorial formula related to the summation formula occurs in connection with $\omega(n)$, defined as the number of orderings of n candidates when ties are permitted; for example, $\omega(3) = 13$.

It can be shown (Good, 1975; and, for further citations, Sloane, 1973, item 1191) that

$$(5.20)\qquad \sum_{n=0}^{\infty} \frac{\omega(n)x^n}{n!} = (2 - e^x)^{-1} = \frac{1}{2}\sum_{m=0}^{\infty} 2^{-m}e^{mx} \quad (|x| < \log 2)$$

and therefore

$$(5.21)\qquad \omega(n) = \sum_{m=-\infty}^{\infty} f(m) \quad (n > 0)$$

where

$$(5.22)\qquad f(x) = \begin{cases} \tfrac{1}{2}x^n e^{-x\log 2} & (x > 0) \\ 0 & (x \leq 0). \end{cases}$$

It then follows from (5.4) that

$$(5.23)\qquad \omega(n) = n!(\log_2 e)^{n+1}\left\{\frac{1}{2} + \sum_{m=1}^{\infty} \cos^{n+1}\theta_m \times \cos[(n+1)\theta_m]\right\} \quad (n \geq 1)$$

where θ_m is the angle the tangent of which is $2\pi m \log_2 e$. We can obtain $\omega(n)$ exactly by taking the series as far as the first term for which $m > n/(2\pi e)$. The term $n!(\log_2 e)^{n+1}/2$ is all that's needed when $n \leq 13$; for example, $\omega(13) = 536{,}858{,}348{,}381$.

Poisson's summation formula has a discrete analogue, related to the discrete Fourier transform, and is given an elementary but not easy proof by Good (1962b). The formula is

$$(5.24)\qquad |G|^{-1/2} \sum_{r\in G} a_r = |H|^{-1/2} \sum_{s\in H} a_s^*$$

where G and H are "orthogonal" subgroups of a finite Abelian group, the orders of these subgroups are $|G|$ and $|H|$, and a_s^* is the discrete multidimensional Fourier transform of a_r. (For a full explanation see the reference cited.) I think it may well have application to the design and analysis of fractional factorial experiments. The reason for this conjecture is that the theory of fractional factorial designs is expressible in terms of subgroups of a finite Abelian group (Fisher, 1942, 1945), and the "characters" of these groups are themselves closely related to multidimensional discrete Fourier transforms. Moreover, the interactions in a 2^n factorial experiment are expressible as the components of an n-dimensional modulo 2 discrete Fourier transform (Good, 1958a, 1960). The theory of optimal fractional factorial experiments is complicated (Raktoe et al., 1981) and might profit by a neat and nontrivial formula that, as far as I know, has not yet been used.

Another analogue is the formula

$$(5.25)\qquad \frac{1}{u}\sum_{r=0}^{u-1} f\left(\frac{2\pi r}{u}\right) = \sum_{n=-\infty}^{\infty} c_{nu}$$

if f has the Fourier series

$$(5.26)\qquad f(\theta) = \sum_{n=-\infty}^{\infty} c_n e^{in\theta}.$$

It would be interesting to know whether Poisson obtained his summation formula by analogy with the more elementary formula (5.25), but I have no evidence that he did. The formula (5.25) is a "Fourier series" form of a formula due to Thomas Simpson (1757/1758) for determining the sum of regularly spaced coefficients in a power series. According to Chrystal (1900, p. 417), Simpson's method was used apparently independently by Waring in 1784. It follows at once from (5.25) (Good, 1955) that if f is "band-limited" so that $c_n = 0$ when $n \geq m$, say, then

$$(5.27)\qquad \frac{1}{2\pi}\int_0^{2\pi} f(\theta)\, d\theta = \frac{1}{t}\sum_{r=0}^{t-1} f\left(\frac{2\pi r}{t}\right)$$

whenever $t \geq m$. One of many deductions is that the Legendre polynomial can be computed, often conveniently, as

$$(5.28)\qquad P_N(x) = \frac{1}{t}\sum_{r=0}^{t-1}\left[x + \sqrt{(x^2-1)}\cos\frac{2\pi r}{t}\right]^N \quad (t > N).$$

When $t \to \infty$ this tends to a well known integral due to Laplace, namely

$$(5.29)\qquad P_N(x) = \frac{1}{\pi}\int_0^{\pi} [x + \sqrt{(x^2-1)}\cos\theta]^N\, d\theta$$

(Whittaker and Watson, 1927, pp. 312–314, who cite Laplace's *Mécanique Céleste*, Livre xi, Chap. 2). Thus (5.28) can be regarded as a generalization of Laplace's formula. It is surprising that this generalization was apparently not published before 1954.

