A STATISTICAL TEST INVOLVING A RANDOM NUMBER OF RANDOM VARIABLES[1]

J. L. ALLEN AND J. A. BEEKMAN

Ball State University

0. Summary. In this paper is studied a technique based on samples of the form $N, X_1, X_2, \cdots, X_N$ where N has a Poisson distribution, and each X_i has the same continuous distribution function. Such samples, rather than fixed number samples, are appropriate for fixed time period observations where the number of occurrences is a Poisson variate, and are used in biology, insurance, and telephone engineering. We shall introduce a one-sided Kac statistic which is similar to the one-sided Kolmogorov statistic, derive forms for its finite dimensional and asymptotic distributions, find a lower bound for the power of the test, and prove that the test is "modified" consistent. Tabulations of the distributions will be given.

1. Introduction. Let $N, X_1, X_2, \cdots$ be independent random variables, N having a Poisson distribution with mean λ and each X_i having the same continuous distribution function $F(y)$. Let $\psi_y(x)$ be 0 or 1 according as $x > y$ or $x \leqq y$. The modified empirical distribution function was defined by M. Kac [6] as

$$F_\lambda^*(y) = \lambda^{-1} \sum_{j=1}^{N} \psi_y(X_j), \qquad -\infty < y < \infty \tag{1.1}$$

where the sum is taken to be zero if $N = 0$. Notice that it is possible for $F_\lambda^*(y)$ to exceed one. The statistic analogous to the one-sided Kolmogorov statistic [7] is $\text{l.u.b.}_{-\infty<y<\infty} [F(y) - F_\lambda^*(y)]$. Since $F_\lambda^*(y)$ was first studied by Kac, we shall refer to the statistic as the one-sided Kac statistic. Using Kolmogorov's result, Kac notes that as long as $F(y)$ is continuous, the distribution of the statistic is independent of $F(y)$. Hence we will confine our attention to the simple case $F(x) = x, 0 \leqq x \leqq 1$.

A random sample will determine an upper confidence contour:

$$F_{\lambda,\epsilon}^*(y) = \min [F_\lambda^*(y) + \epsilon, 1]. \tag{1.2}$$

2. The distribution of the one-sided Kac statistic. We will now derive an explicit form for

$$\begin{aligned} P_\lambda(\epsilon) &= P\{F(y) \leqq F_{\lambda,\epsilon}^*(y), -\infty < y < \infty\} \\ &= P\{\text{l.u.b.}_{-\infty<y<\infty} [F(y) - F_\lambda^*(y)] \leqq \epsilon\}. \end{aligned} \tag{2.1}$$

The analysis relies on a result for the one-sided Kolmogorov statistic obtained by Smirnov, [9], and Birnbaum and Tingey, [2]. If n is a positive integer, and $Y_1 \leqq$

Received 16 August 1965; revised 24 January 1966.

[1] This research was partially supported by the National Science Foundation through Grant NSF GP-4570.

$Y_2 \leqq \cdots \leqq Y_n$ are the order statistics corresponding to $X_1, X_2, \cdots, X_n$, let

$$\begin{aligned} F_{n,\lambda}(y) &= 0, \qquad y < Y_1 \\ &= k/\lambda, \qquad Y_k \leqq y < Y_{k+1}, \; k = 1, 2, \cdots, n-1 \qquad (2.2) \\ &= n/\lambda, \qquad y \geqq Y_n. \end{aligned}$$

Thus $F_{n,\lambda}(y) = (n/\lambda)F_n(y)$, where $F_n(y)$ is the ordinary empirical distribution function. We will also let

$$F_{n,\lambda,\epsilon}(y) = \min [F_{n,\lambda}(y) + \epsilon, 1]. \qquad 0 < \epsilon \leqq 1. \tag{2.3}$$

THEOREM 1. *For $N, X_1, X_2, \cdots$ subject to the previous conditions, and* $0 < \epsilon \leqq 1$,

$$P_\lambda(\epsilon) = 1 - \epsilon\lambda \sum_{j=0}^{[\lambda(1-\epsilon)]} [(\lambda\epsilon + j)^{j-1}/j!]e^{-\lambda\epsilon-j}. \tag{2.4}$$

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

$$\begin{aligned} P_\lambda(\epsilon) &= \sum_{n=0}^{\infty} P\{N = n\} \\ &\quad \cdot P\{y \leqq \min (1, \lambda^{-1} \sum_{i=1}^{n} \psi_y(X_i) + \epsilon), 0 \leqq y \leqq 1\} \qquad (2.5) \\ &= \sum_{n=0}^{\infty} [\lambda^n e^{-\lambda}/n!] P\{y \leqq F_{n,\lambda,\epsilon}(y), 0 \leqq y \leqq 1\}. \end{aligned}$$

