The Annals of Mathematical Statistics
1970, Vol. 41, No. 3, 1086–1090

A SUFFICIENT STATISTICS CHARACTERIZATION OF THE NORMAL DISTRIBUTION

By Douglas Kelker[1] and Ted K. Matthes[2]

Washington State University and University of Oregon

The results in this paper are analogous to those of Teicher [7] for maximum likelihood estimators. He showed that if $\sum x_i/n$ ($\sum x_i^2$) is a maximum likelihood estimator for a location (scale) parameter family of distributions, then the family is the normal family of distributions.

We will show that the normal distribution is the only distribution for which $\bar{X}$ is a sufficient statistic for a location parameter. Similar but weaker results are obtained for S^2 sufficient for a scale parameter and for $(\bar{X}, S)$ sufficient for a location and a scale parameter.

Koopman [4] showed that if $\bar{X}$ is sufficient for a location parameter in a differentiable density, then the density is normal. The first theorem in this paper is a direct extension of a result of Basu. Basu [1] showed that $\sum b_i x_i$ is a boundedly complete sufficient statistic for a location parameter θ, based on an independent sample of size $n (n \geqq 2)$, if and only if each x_i is a normal variable with variance a/b_i for some constant a.

For the direction of the proof which is not trivial, the following theorem considerably strengthens Basu's result by dropping bounded completeness from the hypothesis.

Theorem 1. *Let* $x_1, x_2, \cdots, x_n$ $(n \geqq 2)$ *be independent non-degenerate random variables with* cdf's $F_{x_i}(x) = F_i(x-\theta)$, $-\infty < \theta < \infty$. *A necessary and sufficient condition for* $\sum b_i x_i (\prod b_i \neq 0)$ *to be a sufficient statistic for* θ *is that each* x_i *is a normal variable with variance* a/b_i *for some constant* a.

Proof. That the condition is sufficient is clear from the factorization theorem for sufficient statistics.

Conversely, let $a_1, a_2, \cdots, a_n$ satisfy $\sum a_i = 0$, and without loss of generality assume $\sum b_i = 1$. We will show that $\sum b_i x_i$ is stochastically independent of $\sum a_i x_i$; from this, using the result proved by Skitovich [6], the conclusion follows.

Ghurye ([2] page 161) has shown that if $t(x_1, \cdots, x_n)$ is a sufficient statistic for θ which satisfies $t(ax_1+d, \cdots, ax_n+d) = at(x_1, \cdots, x_n)+d$, and if $s(x_1, \cdots, x_n)$ satisfies $s(ax_1+d, \cdots, ax_n+d) = as(x_1, \cdots, x_n)$ with $a > 0$, then t and s are independent. Let $t(x_1, \cdots, x_n) = \sum b_i x_i$ and $s(x_1, \cdots, x_n) = \sum a_i x_i$. Thus $\sum b_i x_i$ is independent of $\sum a_i x_i$, and therefore each x_i has a normal distribution.

Received March 11, 1969.

[1] This forms a part of the author's dissertation submitted in partial fulfillment of the requirements for the Ph.D. degree at the University of Oregon. Prepared with the partial support of NSF Grant GP–4865.

[2] Research sponsored by NSF Grant GP–4865.

From the factorization equation, it is easily seen that the variance of x_i is equal to a/b_i for some constant a. This concludes the proof.

Additional conditions are necessary for a corresponding theorem on scale parameters. Consider a random variable which is the square root of a gamma random variable that has a scale parameter. The statistic $\sum x_i^2$ is sufficient for the scale parameter.

THEOREM 2. *Let* $x_1, x_2, \cdots, x_n$ $(n \geqq 2)$ *be independent random variables involving a common scale parameter; i.e.* $F_{x_i}(x) = F_i(x/\sigma)$ $i = 1, 2, \cdots, n$ $(\sigma > 0)$. *Let* x_i^2 *be non-degenerate. Then a necessary and sufficient condition for* $\sum x_i^2$ *to be sufficient for* σ *is that each* x_i^2 *has a gamma distribution with a common scale parameter and that for each i, either* $P(x_i > 0) = 1$, *or* $P(x_i < 0) = 1$, *or* $P(x_i < -|x|) = cP(x_i > |x|)$ *for some constant c.*

PROOF. That the condition is sufficient is obvious.