As a closely related deduction from (5.27), consider a "trinomial" random walk on the integers with probabilities a, b, c, respectively, of moving, at each stage, one place to the left, staying still, or one place to the right. Then, starting at the origin, the probability of arriving at position s at time n is

$$(5.30)\qquad \frac{1}{t}\sum_{r=0}^{t-1} (a\omega^{-r} + b + c\omega^r)^n \omega^{-rs}$$

whenever $t > |s| + n$, where $\omega = \exp(2\pi i/t)$. Thus the probability distribution is the discrete Fourier transform of $(a\omega^{-r} + b + c\omega^{-r})^n$. If $a = c = p$ and $s = 0$ (return to the origin) the probability is

$$(5.31)\qquad \frac{1}{t}\sum_{r=0}^{t-1}\left[1 - 4p\sin^2\frac{\pi r}{t}\right]^n \quad (t > n),$$

and, by letting $t \to \infty$, we see this is equal to

$$(5.32)\qquad \int_0^1 (1 - 4p\sin^2 \pi x)^n\, dx,$$

but (5.31) is easier to compute. The relationship between Legendre polynomials and trinomial random walks was discussed by Good (1958b).

6. THE POISSON DISTRIBUTION

The Poisson distribution is of course of great importance in modern statistics. Haight (1967, p. 113) cites three authors who credit de Moivre (1718) with the discovery of the distribution although de Moivre did not give the formula $e^{-a}a^n/n!$ explicitly. Stigler (1982) gives a translation of the passage in Poisson (1837, p. 206) which is apparently the only place where Poisson mentions the distribution. He derived it as a limiting form of the negative binomial as did de Moivre. Clearly Poisson was unaware of the importance that the distribution was destined to have. In many textbooks today the distribution is first regarded as an approximation to the binomial.

Was de Moivre unfairly treated from a kudological point of view? ("Kudology" is the science of assigning credit or kudos; cf. Good, 1962c, p. 3.) Well, the formula $e^{i\theta} = \cos\theta + i \sin\theta$ is often called de Moivre's formula, but it seems to be due to Cotes (1714, p. 32), although he did not express it neatly at least in this place. This formula is itself so important that rough justice has been done to de Moivre, although perhaps not to Cotes.

Perhaps the Poisson distribution should have been named after von Bortkiewicz (1898) because he was the first to write extensively about rare events whereas Poisson added little to what de Moivre had said on the matter and was probably aware of de Moivre's work. See also Gumbel (1978).

An early example of the Poisson distribution in real data was provided by the number of Prussian soldiers, in various corps and years, who died of Bortkiewicz's disease, that is, were kicked to death by a horse. Bortkiewicz's disease is always fatal by definition. There are many other examples of the Poisson distribution of greater scientific interest. For example, Fisher (1922, p. 89) pointed out that the multinomial distribution can be regarded as generated by several independent Poisson distributions made conditional on a total sample size which itself has a Poisson distribution. This is because the multinomial probability $N! \prod_{r=1}^{t} (p_r^{n_r}/n_r!)$ can be written in the form

$$(6.1) \qquad \prod_{r=1}^{t} \frac{(\lambda p_r)^{n_r} e^{-n_r}}{n_r!} \bigg/ \frac{\lambda^N e^{-N}}{N!}.$$

An important use of the Poisson distribution occurs in Poisson processes in radioactive decay and in the theory of queues which originated in the theory of telephone traffic. Doob (1953, pp. 404–406) mentions applications of the Poisson process to molecular and stellar distributions. Parzen (1962) mentions applications to particle counters, birth processes, renewal processes, shot noise, and Brownian motion. Other applications can be readily found from the various volumes of the *Current Index to Statistics*. The distribution of the duration of a busy period in a queue with Poissonian input can be neatly obtained from branching theory (the Bienyamé-Galton-Watson process), by regarding customers and time elements as forming the two phases of a species with alternating generations (Good, 1951).

It is reasonable to maintain that even de Moivre was anticipated by de Montmort (1708) who discussed the matching problem (or *treize*). If two packs, each of n cards, the cards being labeled $1, 2, \cdots, n$ in each pack, are shuffled and laid out in two rows, the probability of exactly r matches, when $n \to \infty$, tends to $e^{-1}/r!$, that is, the Poisson distribution with mean 1. In fact $n - r$ need not be at all large to give a good approximation. For example, with $n = 10$ and $r = 3$, the probability is 0.061310, whereas $e^{-1}/3! = 0.061313$. Montmort stated the correct result for the matching problem with $r = 0$ in 1708 and published a proof that he received from Nicolas Bernoulli in 1713. The case $r = 0$ contains the only difficulty of the problem (and involves what was probably the first example of the principle of "inclusion and exclusion"), so the fact that de Moivre (1718 or perhaps 1711) gave the result for general r is not by itself a strong argument for his priority. This is an example where the special case is 95% of the job. Such examples are not at all rare.

