Statistical Science
1998, Vol. 13, No. 4, 368–377

Early Sample Measures of Variability

H. A. David

Abstract. This paper attempts a brief account of the history of sample measures of dispersion, with major emphasis on early developments. The statistics considered include standard deviation, mean deviation, median absolute deviation, mean difference, range, interquartile distance and linear functions of order statistics. The multiplicity of measures is seen to result from constant efforts to strike a balance between efficiency and ease of computation, with some recognition also of the desirability of robustness and theoretical convenience. Many individuals shaped this history, especially Gauss. The main contributors to our story are in chronological order, Lambert, Laplace, Gauss, Bienaymé, Abbe, Helmert and Galton.

Key words and phrases: Measures of dispersion, standard deviation, mean deviation, median absolute deviation, mean difference, range, interquartile distance, order statistics, chi-squared distribution, Gauss, Laplace, Bienaymé, Abbe, Helmert.

1. INTRODUCTION

The purpose of this article is to trace the early history of sample measures of dispersion, including their statistical properties and their use in determining the accuracy of estimators. Interest in estimates of variability was stimulated originally by astronomical data. The basic paper by Gauss (1816) already illustrates the use of several measures of dispersion on data representing the error incurred in a set of 48 astronomical observations. In addition to the sample standard deviation we will be considering the mean deviation, the median absolute deviation, the mean difference, the range, the interquartile distance and linear functions of order statistics.

Why so many measures? As will be seen, Gauss already realized the importance of what we now call efficiency and unbiasedness (especially asymptotic). But he was also aware of the desirability of ease of computation and it is mainly this aspect that explains the multiplicity of estimators. Another factor was concern with lack of robustness against outliers. Of course, theoretical convenience was also relevant. As a result of these different considerations orderly progress can hardly be expected, especially since research was often undertaken in ignorance of previous work.

Stigler (1986) has given an excellent description of the struggles of 18th- and 19th-century astronomers and geodesists to arrive at ways of reconciling observations that, due to errors of measurement, produced different estimates of unknown parameters. It needs to be noted that the early concern was with variation due to errors of measurement, rather than with variation in general.

Foremost, but not first, in dealing with this problem of reconciliation was the method of least squares. Although Gauss (1777–1855) had been using the method earlier, Legendre (1752–1833) in 1805 was the first to publish it and to name it, with instant success. Before him, Boscovich (1711–1787) had proposed in 1757 minimizing the sum of the absolute (vertical) deviations from a linear regression line and had provided a geometric solution (Eisenhart, 1961; Sheynin, 1973). Both Legendre and Boscovich were content with advancing methods that had intuitive appeal and that were capable of implementation, at least in simple situations.

What was still needed was a probabilistic framework for these two methods. Without such a basis there is no satisfactory way of assessing the accuracy of the estimates produced. Early attempts in this direction go back to Galileo, Simpson, Mayer and especially Lambert. According to Sheynin (1971), Lambert (1728–1777) set down in 1760 the

H. A. David is Professor Emeritus, Statistical Laboratory and Department of Statistics, Snedecor Hall, Iowa State University of Science and Technology, Ames, Iowa 50011-1210 (e-mail: jeanette@iastate.edu).

following three criteria which still seem reasonable for a theory of errors:

1. The absolute values of errors are finite.
2. The quantity of errors with a given absolute value decreases with increase of this absolute value.
3. The probability of an error depends on its absolute value, not its sign.

The real starting point of the present account is the work of Gauss, who, no doubt inspired by his great older contemporary Laplace (1749–1827), raised the crude early attempts to astonishing heights in his study of the method of least squares. We also describe the relevant contributions of Laplace as well as of later 19th-century investigators such as Bienaymé (1796–1878), Abbe (1840–1905), Helmert (1843–1917) and others. The 20th-century will be visited only to note some developments and extensions of the various statistics put forward earlier. See also the integrated historical account of the theory of errors by Sheynin (1995) and the extensive general historical text by Hald (1998).

2. STANDARD DEVIATION

2.1 Mean Known

In his famous "first proof" (Gauss, 1809) of the method of least squares Gauss, in modern terminology, took the distribution underlying n independent errors of observation to be normal with pdf

$$f(x) = \frac{h}{\sqrt{\pi}} \exp(-h^2x^2), \quad -\infty < x < \infty. \tag{2.1}$$

He pointed out that the positive constant h could be viewed as a measure of accuracy of the observations and he named it the *precision*. This term remained in common use for the next 100 years, until h was gradually replaced by the standard deviation $\sigma = 1/h\sqrt{2}$.

We focus here on Gauss (1816), where he considers the estimation of h. To introduce this work Gauss writes almost defensively:

> In the application of the method of least squares to the determination of the most probable values of those quantities [parameters] on which the observations depend, there is no need at all to know the value of h. Also, the *ratio* of the accuracy of the results to the accuracy of the observations is independent of h. However, knowledge of this quantity is in itself of interest and instructive, and I will therefore show how it is possible to attain such knowledge through the observations themselves.

To effect the estimation, Gauss maximizes the posterior pdf of h under the assumption that its prior is uniform over $(0, \infty)$. Since the posterior pdf is by (2.1) proportional to

$$h^n \exp\Big(-h^2 \sum x_i^2\Big),$$

the maximization gives what Edwards (1974) terms the maximum-probability estimator

$$\hat{h} = \left(\frac{n}{2\sum X_i^2}\right)^{1/2}, \tag{2.2}$$

where, as usual, X_i denotes the random variable corresponding to its realization x_i, $i = 1, \dots, n$. Of course, this is also the maximum likelihood estimator, but the distinction between the methods of maximum probability and maximum likelihood is conceptually important.

Gauss points out that (2.2) holds whether n is large or small, but then obtains some interesting large-sample results for the interval estimation of h. He shows that, for large n, the true value of h lies with probability 1/2 in

$$\big(\hat{h}(1 - cn^{-1/2}), \hat{h}(1 + cn^{-1/2})\big), \tag{2.3}$$

where $c = \Phi^{-1}(0.75)/\sqrt{2} = 0.4769363$, Φ denoting the standard normal cdf. Correspondingly, the long widely used true probable error $r = \Phi^{-1}(0.75)\sigma$, which satisfies $P(-r < X_i < r) = 1/2$, lies with probability 1/2 in

$$\big(\hat{r}(1 - cn^{-1/2}), \hat{r}(1 + cn^{-1/2})\big), \quad \hat{r} = c/\hat{h}. \tag{2.4}$$

We see that (2.3) and (2.4) are in modern terminology large-sample 50% Bayesian confidence intervals, corresponding to (inconsistent!) assumptions of uniform priors, and hence are also ordinary large-sample confidence intervals. The important early results (2.3) and (2.4), ingeniously obtained by Gauss, seem to have received little notice. With our present knowledge that asymptotically $(\sum X_i^2/n)^{1/2}$ has a $N(\sigma, \sigma^2/2n)$ distribution, (2.4) is seen at once.