Assume that $\sum x_i^2$ is sufficient for σ, and let Y be any scale invariant function of $x_1, \cdots, x_n$. Using a procedure similar to that in the proof of Corollary 2.2 of Ghurye [2], we get that Y is independent of $\sum x_i^2$.

Now let $u = x_1^2$, $v = x_2^2 + \cdots + x_n^2$. Then $u+v$ is stochastically independent of u/v. Now a theorem of Lukacs ([5] page 319) states that under the circumstances u, v, and $u+v$ have gamma distributions with a common scale parameter. Thus the density of $w_i = x_i^2$ is $f_i(w) = (\beta^{k_i}\Gamma(k_i))^{-1}w^{k_i-1}\exp(-w/\beta)$, $w > 0$. Let $A = \{(x_1, x_2, \cdots, x_n) \mid x_1 > 0\}$. Let y_1 be a random variable with the conditional distribution of x_1 given that $x_1 > 0$. Let $Y = (y_1, x_2, \cdots, x_n)$ and $X = (x_1, x_2, \cdots, x_n)$. Let B be a Borel subset of A. Then $P_\sigma(Y \in B \mid S = s) = P(X \in B \mid S = s)/P(x_1 > 0)$. The right side is independent of σ, and hence the left is also. Thus S is still sufficient for σ with the distribution of x_1 replaced by its conditional distribution, given that $x_1 > 0 (x_1 < 0)$. It follows that y_1^2 has the gamma density f_1 and that the density of x_1 must be of the form $g(x) = 2c(\beta^k\Gamma(k))^{-1}x^{2k-1}\exp(-x^2/\beta)$, $x > 0$ and $g(x) = 2(1-c)(\beta^{k'}\Gamma(k'))^{-1}x^{2k'-1}\exp(-x^2/\beta)$, $x < 0$ for some c, $0 \leqq c \leqq 1$.

Since $f_1(y_1) = (2y_1^{\frac{1}{2}})^{-1}(g(y_1^{\frac{1}{2}}) + g(-y_1^{\frac{1}{2}}))$, it is seen that $k = k' = k_1$, and the conclusions of the theorem are satisfied.

COROLLARY. *Let* $x_1, x_2, \cdots, x_n$ $(n \geqq 2)$ *be independent random variables involving a common scale parameter. Let* F_i, *the distribution function of* x_i, *be absolutely continuous with respect to Lebesgue measure in a neighborhood of the origin. At the point* $x = 0$, *let* F_i' *be non-zero and continuous. Then if* $\sum x_i^2$ *is sufficient for the scale parameter, each* x_i *has a normal distribution with mean zero.*

PROOF. The conditions imply that $k_1 = c = \frac{1}{2}$ in the function g in Theorem 2.

THEOREM 3. *Let* $x_1, x_2, \cdots, x_n$ $(n \geqq 4)$ *be independent, identically distributed random variables with the distribution of each* x_i *having a location parameter* θ $(-\infty < \theta < \infty)$ *and a scale parameter* $\sigma > 0$. *Let* $\bar{X} = \sum x_i/n$, $S^2 = (1/n)\sum(x_i - \bar{x})^2$. *If* $(\bar{X}, S^2)$ *is a sufficient statistic for* (θ, σ), *then each* x_i *has a normal distribution.*

PROOF. Let Y be a statistic that is invariant under changes of location and scale. Procedures similar to those used in Corollary 2.2 of Ghurye [2] can be used to show that Y is stochastically independent of $\bar{X}$ and S. The result follows from the following lemma.

LEMMA. *Let $x_1, x_2, \cdots, x_n$ $(n \geqq 4)$ be independent, identically distributed random variables with the common* cdf F. *Let $Y = (x_1 - x_2)/S$. If Y is stochastically independent of the pair $(\bar{X}, S)$, then each x_i has a normal distribution.*

PROOF. The proof is along the lines of the proof of Kawata and Sakamoto [3] for characterizing the normal distribution in terms of the independence of $\bar{X}$ and S.