An interesting application of the Poisson distribution relates to the generation of random digits. If a Poisson variable of mean c is reduced modulo n, let $p_0, p_1, \cdots, p_{n-1}$ be the probabilities of the "digits" 0, $1, \cdots, n-1$ so obtained and let $\rho = p_0^2 + p_1^2 + \cdots + p_{n-1}^2$, the "repeat rate." Then it can be proved that

$$(6.2) \qquad n\rho - 1 = \sum_{r=1}^{n-1} \exp\left[-4c \sin^2\left(\frac{\pi r}{n}\right)\right]$$

which when $n = 2$ reduces to the more easily proved result e^{-4c}.

Since we are commemorating Poisson rather than the Poisson distribution I have perhaps said too much about this distribution although much more could be said. The inclusion of the discussion of this distribution here can be largely justified from the point of view of information retrieval, because statisticians who are interested in the Poisson distribution, and most of them are, might consult the present paper for such a discussion. Then again, Poisson was so great a scientist that we can continue to use the name "Poisson distribution" to help to perpetuate his name. We don't have much choice.

7. THE CAUCHY DISTRIBUTION

Although Poisson did not invent the Poisson distribution, he invented other things not usually ascribed to him. For example, he invented Abel summation in which the sum of a series $\sum a_n$ that is not convergent can sometimes be usefully defined as

$$\lim_{x \to 1-0} \sum a_n x^n. \tag{7.1}$$

In particular, $1 - 1 + 1 - 1 + \cdots$ is equal to ½ with this, and with other sensible definitions; for example, White's advantage in a game of chess is half a move on the average and also by the definition (7.1). Poisson also invented the Cauchy distribution (Stigler, 1974; Poisson, 1824). His purpose was to show that least squares was not necessarily the best estimation procedure; but, as Stigler says, Poisson did not attach much importance to this distribution because measurement "errors" do not have such thick tails if mistakes rather than "errors" are excluded. The Cauchy distribution arises as a ratio of two normal distributions and as the distribution of tan θ where θ is uniform in $(-\pi/2, \pi/2)$. The distribution is now seen as a very special case of Student's t, the square of which in its turn is a special case of Fisher's F which is basic to the Analysis of Variance. The Cauchy and normal distributions are both special cases of the symmetric stable distributions the characteristic functions of which are of the form $\exp(-a\,|\,t\,|^b)$ $(0 < b \leq 2)$. The case $b = 3/2$, half-way between the Cauchy and the normal, is of interest in physics and astronomy and was first noted by Holtsmark (see, for example, Good, 1961).

A Cauchy prior arose in the theory of invariant priors due to Jeffreys (1961, p. 343) and out of a geometrical invariance argument related to stereographic projection in Good (1962d). Apart from invariance arguments, Cauchy priors might be regarded as sensible because of their thick tails, and have been used, for example, by Tiao and Tan (1965), Rogers (1974), and Zellner and Siow (1980). Both univariate and multivariate Cauchy priors were used by Zellner (1984, p. 293ff). Distributions with very thick tails are apt to be more reasonable as priors than as physical distributions.

In my work and joint work concerned with the hierarchical Bayesian approach to categorical data, I have used log-Cauchy hyperpriors for a positive hyperparameter k (for mixing conjugate priors) because these hyperpriors are almost as noncommittal as possible for large k, while still being "proper." (They are asymptotically proportional to $k^{-1}(\log k)^{-2}$.) See, for example, Good (1981a, 1983b) and Good and Crook (1985). In the application to multinomials, k can be regarded as specifying an inductive procedure in the continuum of inductive procedures developed by Carnap (1952) but in effect anticipated by William Ernest Johnson (1924).

I think Poisson would have found these ideas of some interest, because they use his Cauchy distribution in a modern Bayes-Laplace context that explicitly uses two kinds of probability (even if it doesn't have to).