Gauss goes on to study

$$S_m = \sum_{i=1}^{n} |X_i|^m \tag{2.5}$$

for large n and positive integral m. Evidently using Laplace's central limit theorem, he writes that the most probable value of S_m lies in the interval

$$\Big(n\nu_m - c[2n(\nu_{2m} - \nu_m^2)]^{1/2}, n\nu_m + c[2n(\nu_{2m} - \nu_m^2)]^{1/2}\Big),$$

with probability 1/2, where $\nu_m = E(|X|^m)$ and, as before, $c = \Phi^{-1}(0.75)/\sqrt{2}$. This result, he states, holds for any parent distribution (presumably with finite ν_{2m}). For the normal distribution (2.1), for which $\nu_m = \Gamma((m+1)/2)/h^m\sqrt{\pi}$, Gauss proceeds to demonstrate numerically, for $m \leq 6$, that the length of the interval for the probable error r decreases from $m = 1$ to $m = 2$ and then increases. Of course, a much stronger result can now be stated in view of the sufficiency of S_2 for σ or h.

Also in this paper Gauss suggests the possible use of the mean deviation S_1/n and, surprisingly, of the median absolute deviation, med$|X_i|$, because of the greater computational simplicity of these measures (see Sections 3 and 4).

Gauss was not comfortable with the assumption that had led him to normality of the observations, namely, that for iid observations from a pdf $f(x-\mu)$, the sample mean is the maximum-probability estimator of μ. In a series of stellar papers (Gauss, 1821, 1823, 1826) he developed what has come to be known as the Gauss linear model. These have been translated, with some commentary, by Stewart (1995). An informal English version of all of Gauss's work on least squares, based on Bertrand's (1855) French translation from the original Latin and German, has been prepared by Trotter (1957). Gauss's contributions to least squares have been summarized by, for example, Seal (1967) and Sprott (1978), both using matrix notation. A valuable critical review is given by Sheynin (1979). What concerns us here is just the estimation of the variability of the observations.

Essentially, Gauss assumes the model $X_i = \mu_i + Z_i$, $i = 1, \ldots, n$, where the Z_i are independent with mean zero, unknown variance σ^2 (possibly after appropriate weighting) and a common unknown distribution; the μ_i are linear functions of parameters $\theta_1, \ldots, \theta_k$, $k < n$. Apart from no longer requiring normality of the Z_i, Gauss chooses $\sigma^2 = E(Z_i^2)$ as measure of uncertainty of the deviations $X_i - \mu_i$. The arbitrariness of this measure is fully recognized by him, and he points out that Laplace's use of $E|Z_i|$ is equally arbitrary and "less suited to analytic study" (Gauss, 1821, Section 6). Moreover, a small error incurred twice seems preferable to one twice its size.

In spite of the radically new aim of estimating $\theta_1, \ldots, \theta_k$ so as to minimize σ^2, without distributional assumptions, Gauss arrives again at $\sum Z_i^2/n$ as the appropriate estimator of σ^2. Although aware of the need for a better approach in small samples, he still follows the then standard practice of regarding the residuals $E_i = X_i - \hat{\mu}_i$ as if they were $Z_i = X_i - \mu_i$.

2.2 Mean Unknown

The breakthrough comes in 1823 when Gauss realizes that the estimation of $\theta_1, \ldots, \theta_k$ imposes k linear constraints on the E_i. This enables him to show that the sum of squares of the residuals, $\sum E_i^2$, has to be divided by $n - k$ for an unbiased (not Gauss's term) estimator S^2 of σ^2. However, small-sample unbiasedness is not crucial to Gauss, who advocates S as an estimator of σ. He then undertakes the intricate task of finding the standard error of S^2 for any population (with finite fourth moment).

2.3 Distribution under Normality

We return to the case of independent $N(\mu, \sigma^2)$ variates $X_1, \ldots, X_n$. Gauss did not consider the distribution of S^2. There are fascinating antecedents to Karl Pearson's "discovery" and naming of the χ^2 distribution (Pearson, 1900). See also Hald (1998, pages 633–645).

Ernst Abbe (1840–1905), a man far ahead of his time, made a brilliant meteoric appearance on the statistical scene. His main astounding contribution to statistics was largely forgotten until rediscovered by Sheynin (1966).

Abbe obtained a Ph.D. at 21 from Göttingen with a dissertation on thermodynamics. Two years later he submitted his "habilitation" dissertation (Abbe, 1863), which enabled him to join the faculty at Jena. Referring to Gauss's work on the method of least squares, he proceeds in this paper to derive the distributions of both

$$S_2 = \sum_{i=1}^{n} X_i^2 \quad \text{and} \quad T = \sum_{i=1}^{n}(X_i - X_{i+1})^2,$$

where $X_{n+1} = X_1$, when the X_i are independent variates with pdf (2.1). Abbe's aim is to use S_2 and T as gauges of how compatible the X_i are with the model for specified h. He goes on to obtain the distribution of S_2/T, essentially the first circular serial correlation coefficient.

Our presentation of Abbe's derivations can be brief in view of a fine summary by Kendall (1971). Abbe's approach for S_2 is of interest and facilitated the more difficult arguments needed for T and S_2/T. An English translation of Abbe's paper, with a short introduction, is available from the present writer.