Now $P\{y \leqq F_{n,\lambda,\epsilon}(y), 0 \leqq y \leqq 1\}$ is the probability that the ordered sample $Y_1, Y_2, \cdots, Y_n$ is in the region $Y_j \leqq \epsilon + (j-1)/\lambda, j = 1, 2, \cdots, k+1$, $Y_j \leqq 1, j = k+2, \cdots, n$ where $k = \max \{[\lambda(1-\epsilon)], n-1\}$. This probability may be calculated by slightly changing the analysis in [2] to handle the extra parameter λ, and one obtains

$$\begin{aligned} (2.6) \quad & P\{y \leqq F_{n,\lambda,\epsilon}(y), 0 \leqq y \leqq 1\} \\ &= \{1 - \epsilon \sum_{j=0}^{[\lambda(1-\epsilon)]} \tbinom{n}{j} (1 - \epsilon - (j/\lambda))^{n-j} (\epsilon + (j/\lambda))^{j-1}\}. \end{aligned}$$

After substituting the resulting expression in (2.5), and interchanging order of summation and summing on n, we obtain (2.4).[2]

3. The asymptotic distribution. The asymptotic distribution of the one-sided Kolmogorov statistic has been found by Smirnov [8] to be

$$\lim_{n\to\infty} P\{\text{l.u.b.}_{-\infty<y<\infty} (F(y) - F_n(y)) \leqq \alpha/n^{\frac{1}{2}}\} = 1 - e^{-2\alpha^2}.$$

The analogous theorem for the one-sided Kac statistic is:

THEOREM 2. *For $N, X_1, X_2, \cdots$ subject to the previous conditions,*

[2] The authors gratefully acknowledge the help of the referee in simplifying earlier forms of (2.4) and (4.1). An earlier and completely different derivation of (2.4) appears as Theorem 5 in Takács, L. (1965), Applications of a ballot theorem in physics and in order statistics, *J. Roy. Statist. Soc. Ser. B* **27** 130–137. The present method is also used in Section 4.

$$\lim_{\lambda\to\infty} P\{\text{l.u.b.}_{-\infty<y<\infty} (F(y) - F_\lambda^*(y)) \leqq \alpha/\lambda^{\frac{1}{2}}\}$$

$$(3.1) \qquad = [2/(2\pi)^{\frac{1}{2}}] \int_0^\alpha e^{-u^2/2}\, du, \quad \alpha \geqq 0$$

$$= 0, \qquad \alpha < 0.$$

Proof. One can slightly modify the analysis of Kac [6] for the asymptotic distribution of the two-sided statistic. Since the probability on the left side of (3.1) is independent of $F(y)$, it will be sufficient to let $F(y) = y$, $0 \leqq y \leqq 1$. Consider the process

$$(3.2) \qquad x_\lambda(y) = \lambda^{\frac{1}{2}}\{y - \lambda^{-1} \textstyle\sum_{j=1}^N \psi_y(X_j)\}, \qquad 0 \leqq y \leqq 1,$$

with independent increments. Without loss of generality, we will consider the separable version of the process. Using the distribution-free property, (3.2), separability, and the monotonicity of the sets, we obtain

$$\lim_{\lambda\to\infty} P\{\text{l.u.b.}_{-\infty<y<\infty} (F(y) - F_\lambda^*(y)) \leqq \alpha/\lambda^{\frac{1}{2}}\}$$

$$(3.3) \qquad = \lim_{\lambda\to\infty} P\{\lim_{r\to\infty} \text{l.u.b.}_{1\leqq k\leqq 2^r}\, x_\lambda(k/2^r) \leqq \alpha\}$$

$$= \lim_{\lambda\to\infty} \lim_{r\to\infty} P\{\text{l.u.b.}_{1\leqq k\leqq 2^r}\, x_\lambda(k/2^r) \leqq \alpha\}.$$

Using the fact that the $x_\lambda(y)$ process has independent increments, by similar reasoning to that of Erdös and Kac, [4], one obtains

$$(3.4) \quad \lim_{\lambda\to\infty} \lim_{r\to\infty} P\{\text{l.u.b.}_{1\leqq k\leqq 2^r}\, x_\lambda(k/2^r) \leqq \alpha\}$$

$$= P\{\text{l.u.b.}_{0\leqq u\leqq 1}\, x(u) \leqq \alpha\}$$

where $\{x(u), 0 \leqq u \leqq 1\}$ is the Wiener process. The conclusion follows from a well-known result about that process. (See [3], p. 392.)

