MAXIMUM LIKELIHOOD CHARACTERIZATION OF DISTRIBUTIONS

By Henry Teicher

New York University and Purdue University

1. Introduction. It is a commonplace observation that the sample mean and sample variance from a normal population (based on a random sample) are stochastically independent. Considerably less prosaic is the converse proposition, first proved in 1936 by R. C. Geary [2] (under superfluous restrictions), to the effect that the independence of these two statistics entails normality of the underlying population. This, plus the theorem that if two linear combinations (nonzero coefficients) of a pair of independent random variables are themselves independent, the variables are normally distributed, which was proved by Kac in 1939 [4], are harbingers of what are today referred to as characterization theorems. An extensive bibliography of such theorems appears in [5]. Most of these results have the format: if such-and-such statistics are independent (alternatively, if the distribution of such-and-such a statistic is thus-and-so), the underlying population is so-and-so.

The ensuing theorems belong to this genre but adopt a maximum likelihood posture. The first deals with translation (location) parameter and the latter with scale parameter families of distributions.

2. Preliminaries. Since the results expounded here concern maximum likelihood estimators, it would seem appropriate to say a few words concerning these. It is somewhat surprising that major treatises on mathematical statistics and estimation do not define maximum likelihood *estimators* per se but merely a maximum likelihood *estimate.* (Pitman's terminological demarcation between these notions will be made explicit shortly.) The definitions of [8], [9] are closest in spirit to that given here.

In order to pave the way for a discussion of these questions, let $F(x;\theta)$, $-\infty < x < \infty$, $\theta \varepsilon \Omega \subset R^1$ denote a one parameter family of probability distributions on the real line R^1 with spectra S_θ. Define $S = \bigcup_{\theta\varepsilon\Omega} S_\theta$ and $S^n = S \times S \times \cdots \times S$, the n-fold cartesian product of S with itself.[1] If, for each $\theta \varepsilon \Omega$, $F(x;\theta)$ is absolutely continuous, designate its probability density function (p.d.f.) by $f(x;\theta)$; if, for each θ, $F(x;\theta)$ is a step function, the same notation $f(x;\theta)$ will be used to specify the so-called discrete p.d.f., that is, the mass function of the corresponding distribution (positive at the countable set of points constituting S_θ and zero elsewhere).

The customary definition of a maximum likelihood *estimate* of a parameter θ of a population (family of distributions generally restricted to the aforementioned types), based on a (random) sample of n observations $x_1, x_2, \cdots, x_n$, is a value of θ, say $\hat{\theta}_n$, which renders $\prod_{i=1}^n f(x_i;\theta)$ a maximum. A maximum likelihood

Received August 23, 1960; revised January 20, 1961.

[1] When $\bigcup_\theta S_\theta = [a, \infty)$ or $(-\infty, b]$, the points a, b will be deleted in defining S (so as to avoid a special treatment of the origin in Theorem 2).

estimator (M.L.E.) is presumably a function $\hat{\theta}_n = \hat{\theta}_n(x_1, x_2, \cdots, x_n)$ from S^n into Ω which, for every choice of $x_1, \cdots, x_n$, is a maximum likelihood estimate. (It is by no means apparent that a M.L.E., so defined, is a bona fide random variable and it would seem of interest to give minimal conditions under which it is measurable, [7]. Measurability is, however, tangential to the problem treated here.)

Unfortunately, from the theoretical standpoint, such a definition harbors ambiguities of a petty but disconcerting nature. The fact that these annoyances crop up on sets of measure zero does not seem sufficient reason to ignore their existence.

First, (consider the absolutely continuous case) as a consequence of the fact that for each $\theta \varepsilon \Omega$, $f(x; \theta)$ is defined only to within a set of measure zero, it is possible to change a M.L.E. by altering $f(x; \theta)$ at one value of x for each θ or even prevent its existence by a perverse choice of $f(x; \theta)$. If scale and translation parameter families are involved, $f(x; \theta)$ is a function of one variable only and the scope for tampering is greatly diminished.