To arrive at what is effectively the χ^2 distribution with n degrees of freedom, Abbe notes that

$$\begin{aligned} &P(S_2 < \Delta) \\ (2.6)\qquad &= \frac{h^n}{\pi^{n/2}} \int \cdots \int \exp(-h^2 y)\, dx_1 \cdots dx_n, \end{aligned}$$

where $y = \sum x_i^2$ and the integration extends over $0 < y < \Delta$. To deal with this integral he introduces the "discontinuity factor"

$$(2.7)\qquad \frac{1}{2\pi}\int_{-\infty}^{\infty}\frac{e^{c(a+\phi i)}}{a+\phi i}\,d\phi, \quad i=\sqrt{-1},$$

which, by contour integration, equals 1 or 0 according as $c > 0$ or $c < 0$. The discontinuity factor, a technique which Abbe acknowledges having learned from Riemann at Göttingen, enables him, upon setting $c = \Delta - y$, to replace (2.6) by

$$\begin{aligned}
P(S_2 < \Delta) &= \frac{1}{2\pi}\frac{h^n}{\pi^{n/2}}\int_{-\infty}^{\infty}\frac{\exp(\Delta(a+\phi i))}{a+\phi i}\,d\phi \\
&\quad\cdot\prod_{j=1}^{n}\left(\int_{-\infty}^{\infty}\exp(-(h^2+a+\phi i)x_j^2)\,dx_j\right) \\
&= \frac{h^n}{\Gamma(n/2)}\int_0^{\Delta}\exp(-h^2 y)y^{(n/2)-1}\,dy,
\end{aligned}$$

after further reduction.

Abbe went on to make his name in optics, working with Carl Zeiss. As noted by Seneta (1983), he wrote only two more papers in statistics, in 1878 and 1895. These were short articles on the properties of counts taken with the new Zeiss haemacytometer. Although immediately useful and anticipating Student (1907), these articles were overlooked by statisticians.

In a way, Abbe was scooped by I. J. Bienaymé (1796–1878) in the discovery of the χ^2 distribution. Like Abbe, Bienaymé was long forgotten by statisticians, although unlike Abbe he was well known as a probabilist in his time (Heyde and Seneta, 1977). In the noteworthy paper that concerns us (Bienaymé, 1852), he carries the title Inspecteur général des Finances. Contrasting with Abbe's deliberate approach, Bienaymé arrives quite incidentally at the integral

$$p = \frac{1}{\pi^{n/2}}\int\prod_{i=1}^{n}\left(\exp(-x_i^2)\,dx_i\right),$$

where the integration extends over $\sum x_i^2 \le \gamma^2$. Clearly, $p = P(\sum Z_i^2 \le 2\gamma^2)$, where the Z_i are independent $N(0,1)$ variates.

Bienaymé's evaluation of the integral is of interest in that he uses only elementary methods, whereas modern classroom proofs of $\sum X_i^2 \sim \chi_n^2$ typically ask the student to accept the uniqueness theorem of moment generating functions; but see also Kruskal (1946).

Noting that, by symmetry about 0, p may be written

$$p = \frac{2n}{\pi^{n/2}}\int_0\prod_{i=1}^{n}\left(\exp(-x_i^2)\,dx_i\right),$$

over $\sum x_i^2 \le \gamma^2$, with $x_1 \ge 0, \ldots, x_n \ge 0$, make the transformation from x_n to $u = (\sum x_i^2)^{1/2}$:

$$x_n = \left(u^2 - x_1^2 - \cdots - x_{n-1}^2\right)^{1/2}.$$

This gives

$$\begin{aligned}
p &= \frac{2^n}{\pi^{n/2}}\int_0^{\gamma}u\exp(-u^2)\,du \\
&\quad\cdot\int_0\frac{dx_1\cdots dx_{n-1}}{(u^2-x_1^2-\cdots-x_{n-1}^2)^{1/2}},
\end{aligned}$$

where $\int_0$ now extends over $x_1^2+\cdots+x_{n-1}^2 \le u^2$, with $x_1 \ge 0, \ldots, x_{n-1} \ge 0$.

To deal with the inner integral, Bienaymé makes repeated use of the beta-integral-type result

$$\int_0^{\sqrt{a}}(a-x^2)^{p/2}\,dx = \frac{\sqrt{\pi}}{2}a^{(p+1)/2}\frac{\Gamma(p/2+1)}{\Gamma(p/2+3/2)}$$

by successively setting

$$a = u^2 - x_1^2 - \cdots - x_i^2 \quad \text{for } i = n-2, n-3, \ldots, 1, 0.$$

This reduces the inner integral to

$$u^{n-2}\cdot\frac{\pi^{n/2}}{2^n}\cdot\frac{2}{\Gamma(n/2)}$$

and gives Bienaymé's final result

$$p = \frac{2}{\Gamma(n/2)}\int_0^{\gamma}u^{n-1}\exp(-u^2)\,du,$$

equivalent to showing that

$$p = P\left(\chi_n^2 \le 2\gamma^2\right).$$

As Lancaster (1966) points out, the series expansions for p developed by Bienaymé are essentially those found independently by Pearson (1900).

The prominent German geodesist F. R. Helmert (1843–1917) was evidently unaware of both Abbe's and Bienaymé's derivations and produced yet another proof, by induction (Helmert, 1876a). Much more important, however, is his proof (Helmert, 1876b) that if $X_1, \ldots, X_n$ are independent $N(\mu, \sigma^2)$ variates, then $\sum(X_i - \overline{X})^2/\sigma^2 \sim \chi_{n-1}^2$.

Helmert begins with a transformation from $x_1, \ldots, x_n$ to $u_j = x_j - \overline{x}$, $j = 1, \ldots, n-1$, and $\overline{x}$, and then makes the further transformation

$$v_j = \sqrt{\frac{j+1}{j}}\left[u_j + \frac{1}{j+1}(u_{j+1}+\cdots+u_{n-1})\right], \qquad j = 1, \ldots, n-2,$$

$$v_{n-1} = \sqrt{\frac{n}{n-1}}u_{n-1}.$$

It is the combination of these two transformations, together with $v_n = \sqrt{n}\bar{x}$ and the relabeling $x_n \to x_1$, $x_j \to x_{j+1}$, $j = 1, \ldots, n-1$, that results in the useful orthogonal transformation involving $n-1$ successive linear contrasts

$$v_j = [j(j+1)]^{-1/2}(jx_{j+1} - x_1 - \cdots - x_j), \qquad j = 1, \ldots, n-1$$

and

$$v_n = (x_1 + \cdots + x_n)/\sqrt{n},$$

that has become known as a Helmert transformation. Since $V_1, \ldots, V_{n-1}$ are independent $N(0, \sigma^2)$ variates and

$$\sum_1^n (X_i - \overline{X})^2 = \sum_1^{n-1} V_j^2,$$

the result follows at once.

