DISTRIBUTIONS OF KAC-STATISTICS

GIITIRO SUZUKI

(Received April 19, 1972)

1. Let $N=N_\lambda$, $X_1, X_2, \cdots, X_N$ be independent random variables, N having a Poisson distribution with mean λ and each X_i having the same continuous distribution function F and being independent of N. Furthermore the modified empirical distribution function is defined by

$$F_\lambda^*(x)=\frac{1}{\lambda}\sum_{i=1}^{N}\Delta(X_i, x) ,$$

where $\Delta(y, x)$ equals 0 or 1 according as $y>x$ or $y\leqq x$ respectively.

The purpose of this note is to give a systematic computational method of the probability

$$(1) \qquad \alpha_\lambda^*(\beta, \gamma)=\Pr\{\beta[F(x)]\leqq F_\lambda^*(x)\leqq\gamma[F(x)],\ \text{for all } x\} .$$

Here β and γ are monotone non-decreasing functions on $[0, 1]$ with β continuous to the left and γ continuous to the right. When β and γ are the linear functions

$$\beta(t)=t-\xi , \qquad \gamma(t)=t+\xi ,$$

$\alpha_\lambda^*(\beta, \gamma)$ represents the probability distribution of the Kac-statistic

$$(2) \qquad K(\lambda)=\sup_{-\infty<x<\infty}\left|F_\lambda^*(x)-F(x)\right| .$$

2. Let $U_1, U_2, \cdots$ be the independent random variables having the uniform distribution on $[0, 1]$. Since

$$\begin{aligned}
&\Pr\{\beta[F(x)]\leqq F_\lambda^*(x)\leqq\gamma[F(x)],\ -\infty<x<\infty\,|\,N=n\}\\
&=\Pr\left\{\beta[F(x)]\leqq\frac{1}{\lambda}\sum_{i=1}^{n}\Delta(X_i, x)\leqq\gamma[F(x)],\ -\infty<x<\infty\right\}\\
&=\Pr\left\{\beta[F(x)]\leqq\frac{1}{\lambda}\sum_{i=1}^{n}\Delta(F(X_i), F(x))\leqq\gamma[F(x)],\ -\infty<x<\infty\right\}\\
&=\Pr\left\{\beta(t)\leqq\frac{1}{\lambda}\sum_{i=1}^{n}\Delta(U_i, t)\leqq\gamma(t),\ 0<t<1\right\}\equiv\alpha_n^{(\lambda)}(\beta, \gamma) ,
\end{aligned}$$

then we can express the probability (1) as

$$\begin{aligned}\alpha_\lambda^*(\beta, \gamma) &= \sum_{n=0}^{\infty} \Pr\{N=n\}\alpha_n^{\langle\lambda\rangle}(\beta, \gamma) \\ &= \sum_{n=0}^{\infty} \frac{\lambda^n e^{-\lambda}}{n!}\alpha_n^{\langle\lambda\rangle}(\beta, \gamma) .\end{aligned}$$

It is easily seen that $\alpha_n^{\langle\lambda\rangle}(\beta, \gamma)=0$ for $n<\lambda\beta(1)$ or $n>\lambda\gamma(1)$. Then we have

$$\alpha_\lambda^*(\beta, \gamma) = \sum_{n=m_1}^{m_2} \frac{\lambda^n e^{-\lambda}}{n!}\alpha_n^{\langle\lambda\rangle}(\beta, \gamma) , \tag{3}$$

where $m_1=m_1(\beta)$ is the smallest integer not smaller than $\lambda\beta(1)$ and $m_2=m_2(\gamma)$ is the largest integer smaller than $\lambda\gamma(1)$. From the assumptions on β and γ we can define

$$\nu_k^{\langle\lambda\rangle}=\nu_k^{\langle\lambda\rangle}(\beta)=\sup_{0\leqq t\leqq 1}\left\{t : \beta(t)\leqq\frac{k-1}{\lambda}\right\} , \qquad k=1, 2, \cdots, m_1 \tag{4}$$

$$\mu_k^{\langle\lambda\rangle}=\mu_k^{\langle\lambda\rangle}(\gamma)=\inf_{0\leqq t\leqq 1}\left\{t : \gamma(t)\geqq\frac{k}{\lambda}\right\} , \qquad k=1, 2, \cdots, m_2 \tag{5}$$

