DISTRIBUTION RESULTS AND POWER FUNCTIONS FOR KAC STATISTICS*

MIKLÓS CSÖRGÖ AND MAYER ALVO

(Received Nov. 12, 1968; revised Jan. 5, 1970)

This note is concerned with random samples of the form $N_\lambda, X_1, \cdots, X_{N_\lambda}$ defined on some probabilitiy space $(\Omega, \mathfrak{A}, P)$, where N_λ is a Poisson random variable with mean λ and the X_i have some continuous distribution function F. Following M. Kac [8] we define the modified empirical distribution function

$$(1) \qquad F_\lambda^*(y)=\lambda^{-1}\sum_{j=1}^{N_\lambda}\Psi_y(X_j)\,, \qquad -\infty<y<+\infty\,,$$

where $\Psi_y(x)$ is 0 or 1 according as $x>y$ or $x\leqq y$ and the sum is taken to be zero if $N_\lambda=0$. Analogous one and two sided Kac statistics of the original one and two sided Kolmogorov statistics are

$$\text{l.u.b.}_{-\infty<y<+\infty}[F(y)-F_\lambda^*(y)] \qquad \text{and} \qquad \text{l.u.b.}_{-\infty<y<+\infty}|F(y)-F_\lambda^*(y)|$$

respectively. The exact and limiting distribution of the first one of these random variables was studied by J. L. Allen and J. A. Beekman [1], and they also studied the exact distribution of the two sided Kac statistic [2] whose asymptotic distribution was found by M. Kac [8]. As long as F is continuous, the distribution of the Kac statistics is independent of F and we can therefore confine our attention to the simple case $F(y)=x$, $0\leqq x\leqq 1$.

Let n be a positive integer and $Y_1<Y_2<\cdots<Y_n$ be the order statistics corresponding to $X_1, X_2, \cdots, X_n$. Define

$$(2) \qquad F_{n,\lambda}(y)=\begin{cases} 0\,, & y<Y_1 \\ k/\lambda\,, & Y_k\leqq y<Y_{k+1},\ k=1,2,\cdots,n-1 \\ n/\lambda\,, & y\geqq Y_n\,. \end{cases}$$

Thus $F_{n,\lambda}(y)=(n/\lambda)F_n(y)$, where $F_n(y)$ is the ordinary empirical distribu-

* Research supported in part by operating grants of the National Research Council of Canada.

tion function. $F_{n,\lambda}(y)$ as defined here will be used in the sequel. Let y_b be a real number with $F(y_b)=b$. We will now derive an explicit form for

$$(3)\qquad P_\lambda(\varepsilon, b)=P\{\text{l.u.b.}_{-\infty<y\leq y_b}[F(y)-F_\lambda^*(y)]\leqq\varepsilon\}$$

and for

$$(4)\qquad M_\lambda(\varepsilon, b)=P\left\{\text{l.u.b.}_{-\infty<y\leq y_b}\frac{F(y)-F_\lambda^*(y)}{1-F(y)}\leqq\varepsilon\right\}.$$

THEOREM 1. *For* $N_\lambda, X_1, X_2, \cdots$ *subject to the previous conditions, and* $0<\varepsilon\leqq b\leqq 1$,

$$(5)\qquad \begin{aligned} P_\lambda(\varepsilon, b)&=1-\varepsilon\lambda\sum_{j=0}^{[\lambda(b-\varepsilon)]}[(\lambda\varepsilon+j)^{j-1}/j!]e^{-\lambda\varepsilon-j}\\ &=\varepsilon\lambda\sum_{j=[\lambda(b-\varepsilon)]+1}^{\infty}[(\lambda\varepsilon+j)^{j-1}/j!]e^{-\lambda\varepsilon-j}. \end{aligned}$$

PROOF. By the independence of $N_\lambda, X_1, X_2, \cdots$ and the distribution free property,

$$(6)\qquad P(\varepsilon, b)=\sum_{n=0}^{\infty}(\lambda^n e^{-\lambda}/n!)P_r\{\text{l.u.b.}_{0<y\leq b}(y-F_{n,\lambda}(y))\leqq\varepsilon\}.$$

Following the proof of Theorem 1 of [7] and slightly changing the analysis to handle the extra parameter λ, one obtains

$$(7)\qquad \begin{aligned} &P_r\{\text{l.u.b.}_{0<y\leq b}(y-F_{n,\lambda}(y))\leqq\varepsilon\}\\ &=1-\varepsilon\sum_{j=0}^{[\lambda(b-\varepsilon)]}\binom{n}{j}(1-\varepsilon-(j/\lambda))^{n-j}(\varepsilon+(j/\lambda))^{j-1}, \end{aligned}$$

and substituting this expression in (6) and interchanging order of summation and summing on n, we obtain the first equality of (5). The second equality of (5) follows from the fact that

$$\sum_{j=0}^{n}\binom{n}{j}(1-\varepsilon-(j/\lambda))^{n-j}(\varepsilon+(j/\lambda))^{j-1}\varepsilon=1,$$

which statement, in turn, follows immediately from Lemma 3 of [6].