Although almost there, Helmert fails to note the independence of $\overline{X}$ and S^2. This independence is explicitly pointed out in Fisher (1920), where Fisher provides a swift geometric proof of the distribution of $(n-1)S^2$. Pearson (1931) apologizes for Helmert's work having been overlooked by the "English school of statisticians" and gives a detailed exposition of Helmert's argument, with a slight geometric addition. He suggests that the distribution should in future be named for Helmert.

Fisher first shows awareness of Helmert's work in Fisher (1934), but there and subsequently cites only Helmert (1875), a note that merely announces that, essentially,

$$\sum (X_i - \mu)^2/\sigma^2 \sim \chi_n^2.$$

Fisher's first-hand knowledge of the relevant Helmert (1876b) must be in doubt, especially as in Fisher (1939) he is clearly unaware that Helmert had found the joint distribution of S^2 and $\overline{X}$ on the way to finding the distribution of S^2.

Helmert continues to use S as the "most favorable" (*günstigster*) estimator of σ. His considerable contributions to statistics have been extensively reviewed by Sheynin (1995).

3. MEAN DEVIATION

3.1 Mean and Median

Given observations $x_1, \ldots, x_n$ we can distinguish two mean deviations: the mean absolute deviation from the mean $\bar{x}$ and from the median m, given by $nd = \sum |x_i - \bar{x}|$ and $nd' = \sum |x_i - m|$. In a way to be made more precise shortly, m is the value of μ minimizing $\sum |x_i - \mu|$ and, of course, $\bar{x}$ is the value of μ minimizing $\sum (x_i - \mu)^2$. These are very special cases of L_1- and L_2-estimation, respectively.

Laplace (1818) compares the large-sample behavior of L_1- and L_2-estimation in a situation that includes as a special case the estimation of the center of a symmetric distribution by the median and by the mean.

Assuming that his central-limit theorem applies, Laplace uses an order-statistics-type argument to show that in large samples the median, like the mean, is approximately normally distributed. By comparing their two large-sample densities Laplace shows that the median is superior to the mean (i.e., the pdf of M is more concentrated about μ than the pdf of $\overline{X}$) if, in modern notation,

$$f(0) > 1/(2\sigma), \tag{3.1}$$

where the parent density $f(x)$ is taken to be symmetric about zero, with variance σ^2. Clearly, Laplace has come impressively close to the notion of asymptotic relative efficiency. He goes on to point out that for a normal distribution $\overline{X}$ is superior. Although Laplace provides no other example, his language implies that he regards (3.1) as quite possible. Nevertheless, he continues to call L_2-estimation "the most advantageous method," leaving the reader to infer that his earlier claims of optimality for L_2 and his choice of terminology were unfortunate. See also Stigler (1973a).

3.2 Mean Absolute Deviation from the Median

As a special case of a more general result in Laplace (1799, Section 40) and repeated in Laplace (1818), it follows that $\sum |x_i - \mu|$ is minimized by $\mu = m = x_{((n+1)/2)}$, n odd, and by m in $[x_{(n/2)}, x_{(n/2+1)}]$, n even.

It seems natural therefore to regard $D' = \sum |X_i - M|/n$ as a basis for estimating variability. But D' was difficult to handle in the 19th century and in any case turns out to be not very promising.

Since

$$nD' = \begin{cases} X_{(n)} + \cdots + X_{((n+3)/2)} \\ \quad - X_{((n-1)/2)} - \cdots - X_{(1)}, & n \text{ odd}, \\ X_{(n)} + \cdots + X_{(n/2+1)} \\ \quad - X_{(n/2)} - \cdots - X_{(1)}, & n \text{ even}, \end{cases}$$

the mean and variance of D' can be found from tables of the first two moments of the order statistics. In the normal case D' was found to be less efficient than $D = \sum |X_i - \overline{X}|/n$ (e.g., Godwin, 1949). On the other hand, one would expect D' to do well for a

Laplacian pdf

$$f(x) = \frac{1}{2\theta}\exp\left(-\left|\frac{x-\mu}{\theta}\right|\right), \quad \theta > 0,$$

since M and D' are MLE's for μ and θ.

However, even here D' is not optimal and better estimators of θ that are linear in the order statistics can be found (Govindarajulu, 1966).

3.3 Mean Absolute Deviation from the Mean

Commonly known simply as the mean deviation, $d = \sum|x_i - \overline{x}|/n$ was long motivated primarily by being easier to compute than the standard deviation s. Gauss (1816) already made essentially this point when comparing $\sum|x_i - \mu|/n$ and $[\sum(x_i - \mu)^2/n]^{1/2}$ (Section 2.1) and Pearson (1945) still cites ease of computation of d in routine situations when introducing tables of the cdf of D in normal samples obtained by Godwin and Hartley (1945).

The editor of *Astronomische Nachrichten* seems to have been the first to find, in present terminology, that (Peters, 1856)

$$E(D) = \sigma\sqrt{\frac{n-1}{n}}\cdot\sqrt{\frac{2}{\pi}}.$$

Writing $X_i = \mu + \varepsilon_i$, so that, for example,

$$X_1 - \overline{X} = \frac{n-1}{n}\varepsilon_1 - \frac{\varepsilon_2}{n} - \cdots - \frac{\varepsilon_n}{n},$$

he was lucky to have obtained the correct answer, tacitly assuming that $X_i - \overline{X}$ is normally distributed. The normality of a sum of independent normal variates was laboriously proved by Helmert (1876b), although this result cannot have been unknown (see Hald, 1998, page 634). In the same paper Helmert also succeeds in deriving the variance of D by use of a discontinuity factor (cf. Section 5 and see Hald, 1998, page 642).

It is of interest to return here to Fisher (1920), already cited in Section 3. This early paper of Fisher's is important for introducing an example of a sufficient statistic. Fisher shows that if in a sample from a $N(\mu, \sigma^2)$ population the value of S is known, then D or any other statistic can shed no further light on σ. Preceding this result Fisher expeditiously derives the distribution of $(n-1)S^2$ and the variance of D. Later (Fisher, 1950) he acknowledges that both these results, although unknown to him in 1920, were not new then. However, he manages not to mention Helmert's name at this point. In Fisher (1939) he makes clear his low opinion of D for which "the only recommendation seems to be that for some types of work it is more expeditious than the use of squares." Fisher continues that in preoccupation with D "Helmert seems to have lost sight of the value of his discovery respecting the mean square."