Remark. Thus for large λ, we can evaluate $P_\lambda(\epsilon) \fallingdotseq 2F(\alpha) - 1$ where $\epsilon = \alpha/\lambda^{\frac{1}{2}}$, and $F(\alpha)$ is the area to the left of α $(\alpha \geqq 0)$ under a standardized normal curve. We shall see in Table 2 that even for $\lambda = 25$, the true and asymptotic values are close together. In many practical situations, λ would exceed 50 for an appropriate observation period.

4. The power and consistency of the one-sided test. Let $F(x)$ be a continuous cumulative distribution function of a random variable X, and $F_\lambda^*(x)$ be the modified empirical distribution function. We shall test the hypothesis that $F(x) = \theta(x)$ against the alternative $F(x) = \phi(x)$, and obtain a lower bound for the power for the alternative such that $\sup_{-\infty<x<\infty} (\theta(x) - \phi(x)) = \delta$ where δ is preassigned and positive.

We consider the class C of continuous cumulative distribution functions $F(x)$. Assume that $\phi(x) \varepsilon C$ and that $\theta(x) \varepsilon C$; then a test of size α is made by choosing, for a given λ, the ϵ corresponding to $1 - \alpha$ in Table 1, obtaining an ordered sample $Y_1, Y_2, \cdots, Y_N$, and rejecting $\theta(x)$ if and only if $\theta(x) \leqq F_\lambda^*(x) + \epsilon$ fails to hold for all x. The power of this test is the complementary probability to

$$P\{\theta(x) \leqq F_\lambda^*(x) + \epsilon \text{ for all } x;\ \phi(x)\} = \sum_{n=0}^{\infty} P\{\theta(x) \leqq F_{n,\lambda}(x) + \epsilon \text{ for all } x;\ \phi(x)\} P\{N = n\}.$$

Modifying the analysis of Birnbaum [1], it is easy to show that

$$\text{Power} \geqq \sum_{n=0}^{\infty} \sum_{i=0}^{k} (\lambda^n e^{-\lambda}/n!) \binom{n}{i} u_0^i (1 - u_0)^{n-i} = \sum_{i=0}^{k} [e^{-\lambda u_0} (\lambda u_0)^i]/i! \tag{4.1}$$

where $u_0 = \theta(x_0) - \delta$, x_0 is determined by $\theta(x_0) - \phi(x_0) = \delta$, and $k = [\lambda(\theta(x_0) - \epsilon)]$.

The test based on the one-sided Kolmogorov statistic is consistent. (See Birnbaum [1], and Wilks [10], p. 440). If we use the phrase "modified consistent" to indicate the limiting value of the power as $\lambda \to \infty$, the test based on the one-sided Kac statistic is modified consistent.

THEOREM 3. *For $N, X_1, X_2, \cdots, \phi(x), \theta(x), \delta, \alpha$, and ϵ_λ subject to the previous conditions,* $\lim_{\lambda\to\infty} P\{\theta(x) > F_\lambda^*(x) + \epsilon_\lambda$ *for some* $x;\ \phi(x)\} = 1$.

PROOF. As in [1], and [10], we shall assume $u_0 = (\theta(x_0) - \delta) > 0$. Let $\beta > 0$. For fixed α, with $P_\lambda(\epsilon_\lambda) = 1 - \alpha$, $\lim_{\lambda\to\infty} \epsilon_\lambda = 0$. Hence it is easy to find Λ_1 and a positive constant c independent of λ such that $\lambda > \Lambda_1$ implies that

$$[\lambda(\theta(x_0) - \epsilon_\lambda)]/\lambda u_0 \geqq 1 + c. \tag{4.2}$$

By (4.1) and (4.2), for $\lambda > \Lambda_1$,

$$P\{\theta(x) > F_\lambda^*(x) + \epsilon_\lambda \text{ for some } x;\ \phi(x)\} \geqq P\{X \leqq (1 + c)\lambda u_0\} \tag{4.3}$$

where X has a Poisson distribution with mean λu_0. There exists Λ_2 such that $I = \int_{-\infty}^{c(\Lambda_2 u_0)^{1/2}} (2\pi)^{-1/2} e^{-x^2/2}\,dx \geqq 1 - \beta/2$. By the limiting distribution for X, (see, for example, [5], p. 230), there exists Λ_3 such that $\lambda > \max(\Lambda_2, \Lambda_3)$ implies that

$$P\{X \leqq (1 + c)\lambda u_0\} \geqq P\{(X - \lambda u_0)/(\lambda u_0)^{1/2} \leqq c(\Lambda_2 u_0)^{1/2}\} \geqq I - \beta/2. \tag{4.4}$$

Hence for $\lambda > \max(\Lambda_1, \Lambda_2, \Lambda_3)$, $P\{\theta(x) > F_\lambda^*(x) + \epsilon_\lambda$ for some $x;\ \phi(x)\} \geqq 1 - \beta$. Since β was arbitrary, the result follows.