Suppose that a suitable version of $f(x; \theta)$ has been singled out. Then, a M.L.E. $\hat{\theta}_n$ will be interpreted as a function from S^n into Ω satisfying

$$\prod_{i=1}^{n} f(x_i ; \hat{\theta}_n) \geqq \prod_{i=1}^{n} f(x_i ; \theta) \tag{0}$$

for *all* $\theta \varepsilon \Omega$ and *all* $(x_1, \cdots, x_n) \varepsilon S^n$. (Note that if $R^1 - S$ is non-empty and $\hat{\theta}_n$ were assigned any value in Ω for $(x_1, \cdots, x_n) \varepsilon R^n - S^n$, (0) would hold in the degenerate form $0 = 0$.)

Secondly, if all $(x_1, \cdots, x_n)$ in S^n are pertinent to the definition of a M.L.E., how is one to interpret $0 \cdot \infty$ if it occurs in (0)? (A p.d.f. may be infinite on a set of measure zero.) The conventional interpretation of this product as zero seems mandatory when the value x_o for which $f(x_o ; \theta) = 0$ belongs to $R^1 - S$ (if this set is non-empty) and will be adopted for $x_o \varepsilon S$ as well.

3. Characterization theorems. Theorem 1 which deals with translation parameter families emerges as a generalization and modernization of a result of Gauss [1] when the latter is suitably interpreted and rescued from its context of least squares.[2]

THEOREM 1: *Let* $\{F(x - \theta), \theta \varepsilon R^1\}$ *be a translation parameter family of absolutely continuous distributions on the real line and let the version of the* p.d.f. $f(x)$ *be lower semi-continuous at* $x = 0$. *If, for all (random) samples of sizes two and three, a maximum likelihood estimate of* θ *is the sample arithmetic mean, then* $F(x)$ *is a normal distribution with mean zero.*

[2] There is a vast literature consisting of discussions, proofs and reproofs of Gauss' result. In view of the fact that the latter was formulated in a least squares context and further that many notions and distinctions which are today commonplace were then only dimly (if at all) perceived, many of the disquisitions are heuristic and unrigorous by modern standards. Among the multitude of commentaries, three ([10], [11], [12] p. 169) are singled out for reference.

All prior proofs which have come to the writer's attention assume implicitly or explicitly that the density function $f(x)$ is differentiable.

PROOF: If $\bar{x} = (n + 1)^{-1} \sum_{i=1}^{n+1} x_i$, it follows from (0) and the hypothesis that for $n = 1, 2$ and all real $x_1, x_2, \cdots, x_{n+1}$ and θ,

$$\prod_{i=1}^{n+1} f(x_i - \bar{x}) \geqq \prod_{i=1}^{n+1} f(x_i - \theta);$$

hence,

$$\prod_{i=1}^{n+1} f(y_i) \geqq \prod_{i=1}^{n+1} f(y_i - \theta) \tag{0$'$}$$

for $n = 1, 2$ and all real θ and $y_1, \cdots, y_{n+1}$ satisfying $\sum_{i=1}^{n+1} y_i = 0$.

Set $n = 1$, $y_1 = y = -y_2$ in (0)′ to obtain

$$f(y)f(-y) \geqq f(y - \theta)f(-y - \theta), \qquad \text{all real } y, \theta. \tag{1}$$

Note that $f(0) = 0$ implies $f(y)$ vanishes identically. Suppose that $f(a) = \infty$ for some real number a. Then, according to (1), for each $y \varepsilon R^1$ either $f(y)f(-y) = \infty$ or $f(2y + a) = 0$. For a p.d.f., the former cannot hold on a set of positive (Lebesgue) measure, while the latter cannot hold almost everywhere. Thus, every p.d.f. satisfying (0)′ is positive at the origin and everywhere finite (so the product $0 \cdot \infty$ will not arise in (0)′).

Let $h(x) = \log_e f(x)$ where $h(x)$ may possibly assume the extended real value $-\infty$. Then it follows from (0)′ that for $n = 1, 2$ and all real $y_1, y_2, \cdots, y_n, \theta$,

$$\sum_{i=1}^{n} h(y_i) + h\left(-\sum_{i=1}^{n} y_i\right) \geqq \sum_{i=1}^{n} h(y_i - \theta) + h\left(-\sum_{i=1}^{n} y_i - \theta\right). \tag{2}$$

As will be seen, under the meager assumptions contained in the probabilistic framework, the functional inequality (2) determines $h(x)$ and therefore $f(x)$.