Pace Fisher and his strong brief for the normal distribution (e.g., Fisher, 1939, page 2), it has also long been recognized that D is less sensitive to outliers than S, a point very forcefully made by Tukey (1960). Whereas for a sample from a $N(\mu, \sigma^2)$ population D has asymptotic efficiency 0.88 relative to S in estimating σ, the situation changes drastically if some contamination by a wider normal, say $N(\mu, 9\sigma^2)$, is present: as little as 0.008 of the wider population will render D asymptotically superior.

4. MEDIAN ABSOLUTE DEVIATION

As already mentioned, Gauss (1816) briefly treats $M = \text{med}|X_i|$, where $X_1, \ldots, X_n$ is a random sample from (2.1). In fact, Gauss states without proof that, for large n, the probable error $r = \Phi^{-1}(0.75)\sigma$ ($\sigma = 1/h\sqrt{2}$) is the "most probable" value of M and that M lies in $r(1 \pm \exp(\rho^2)\sqrt{\pi/(8n)}))$ with probability 1/2, where $\rho = rh = \Phi^{-1}(0.75)/\sqrt{2}$.

A detailed proof, by Dirichlet (1805–1859), is provided in Encke (1834). Writing $\psi(x)$ for the pdf of $|X|$, where X has an unspecified distribution, Dirichlet shows that the likelihood of M is still maximized asymptotically when M equals the probable error r. Here $r = \Psi^{-1}(1/2)$, the median of the cdf of $|X|$. Dirichlet goes on to establish the asymptotic normality of M, without being aware that this follows from Laplace (1818) (Section 3.1). He then obtains the interval containing M with probability 1/2 as $(r - \delta, r + \delta)$, where $\delta = rh/(\sqrt{2n}\psi(r))$. In the normal case, $\psi(x) = (2h/\sqrt{\pi})\exp(-h^2x^2)$, Gauss's formula results.

M requires 272 observations to achieve the same efficiency (interval) as does S for 100 observations. See also Harter (1978) on Gauss (1816).

5. MEAN DIFFERENCE

Some 50 years after the path-breaking statistical researches of Gauss a lively sequence of papers on the estimation of dispersion begins in the German journal *Astronomische Nachrichten* (*Astronomical News*). Triggering this activity is the bumbling figure of W. Jordan, professor at Carlsruhe. He proposes to improve on Gauss by basing the estimation not on the n observations $x_1, \ldots, x_n$ themselves, but rather on their $n(n-1)/2$ absolute differences $|x_i - x_j|$, $i, j = 1, \ldots, n$, $i \neq j$. Specifically he proposes, in modern notation, the statistic (Jordan, 1869)

$$G_k = \frac{\sum_{i<j}^n |X_i - X_j|^k}{n(n-1)/2},$$

for $k = 1$ or 2. Having in mind an underlying normal distribution, Jordan thinks he has improved precision on Gauss's root-mean-square estimator essentially by a factor of $[n(n-1)/2]^{1/2}$ to $(n-1)^{1/2}$, in view of the respective divisors!

Jordan's paper is dated July 4, 1869. It appears that Herr von Andrae, privy counselor in Kopenhagen, cannot wait to register his protest in German and on August 11 fires off a letter to the editor in Danish (von Andrae, 1869): Jordan has overlooked the dependence of the differences; moreover, G_2 is just $2S^2$.

Nevertheless, Jordan has started something: G_1, henceforth just G, seems to deserve further attention. Von Andrae (1872) makes a detailed study of G, obtaining its mean and standard error. He notes that g may be conveniently calculated from

$$g = \sum(n-2i+1)(x_{(n+1-i)} - x_{(i)}), \tag{5.2}$$

where the summation runs from 1 to the integral part of $n/2$. The computational convenience misleads von Andrae, however, in using (5.2) for theoretical purposes, resulting in long and not quite complete proofs of the results for a normal distribution with s.d. σ:

$$E(G) = \frac{2\sigma}{\sqrt{\pi}}, \tag{5.3}$$

$$\begin{aligned}\text{s.d. } (G) = {} & \frac{2\sigma}{[n(n-1)\pi]^{1/2}} \\ & \cdot \left[n\left(\frac{1}{3}\pi + 2\sqrt{3} - 4\right)\right. \\ & \left. + \left(6 - 4\sqrt{3} + \frac{1}{3}\pi\right)\right]^{1/2}.\end{aligned} \tag{5.4}$$

From (5.4) von Andrae shows essentially that the asymptotic efficiency of G (w.r.t. S) is 97.8% and writes that therefore G, given its much simpler calculation, is a serious competitor to the usually preferred S.

Helmert (1876b) firmly establishes (5.3) and (5.4) (apart from two canceling minor errors in his proof!) by use of (5.1) and the elegant representation (compare (2.7))

$$|x_i - x_j| = \frac{2}{\pi}\int_0^\infty (x_i - x_j)\sin[u(x_i - x_j)]\frac{du}{u}.$$

These results in *Astronomische Nachrichten* have often been overlooked by later writers. In fact, G is generally known as Gini's mean difference, Gini (1912) being an influential paper in which Gini introduced G as an index of variability in a *population* consisting of $x_1, \ldots, x_n$. Actually, Gini states that he became aware of the earlier work after completing his own. See also David (1968).

6. RANGE

The range $W_n = X_{(n)} - X_{(1)}$ is the most obvious estimator of variability. According to Ptolemy (Harter, 1978), it was used already in the second century B.C. by Hipparchus, who estimated the maximum variation in his observations on the length of the year by half their range. However, the range remained only a descriptive statistic until the 20th century. In view of its importance, especially in quality control, we now trace its rather gradual development.

