A note on characterization by sufficiency of the sample mean

By W. EBERL, jr. (Hagen)

KAGAN [2] (see also KAGAN—LINNIK—RAO [4], Theorem 8.5.3) proved that the sample mean is a sufficient statistic with respect to a translation class of distributions iff the distributions are normal, thereby assuming the translation class to be dominated by the Lebesgue measure. Since any dominated translation class is dominated by the Lebesgue measure, this theorem characterizes the normal distribution within all dominated translation classes. BÁRTFAI [1] succeeded in proving this theorem without the assumption of dominatedness. Basing on BÁRTFAI's ideas we shall give a shorter proof, which is somewhat more straight forward and at the same time does not use characteristic functions. More precisely, it is shown that except the translation classes of (non degenerate) normal distributions the only translation class with a sufficient sample mean is the degenerate translation class, i.e. the class of point masses, which is not dominated. (This degenerate case seems to have been excluded tacitly by BÁRTFAI.) The corresponding result for scaling classes (KAGAN [3] (see also KAGAN—LINNIK—RAO [4], Theorem 8.5.4), BÁRTFAI [1]) can be verified by analogous arguments.

I. Sufficiency of the sample mean with respect to translation classes

Let P denote any probability measure on $\mathfrak{B}$, the system of the Borel sets in $\mathbf{R}$, and $X_j\colon \mathbf{R}^n \to \mathbf{R}$ the j-th projection $(1 \leqq j \leqq n)$; for $\gamma \in \mathbf{R}$ let T_γ be the translation given by $T_\gamma(x) = (x_1+\gamma, \dots, x_n+\gamma)$ $(x=(x_1, \dots, x_n) \in \mathbf{R}^n)$.

Theorem 1. *For* $n \geqq 2$ *let* $\mathfrak{W} = \left\{ P_\gamma = \left(\bigotimes_1^n P \right)_{T_\gamma} \middle| \gamma \in \mathbf{R} \right\}$ *be a translation class.*[1] *Then* $S = \overline{X}$ *is sufficient with respect to* $\mathfrak{W}$ *iff* P *is a normal distribution or a point mass.*

PROOF. The sufficiency of the sample mean in the case of (non degenerate) normal distributions is well known. Now, if P is w.l.o.g. the point mass in 0, let $h\colon \mathbf{R} \to \mathbf{R}^n$ be the (measurable) mapping with $h(x) = (x, \dots, x)$ $(x \in \mathbf{R})$. Let $B \in \mathfrak{B}^n$ and consider the mapping $g = 1_{(h \circ S)^{-1}(B)}$. Then we have for $D \in \mathfrak{B}$ and $\gamma \in \mathbf{R}$

(i) $$\int_{S^{-1}(D)} 1_B \, dP_\gamma = 1_B(\gamma, \dots, \gamma) 1_{S^{-1}(D)}(\gamma, \dots, \gamma)$$

(ii) $$\int_{S^{-1}(D)} g \, dP_\gamma = g(\gamma, \dots, \gamma) \cdot 1_{S^{-1}(D)}(\gamma, \dots, \gamma).$$

[1] $\left(\bigotimes_1^n P \right)_{T_\gamma}$ denotes the distribution of T_γ under $\bigotimes_1^n P$

Since obviously $(\gamma, \ldots, \gamma)\in B$ holds iff $(\gamma, \ldots, \gamma)\in(h\circ S)^{-1}(B)$, the right hand sides of (i) and (ii) coincide. Therefore, the same is true for the left hand sides, which together with the $S^{-1}(\mathfrak{B})-\mathfrak{B}$-measurability of g yields the sufficiency of S.

For the only if part, let $z\in\mathbf{R}$ and $B:=\{X_2-X_1<z\}$. According to the assumed sufficiency of S there exists a (measurable) function $g\colon(\mathbf{R},\mathfrak{B})\to(\mathbf{R},\mathfrak{B})$ with $g\circ S=E_\gamma^S(1_B)$ $(\gamma\in\mathbf{R})$, where E_γ^S denotes a conditional expectation with respect to P_γ under S. Taking into account the translation invariance of B we obtain

$$\text{(iii)}\quad \begin{aligned} P_0(B\cap\{S<t\}) &= P_\gamma(B\cap\{S<t+\gamma\}) = \int_{\{S<t+\gamma\}} 1_B\,dP_\gamma = \\ &= \int_{\{S<t+\gamma\}} g\circ S\,dP_\gamma = \int_{\{S<t\}} g(S+\gamma)\,dP_0 \quad (t\in\overline{\mathbf{R}};\ \gamma\in\mathbf{R}). \end{aligned}$$