In particular, (2) implies

$$nh(y) + h(-ny) \geqq nh(y - \theta) + h(-ny - \theta),$$

which, for $n = 1$, becomes

$$h(y) + h(-y) \geqq h(y - \theta) + h(-y - \theta). \tag{3}$$

Note that if in (2), θ is replaced by $-\theta$ and y_i by $-y_i$, the resulting inequality when added to (2) reveals that, if $g(y)$ is a solution of (2), so is $h(y) = g(y) + g(-y)$ and we therefore confine attention at first to symmetric solutions of (2), which then takes the form

$$\sum_{i=1}^{n} h(y_i) + h\left(\sum_{i=1}^{n} y_i\right) \geqq \sum_{i=1}^{n} h(y_i - \theta) + h\left(\sum_{i=1}^{n} y_i + \theta\right). \tag{2$'$}$$

Similarly, (3) becomes

$$2h(y) \geqq h(y - \theta) + h(y + \theta), \qquad \text{all } y, \theta. \tag{3$'$}$$

Suppose that $h(y) = -\infty$ for some $y > 0$ and let c be the infimum of the set A of all such positive values y. Taking $n = 2$, $y_1 = y_2 = \frac{1}{2}c_m$, $\theta = -\frac{1}{4}c_m$ in (2)′ yields $2h(\frac{1}{2}c_m) + h(c_m) \geqq 3h(\frac{3}{4}c_m)$. Choose $c_m \searrow c$, $c_m \in A$, $m = 1, 2, \cdots$, implying $h(\frac{3}{4}c_m) = -\infty$, $m \geqq 1$. If $c > 0$, a contradiction ensues, while if $c = 0$, $f(x)$ is not lower semi-continuous at zero.

Thus, $h(y)$ is everywhere finite and according to (3)′ concave. Since any p.d.f. $f(x)$ is necessarily measurable, $h(x)$ is likewise, whence [3, 6], h is a continuous concave function.

Let D denote the complement of the (at most countable) set of points at which $h(x)$ is not differentiable and denote by $q(x)$ the derivative of $h(x)$. Then $q(x)$ is monotone and defined for all $x \in D$.

The fact, as expressed by (2)′, that $\theta = 0$ maximizes $\sum_{i=1}^n h(y_i - \theta) + h(\sum_{i=1}^n y_i + \theta)$ now requires that

$$(4)' \qquad -\sum_{i=1}^{n} q(y_i) + q\left(\sum_{i=1}^{n} y_i\right) = 0$$

for all $y_i \in D$ such that $\sum_{i=1}^n y_i \in D$, $n = 1, 2$. For $n = 2$, (4)′ becomes

$$(5)' \qquad q(y_1) + q(y_2) = q(y_1 + y_2) \quad \text{for } y_1, y_2, y_1 + y_2 \in D.$$

Let $C = \{f\}$ be the class of non-negative measurable functions on R^1 which are everywhere finite, lower semi-continuous at zero, and do not vanish almost everywhere. Let C' be the subclass of functions in C which do not vanish anywhere. Since the only monotone solution of Cauchy's functional equation (5)′ is $q(y) = c_1 y$, it follows that the only *symmetric* functions of C satisfying (0)′ (which are necessarily in C') are given by $h(y) = \log_e f(y) = -cy^2 + d$ for $y \in D$ and therefore by continuity for all real y.

Suppose next that $f(y)$ is *any* element of C satisfying (0)′. According to (1), $f(y) \cdot f(-y)$ does not vanish almost everywhere (take $\theta = -y$); it is readily seen that $f(y) \cdot f(-y)$ is a symmetric function in C and, as previously noted, a solution of (0)′. Thus $f(y) \cdot f(-y) \in C'$ implying $f(y) \in C'$.

In fact, necessarily for some real constants c and d,

$$g(y) = \log_e f(y) = -\tfrac{1}{2}(cy^2 - d) + b(y)$$

where $b(y)$ is an odd function. For, by the preceding, $g(y) + g(-y) = -cy^2 + d$ for some c, d and this implies that $b(y) = g(y) + \frac{1}{2}(cy^2 - d)$ satisfies $b(y) = -b(-y)$.

Substituting for $g(y)$ in (3) yields

$$(4) \qquad c\theta^2 \geqq b(y - \theta) - b(y + \theta), \qquad \text{all } y, \theta.$$

Replacing y by $-y$ and θ by $-\theta$ in (4) and combining the result with (4), produces

$$(5) \qquad |b(y - \theta) - b(y + \theta)| \leqq c\theta^2, \qquad \text{all } y, \theta,$$

which, in turn, necessitates $c \geqq 0$ and implies that $b(y)$ is differentiable and constant, hence identically zero.