8. THE SEQUENTIAL USE OF BAYES FACTORS

Poisson (1837) was much concerned with legal issues and included a discussion of the effect of several independent witnesses. Suppose that the initial or prior probability of some event is p, and that n independent witnesses (or jurymen) are asked whether it occurred or not, and they all said that it did. Let the probabilities that these witnesses speak the truth be $q_1, q_2, \cdots, q_n$. Then the final or posterior probability of the event is given by Poisson as

$$\frac{p}{p + (1 - p)\rho_1\rho_2 \ldots \rho_n} \tag{8.1}$$

where $\rho_i = (1 - q_i)/q_i$. This formula depends on the over-simplified assumption that witness i will tell the truth with probability q_i whether or not the event occurred.

If the formula is rewritten in terms of odds it assumes the form

$$\text{Final odds} = \text{initial odds} \div (\rho_1\rho_2 \cdots \rho_n) \tag{8.2}$$

and is then seen to be a special case of the principle that independent "Bayes factors" are multiplicative. A Bayes factor in favor of a hypothesis H, provided by evidence E, is defined as the ratio of the final odds of H to its initial odds, that is, $O(H\,|\,E)/O(H)$. ("Odds" means $p/(1 - p)$ where p is a probability.) This Bayes factor is equal to

$$P(E\,|\,H)/P(E\,|\,\bar{H}) \tag{8.3}$$

where $\bar{H}$ denotes the negation of H. When H and $\bar{H}$ are both simple statistical hypotheses, (8.3) reduces to a simple likelihood ratio, but even in this case the Bayesian interpretation has more intuitive appeal.

Legal logic is largely Bayesian and when the evidence becomes overwhelming optional stopping is permissible, that is, it is not necessary for the law to specify in advance exactly how much effort should be expended in collecting the evidence. A simple statistical analogy is the use of sequential sampling for deciding whether a die is loaded. The Bayesian can stop sampling when the final odds that the die is loaded reaches some convincing level. If instead a non-Bayesian decides to sample until some tail probability becomes smaller than some "convincingly small" value, then even a fair die, with probability 1, will be judged to be biased. This is sometimes called "sampling to a foregone conclusion." (See, for example, Greenwood, 1938; Jeffreys, 1939, pp. 359–360; Feller, 1940; Robbins, 1952; Anscombe, 1954; and, for some more history, Good, 1982.) Thus optional stopping is not permitted in non-Bayesian statistics, and this may be regarded as support for the view stated above that the Law is largely Bayesian.

Formula (8.2) is of course not restricted to legal applications; it is applicable whenever evidence arrives in several independent pieces. It is then convenient to use logarithms. A log factor is sometimes called a weight of evidence, a term that is always appropriate but especially so in a legal context (Good, 1985). One notation for it is $W(H:E)$, the weight of evidence in favor of H provided by E. It has the additive property

$$W[H:(E\&F)] = W(H:E) + W(H:E \mid F) \tag{8.4}$$

which does not require that E and F should be independent. The concept of weight of evidence, for the case where numerous independent pieces of evidence are sequentially combined, was of great help in cryptanalysis in World War II and therefore in the destruction of Hitler. The original application was suggested by A. M. Turing in relation to an attack on the engima known as Banburismus. This was a refinement of Rozycki's "clock method" that is described by Rejewski (1981, p. 223). Turing introduced a name, "deciban," for a unit for weight of evidence, and this had immediate intuitive appeal to the Banburists even if they had had no previous statistical training. A Bayes factor F corresponds to 10 $\log_{10}F$ decibans, and a deciban is about the smallest unit of weight of evidence perceptible to the human judgment. Simple concepts, like weights of evidence, decibans, the Poisson distribution, and the distinction between epistemic and physical probability, are often valuable, partly because of their very simplicity. Obversely, as J. E. Littlewood wittily commented, in complicated doctoral theses there is often "less than meets the eye."

9. INDIRECT EFFECTS

In addition to Poisson's direct effect on statistics and probability, some of his work in mathematics and physics has had indirect effects. His summation formula is one example. Another is the Poisson bracket which is important in classical mechanics (Poisson, 1809, p. 281; Goldstein, 1950, pp. 250–272). Its analogue $[u, v] = uv - vu$ in quantum mechanics is even more important especially in Dirac's formulation (Dirac, 1947, pp. 84–89, 112–113). Since quantum mechanics is a statistical theory—at least that is the usual assumption—we can regard the Poisson bracket as having had an indirect but important effect on statistics if "statistics" is broadly interpreted.

In another direction, Stigler (1978, p. 294) states that Edgeworth got his start on the Edgeworth series from Poisson.