Theoretical research on the range was stimulated by pioneering work of von Bortkiewicz (1922a, b), who concentrated on $E(W_n)$ in normal samples (see Harter, 1978). Von Mises (1923) soon pointed out that, for any distribution, $E(W_n) = E(X_{(n)}) - E(X_{(1)})$ and that, in generalization of von Bortkiewicz's result, $E(W_n)$ is given by the Stieltjes integral

$$E(W_n) = \int_{-\infty}^{\infty} x\{dF^n(x) - d[1-F(x)]^n\}. \tag{6.1}$$

Moreover, von Mises initiated asymptotic theory, proving that for a continuous distribution for which $E|X| < \infty$ and

$$\lim_{n\to\infty} \frac{1-F(x+c)}{1-F(x)} = 0 \quad (c \text{ any positive constant}),$$

one has

$$\lim_{n\to\infty} \frac{E(X_{(n)})}{F^{-1}(1-1/n)} = 1.$$

Important progress was made by Tippett (1925), who using quadrature on the formula

$$E(W_n) = \int_{-\infty}^{\infty} \{1 - [1-F(x)]^n - F^n(x)\}\,dx, \tag{6.2}$$

computed $E(W_n)$ when $F(x) = \Phi(x)$, the standard normal cdf, for $n = 2(1)1{,}000$. Equation (6.2) follows from (6.1), but was obtained otherwise by Tippett, who also developed formulae for the higher moments of W_n and obtained some approximate numerical values. With $d_n = E(W_n)$, one now has an unbiased range estimator σ^* of σ, namely, $\sigma^* = \overline{W}_n/d_n$, where $\overline{W}_n$ is the mean of k samples, each of size n. In Shewhart style, control charts for the mean could now use control limits at, for level α,

$$\bar{\bar{x}} \pm \frac{3\sigma^*}{\sqrt{n}} = \bar{\bar{x}} \pm A_\alpha \overline{w},$$

where $\bar{\bar{x}}$ is the grand mean of k samples of n, mean range replaces standard deviation and A_α is a widely tabulated constant.

Dealing with the distribution of W_n still took some effort. Supplementing Tippett's results, Pearson (1932) was able to obtain approximate percentage points of W_n in normal samples by fitting his father's curves. This made possible approximate control charts for the range. Publication of constants needed in the construction of control charts for both μ and σ ensued in the influential manual by Pearson (1935).

The exact pdf of W_n for samples from any continuous distribution was first derived by McKay and Pearson (1933), but the more convenient formula for the cdf

$$P(W_n \le w) = n \int_{-\infty}^{\infty} [F(x+w) - F(x)]^{n-1} f(x)\, dx$$

had to wait for Hartley (1942). Detailed tables of the cdf for $F = \Phi$ were then developed in Pearson and Hartley (1942). Also given were percentage points for $n \le 12$ which made it possible to construct exact control limits for the range.

7. INTERQUARTILE RANGE AND LINEAR FUNCTIONS OF ORDER STATISTICS

The interquartile range (IQR) or interquartile distance of a population with cdf $F(x)$ is $F^{-1}(0.75) - F^{-1}(0.25)$. Since for a symmetric distribution with mean μ this is twice the probable error $r = F^{-1}(0.75) - \mu$, the IQR can be regarded as having a very long history (see Hald, 1998, page 360). If $X \sim N(0, \sigma^2)$, Gauss (1816) estimated r from a sample $x_1, \ldots, x_n$ as in (2.4). However, nonparametric estimation of r, directly from the (grouped) frequency distribution of the x's, seems to have been first attempted by the Belgian "father of biometry" Adolphe Quetelet (Quetelet, 1846, Letter 18).

Direct consideration of the interquartile range and the interdecile range began with Galton (1882), who coined both terms. His approach is to arrange the observations in increasing order of magnitude and then to remove the desired fraction from each end, using interpolation to obtain the exact cutoff points.

Although Galton recognized the greater stability of the more central quantiles, his "interquantile" ranges were necessarily purely descriptive statistics. From formulae for the variances and covariances of sample quantiles $X_{(n\lambda)}$, $0 < \lambda < 1$, Pearson (1920) points out that while (1/2)IQR has the advantage of directly estimating the probable error, other pairs of symmetrically spaced sample quantiles can provide more efficient estimators of variability in the normal case. Specifically, he recommends, as estimator of σ,

$$(X_{(n\lambda_2)} - X_{(n\lambda_1)})/2.93050,$$

with $\lambda_1 = 1 - \lambda_2 = 1/14$. This has asymptotic efficiency (AE) 65% as against

$$(X_{((3/4)n)} - X_{((1/4)n)})/1.34898,$$

with AE 37%.

The asymptotic estimation of both location and scale parameters by general linear functions of the order statistics was considered in a pioneering paper by Daniell (1920); see Stigler (1973b).

Such investigations were greatly advanced when Mosteller (1946) showed that normalized sample quantiles are asymptotically multivariate normal (see also David, 1992). Much more efficient estimators of σ that are linear functions of a larger number of sample quantiles are developed by Ogawa (1951).

Major progress in the small-sample estimation by linear functions of order statistics was made possible by the advent of high-speed computers. Tables of means, variances and covariances of order statistics were prepared for many (standardized) distributions, typically for $n \le 20$. With the help of these it has become easy to estimate the parameters of distributions which, like the normal, depend only on parameters of location and scale. Using a generalized least-squares approach, Lloyd (1952) was able to give expressions for the linear functions of order statistics that are of minimum variance for σ within the class of unbiased estimators that are linear in the order statistics; likewise for the estimation of μ.

For these and related developments see, for example, Sarhan and Greenberg (1962), Harter (1970), David (1981), Hoaglin, Mosteller and Tukey (1983), Balakrishnan and Cohen (1991) and Arnold, Balakrishnan and Nagaraja (1992).

ACKNOWLEDGMENTS

Although in nearly all cases basing my account on original papers (occasionally in translation), I am indebted to all previous writers on the history of statistics cited in the references. This applies especially to Harter (1978), a remarkable labor of love. I am also grateful to Oscar Sheynin, the referees and the Editor for very helpful comments on earlier versions of this paper. The extraordinary recent historical text by Hald (1998) has allowed me to refer the reader to this account for further details on several topics. More information on many of the individuals mentioned can be found there and in Johnson and Kotz (1997). Typically longer accounts of historic figures are given in the *Dictionary of Scientific Biography*.

REFERENCES

ABBE, E. (1863). Ueber die Gesetzmässigkeit in der Vertheilung der Fehler bei Beobachtungsreihen. In *Gesammelte Abhandlungen* **2**. Fischer, Jena.

ARNOLD, B. C., BALAKRISHNAN, N. and NAGARAJA, H. N. (1992). *A First Course in Order Statistics*. Wiley, New York.

BALAKRISHNAN, N. and COHEN, A. C. (1991). *Order Statistics and Inference. Estimation Methods*. Academic Press, New York.

BERTRAND, J. (1855). *Méthode des Moindres Carrés. Mémoire sur la Combinaison des Observations, par Ch.-Fr. Gauss*. Mallet-Bachelier, Paris.