Consequently, the only solutions of $(0)'$ in C are given implicitly by $h(x) = -\frac{1}{2}cx^2 + d$, $c \geqq 0$ and thus the only p.d.f.'s satisfying the conditions of the theorem are $f(x) = (c/2\pi)^{\frac{1}{2}}e^{-\frac{1}{2}cx^2}$, that is, normal density functions with mean zero.

REMARK 1: The integers two and three of the theorem may clearly be replaced by other pairs, e.g., $2k$, $3k$, $k > 1$. It seems most desirable, however, to state the result with minimal n.

REMARK 2: If $\int xf(x)\,dx$ exists and is zero, the translation parameter θ is the mean of the distribution $F(x - \theta)$. In such cases (excluding the normal), the theorem implies that the sample mean is not (for samples of sizes both two and three, a fortiori, for all n) a maximum likelihood estimator of the population mean. This is readily seen, for example, if

$$f(x; \theta) = C_\alpha \cdot \exp\{-|x - \theta|^\alpha\}, \qquad \alpha \neq 2.$$

REMARK 3: It seems of interest to note in the case where $F(x - \theta)$ is a rectangular distribution with mean θ, i.e., $f(x) = 1$ for $|x - \theta| \leqq \frac{1}{2}$ and zero otherwise, that, whereas $\bar{x}$ is a M.L.E. of θ for $n = 2$, its numerical value is not a maximum likelihood estimate of θ for *all* random samples of size three.

4. Scale parameter families. Consider a scale parameter family[3] of absolutely continuous distributions $\mathfrak{F} = \{F(x/\sigma), \sigma > 0\}$. The joint density function of n independent random variables, each distributed as $F(x/\sigma)$, is $\sigma^{-n}\prod_{i=1}^n f(x_i/\sigma)$ where $F(x) = \int_{-\infty}^x f(u)\,du$. To say that $\hat{\sigma} = \hat{\sigma}(x_1, \cdots, x_n)$ is a maximum likelihood estimator of σ is to say for all $\sigma > 0$ and $x_1, \cdots, x_n$ in S^n that

$$\hat{\sigma}^{-n} \prod_{i=1}^n f(x_i/\hat{\sigma}) \geqq \sigma^{-n} \prod_{i=1}^n f(x_i/\sigma).$$

Let $y_i = x_i/\hat{\sigma}$, $\lambda = \hat{\sigma}/\sigma$. Then, if $\hat{\sigma}$ is a homogeneous function of degree one in $x_1, \cdots, x_n$, the preceding implies

$$\prod_{i=1}^n f(y_i) \geqq \lambda^n \prod_{i=1}^n f(\lambda y_i) \tag{6}$$

for all $\lambda > 0$ and $y_1, \cdots, y_n$ satisfying

$$\hat{\sigma}(y_1, y_2, \cdots, y_n) = 1. \tag{7}$$

If $h(y) = \log_e f(y)$ is finite valued, (6) may be transcribed as

$$\frac{1}{n} \sum_{i=1}^n [h(y_i) - h(\lambda y_i)] \geqq \log_e \lambda. \tag{8}$$

Inspection shows that $h(y) = -\log_e y + \text{const.}$ satisfies (8), with equality holding for all choices of $y_1, \cdots, y_n$ and *a fortiori* for y_i's satisfying (7). How-

[3] The standard device of reducing a scale parameter family to a translation parameter family (so as to utilize Theorem 1) appears fruitless here.

ever, $f(y) = cy^{-1}$ is not integrable on $(0, \infty)$ and truncation to a finite interval will be precluded by the conditions of Theorems 2 and 3.

In the following theorems the indispensable absolute continuity assumption is augmented by a possibly dispensable continuity assumption of the density function. The seemingly ad hoc condition (ii) on the other hand, appears to be crucial.

