Ann. Inst. Statist. Math.
35 (1983), Part A, 91-94

ON A CHARACTERISTIC PROPERTY OF THE UNIFORM DISTRIBUTION

R. SHIMIZU AND J. S. HUANG

(Received Sept. 9, 1982)

Summary

Let $X_{1,n}\leqq X_{2,n}\leqq\cdots\leqq X_{n,n}$ be the order statistics based on a sample from an absolutely continuous distribution F. It is proved that the uniform distribution on the interval $[0, a]$ is characterized by the property that $X_{2,n}-X_{1,n}$ and $X_{1,n}$ are identically distributed.

Let $X_1,\cdots, X_n$ be i.i.d. random variables from a distribution F, with the corresponding order statistics denoted by $X_{1,n}\leqq\cdots\leqq X_{n,n}$. If F is the uniform distribution on the interval $[0, a]$, $a>0$, the spacings $X_{k+1,n}-X_{k,n}$, $k=1,\cdots, n-1$, are all identically distributed as $X_{1,n}$. The converse is true under various regularity conditions ([6], [4], [5]), thus characterizing the uniform distribution. The problem first arose in the context of characterizing the exponential distribution and using it to test the composite hypothesis of exponential distribution with unspecified scale parameter ([6]), and was later extended to goodness of fit test for the normal distribution ([1], [2], [3]). For a discussion of how these results are related to the characterization of the uniform distribution, see [5] (Section 3). The main result of [5] is their Theorem 2: If for some n, $X_{2,n}-X_{1,n}$ and $X_{1,n}$ are identically distributed, if the support of F is a finite interval, and if its density is continuous with finite limits at the end points of the supporting interval, then F is uniform. In this note we present a refinement of the above theorem. Here we are able to dispense with the conditions of finite support and the continuity of the density.

THEOREM. *Uniform distribution is the only absolutely continuous distribution whose $X_{2,n}-X_{1,n}$ and $X_{1,n}$ are identically distributed.*

PROOF. Let $\bar{F}\equiv 1-F$. Identical distribution of $X_{1,n}$ and $X_{2,n}-X_{1,n}$ means

$$\bar{F}^n(x)=n\int_0^\infty \bar{F}^{n-1}(x+y)dF(y)\ , \qquad x\geqq 0\ . \tag{1}$$

Differentiating the above, with f being any version of the density of F, we get

$$(2)\qquad \bar{F}^{n-1}(x)f(x)=(n-1)\int_0^\infty \bar{F}^{n-2}(x+y)f(x+y)f(y)dy\ ,$$

for $x\geqq 0$ a.e. Lebesgue. Let Ω be the set of x values satisfying (2). According to Lemma 1 of [5], the support of F is an interval of the form $[0, a]$, where a may be infinite. From (2) it follows that for each $x\in\Omega$ and $0\leqq x<a$,

$$(3)\qquad f(x)=\int_0^{a-x} f(y)dG_x(y)\ ,$$

where

$$G_x(y)\equiv 1-\left[\frac{\bar{F}(x+y)}{\bar{F}(x)}\right]^{n-1}$$

is a probability distribution function on $[0, a-x]$, and is strictly increasing therein ([5], Lemma 1). Let

$$c\equiv \operatorname*{ess\,inf}_{0\leqq x<a} f(x)\ .$$

Our plan is to show the finiteness of a (Lemmas 1 and 2), then the constancy of f over $[0, a]$.

LEMMA 1. *There exists* $\varepsilon>0$ *and* $\delta>0$ *such that* $f(x)\geqq\varepsilon$ *for all* $x\in\Omega$ *and* $0<x<\delta$.

PROOF. Suppose not. Then there exists in Ω a sequence $x_1, x_2, \cdots \to 0$ such that $f(x_k)\to 0$. Abbreviating G_{x_k} by G_k, we see that as $k\to\infty$, $G_k\to G_0\equiv 1-\bar{F}^{n-1}$. Put

$$A(\lambda)\equiv\{x\colon f(x)\geqq\lambda\}\ ,\qquad \lambda>0\ .$$

Then from (3) we have

$$(4)\qquad f(x_k)=\int_0^\infty f(y)dG_k(y)\geqq\int_{A(\lambda)} f(y)dG_k(y)\geqq\lambda\int_{A(\lambda)} dG_k(y)\ .$$

Letting $k\to\infty$ in (4), we get

$$0\geqq\lambda\int_{A(\lambda)} dG_0(y)\ ,$$

or, $A(\lambda)$ has G_0-measure zero for each $\lambda>0$. Since G_0 dominates F, $A(\lambda)$ also has F-measure zero, $\lambda>0$. Namely, $f\equiv 0$ a.e. F, which is impossible.

LEMMA 2. $a<\infty$.