5. Distribution tables. Tables 1 and 2, except for the asymptotic values of Table 2, were computed on an IBM 1620 digital computer using Formula (2.4). Twenty-five places were kept in the calculations. The results were then rounded to five places, giving an error estimate $\leqq 5 \cdot 10^{-5}$.

Table 1 also serves for lower confidence contours since

$$P\{\text{g.l.b.}_{-\infty<y<\infty} (F(y) - F_\lambda^*(y)) \geqq -\epsilon\} = P_\lambda(\epsilon), \quad 0 < \epsilon \leq 1.$$

Table 2 indicates the convergence of the true distribution to the asymptotic distribution. The oscillatory nature of the convergence is caused by the jumps occasioned by the upper limit of summation, $[\lambda(1 - \epsilon/\lambda^{1/2})]$.

These tables can be used to determine an upper or lower confidence contour for an unknown $F(y)$, or in testing the hypothesis that the sample is from a specified distribution.

Two-sided statistic. The authors are currently working on the two-sided Kac statistic: $\text{l.u.b.}_{-\infty<y<\infty} |F(y) - F_\lambda^*(y)|$. The corresponding asymptotic distribution was derived by Kac in [6].

REFERENCES

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

[2] BIRNBAUM, Z. W. and TINGEY, F. H. (1951). One-sided confidence contours for probability distribution functions. *Ann. Math. Statist.* **22** 592–596.

[3] DOOB, J. L. (1953). *Stochastic Processes*. Wiley, New York.

[4] ERDÖS, P. and KAC, M. (1946). On certain limit theorems of the theory of probability. *Bull. Amer. Math. Soc.* **52** 292–302.

[5] FELLER, W. (1957). *An Introduction to Probability Theory and its Applications*, 2nd ed. Wiley, New York.

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

[7] KOLMOGOROV, A. (1933). Sulla determinazione empirica di una legge di distribuzione. *Giorn. Ist. Ital. Attuari.* **4** 1–11.

[8] SMIRNOV, N. (1939). Sur les ecarts de la courbe de distribution empirique. *Mat. Sb.* **6** 3–26.

[9] SMIRNOV, N. (1944). Approximation of a distribution function from a sample. *Uspehi Mat. Nauk.* **10** 179–206.

[10] WILKS, S. S. (1962). *Mathematical Statistics*. Wiley, New York.

TABLE 1

$P_\lambda(\epsilon) = \beta$

ϵ	λ						
	5	10	15	20	25	30	35
.001	.00187	.00258	.00314	.00361	.00402	.00440	.00475
.01	.01859	.02571	.03125	.03595	.04010	.04386	.04732
.025	.04621	.06388	.07763	.08926	.09953	.10881	.11735
.05	.09146	.12629	.15325	.17597	.19982	.21739	.23353
.075	.13564	.18692	.23386	.26555	.29345	.32366	.34601
.1	.17868	.24551	.30583	.34623	.38859	.41901	.45241
.125	.22051	.31669	.37402	.43159	.47977	.51355	.55066
.15	.26106	.37256	.45177	.50270	.55574	.60056	.63914
.175	.30029	.42553	.51263	.57973	.63390	.67883	.71677
.2	.33817	.47549	.56876	.63892	.69423	.73903	.77598
.225	.41218	.54587	.63697	.70454	.75677	.79816	.83151
.25	.44931	.59032	.68364	.75090	.81109	.84775	.87658
.275	.48465	.63142	.74303	.80348	.84804	.88820	.91209
.3	.51823	.66928	.77969	.83753	.88681	.91450	.93921
.325	.55009	.73045	.81194	.87702	.91814	.94008	.95927
.35	.58025	.76137	.85553	.90035	.93672	.95931	.97360
.375	.60876	.78927	.87877	.92802	.95644	.97330	.98348
.4	.63568	.81436	.89869	.94287	.96714	.98085	.98873
.425	.71636	.86046	.92720	.96086	.97857	.98812	.99335
.45	.73916	.87889	.94024	.96958	.98659	.99292	.99623
.475	.76032	.89512	.95976	.98038	.99030	.99596	.99795
.5	.77993	.90936	.96758	.98508	.99431	.99729	.99894
.55	.81489	.94805	.98407	.99331	.99780	.99927	.99975
.6	.84473	.96265	.99015	.99729	.99924	.99978	.99994
.65	.91487	.98196	.99588	.99902	.99984	.99996	.99999
.7	.93092	.98756	.99851	.99969	.99996	.99999	1.00000
.75	.94404	.99525	.99956	.99992	.99999	1.00000	1.00000
.8	.95473	.99686	.99976	.99998	1.00000	1.00000	1.00000
.85	.98574	.99916	.99995	1.00000	1.00000	1.00000	1.00000
.9	.98889	.99947	.99999	1.00000	1.00000	1.00000	1.00000
.95	.99135	.99993	1.00000	1.00000	1.00000	1.00000	1.00000
.99	.99292	.99995	1.00000	1.00000	1.00000	1.00000	1.00000