THEOREM 2: *Let* $\{F(x/\sigma), \sigma > 0\}$ *constitute a scale parameter family of absolutely continuous distributions with the version of the* p.d.f. $f(x)$ *satisfying*

(i) $f(x)$ *is continuous on* $(0, \infty)$

(ii)[4] $$\lim_{y \downarrow 0} \frac{f(\lambda y)}{f(y)} = 1, \qquad \textit{all } \lambda > 0.$$

If, for all sample sizes, a maximum likelihood estimator of σ *is the sample arithmetic mean, then* F *is the exponential distribution, i.e.,*

$$f(x) = e^{-x}, \quad x > 0, \qquad f(x) = 0, \quad x \leqq 0.$$

PROOF: Since $\Omega = \{\sigma : \sigma > 0\}$ and $\bar{x}$ is a posited M.L.E., necessarily $S \subset S' = \{x : x > 0\}$, whence $f(x) \equiv 0$ in $R^1 - S'$. It suffices, therefore, to consider $f(x)$, $x \in S'$, noting that $f(x) \not\equiv 0$ in S'. Infinite values of f are precluded by continuity and it will now be shown that $f(x) > 0$ in S', i.e., $S = S'$.

From prior remarks, (6) obtains with (7) becoming

$$\sum_{i=1}^{n} y_i = n. \tag{7.1}$$

In (6), choose $y_i = k/m$, $1 \leqq i \leqq m$ and $y_i = [(n-k)/(n-m)]$, $m + 1 \leqq i \leqq n$, where k, m, n are positive integers satisfying $k < m < n$; this yields

$$f^m\left(\frac{k}{m}\right) f^{n-m}\left(\frac{n-k}{n-m}\right) \geqq \lambda^n f^m\left(\frac{\lambda k}{m}\right) f^{n-m}\left(\frac{\lambda(n-k)}{n-m}\right). \tag{9}$$

Let $k/m \to \alpha$, $m/n \to c$. Then for all positive λ and all c, α in $(0, 1)$

$$f^c(\alpha) f^{1-c}\left(\frac{1-\alpha c}{1-c}\right) \geqq \lambda f^c(\lambda\alpha) f^{1-c}\left(\frac{\lambda(1-\alpha c)}{1-c}\right). \tag{10}$$

Now, if there exists a sequence $\alpha_n \to 0$ with $f(\alpha_n) > 0$, $n = 1, 2, \cdots$, it follows from (10) and (ii) that

$$f(y) \geqq \lambda^y f(\lambda y) \quad \text{for} \quad y \geqq 1, \qquad \lambda > 0. \tag{11}$$

Thus, if f vanished for some $y \geqq 1$, it would vanish identically. Further, from (6), if $f(y)$ were zero for some y in $(0, 1)$, $f(y)$ would have zeros in $(1, \infty)$.

Alternatively, suppose that for some $\delta > 0$, $f(y) \equiv 0$ in $(0, \delta)$. Since $f(y) \not\equiv 0$

[4] This condition is automatically fulfilled if $0 < \lim_{x \searrow 0} f(x) < \infty$. Also, it is reiterated that only random samples are under consideration.

in S', (6) insures ($y_i \equiv 1$) that $f(1) > 0$, whence $\delta < 1$ and there is no loss of generality in supposing that $(0, \delta)$ is the maximal interval in which f vanishes identically. Then, from (10) follows

$$0 = \lambda f^c(\lambda\delta) f^{1-c}\left(\frac{\lambda(1 - \delta c)}{1 - c}\right) \tag{12}$$

for all $\lambda > 0$ and c in $(0, 1)$. By continuity, $f(y) > 0$ in $(1 - \epsilon, 1 + \epsilon)$ for all sufficiently small $\epsilon > 0$. Hence, taking $\lambda = (1 - \epsilon)/\delta$ in (12),

$$f\{[(1 - \epsilon)(1 - \delta c)]/\delta(1 - c)\} = 0$$

for $0 < c < 1$ implying $f[(1 - \epsilon)/\delta] = 0$, all sufficiently small $\epsilon > 0$.

Now, let $k + 1$ be an integer greater than $(1 - \epsilon)/\delta$. From (6),

$$0 \geqq f[\lambda(1 - \epsilon)/\delta]\cdot f^k\{\lambda[k + 1 - (1 - \epsilon)\delta^{-1}]/k\}.$$

Hence, $f\{[\delta/k(1 - \epsilon)][k + 1 - (1 - \epsilon)\delta^{-1}]\} = 0$ for all sufficiently large k and all sufficiently small ϵ, implying $f[\delta/(1 - \epsilon)] = 0$ for all sufficiently small $\epsilon > 0$, which contradicts the maximality of δ.