10. CONCLUDING COMMENTS

Another French mathematical genius of the early nineteenth century was Evariste Galois who was tragically killed by an expert swordsman in a duel at the age of 20, and might have been tricked into it by the French secret service because of his anti-Royalist activities. A recent article by Rothman (1982) indicates that Galois was psychologically somewhat unbalanced and had been whitewashed in some other accounts in which Poisson had also been blackwashed for having rejected a path-breaking paper by Galois on permutation groups. It seems that Poisson merely demanded that the paper should be rewritten so as to be intelligible, but Galois was too proud or paranoid to accept this demand at its face value. (Two earlier versions of his work had been lost by Cauchy, according to Kline, 1972, p. 756, so some paranoia was justified.) It seems therefore that Rothman's historical research has cleared Poisson's name as far as this incident is concerned.

In conclusion, I would like to quote a comment from the oration of Arago (1854) at Poisson's funeral. Arago said that a genius does not die, he survives in his work. 130 years later we know that Poisson's work and his name will survive as long as civilization does.

ACKNOWLEDGMENTS

I am indebted to S. M. Stigler for some of the references. This work was supported in part by a grant from the National Institutes of Health.

REFERENCES

AITCHISON, J. (1959). A statistical theory of remnants. *J. Roy. Statist. Soc. Ser. B* **21** 158–168.

ANSCOMBE, F. J. (1954). Fixed sample-size analysis of sequential observations. *Biometrics* **10** 89–100.

ARAGO, D. F. J. (1854). Discours prononcé aux funerailles de Poisson, le jeudi 30 avril 1840. In *Oevres complètes de Francois Arago* (D. F. J. Arago, ed.) **2** 690–698. Gide and Baudry, Paris.

BENFORD, F. (1938). The law of anomalous numbers. *Proc. Amer. Philos. Soc.* **78** 551–572.

BORTKIEWICZ, L. VON (1898). *Das Gesetz der Kleinen Zahlen.* Teubner, Leipzig.

CARNAP, R. (1950). *Logical Foundations of Probability.* Univ. of Chicago Press.

CARNAP, R. (1952). *The Continuum of Inductive Methods.* Univ. of Chicago Press.

CAUCHY, A. L. (1817). Sur une loi de reciprocité qui existe entre certaines fonctions. *Bull. Sci. Soc. Philomat. Paris* 121–124.

CHRYSTAL, G. (1900). *Textbook of Algebra, part two.* Adam and Charles Black. Reprinted by Dover, New York, 1961.

CHURCHILL, W. S. (1951). In *The Observer*, August 5; quoted in Cohen and Cohen, 1971, p. 46, quote 67.

COHEN, J. M. and COHEN, M. J. (1971). *A Dictionary of Modern Quotations.* Penguin Books.

COTES, R. (1714). Logometria. *Philos. Trans. Roy. Soc.* **29** 5–54.

COURANT, R. (1962). *Methods of Mathematical Physics* **2**. Interscience, New York.

CZUBER, E. (1899). Die Entwicklung der Wahrscheinlichkeitstheorie und ihre Anwendungen. *Jber. Deutsch. Math.-Verein.* **7** (second part) 1–279. (Cited by Heyde and Seneta, 1977.)

DE FINETTI, B. (1974). *Theory of Probability.* Wiley, New York.

DE MOIVRE, A. (1718). *The Doctrine of Chances.* W. Pearson, London. Based on a version in the *Philos. Trans.* **27** (1711) 213–264 according to Todhunter (1865/1965, p. 136). The third edition of 1756 was reprinted by Chelsea, New York in 1967.

DE MONTMORT, P. R. (1708). *Essai d'Analyse sur les Jeux de Hazards.*

DIRAC, P. A. M. (1947). *The Principles of Quantum Mechanics.* Clarendon, Oxford.

DOOB, J. L. (1953). *Stochastic Processes.* Wiley, New York.

FELLER, W. (1940). Statistical aspects of ESP. *J. Parapsychol.* **4** 271–298.

FELLER, W. (1968). *An Introduction to Probability Theory and Its Applications* **1**, 3rd ed. Wiley, New York.

FELLER, W. (1971). *An Introduction to Probability Theory and Its Applications* **2**, 2nd ed. Wiley, New York.

FÉRON, R. (1978). Poisson, Siméon Denis. *Internat. Encycl. Statist.* **2** 704–706.

FINE, T. L. (1973). *Theories of Probability.* Academic, New York.

FISHER, R. A. (1922). On the interpretation of χ^2 from contingency tables, and the calculation of *P*. *J. Roy. Statist. Soc.* **85** 87–94.

FISHER, R. A. (1942). The theory of confounding in factorial experiments in relation to the theory of groups. *Ann. Eugenics* **11** 341–353.