and for each n $(m_1\leqq n\leqq m_2)$,

$$\begin{aligned}\alpha_n^{\langle\lambda\rangle}(\beta, \gamma) &= \Pr\{\mu_1^{\langle\lambda\rangle}\leqq U_1^{(n)}\leqq\nu_1^{\langle\lambda\rangle}, \cdots, \mu_{m_1}^{\langle\lambda\rangle}\leqq U_{m_1}^{(n)}\leqq\nu_{m_1}^{\langle\lambda\rangle}, \\ &\qquad \mu_{m_1+1}^{\langle\lambda\rangle}\leqq U_{m_1+1}^{(n)}, \cdots, \mu_n^{\langle\lambda\rangle}\leqq U_n^{(n)}\} \\ &\equiv g_n^{\langle\lambda\rangle}(\nu_1^{\langle\lambda\rangle}, \cdots, \nu_{m_1}^{\langle\lambda\rangle}; \mu_1^{\langle\lambda\rangle}, \cdots, \mu_n^{\langle\lambda\rangle}) ,\end{aligned} \tag{6}$$

where $U_1^{(n)}, \cdots, U_n^{(n)}$ are the order statistics from a sample of n independent uniform random variables. Thus we have

THEOREM 1. *The probability* (1) *can be calculated by* (3) *and* (6), *where the values of* $g_n^{\langle\lambda\rangle}$ *are obtained by Theorem* 4 *of Suzuki* [7] *or Theorem of Steck* [6].

3. Next we proceed to the one-sided case

$$\alpha_\lambda^*(\beta, \infty)=\Pr\{\beta[F(x)]\leqq F_\lambda^*(x), \text{ for all } x\} , \tag{7}$$

$$\alpha_\lambda^*(0, \gamma)=\Pr\{F_\lambda^*(x)\leqq\gamma[F(x)], \text{ for all } x\} , \tag{8}$$

which are the generalized forms of the one-sided Kac-statistics

$$K^+(\lambda)=\sup\,[F(x)-F_\lambda^*(x)] , \tag{9}$$

$$K^-(\lambda)=\sup\,[F_\lambda^*(x)-F(x)] . \tag{10}$$

when $\gamma=\infty$ we have $m_2=m_2(\gamma)=\infty$ and

$$\alpha_\lambda^*(\beta, \infty)=\frac{\lambda^{m_1}e^{-\lambda}}{m_1!}\sum_{j=0}^{\infty}\lambda^j\alpha_{\lambda,j}\ , \tag{11}$$

where

$$\alpha_{\lambda,j}=\frac{m_1!}{(m_1+j)!}\alpha_{m_1+j}^{\langle\lambda\rangle}(\beta, \infty)=\frac{m_1!}{(m_1+j)!}\Pr\{U_1^{(m_1+j)}<\nu_1^{\langle\lambda\rangle},\cdots, U_{m_1}^{(m_1+j)}<\nu_{m_1}^{\langle\lambda\rangle}\}\ .$$

We shall now define the following polynomials for any $n\geqq 1$ and $\mu_1,\cdots,\mu_n$ $(0\leqq\mu_1\leqq\cdots\leqq\mu_n\leqq 1)$,

$$Q_0=1$$

$$Q_k=Q_k(\mu_1,\cdots,\mu_k)=-\sum_{i=0}^{k-1}\binom{k}{i}\mu_k^{k-i}Q_i\ ,\qquad k=1, 2,\cdots, n\ .$$

Then we have from Theorem 3 of Suzuki [7]

$$\Pr\{U_1^{(n)}\geqq\mu_1,\cdots, U_n^{(n)}\geqq\mu_n\}=\sum_{k=0}^{n}\binom{n}{k}Q_k\equiv f_n(\mu_1,\cdots,\mu_n)\ .$$