THEOREM 2. *For* $N_\lambda, X_1, X_2, \cdots$ *subject to the previous conditions,* $\varepsilon>0$ *and* b *such that* $0<\varepsilon/(1+\varepsilon)\leqq b<1$,

$$(8)\qquad \begin{aligned} M(\varepsilon, b)=1-\frac{\varepsilon}{1+\varepsilon}\lambda\sum_{j=0}^{[\lambda\{(1+\varepsilon)b-\varepsilon\}]}&\left[\left(\frac{\lambda\varepsilon}{1+\varepsilon}+\frac{j}{1+\varepsilon}\right)^{j-1}\Big/ j!\right]\\ &\cdot\exp\left(-\frac{\lambda\varepsilon}{1+\varepsilon}-\frac{j}{1+\varepsilon}\right) \end{aligned}$$

$$=\frac{\varepsilon}{1+\varepsilon}\lambda\sum_{j=[\lambda\{(1+\varepsilon)b-\varepsilon\}]+1}^{\infty}\left[\left(\frac{\lambda\varepsilon}{1+\varepsilon}+\frac{j}{1+\varepsilon}\right)^{j-1}\Big/ j!\right]$$
$$\cdot\exp\left(-\frac{\lambda\varepsilon}{1+\varepsilon}-\frac{j}{1+\varepsilon}\right).$$

PROOF. By the independence of $N_\lambda, X_1, X_2, \cdots$ and the distribution free property,

$$M_\lambda(\varepsilon, b)=\sum_{n=0}^{\infty}(\lambda^n e^{-\lambda}/n!)P_r\left\{\text{l.u.b.}_{0<y<b}\frac{y-F_{n,\lambda}(y)}{1-y}\leqq\varepsilon\right\}. \tag{9}$$

Following the proof Theorem 1 of [6] and slightly changing the analysis to handle the extra parameter λ, one obtains

$$P_r\left\{\text{l.u.b.}_{0<y<b}\frac{y-F_{n,\lambda}(y)}{1-y}\leqq\varepsilon\right\} \tag{10}$$
$$=1+\frac{\varepsilon}{1+\varepsilon}\sum_{j=0}^{[\lambda\{(1+\varepsilon)b-\varepsilon\}]}\binom{n}{j}\left(1-\frac{j}{\lambda(1+\varepsilon)}-\frac{\varepsilon}{1+\varepsilon}\right)^{n-j}$$
$$\cdot\left(\frac{j}{\lambda(1+\varepsilon)}+\frac{\varepsilon}{1+\varepsilon}\right)^{j-1}$$
$$=\frac{\varepsilon}{1+\varepsilon}\sum_{j=[\lambda\{(1+\varepsilon)b-\varepsilon\}]+1}^{n}\binom{n}{j}\left(1-\frac{j}{\lambda(1+\varepsilon)}-\frac{\varepsilon}{1+\varepsilon}\right)^{n-j}$$
$$\cdot\left(\frac{j}{\lambda(1+\varepsilon)}+\frac{\varepsilon}{1+\varepsilon}\right)^{j-1}.$$

Substituting these two expressions in (9) and interchanging order of summation and summing on n we obtain both forms of (8).

In our proofs we have used the trivial fact that, if N_λ is independent of the X_i, for any arbitrary measurable function φ we have

$$P\{\varphi[F_\lambda^*(y), F(y)]<\varepsilon\}=\sum_{n=0}^{\infty}(\lambda^n e^{-\lambda}/n!)P\{\varphi[F_{n,\lambda}(y), F(y)]<\varepsilon\}.$$

Using results of [3] and [5] for $P\{\varphi[F_{n,\lambda}(y), F(y)]<\varepsilon\mid F=K\}$ and this relation one can determine computational methods for exact power functions for the general hypothesis testing problem

$$H_0: F=H \text{ versus } H_1: F=K$$

where H and K are continuous distribution functions.

McGILL UNIVERSITY, MONTREAL

REFERENCES

[1] Allen, J. L. and Beekman, J. A. (1966). A statistical test involving a random number of random variables, *Ann. Math. Statist.*, **37**, 1305-1309.

[2] Allen, J. L. and Beekman, J. A. (1967). Distribution of a M. Kac statistic, *Ann. Math. Statist.*, **38**, 1919-1923.

[3] Alvo, M. (1968). *Power functions of One-sided Kolmogorov-Smirnov statistics*, To be published.

[4] Birnbaum, Z. W. and Tingey, F. H. (1951). One sided confidence contours for probability distribution function, *Ann. Math. Statist.*, **22**, 592-596.

[5] Birnbaum, Z. W. (1953). On the power of a one sided test of fit for continuous probability functions, *Ann. Math. Statist.*, **24**, 484-489.

[6] Csörgö, M. (1965). Exact probability distribution functions of some Rényi type statistics, *Proc. Amer. Math. Soc.*, **16**, 1158-1167.

[7] Csörgö, M. (1965). Exact and limiting probability distribution of some Smirnov type statistics, *Canadian Math. Bull.*, 8, 93-103.

[8] Kac, M. (1949). On deviations between theoretical and empirical distributions, *Proc. Nat. Acad. Sci.*, U.S.A., **35**, 252-257.

[9] Rényi, A. (1953). On the theory of order statistics, *Acta. Math. Acad. Sci. Hungar.*, **4**, 191-231.