BIENAYMÉ, I.-J. (1852). Sur la probabilité des erreurs d'aprés la méthode des moindres carrés. *J. Math. Pures Appl.* **17** 33–78.

DANIELL, P. J. (1920). Observations weighted according to order. *Amer. J. Math.* **42** 222–236.

DAVID, H. A. (1968). Gini's mean difference rediscovered. *Biometrika* **55** 573–575.

DAVID, H. A. (1981). *Order Statistics*, 2nd ed. Wiley, New York.

DAVID, H. A. (1992). Introduction to Mosteller (1946), "On some useful 'inefficient' statistics." In *Breakthroughs in Statistics* (S. Kotz and N. L. Johnson, eds.) **2** 203–208. Springer, New York.

DAW, R. H. and PEARSON, E. S. (1972). Studies in the history of probability and statistics. XXX. Abraham De Moivre's 1733 derivation of the normal curve: a bibliographical note. *Biometrika* **59** 677–680.

EDWARDS, A. W. F. (1974). The history of likelihood. *Internat. Statist. Rev.* **42** 9–15.

EISENHART, C. (1961). Boscovich and the combination of observations. In *Roger Joseph Boscovich* (L. L. Whyte, ed.) 200–212. Allen and Unwin, London. [Reprinted in Kendall and Plackett (1977) 88–100.]

ENCKE, J. F. (1834). Über die Methode der kleinsten Quadrate. *Berliner Astronomisches Jahrbuch* **1834** 249–312.

FISHER, R. A. (1920). A mathematical examination of the methods of determining the accuracy of an observation by the mean error, and by the mean square error. *Monthly Notices of the Royal Astronomical Society* **80** 758–770. [Reprinted as Paper 2 in Fisher (1950).]

FISHER, R. A. (1934). *Statistical Methods for Research Workers*, 5th ed. Oliver and Boyd, Edinburgh.

FISHER, R. A. (1939). "Student." *Annals of Eugenics* **9** 1–9.

FISHER, R. A. (1950). *Contributions to Mathematical Statistics*. Wiley, New York.

GALTON, F. (1882). Report of the Anthropometric Committee. In *Report of the 51st Meeting of the British Association for the Advancement of Science*, *1881* 245–260. John Murray, London.

GAUSS, C. F. (1809). *Theoria Motus Corporem Coelestium*. [English Translation (1857) by C. H. Davis. Little, Brown, Boston.]

GAUSS, C. F. (1816). Bestimmung der Genauigkeit der Beobachtungen. [Reprinted in (1880) *Carl Friedrich Gauss Werke* **4** 109–117. Königliche Gesellschaft der Wissenschaften, Göttingen.]

GAUSS, C. F. (1821, 1823, 1826). Theoria combinationis observationum erroribus minimis obnoxiae. [Reprinted in (1880) *Carl Friedrich Gauss Werke* **4** 3–93. Königliche Gesellschaft der Wissenschaften, Göttingen. For English translation see Stewart (1995).]

GINI, C. (1912). Variabilità e mutabilità, contributo allo studio delle distribuzioni e delle relazione statistiche. *Studi Economico-Giuredici dell' Università di Cagliari* **3**(part 2) i–iii, 3–159.

GODWIN, H. J. (1949). On the estimation of dispersion by linear systematic statistics. *Biometrika* **36** 92–100.

GODWIN, H. J. and HARTLEY, H. O. (1945). On the distribution of the estimate of the mean deviation obtained from samples from a normal population. *Biometrika* **32** 254–265.

GOVINDARAJULU, Z. (1966). Best linear estimates under symmetric censoring of the parameters of a double exponential population. *J. Amer. Statist. Assoc.* **61** 248–258. [Correction (1976) **71** 255.]

HALD, A. (1998). *A History of Mathematical Statistics from 1750 to 1930*. Wiley, New York.

HARTER, H. L. (1970). *Order Statistics and Their Uses in Testing and Estimation* **1**, **2**. U.S. Government Printing Office, Washington, D.C.

HARTER, H. L. (1978). *A Chronological Annotated Bibliography of Order Statistics* **1**. *Pre-1950.* U.S. Government Printing Office, Washington, DC. [Reprinted by American Sciences Press, Syracuse, NY (1983)].

HARTLEY, H. O. (1942). The range in random samples. *Biometrika* **32** 334–348.

HELMERT, F. R. (1875). Ueber die Berechnung des wahrscheinlichen Fehlers aus einer endlichen Anzahl wahrer Beobachtungsfehler. *Zeitschrift für Mathematik und Physik* **20** 300–303.

HELMERT, F. R. (1876a). Ueber die Wahrscheinlichkeit der Potenzsummen der Beobachtungsfehler und über einige damit im Zusammenhange stehende Fragen. *Zeitschrift für Mathematik und Physik* **21** 192–218.

HELMERT, F. R. (1876b). Die Genauigkeit der Formel von Peters zur Berechnung des wahrscheinlichen Fehlers directer Beobachtungen gleicher Genauigkeit. *Astronom. Nachr.* **88** 113–132.

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

HOAGLIN, D. C., MOSTELLER, F. and TUKEY, J. W., eds. (1983). *Understanding Robust and Exploratory Data Analysis*. Wiley, New York.

JOHNSON, N. L. and KOTZ, S., eds. (1997). *Leading Personalities in Statistical Sciences*. Wiley, New York.

JORDAN, W. (1869). Ueber die Bestimmung der Genauigkeit mehrfach wiederholter Beobachtungen einer Unbekannten. *Astronom. Nachr.* **74** 209–226.

KENDALL, M. G. (1971). Studies in the history of probability and statistics. XXVI. The work of Ernst Abbe. *Biometrika* **58** 369–373.

KENDALL, M. G. and PLACKETT, R. L., eds. (1977). *Studies in the History of Statistics and Probability* **2**. Griffin, London.

KRUSKAL, W. H. (1946). Helmert's distribution. *Amer. Math. Monthly* **53** 435–438.

KRUSKAL, W. H. and STIGLER, S. M. (1997). Normative terminology: "normal" in statistics and elsewhere. In *Statistics and Public Policy* (B. D. Spencer, ed.) 85–111. Clarendon, Oxford.

LANCASTER, H. O. (1966). Forerunners of the Pearson χ^2. *Austral. J. Statist.* **8** 117–126.