Therefore putting $\mu_1=1-\nu_{m_1}^{\langle\lambda\rangle},\cdots,\mu_{m_1}=1-\nu_1^{\langle\lambda\rangle}$,

$$\begin{aligned}\alpha_{\lambda,j}&=\frac{m_1!}{(m_1+j)!}\Pr\{U_1^{(m_1+j)}\leqq\nu_1^{\langle\lambda\rangle},\cdots, U_{m_1}^{(m_1+j)}\leqq\nu_{m_1}^{\langle\lambda\rangle}\}\\&=\frac{m_1!}{(m_1+j)!}f_{m_1+j}(\overbrace{0,\cdots,0}^{j},\mu_1,\cdots,\mu_{m_1})\\&=\frac{m_1!}{(m_1+j)!}\left[1+\sum_{i=1}^{m_1}\binom{m_1+j}{i+j}Q_{i+j}(\overbrace{0,\cdots,0}^{j},\mu_1,\cdots,\mu_i)\right]\\&=\sum_{i=0}^{m_1}\binom{m_1}{i}Q_i^{(j)}\ ,\end{aligned}$$

where

$$\begin{aligned}&Q_0^{(j)}=\frac{m_1!}{(m_1+j)!}\\&Q_i^{(j)}=Q_i^{(j)}(\mu_1,\cdots,\mu_i)=\frac{i!}{(i+j)!}Q_{i+j}(\overbrace{0,\cdots,0}^{j},\mu_1,\cdots,\mu_i)\ .\end{aligned}\tag{12}$$

Consequently from (11)

$$\alpha_\lambda^*(\beta, \infty)=\frac{\lambda^{m_1}e^{-\lambda}}{m_1!}\sum_{j=0}^{\infty}\lambda^j\sum_{i=0}^{m_1}\binom{m_1}{i}Q_i^{(j)}\ , \tag{13}$$

It is easily shown by induction that

(14) $$Q_k=\sum_{i=0}^{k-1}(-1)^i\binom{k}{i}f_{k-i}+(-1)^kQ_0 .$$

Noting that f_k represents probability, we have

$$|Q_i^{(j)}|\leqq\frac{i!}{(i+j)!}2^{i+j}$$

and then $\sum_{j=0}^{\infty}\lambda^j|Q_i^{(j)}|\leqq 2^i i! e^{2\lambda}$. Thus we can express the relation (13) as

(15) $$\alpha_\lambda^*(\beta, \infty)=\frac{\lambda^{m_1}e^{-\lambda}}{m_1!}\sum_{i=0}^{m_1}\binom{m_1}{i}Q_i^{\langle\lambda\rangle} ,$$

where

(16) $$Q_i^{\langle\lambda\rangle}=\sum_{j=0}^{\infty}\lambda^jQ_i^{(j)} , \qquad i=0, 1,\cdots, m_1 .$$

It should be more convenient that the relation (16) is expressed by some recurrence formula. From the definition of Q_h

$$\sum_{h=0}^{k+j}\binom{k+j}{h}\mu_{k+j}^{k+j-h}Q_h(\mu_1', \cdots, \mu_h')=0 .$$

Putting $\mu_1'=\cdots=\mu_j'=0$, $\mu_{j+1}'=\mu_1,\cdots, \mu_{k+j}'=\mu_k$, we have

$$\mu_k^{k+j}+\sum_{i=1}^{k}\binom{k+j}{i+j}\mu_k^{k-i}Q_{i+j}(\overbrace{0,\cdots, 0}^{j}, \mu_1, \cdots, \mu_i)=0 ,$$

i.e.

$$\mu_k^{k+j}+\frac{(k+j)!}{k!}\sum_{i=1}^{k}\binom{k}{i}\mu_k^{k-i}Q_i^{(j)}(\mu_1, \cdots, \mu_i)=0 .$$

Multiplying $k!\lambda^j/(k+j)!$ and summing over j, we have

PROOF OF (14). Since relation (14) holds for $k=1$, we suppose the relation (14) is true for $k\leqq k_0$. Then denoting $f_0\equiv 0$ we have