Let $t = \sum x_i$, $q = \sum x_i^2$, and let $f(t, q)$ be any measurable function such that $Ef(t, q)$ is finite. The distribution of Y has moments of all orders since $|Y| \leqq (2n)^{\frac{1}{2}}$. Thus for any positive integer r,

$$EY^r f(t, q) = EY^r Ef(t, q), \tag{1}$$

since the assumptions of the lemma imply that Y is independent of $f(t, q)$. The equations (1) will help define the distribution if the quantities involved can be evaluated in terms of some characteristic functional. This is achieved by taking r even $(= 2k)$ and $f(t, q) = S^{2k} p(t, q) \exp(iut - bq)$, with p a polynomial and $b > 0$.

Define $G(x) = k_0 \int_{-\infty}^{x} \exp(-bv^2)\, dF((v - \theta_0)/\sigma_0)$, with k_0 a normalizing constant and with b, θ_0, σ_0 such that the mean of G is zero and the variance of G is one. This can be done if b is small enough. In particular, b must be less than $\frac{1}{2}$. G has moments of all orders. $F(x)$ could just as well have been $F((x - \theta_0)/\sigma_0)$. This leads to the equations

$$E_G\{(x_1 - x_2)^{2k} p(t, q) \exp(iut)\} = (E_F Y^{2k}) E_G\{S^{2k} p(t, q) \exp(iut)\}. \tag{2}$$

Now, it turns out that $k = 1$ results in an identity which gives us no information; hence, we must try $k \geqq 2$. Another point to note is that all equations (2) obtained by taking $p = t^r g(t, q)$, with g a polynomial, are derivatives of the equations obtained by taking $p = g$. Thus it is the powers of q (or equivalently, of S) in p that give additional information. Let $K = EY^4$ and let h be the logarithm of the characteristic function of G. Taking $k = 2$ and $p = 1$ in (2) gives us

$$(2n - K(n-1)^2) h^{(4)} = n((n^2 - 1)K - 12)(h'')^2.$$

Since $n \geqq 4$, both coefficients cannot be zero; thus $2n - K(n-1)^2 \neq 0$, since $h''(0) \neq 0$. Let $c = n((n^2 - 1)K - 12)/(2n - K(n-1)^2)$. We now have

$$h^{(4)} = c(h'')^2. \tag{3}$$

We propose to show that $c = 0$. For if $c = 0$, then from equation (3) and the conditions on G we get that $h(u) = -u^2/2$; and since

$$h(u) = \log\{k_0 \int \exp(iux - bx^2)\, dF((x - \theta_0)/\sigma_0)\},$$

it follows that $k_0 \exp(-bx^2)\, dF((x - \theta_0)/\sigma_0) = (2\pi)^{-\frac{1}{2}} \exp(-x^2/2)$. Since $b < \frac{1}{2}$, we conclude that F is a normal distribution.

The remainder of the proof consists of showing that $c = 0$. Let $\alpha_r = E_G(x_1)^r$. The function G was picked so that $\alpha_1 = 0$ and $\alpha_2 = 1$. Evaluating equation (3) at $u = 0$, we get $h^{(4)}(0) = c = \alpha_4 - 3$.

Let $u = 0$, $k = 2$, and $p = S^2$ in equation (2). Straightforward but time-consuming computations lead to

$$(n-1)\alpha_6 - 2(n-1)\alpha_4{}^2 + (n+3)\alpha_4 + 2\alpha_4\alpha_3{}^2 - 4(n-1)\alpha_3{}^2 - 12 = 0. \tag{4}$$

Taking the derivative of equation (3) and letting $u = 0$, we get $\alpha_5 = (4+2\alpha_4)\alpha_3$. Two derivatives of equation (3) give us

$$\alpha_6 = 2\alpha_4{}^2 + 3\alpha_4 + 2\alpha_4\alpha_3{}^2 + 4\alpha_3{}^2 - 12. \tag{5}$$