FISHER, R. A. (1945). A system of confounding for factors with more than two alternatives, giving completely orthogonal cubes and higher powers. *Ann. Eugenics* **12** 283–290.

FOURIER, J. B. J. (1812). Théorie des mouvements de la chaleur dans les corps solides, a memoir awarded a prize by the French Académie des Sciences. Reprinted in Fourier's *La Théorie analytique de la chaleur* (1822).

GAVARRET, J. (1840). *Principes généraux de statistique médicale, ou Developpment des règles qui doivent présider a son emploi.* Bechet, Paris.

GELFAND, A. E. and SOLOMON, H. (1973). Study of Poisson's models for jury verdicts in criminal and civil trials. *J. Amer. Statist. Assoc.* **68** 271–278.

GELFAND, A. E. and SOLOMON, H. (1974). Modeling jury verdicts in the American jury system. *J. Amer. Statist. Assoc.* **69** 32–37.

GELFAND, A. E. and SOLOMON, H. (1975). Analyzing the decision-making process of the American jury. *J. Amer. Statist. Assoc.* **70** 305–310.

GOLDSTEIN, H. (1950). *Classical Mechanics.* Addison-Wesley, Reading, Mass.

GOOD, I. J. (1951). Contribution to the discussion of a paper by D. G. Kendall on the theory of queues. *J. Roy. Statist. Soc. Ser. B* **13** 182–183.

GOOD, I. J. (1955). A new finite series for Legendre polynomials. *Proc. Cambridge Philos. Soc.* **51** 385–388.

GOOD, I. J. (1958a). The interaction algorithm and practical Fourier analysis. *J. Roy. Statist. Soc. Ser. B* **20** 361–372.

GOOD, I. J. (1958b). Legendre polynomials and trinomial random walks. *Proc. Cambridge Philos. Soc.* **54** 39–42.

GOOD, I. J. (1959). Kinds of probability. *Science* **129** 443–447. Reprinted in Peters (1969, 28–37) and in Good (1983a).

GOOD, I. J. (1960). The interaction algorithm and practical Fourier analysis: an addendum. *J. Roy. Statist. Soc. Ser. B* **22** 372–375.

GOOD, I. J. (1961). The real stable characteristic functions and chaotic acceleration. *J. Roy. Statist. Soc. Ser. B* **23** 180–183.

GOOD, I. J. (1962a). Contribution to the discussion of a paper by A. R. Thatcher on inventory control. *J. Roy. Statist. Soc. Ser. B* **24** 35–37.

GOOD, I. J. (1962b). Analogues of Poisson's summation formula. *Amer. Math. Monthly* **69** 259–266.

GOOD, I. J. (1962c). pbis. In *The Scientist Speculates.* Heinemann, London. Paperback, Capricorn Books, New York, 1965, 1–3.

GOOD, I. J. (1962d). Contribution to the discussion of a paper by Charles Stein on confidence sets for the mean of a multivariate distribution. *J. Roy. Statist. Soc. Ser. B* **24** 289–291.

GOOD, I. J. (1965a). Letter concerning Tippett's random numbers. *Amer. Statist.* **19** 43.

GOOD, I. J. (1965b). *The Estimation of Probabilities: An Essay on Modern Bayesian Methods.* MIT Press, Cambridge, Mass.

GOOD, I. J. (1966). How to estimate probabilities. *J. Inst. Math. Appl.* **2** 364–383.

GOOD, I. J. (1975). The number of orderings of *n* candidates when ties are permitted. *Fibonacci Quart.* **13** 11–18.

GOOD, I. J. (1981a). Some history of the hierarchical Bayesian methodology. In *Bayesian Statistics* (J. M. Bernardo, M. H. DeGroot, D. V. Lindley, and A. F. M. Smith, eds.) 489–510 and 512–519 (with discussion). Univ. of Valencia Press.

GOOD, I. J. (1981b). The autocorrelation function $\rho^{|\tau|^c}$. C90 in *J. Statist. Comp. Simulation* **12** 148–152.

GOOD, I. J. (1982). Comment. *J. Amer. Math. Soc.* **77** 342–344.

GOOD, I. J. (1983a). *Good Thinking: the Foundations of Probability and Its Applications.* Univ. of Minnesota Press, Minneapolis.

GOOD, I. J. (1983b). The robustness of a hierarchical model for multinomials and contingency tables. In *Scientific Inference, Data Analysis, and Robustness* (G. E. P. Box, T. Leonard, and C.-F. Wu, eds.) 191–211. Academic, New York.

GOOD, I. J. (1985). Weight of evidence: a brief survey. In *Bayesian Statistics* **2** (J. M. Bernardo, M. H. DeGroot, D. V. Lindley, and A. F. M. Smith, eds.) 249–269. North Holland, Amsterdam.