$$\begin{aligned}
Q_{k_0+1}&=f_{k_0+1}-\sum_{j=0}^{k_0}\binom{k_0+1}{j}Q_i\\
&=f_{k_0+1}-\sum_{j=1}^{k_0}\binom{k_0+1}{j}\left[\sum_{i=0}^{j-1}(-1)^i\binom{j}{i}f_{j-i}+(-1)^jQ_0\right]-Q_0\\
&=f_{k_0+1}-\sum_{h=1}^{k_0}\sum_{i=0}^{k_0-h}(-1)^i\binom{k_0+1}{i}\binom{k_0+1-i}{h}f_h+(-1)^{k_0+1}Q_0\\
&=f_{k_0+1}-\sum_{h=1}^{k_0}\binom{k_0+1}{k_0+1-h}f_h\sum_{i=0}^{k_0-h}\binom{k_0+1-h}{i}(-1)^i+(-1)^{k_0+1}Q_0\\
&=f_{k_0+1}+\sum_{h=0}^{k_0}(-1)^{k_0+1-h}\binom{k_0+1}{k_0+1-h}f_h+(-1)^{k_0+1}Q_0\\
&=f_{k_0+1}+\sum_{i=1}^{k_0}(-1)^i\binom{k_0+1}{i}f_{k_0+1-i}+(-1)^{k_0+1}Q_0\\
&=\sum_{i=0}^{k_0}(-1)^i\binom{k_0+1}{i}f_{k_0+1-i}+(-1)^{k_0+1}Q_0 .
\end{aligned}$$

$$\sum_{j=0}^{\infty}\frac{k!}{(k+j)}\lambda^j\mu_k^{k+j}+\sum_{i=1}^{k}\binom{k}{i}\mu_k^{k-i}Q_i^{\langle\lambda\rangle}(\mu_1,\cdots,\mu_i)=0\ .$$

Consequently

$$Q_0^{\langle\lambda\rangle}=\frac{m_1!}{\lambda^{m_1}}\sum_{i=m_1}^{\infty}\frac{\lambda^i}{i!}$$

(17)
$$Q_1^{\langle\lambda\rangle}=Q_1^{\langle\lambda\rangle}(\mu_1)=-\frac{1}{\lambda}\sum_{i=1}^{\infty}\frac{(\lambda\mu_1)^i}{i!}$$

$$Q_k^{\langle\lambda\rangle}=Q_k^{\langle\lambda\rangle}(\mu_1,\cdots,\mu_k)=-\frac{k!}{\lambda^k}\sum_{i=k}^{\infty}\frac{(\lambda\mu_k)^i}{i!}-\sum_{i=1}^{k-1}\binom{k}{i}\mu_k^{k-i}Q_i^{\langle\lambda\rangle}\ ,\qquad k=2,\cdots,m_1\ .$$

Summalizing these we have

THEOREM 2. *The probability* (7) *can be calculated by* (14) *using the relation* (17).

4. We finally propose the computational method for the probability (8).

THEOREM 3. *The probability* (8) *can be calculated by*

$$\alpha_\lambda^*(0,\gamma)=\sum_{k=0}^{m_2}c_k(\lambda)Q_k(\mu_1^{\langle\lambda\rangle},\cdots,\mu_k^{\langle\lambda\rangle})\ ,$$

where $c_k(\lambda)=e^{-\lambda}\frac{\lambda^k}{k!}\sum_{l=0}^{m_2-k}\frac{\lambda^l}{l!}$.

PROOF. When $\beta=0$, $m_1=0$ and we have

$$\begin{aligned}\alpha_\lambda^*(0,\gamma)&=e^{-\lambda}+\sum_{n=1}^{m_2}\frac{\lambda^n e^{-\lambda}}{n!}\alpha_n^{\langle\lambda\rangle}(0,\gamma)\\&=e^{-\lambda}+\sum_{n=1}^{m_2}\frac{\lambda^n e^{-\lambda}}{n!}\Pr\{U_1^{(n)}\geqq\mu_1^{\langle\lambda\rangle},\cdots,U_n^{(n)}\geqq\mu_n^{\langle\lambda\rangle}\}\\&=e^{-\lambda}\left\{1+\sum_{n=1}^{m_2}\frac{\lambda^n}{n!}f_n(\mu_1^{\langle\lambda\rangle},\cdots,\mu_n^{\langle\lambda\rangle})\right\}\\&=e^{-\lambda}\left\{1+\sum_{n=1}^{m_2}\frac{\lambda^n}{n!}\sum_{k=0}^{n}\binom{n}{k}Q_k\right\}\\&=e^{-\lambda}\sum_{k=0}^{m_2}Q_k\sum_{n=k}^{m_2}\frac{\lambda^n}{k!(n-k)!}\\&=e^{-\lambda}\sum_{k=0}^{m_2}Q_k\frac{\lambda^k}{k!}\sum_{h=0}^{m_2-k}\frac{\lambda^h}{h!}\\&=\sum_{k=0}^{m_2}c_k(\lambda)Q_k(\mu_1,\cdots,\mu_k)\ .\end{aligned}$$