Therefore, if $T\colon\mathbf{R}^n\to\mathbf{R}$ is the statistic given by $T(y)=\frac{1}{n}\sum y_i$ $(y=(y_1, \ldots, y_n)\in\mathbf{R}^n)$, we get with the aid of FUBINI's Theorem

$$\text{(iv)}\quad \begin{aligned} P_0(B\cap\{S<t\}) &= \int\Big[\int_{\{S<t\}} g\big(S(x)+T(y)\big)\,dP_0(x)\Big]\,dP_0(y) = \\ &= \int_{\{S<t\}}\Big[\int g\big(S(x)+T(y)\big)\,dP_0(y)\Big]\,dP_0(x) = P_0(B)\cdot P_0\{S<t\} \quad (t\in\overline{\mathbf{R}}). \end{aligned}$$

Recalling the definition of B (iv) yields the independence of $X_1+\ldots+X_n$ and X_2-X_1 under P_0. Since $X_1, \ldots, X_n$ are iid with distribution P under P_0, it follows by the well known DARMOIS—SKITOVICH Theorem that P is a normal distribution or a point mass.

II. Sufficiency of the sample mean with respect to scaling classes

Let P denote any probability measure on $\mathfrak{B}$ with $P(\mathbf{R}_+-\{0\})=1$, i.e., P is concentrated on the positive real axis; for $\gamma>0$ let S_γ be the statistic given by $S_\gamma(x)=(\gamma x_1, \ldots, \gamma x_n)$ $(x=(x_1, \ldots, x_n)\in\mathbf{R}^n)$.

Theorem 2. *For* $n\geqq 2$ *let* $\mathfrak{W}=\Big\{P_\gamma=\Big(\bigotimes_1^n P\Big)_{S_\gamma}\,\Big|\,\gamma\in\mathbf{R}_+-\{0\}\Big\}$ *be a scaling class. Then* $S=\overline{X}$ *is sufficient with respect to* $\mathfrak{W}$ *iff* P *is a* Γ*-distribution or a point mass.*

PROOF. This proof is quite similar to the first one. The if part is again well known in the non degenerate case and follows from Theorem 1 in the degenerate case. For the only if part, let $z\in\mathbf{R}$ and

$$B:=\{X_1\cdot[X_2+\ldots+X_n]^{-1}<z,\quad X_j>0\quad(1\leqq j\leqq n)\};$$

let again $g\colon(\mathbf{R},\mathfrak{B})\to(\mathbf{R},\mathfrak{B})$ be a function with $g\circ S=E_\gamma^S(1_{\mathfrak{B}})$ $(\gamma>0)$. With the same arguments as in (iii) we obtain

$$P_1(B\cap\{S<t\}) = \int_{\{S<t\}} g(S\gamma)\,dP_1 \quad (t\in\overline{\mathbf{R}};\ \gamma>0);$$

therefore the same conclusions, which yielded (iv), lead to

$$P_1(B\cap\{S<t\})=P_1(B)\cdot P_1\{S<t\}\quad (t\in\overline{\mathbf{R}}).$$

In view of the definition of B this relation indicates the independence of $X_1\cdot\cdot[X_2+\ldots+X_n]^{-1}$ and $X_1+\ldots+X_n$. By a theorem of LUKACS [5] this implies the statement of Theorem 2.

Acknowledgment

I point aut that meanwhile the book "Mathematische Statistik" by Eberl-Moeschlin (de Gruyter, 1981) has been publis hed which contains Theorem 1 as Example 2.3.14 and Theorem 2.3.15

References

[1] P. BÁRTFAI, Characterizations by sufficient statistics; in Colloquia Mathematica Societatis János Bolyai 21. *Analytic Function Methods in Probability Theory (Debrecen)*, edited by B. GYIRES (1980).

[2] A. M. Каган, Теория оценивания для семейств с параметрами сдвига, масштаба и экспонентных; *Труды матем. ин.-та им. стеклова ан СССР* **104** (1968)

[3] Ф. M. Каган, Некоторые теоремы о характеризации гамма-распределения и близких к к нему; *литовкий матем. сборник VIII*, **2** (1968)

[4] A. M. KAGAN, YU. V. LINNIK and C. R. RAO, Characterization Problems in Mathematical Statistics, *John Wiley, New York* (1973) *(translation from the Russian original)* Характеризационные задачи математической статистики; изд.-во наука, *Москва* (1972)

[5] E. LUKACS, A characterization of the Gamma distribution, *AMS* **26** (1955).

WALTHER EBERL, JR.
MATHEMATICAL DEPARTMENT
UNIVERSITY OF HAGEN
LÜTZOWSTRASSE 125
D-5800 HAGEN, FRG

(Received July 5, 1980)