LAPLACE, P. S. (1799). *Traité de Mécanique Celeste* **3**. Duprat, Paris.

LAPLACE, P. S. (1818). *Théorie Analytique des Probabilités, Deuscième Supplément*. [Included in Laplace (1820).]

LAPLACE, P. S. (1820). *Théorie Analytique des Probabilités*, 3rd ed. Courcier, Paris. [Reprinted as (1847) *Oeuvres de Laplace* **7**. Imprimerie Royale, Paris.]

LLOYD, E. H. (1952). Least-squares estimation of location and scale parameters using order statistics. *Biometrika* **39** 88–95.

MCKAY, A. T. and PEARSON, E. S. (1933). A note on the distribution of range in samples of n. *Biometrika* **25** 415–420.

MOSTELLER, F. (1946). On some useful "inefficient" statistics. *Ann. Math. Statist.* **17** 377–408.

OGAWA, J. (1951). Contributions to the theory of systematic statistics, I. *Osaka Math. J.* **3** 175–213.

PEARSON, E. S. (1932). The percentage limits for the distribution of range in samples from a normal population. *Biometrika* **24** 404–417.

PEARSON, E. S. (1935). The application of statistical methods to industrial standardization and quality control. No. 600, British Standards Institution.

PEARSON, E. S. (1945). The probability integral of the mean deviation. Editorial note. *Biometrika* **32** 252–253.

PEARSON, E. S. and HARTLEY, H. O. (1942). The probability integral of the range in samples of n observations from a normal population. *Biometrika* **32** 301–310.

PEARSON, K. (1900). On the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling. *Philosophical Magazine, 5th Series* **50** 157–175.

PEARSON, K. (1920). On the probable errors of frequency constants, III. *Biometrika* **13** 113–132.

PEARSON, K. (1931). Historical note on the distribution of the standard deviations of samples of any size drawn from an indefinitely large normal parent population. *Biometrika* **23** 416–418.

PETERS, C. A. F. (1856). Über die Bestimmung des wahrscheinlichen Fehlers einer Beobachtung aus den Abweichungen der Beobachtungen von ihrem arithmetischen Mittel. *Astronom. Nachr.* **44** 29–32.

QUETELET, A. (1846). *Lettres à S.A.R. le Duc Regnant de Saxe-Cobourg et Gotha, sur la Théorie des Probabilités, Appliquée aux Sciences Morales et Politiques*. Hayez, Bruxelles. [English translation (1849) by O. G. Downes. Layton, London.]

SARHAN, A. E. and GREENBERG, B. G. (eds.) (1962). *Contributions to Order Statistics*. Wiley, New York.

SEAL, H. L. (1967). Studies in the history of probability and statistics. XV. The historical development of the Gauss linear model. *Biometrika* **54** 1–24.

SENETA, E. (1983). Modern probabilistic concepts in the work of E. Abbe and A. de Moivre. *Math. Sci.* **8** 75–80.

SHEYNIN, O. B. (1966). Origin of the theory of errors. *Nature* **211** 1003–1004.

SHEYNIN, O. B. (1971). J. H. Lambert's work on probability. *Arch. Hist. Exact Sci.* **7** 244–256.

SHEYNIN, O. B. (1973). R. J. Boscovich's work on probability. *Arch. Hist. Exact Sci.* **9** 306–324.

SHEYNIN, O. B. (1979). C.F. Gauss and the theory of errors. *Arch. Hist. Exact Sci.* **20** 21–72.

SHEYNIN, O. (1995). Helmert's work in the theory of errors. *Arch. Hist. Exact Sci.* **49** 73–104.

SHEYNIN, O. (1996). *The History of the Theory of Errors.* Hänsel-Hohenhausen, Egelsbach, Germany.

SPROTT, D. A. (1978). Gauss's contributions to statistics. *Historia Math.* **5** 183–203.

STEWART, G. W. (1995). *Theory of the Combination of Observations Least Subject to Errors*. SIAM, Philadelphia. [Translation of Gauss (1821, 1823, 1826).]

STIGLER, S. M. (1973a). Laplace, Fisher, and the discovery of the concept of sufficiency. *Biometrika* **60** 439–445. [Reprinted in Kendall and Plackett (1977) 271–277.]

STIGLER, S. M. (1973b). Simon Newcomb, Percy Daniell, and the history of robust estimation, 1885–1920. *J. Amer. Statist. Assoc.* **68** 872–879. [Reprinted in Kendall and Plackett (1977) 410–417.]

STIGLER, S. M. (1986). *The History of Statistics: The Measurement of Uncertainty before 1900*. Harvard Univ. Press.

STUDENT (1907). On the error of counting with a haemacytometer. *Biometrika* **5** 351–360.

TIPPETT, L. H. C. (1925). On the extreme individuals and the range of samples taken from a normal population. *Biometrika* **17** 364–387.

TROTTER, H. F. (1957). *Gauss's Work (1803–1826) on the Theory of Least Squares, an English Translation*. Princeton Univ.

TUKEY, J. W. (1960). A survey of sampling from contaminated distributions. In *Contributions to Probability and Statistics* (I. Olkin, S. G. Ghurye, W. Hoeffding, W. G. Madow and H. B. Mann, eds.) 448–485. Stanford Univ. Press.

VON ANDRAE, C. G. (1869). Schreiben des Herrn Geheimen Etatsraths von Andrä an den Herausgeber. *Astronom. Nachr.* **74** 283–284.

VON ANDRAE, C. G. (1872). Ueber die Bestimmung des wahrscheinlichen Fehlers durch die gegebenen Differenzen von m gleich genauen Beobachtungen einer Unbekannten. *Astronom. Nachr.* **79** 257–272.

VON BORTKIEWICZ, L. (1922a). Variationsbreite und mittlerer Fehler. *Sitzungsberichte der Berliner Mathematischen Gesellschaft* **21** 3–11.

VON BORTKIEWICZ, L. (1922b). Die Variationsbreite beim Gauss'schen Fehlergesetz. *Nordisk Statistisk Tidskrift* **1** 11-38, 193–220.

VON MISES, R. (1923). Über die Variationsbreite einer Beobachtungsreihe. *Sitzungsberichte der Berliner Mathematischen Gesellschaft* **22** 3–8. [Reproduced in (1964) *Selected Papers of Richard von Mises* **2** 129–134. Amer. Math. Soc., Providence, RI.]