5. Finally we shall briefly sketch some historical results on distribution of Kac-statistics. We define the following functions:

$$\beta_1(t;\varepsilon,a,b)=\begin{cases}\min(t-\varepsilon,0) & 0\leqq t\leqq a\\ \max(t-\varepsilon,0) & a\leqq t\leqq b\\ b-\varepsilon & b\leqq t\leqq 1\end{cases}\qquad (0\leqq a<b\leqq 1)$$

$$\beta_2(t;\varepsilon,b)=\begin{cases}0 & 0\leqq t\leqq \varepsilon/(1+\varepsilon)\\ (1+\varepsilon)t-\varepsilon & \varepsilon/(1+\varepsilon)\leqq t\leqq b\\ (1+\varepsilon)b-\varepsilon & b\leqq t\leqq 1\end{cases}\qquad (0<\varepsilon/(1+\varepsilon)\leqq b\leqq 1)$$

$$\gamma_1(t;\varepsilon,a,b)=\begin{cases}a+\varepsilon & 0\leqq t\leqq a\\ t+\varepsilon & a\leqq t\leqq b\\ \infty & b\leqq t\leqq 1\end{cases}\qquad (0\leqq a<b\leqq 1)\,.$$

Then we can summarize various results on Kac-statistics as follows.

Type of statistics	Finite form	Limit form ($\lambda\to\infty$)
One-sided Kolmogorov type		
$\beta_1(\cdot\,;\varepsilon,a,b)$		Csörgő [3], Theorem 2
$\gamma=\infty$		
$a=0$	Csörgő and Alvo [4], Theorem 1	Csörgő [3], Theorem 1
$\varepsilon=0$		Csörgő [3], Corollary 2
$a=0,\ b=1$	Takács [8], Theorem 5 Allen and Beekman [1], Theorem 1	Allen and Beekman [1], Theorem 2
Adjoint form		
$\beta=0$	Takács [8], Theorem 4	
$\gamma_1=(\cdot\,;\varepsilon,0,1)$		
One-sided Reny type		
$\beta_2(\cdot\,;\varepsilon,b)$	Csörgő and Alvo [4], Theorem 2	
$\gamma=\infty$		

(Continued)

Type of statistics	Finite form	Limit form ($\lambda\to\infty$)
Two-sided Kolmogorov type		
$\beta_1(\cdot\,;\,\varepsilon, a, b)$		
$\gamma_1(\cdot\,;\,\varepsilon, a, b)$		
$a=a'=0$		Csörgő [3], Theorem 5
$a=a',\ b=b'$		Csörgő [3], Theorem 4
$a=a'=0,\ b=b'$		Csörgő [3], Theorem 3
$a=a'=0,\ b=b'=1$	Allen and Beekman [2], Theorem 1	Kac [5]

INSTITUTE OF STATISTICAL MATHEMATICS

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). Distributions of an M. Kac statistic, *Ann. Math. Statist.*, **38**, 1919-1923.
[3] Csörgő, M. (1972). Distribution results for distance functions based on the modified empirical distribution function of M. Kac, *Ann. Inst. Statist., Math.*, **24**, 101-110.
[4] Csörgő, M. and Alvo, M. (1970). Distribution results and power functions for Kac statistics, *Ann. Inst. Statist. Math.*, **22**, 257-260.
[5] Kac, M. (1949). On deviations between theoretical and empirical distributions, *Proc. Nat. Acad. Sci.*, **35**, 252-257.
[6] Steck, G. P. (1971). Rectangle probabilities for uniform order statistics and the probability that the empirical distribution function lies between two distribution functions, *Ann. Math. Statist.*, **42**, 1-11.
[7] Suzuki, G. (1967). On exact probabilities of some generalized Kolmogorov's D-statistics, *Ann. Inst. Statist. Math.*, **19**, 373-388.
[8] Takács, L. (1965). Applications of a ballot theorem in physics and in order statistics, *J. Roy. Statist. Soc. Ser. B*, **27**, 130-137.