PROOF. Suppose $a=\infty$. Then there exists a sequence $\{x_k\}$ of positive numbers such that $x_k \in \Omega$, $x_k \to \infty$ and $f(x_k)\to 0$. Again letting G_k be defined by G_{x_k}, we can assume without loss of generality that G_k converges weakly to a monotone function, G_0, say, such that $G_0(x)=0$ for $x<0$ and $G_0(\infty)\leqq 1$. It follows from (1)

$$\bar{F}(x_k)=n\int_0^\infty \left[\frac{\bar{F}(x_k+y)}{\bar{F}(x_k)}\right]^{n-1} dF(y)=n\int_0^\infty (1-G_k(y))dF(y) .$$

Letting $k\to\infty$, we obtain

$$0=\lim \bar{F}(x_k)=n\int_0^\infty (1-G_0(y))dF(y)$$

which implies $G_0(x)=1$ a.e. F. But as F is strictly monotone ([5], Lemma 1) and G_0 is also monotone, $G_0(x)=1$ for all $x>0$. On the other hand, by Lemma 1, there exist $\varepsilon>0$ and $\delta>0$ such that $f(x)\geqq\varepsilon$ for all $x\in\Omega\cap(0,\delta)$. Thus by (3), using the above x's we have

$$f(x_k)=\int_0^\infty f(y)dG_k(y)\geqq\int_0^\delta f(y)dG_k(y)\geqq\varepsilon\int_0^\delta dG_k(y) . \tag{5}$$

Letting $k\to\infty$ in (5), we arrive at a contradiction: $0\geqq\varepsilon$.

PROOF OF THE THEOREM (Continued). Let $x\in\Omega$, $a/2<x<a$. By (3), $f(x)$ is an average of f values over $[0, a-x]$, a subinterval of $[0, a/2]$. This implies that

$$f(x)\geqq \operatorname*{ess\,inf}_{0\leqq t\leqq a-x} f(t)\geqq \operatorname*{ess\,inf}_{0\leqq t\leqq a/2} f(t) ,$$

and hence

$$\operatorname*{ess\,inf}_{0\leqq t\leqq a/2} f(t)=\operatorname*{ess\,inf}_{0\leqq t\leqq a} f(t) \qquad (=c) .$$

Let $\{x_k\}$ be a sequence in $\Omega\cap[0, a/2]$, converging to, say z ($\leqq a/2$), with $f(x_k)\to c$. For each $\varepsilon>0$ put

$$B=B(\varepsilon)=\{x:\ f(x)>c+\varepsilon\} ,$$

and notice

$$f(x_k)=\int_B f(y)dG_k(y)+\int_{\bar{B}} f(y)dG_k(y)\geqq c+\varepsilon\int_B dG_k(y) .$$

Taking limit $k\to\infty$ again, we get

$$c\geqq c+\varepsilon\int_{B(\varepsilon)} dG_z(y) .$$

Thus $B(\varepsilon)$ has G_z-measure zero for each $\varepsilon>0$. But $G_z(y)=1-[\bar{F}(z+y)/$

$\bar{F}(z)]^{n-1}$ concentrates on the interval $[0, a-z]$, and dominates the shifted-F. Thus

$$\int_{(B(\varepsilon)+z)\cap(z,a)} dF(y)=0 .$$

This means that the set $\{z: f(x-z)>c+\varepsilon,\ x>z\}$ has F-measure zero for each $\varepsilon>0$, or,

$$f(x)=c , \qquad 0<x<a-z \text{ a.e. } F .$$

Now, for each x in $(a-z)\cap\Omega$, since $f(x)$ is an average of f values over $(0, a-x)$, a subinterval of $(0, a-z)$, and since f is constant a.e. there, it follows that f is also constant $(=c)$ a.e. over $(a-z, a)$. Hence $f=c$ a.e. over the entire interval $(0, a)$.

THE INSTITUTE OF STATISTICAL MATHEMATICS
UNIVERSITY OF GUELPH

REFERENCES

[1] Csörgő, M. and Seshadri, V. (1971). Characterizing the Gaussian and exponential laws via mappings onto the unit interval, *Zeit. Wahrscheinlichkeitsth.*, **18**, 333-339.

[2] Csörgő, M. and Seshadri, V. (1971). Characterizations of the Behrens-Fisher and related problems (A goodness of fit point of view), *Theory Prob. Appl.*, **16**, 23-35.

[3] Csörgő, M., Seshadri, V. and Yalovski, M. (1973). Some exact tests for normality in the presence of unknown parameters, *J. R. Statist. Soc.*, B, **35**, 507-522.

[4] Csörgő, M., Seshadri, V. and Ylaovski, M. (1975). Applications of characterizations in the area of goodness of fit, *Statist. Dist. in Scientific Work*, **2**, (eds. G. P. Patil, S. Kotz and J. K. Ord), Reidel, Boston, 79-90.

[5] Huang, J. S., Arnold, B. C. and Ghosh, M. (1979). On characterizations of the uniform distribution based on identically distributed spacings, *Sankhyā*, B, **41**, 109-115.

[6] Seshadri, V., Csörgő, M. and Stephens, M. (1969). Tests for the exponential distribution using Kolmogorov-type statistics, *J. R. Statist. Soc.*, B, **31**, 499-509.