Thus, any p.d.f. satisfying (6) is non-zero in S'. Consequently, (11) holds unconditionally and may be rewritten as

$$y^{-1}[h(y) - h(\lambda y)] \geqq \log_e \lambda, \qquad y \geqq 1, \lambda > 0 \tag{13}$$

where, as before, $h(y) = \log_e f(y)$.

Replace λ by λ^{-1} in (13) and combine the result with (13) to obtain

$$0 \geqq h(\lambda y) - 2h(y) + h(y/\lambda), \qquad y \geqq 1, \lambda > 0. \tag{14}$$

This asserts that $H(y) = h(e^y)$ is concave for $y \geqq 0$ and hence that $h(y)$ is differentiable in $(1, \infty)$ except perhaps for a countable subset D thereof. From (13), for $\lambda < 1$ and $y \geqq 1$,

$$\frac{h(\lambda y) - h(y)}{y(\lambda - 1)} \geqq \frac{\log_e \lambda}{1 - \lambda} \geqq \frac{h(y/\lambda) - h(y)}{\lambda y[(1/\lambda) - 1]}$$

whence ($\lambda \nearrow 1$), $h'(y) = -1$ on $(1, \infty) - D$. Then by continuity,

$$h(y) = -y + c, \qquad y \geqq 1. \tag{15}$$

Next, choosing $y_i < 1$, $i = 1, \cdots, r < n$ and $y_j > 1$, $j > r$, in (6) and employing (15) and (7.1), we find for all $\lambda > 0$, $r < n$ and y_i in $(0, 1)$ that

$$\sum_{i=1}^{r} [h(y_i) - h(\lambda y_i) + (1 - \lambda)y_i] \geqq n\,[\log_e \lambda + 1 - \lambda]. \tag{16}$$

For $0 < y < 1$, (16) asserts ($r = 1$) that

$$1/n[h(y) - h(\lambda y)] \geqq \log_e \lambda + (1 - \lambda)(1 - (y/n)). \tag{17}$$

But for $0 < x < 1$, $\log_e \lambda + x(1 - \lambda) > 0$ for all λ sufficiently close to and larger than unity. Thus, from (17), h is monotone decreasing in $(0, 1)$.

Set $h = h_1 + h_2$ where h_2 is absolutely continuous and h_1 is singular, i.e., $h_1'(y) = 0$ a.e. in (0, 1) and $h_1(y) \equiv 0$ for $y > 1$.

Again taking $r = 1$ in (16), there follows for $0 < y < 1$ and $\lambda < 1$

$$\frac{h(\lambda y) - h(y) + (\lambda - 1)y}{\lambda - 1} \geqq n\left[\frac{\log_e \lambda}{1 - \lambda} + 1\right]. \tag{18}$$

This implies $y[h_2'(y) + 1] \geqq 0$ almost everywhere in (0, 1). Similarly, replacing λ by λ^{-1} in (16), the inequality is reversed. Hence, $h_2'(y) = -1$, almost everywhere in (0, 1) implying $h_2(y) = -y + c_1$.

Utilizing this result in (18), there follows

$$\limsup_{\lambda \to 1} \frac{h_1(\lambda y) - h_1(y)}{y(\lambda - 1)} \geqq \limsup_{\lambda \nearrow 1} \frac{h_1(\lambda y) - h_1(y)}{y(\lambda - 1)} \geqq 0 \tag{19}$$

for all y in (0, 1); hence $h_1(y) \equiv 0$. Also, by continuity $c_1 = c$.

Thus, for $y \varepsilon S'$, $f(y) = ae^{-y}$ and since f vanishes outside S', $a = 1$.