Combining (4) and (5), we get

$$\alpha_3{}^2 = (6-2\alpha_4)/\alpha_4, \tag{6}$$

Letting $u = 0$, $k = 2$ and $p = S^4$ in equation (2), we have

$$\begin{aligned}(n-1)^2\alpha_8 &- 2(n-1)^2\alpha_6\alpha_4 + 2(n-1)(n^2-2n+11)\alpha_6 \\ &+ 8(n-1)\alpha_5\alpha_4\alpha_3 - 16(n^2-2n+2)\alpha_5\alpha_3 \\ &- 2(n^2-2n+3)\alpha_4{}^3 - (4n^3-17n^2+38n-49)\alpha_4{}^2 \\ &+ 2(n^3-45n+54)\alpha_4 + 4(4n^2-17n+25)\alpha_4\alpha_3{}^2 \\ &- 4(2n^3-15n^2+44n-65)\alpha_3{}^2 - 24(n^2-6n+12) = 0.\end{aligned} \tag{7}$$

Three derivatives of equation (3) lead to

$$\begin{aligned}\alpha_8 - 28\alpha_6 - 56\alpha_5\alpha_3 - 20\alpha_4{}^2\alpha_3{}^2 - 10\alpha_4{}^3 + 55\alpha_4{}^2 \\ + 120\alpha_4\alpha_3{}^2 + 150\alpha_4 + 380\alpha_3{}^2 - 360 = 0.\end{aligned} \tag{8}$$

Multiplying equation (8) by $(n-1)^2$ and subtracting from equation (7) and then writing the result in terms of α_4, we get the equation

$$\begin{aligned}n(n-2)\alpha^4 &- 2n(4n-7)\alpha^3 - (5n^2-10n+8)\alpha^2 \\ &+ 6(16n^2-29n+4)\alpha - 108n(n-2) = 0, \quad \text{or} \\ (\alpha-3)[n(n-2)\alpha^3 &- n(5n-8)\alpha^2 - (20n^2-34n+8)\alpha \\ &+ 36n(n-2)] = 0.\end{aligned} \tag{9}$$

We now show that $\alpha = 3$ is the only root of equation (9). From equation (6) we get that $\alpha_4 \leqq 3$. Using the relation $\alpha_5{}^2 \leqq \alpha_6\alpha_4$, which is valid for the moments of any distribution, we have

$$\begin{aligned}(4+2\alpha_4)^2(6-2\alpha_4)/\alpha_4 &= \alpha_5{}^2 \leqq \alpha_6\alpha_4 \\ &= [2\alpha_4{}^2 + 3\alpha_4 + 2\alpha_4(6-2\alpha_4)/\alpha_4 + 4(6-2\alpha_4)/\alpha_4 - 12]\alpha_4, \quad \text{or} \\ 2\alpha^4 + 7\alpha^3 &- 40\alpha - 96 \geqq 0.\end{aligned}$$

The expression on the left is convex for $\alpha > 0$, and is negative for $\alpha = 0, 2$. Thus we conclude that $2 \leqq \alpha \leqq 3$. The right-hand factor in equation (9) is negative at $\alpha = 2$ and at $\alpha = 3$ and is convex for $2 < \alpha \leqq 3$. Applying this information to equation (9), we conclude that $\alpha = 3$. If $\alpha = 3$, then $c = 0$, and this concludes the proof.

The method of proof used in the lemma did not yield any information about what the conclusion is if $n = 3$. For $n = 2$, $\bar{X}$ and S^2 are equivalent to the order statistics. As a particular counter example for $n = 2$, the order statistics are sufficient for the uniform distribution.

REFERENCES

[1] Basu, D. (1955). On statistics independent of a complete sufficient statistic. *Sankhyā* **15** 277–380.

[2] Ghurye, S. G. (1958). Note on sufficient statistics and two-stage procedures. *Ann. Math. Statist.* **29** 155–166.

[3] Kawata, T. and Sakamoto, H. (1949). On the characterization of the normal population by the independence of the sample mean and the sample variance. *J. Math. Soc. Japan* **1** 111–115.

[4] Koopman, L. H. (1936). On distributions admitting a sufficient statistic. *Trans. Amer. Math. Soc.* **39** 399–409.

[5] Lukacs, E. (1955). A characterization of the gamma distribution. *Ann. Math. Statist.* **26** 319–324.

[6] Skitovich, V. P. (1954). Linear forms of independent random variables and the norma distribution law. *Izv. Acad. Nauk SSSR Ser. Mat.* **18** 185–200.

[7] Teicher, H. (1961). Maximum likelihood characterization of distributions. *Ann. Math. Statist.* **32** 1214–1222.