GOOD, I. J. and CROOK, J. F. (1985). The robustness and sensitivity of the mixed-Dirichlet Bayesian test for 'independence' in contingency tables. Unpublished manuscript.

GOOD, I. J. and TULLOCK, G. (1984). Judicial errors and a proposal for reform. *J. Legal Studies* **13** 289–298.

GOOD, I. J. and TULLOCK, G. (1985). On judicial errors, C221. *J. Statist. Comp. Simulation* **21** 75–76.

GRATTAN-GUINNESS, I. (1972). *Joseph-Fourier 1768–1830.* MIT Press, Cambridge, Mass.

GREENWOOD, J. A. (1938). An empirical investigation of some sampling problems. *J. Parapsychol.* **2** 222–230.

GUMBEL, E. J. (1978). Bortkiewicz, Ladislaus von. *Internat. Encycl. Statist.* **1** 24–27. Reprinted from the *International Encyclopedia of Social Sciences* (1968).

HAIGHT, F. A. (1967). *Handbook of the Poisson Distribution.* Wiley, New York.

HEISS, K. P. (1978). Lexis, Wilhelm. *Internat. Encycl. Statist.* **1** 507–512.

HEWITT, E. and SAVAGE, L. J. (1965). Symmetric measures on Cartesian products. *Trans. Amer. Math. Soc.* **80** 470–501.

HEYDE, C. C. and SENETA, E. (1977). *I. J. Bienaymé: Statistical Theory Anticipated.* Springer, New York.

HIGGINS, J. R. (1985). Five short stories about the cardinal series. *Bull. Amer. Math. Soc. (new series)* **12** 45–89.

JEFFREYS, H. (1939). *Theory of Probability.* 1st ed. Clarendon, Oxford. The 3rd ed. was published in 1961.

JOHNSON, W. E. (1924). Appendix (R. B. Braithwaite, ed.) to Probability: deductive and inductive problems. *Mind* **41** 421–423.

KEMBLE, E. C. (1942). Is the frequency theory of probability adequate for all scientific purposes? *Amer. J. Physics* **10** 6–16.

KEYNES, J. M. (1921). *A Treatise on Probability.* Macmillan, London.

KLINE, M. (1972). *Mathematical Thought from Ancient to Modern Times.* Oxford Univ. Press, New York.

LAPLACE, P. S. DE (1820). *Théorie analytique des probabilités.* 3rd ed. Coucier, Paris. Reprinted as Volume 1 in *Oevres complètes de Laplace.*

LÉVY, P. (1939). L'addition des variables aléatoires définés sur une circonférence. *Bull. Soc. Math. France* **67** 1–41.

LINFOOT, E. H. (1928). A sufficiency condition for Poisson's formula. *J. London Math. Soc.* **4** 54–61.

MORDELL, L. J. (1928). Poisson's summation formula and the Riemann zeta function. *J. London Math. Soc.* **4** 285–291.

OGDEN, C. K. (1932). *Bentham's Theory of Fictions.* Routledge and Kegan Paul, London.

PAIS, A. (1982). *Subtle is the Lord* . . . . Clarendon, Oxford.

PARZEN, E. (1962). *Stochastic Processes.* Holden-Day, San Francisco.

PELZ, W. and GOOD, I. J. (1976). Approximating the lower tail-areas of the Kolmogorov-Smirnov one-sample statistic. *J. Roy. Statist. Soc. Ser. B* **38** 152–156.

PETERS, W. S. (ed.) (1969). *Readings in Applied Statistics.* Prentice-Hall, Englewood Cliffs, N.J.

POINCARÉ, H. (1912). *Calcul des Probabilités.* 2nd ed. Gauthiers-Villars, Paris.

POISSON, S. D. (1809). Mémoire sur la variation des constantes arbitraires dans les questions de mécanique. *J. Ecole Polytech.* **8** (Cahier 15) 266–344.

POISSON, S. D. (1823). Suite du mémoire sur les intégrales définies et sur la sommation des séries: sommation des séries de quantités périodiques. *J. Ecole Polytech.* **12** (Cahier 19) 404–432.

POISSON, S. D. (1824). Sur la probabilité des résultats moyens des observations. *Connaissance des Tems pour l'an 1827*, 273–302. (Reference from Stigler, 1974.)

POISSON, S. D. (1827). Sur le calcul numérique des Intégrales définies. *Mém. Acad. Sci. Inst. France* **6** 571–602.