THEOREM 3: *Let* $\{F(x/\sigma), \sigma > 0\}$ *be a scale parameter family of absolutely continuous distributions with the version of the* p.d.f. $f(x)$ *satisfying*

(i) $f(x)$ *continuous on* $(-\infty, \infty)$

(ii)[4] $\lim_{y \to 0} [f(\lambda y)/f(y)] = 1$, all $\lambda > 0$

If, for all sample sizes, a maximum likelihood estimator of σ *is* $(n^{-1} \sum_{i=1}^n x_i^2)^{\frac{1}{2}}$, *then* $F(x)$ *is the normal distribution with mean zero and variance one.*

PROOF: Here (7) specializes to

$$\sum_{i=1}^{n} y_i^2 = n \tag{7.2}$$

In (6), set $y_i = \pm(k/m)^{\frac{1}{2}}$, $1 \leqq i \leqq m$; $y_i = \pm[(n - k)/(n - m)]^{\frac{1}{2}}$, $m + 1 \leqq i \leqq n$. Analogous to (10), there follows

$$f^c(\alpha)f^{1-c}\left(\pm \sqrt{\frac{(1 - \alpha^2 c)}{(1 - c)}}\right) \geqq \lambda f^c(\lambda\alpha)f^{1-c}\left(\pm \lambda \sqrt{\frac{(1 - \alpha^2 c)}{(1 - c)}}\right) \tag{10$'$}$$

valid for $\lambda > 0$, $|\alpha| \leqq 1$, $0 < c < 1$. An argument akin to that employed in Theorem 2 shows that f is non-vanishing and it follows from (10)$'$ that

$$y^{-2}[h(y) - h(\lambda y)] \geqq \log_e \lambda, \qquad |y| \geqq 1, \lambda > 0.$$

Again, h is differentiable, this time in the region $A: |y| \geqq 1$ except perhaps on a countable subset D' thereof. Proceeding as in the proof of Theorem 2, we find that $h'(y) = -y$ for y in $A - D'$ and hence that $h(y) = -\frac{1}{2}y^2 + c$ for $|y| \geqq 1$. The analogue of (16) is

$$\sum_{i=1}^{i=r} [h(y_i) - h(\lambda y_i) + (1 - \lambda^2)\tfrac{1}{2}y_i^2] \geqq n\,[\log_e \lambda + \tfrac{1}{2}(1 - \lambda^2)] \tag{16$'$}$$

where $|y_i| < 1$ and $r < n$. This implies that h is decreasing in (0, 1) and increasing in $(-1, 0)$. An argument paralleling that of Theorem 2 yields $h(y) = -\frac{1}{2}y^2 + c$, for $|y| < 1$ implying $f(y) = a \exp\{-y^2/2\}$. Finally, $a = (2\pi)^{-\frac{1}{2}}$.

REFERENCES

[1] CARL FRIEDRICH GAUSS, "Theoria motus corporum coelestium," *Werke*, liber II, Sectio III, pp. 240–44. See also "Gauss' work on the theory of least squares" by Hale F. Trotter, pp. 127–33, Statistical Research Group, Princeton, N. J.

[2] R. C. GEARY, "Distribution of Student's ratio for non-normal samples," *J. Roy. Stat. Soc. Suppl.*, Ser. B, Vol. 3 (1936), pp. 178–84.

[3] H. BLUMBERG, "On convex functions," *Trans. Amer. Math. Soc.*, Vol. 20 (1919), pp. 40–44.

[4] M. KAC, "On a characterization of the normal distribution," *Amer. J. Math.*, Vol. 61 (1939), pp. 726–28.

[5] E. LUKACS, "Characterization of populations by properties of suitable statistics," *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability*, University of California Press, Berkeley, 1956, pp. 195–214.

[6] W. SIERPINSKI, "Sur les fonctions convexes mesurables," *Fundamenta Math.*, Vol. 1 (1920), pp. 125–9.

[7] R. R. BAHADUR, "On the asymptotic efficiency of tests," *Sankhyā*, Vol. 22 (1960), pp. 229–52.

[8] J. HANNAN, "Consistency of maximum likelihood estimation of discrete distributions," *Contributions to Probability and Statistics*, Stanford University Press, Stanford, Calif., 1960, pp. 249–57.

[9] LUCIEN LECAM, "On some asymptotic properties of maximum likelihood estimates and related Bayes' estimates," *University of California Publications in Statistics*, Vol. 1, pp. 277–330.

[10] J. GLAISHER, "On the law of facility of errors of observations and on the method of least squares," *Mem. Astron. Soc.*, London, Vol. 39 (1872), pp. 75–124.

[11] MANSFIELD MERRIMAN, "A list of writings relating to the method of least squares with historical and critical notes," *Trans. Connecticut Academy of Arts and Sciences*, Vol. 4 (1877–82), pp. 151–231.

[12] H. POINCARÉ, *Calcul des Probabilités*, Gauthier-Villars, Paris, 1912.