TABLE 2

$P_\lambda(\epsilon/\lambda^{\frac{1}{2}}) = \beta$

ϵ	λ				ϵ	λ			
	25	30	35	$\lim_{\lambda\to\infty}$		25	30	35	$\lim_{\lambda\to\infty}$
.025	.02009	.02006	.02004	.0200	.825	.60751	.60242	.59900	.5906
.05	.04010	.04005	.04002	.0398	.85	.62087	.61582	.61958	.6046
.075	.06002	.05995	.05991	.0598	.875	.63390	.62891	.63273	.6184
.1	.07983	.07976	.07971	.0796	.9	.64661	.64168	.64556	.6318
.125	.09953	.09945	.09941	.0996	.925	.65899	.66265	.65807	.6450
.15	.11910	.11903	.11899	.1192	.95	.67105	.67478	.67025	.6578
.175	.13854	.13847	.14042	.1390	.975	.68280	.68658	.68212	.6706
.2	.15783	.16041	.16000	.1586	1.00	.69423	.69807	.69368	.6827
.225	.18050	.17987	.17943	.1782	1.1	.74696	.74938	.74398	.7286
.25	.19982	.19915	.19869	.1974	1.2	.78442	.78699	.78906	.7698
.275	.21895	.21825	.21778	.2168	1.3	.82667	.82781	.82212	.8064
.3	.23789	.23717	.23669	.2358	1.4	.85466	.85595	.85710	.8384
.325	.25662	.25589	.25540	.2548	1.5	.88681	.88686	.88136	.8664
.35	.27515	.27441	.27776	.2736	1.6	.90663	.90685	.90723	.8904
.375	.29345	.29749	.29628	.2924	1.7	.92979	.92908	.92878	.9108
.4	.31153	.31582	.31458	.3108	1.8	.94304	.94256	.94241	.9282
.425	.33576	.33391	.33265	.3292	1.9	.95881	.95772	.95711	.9426
.45	.35363	.35175	.35047	.3472	2.0	.96714	.96632	.96588	.9544
.475	.37125	.36934	.36805	.3652	2.1	.97724	.97610	.97541	.9642
.5	.38859	.38667	.38538	.3830	2.2	.98215	.98350	.98269	.9722
.525	.40567	.40374	.40784	.4004	2.3	.98821	.98723	.98660	.9786
.55	.42247	.42710	.42480	.4176	2.4	.99092	.99155	.99091	.9836
.575	.43898	.44380	.44148	.4348	2.5	.99431	.99357	.99308	.9876
.6	.45522	.46020	.45788	.4514	2.6	.99569	.99594	.99549	.9906
.625	.47977	.47631	.47399	.4682	2.7	.99745	.99696	.99662	.9930
.65	.49558	.49213	.48980	.4844	2.8	.99811	.99818	.99789	.9948
.675	.51109	.50764	.50532	.5004	2.9	.99895	.99866	.99872	.9962
.7	.52628	.52285	.52708	.5160	3.0	.99924	.99924	.99907	.9974
.725	.54117	.53775	.54209	.5316	3.1	.99961	.99945	.99946	.9980
.75	.55574	.56028	.55679	.5468	3.2	.99972	.99971	.99962	.9986
.775	.57000	.57465	.57117	.5618	3.3	.99987	.99985	.99979	.9990
.8	.58394	.58869	.58524	.5762	3.4	.99991	.99990	.99989	.9994