POISSON, S. D. (1837). *Recherches sur la Probabilité des Jugements.* Bachelier, Paris.

RAIMI, R. A. (1976). The first digit problem. *Amer. Math. Monthly* **83** 521–538.

RAIMI, R. A. (1985). The first digit phenomenon again. *Proc. Amer. Philos. Soc.* **129** 211–219.

RAKTOE, B. L., HEDAYAT, A. and FEDERER, W. T. (1981). *Factorial Designs.* Wiley, New York.

REJEWSKI, M. (1981). How Polish mathematicians deciphered the Enigma. *Ann. Hist. Comput.* **3** 213–234 (with discussion).

ROBBINS, H. E. (1952). Some aspects of the sequential design of experiments. *Bull. Amer. Math. Soc.* **58** 527–535.

ROGERS, J. M. (1974). Some examples of compromises between Bayesian and non-Bayesian statistical methods. Ph.D. thesis, Virginia Polytechnic Institute and State Univ.

ROTHMAN, T. (1982). Genius and biographers: the fictionalization of Evariste Galois. *Amer. Math. Monthly* **89** 84–106.

SAVAGE, L. J. (1954). *The Foundations of Statistics.* 2nd ed. Wiley, New York.

SIMPSON, T. (1757/1758). The invention of a general method for determining the sum of every 2d, 3d, 4th, or 5th etc. term of a series, taken in order; the sum of the whole series being known. *Philos. Trans. Roy. Soc. London* **50** 757–769.

SLOANE, N. J. A. (1973). *A Handbook of Integer Sequences.* Academic, New York.

STIGLER, S. M. (1974). Cauchy and the Witch of Agnessi: an historical note on the Cauchy distribution. *Biometrika* **61** 375–380.

STIGLER, S. M. (1978). Francis Ysidro Edgeworth, Statistician. *J. Roy. Statist. Soc. Ser. A* **141** 287–322 (with discussion).

STIGLER, S. M. (1980). Stigler's law of eponymy. *Trans. N. Y. Acad. Sci. (Ser. II)* **39** 147–158.

STIGLER, S. M. (1982). Poisson on the Poisson distribution. *Statist. Probab. Lett.* **1** 33–35.

STIGLER, S. M. (1983). Who discovered Bayes's theorem? *Amer. Statist.* **37** 290–296.

TIAO, G. C. and TAN, W. Y. (1965). Bayesian analysis of random-effect models in the analysis of variance. I. Posterior distribution of variance-components. *Biometrika* **52** 37–53.

TITCHMARSH, E. C. (1932). *The Theory of Functions.* Clarendon, Oxford.

TITCHMARSH, E. C. (1937). *Introduction to the Theory of Fourier Integrals.* Clarendon, Oxford.

TODHUNTER, I. (1865/1965). *A History of the Mathematical Theory of Probability.* Cambridge; reprinted by Chelsea, New York.

USPENSKY, J. V. (1937). *Introduction to Mathematical Probability.* McGraw-Hill, New York.

VAIHINGER, H. (1911, 1935). *Die Philosophie des Als Ob. System der theoretischen, praktischen und religiosen Fiktionen der Menschheit auf Grund eines idealistischen Positivismus. Mit ein Anhang über Kant und Nietzsche.* 6th ed. F. Meiner, Leipzig. (1920). Parts of this sixth edition were translated by C. K. Ogden in *The Philosophy of 'As if': a System of Theoretical, Practical and Religious Fictions of Mankind.* 2nd ed. Routledge and Kegan Paul, London, 1935.

WALLIS, W. A. and ROBERTS, H. V. (1957). *Statistics: a New Approach.* Methuen, London.

WALSH, W. H. (1967). Kant, Immanuel. *The Encyclopedia Philos.* **4** 305–324.

WESTERGAARD, H. (1932). *Contributions to the History of Statistics.* P. S. King, London. Reprinted by A. M. Kelly, New York, 1969.

WHITTAKER, E. T. (1915). On the functions which are represented by the expansions of the interpolatory theory. *Proc. Roy. Soc. Edinburgh* **35** 181–194.

WHITTAKER, E. T. and WATSON, G. N. (1927). *A Course of Modern Analysis.* 4th ed. Univ. Press, Cambridge.

ZELLNER, A. (1984). *Basic Issues in Econometrics.* Univ. of Chicago Press.

ZELLNER, A. and SIOW, A. (1980). Posterior odds ratios for selected regression hypotheses. In *Bayesian Statistics* (J. M. Bernardo, M. H. DeGroot, D. V. Lindley, and A. F. M. Smith, eds.) 585–603. Univ. of Valencia Press.