For statistics having normal sampling distributions such a ratio would be independent of α and would be equivalent to the ratio of the variances of these sampling distributions. It was found that δ_α^2 is independent of α except for a maximum change of 1 in the second decimal for the values of $\alpha = .005, .01, .025, .05, .10$. These values of δ^2 are presented in Table 3 along with the relative precision of the range as an estimate of σ as given by Mosteller [1].

It is interesting to note that δ^2 corresponds very closely to the relative precision of the range as an estimate of σ.

REFERENCE

[1] F. MOSTELLER, "On some useful 'inefficient' statistics," *Annals of Math. Stat.*, Vol. 17 (1946), pp. 377–408.

A NOTE ON THE ESTIMATION OF A DISTRIBUTION FUNCTION BY CONFIDENCE LIMITS

BY FRANK J. MASSEY, JR.

University of Oregon

Let $F(x)$ be the continuous cumulative distribution function of a random variable X, and let $x_1 < x_2 < x_3 < \cdots < x_n$ be the results of n independent observations on X arranged in order of size. We wish to estimate $F(x)$ by means of the band $S_n(x) \pm \lambda/\sqrt{n}$ where $S_n(x)$ is defined by

$$S_n(x) = \begin{cases} 0 & \text{if } x < x_1, \\ k/n & \text{if } x_k \leq x < x_{k+1}, \\ 1 & \text{if } x \geq x_n. \end{cases}$$

Thus we wish to know the probability, say $P_n(\lambda)$, that the band is such that $S_n(x) - \dfrac{\lambda}{\sqrt{n}} < F(x) < S_n(x) + \dfrac{\lambda}{\sqrt{n}}$ for all x. This problem has been previously studied [1] [2] [3] [4] [5] and a limiting distribution has been obtained [1] [4] [5] and tabled [3] [4]. However apparently no error terms for the limiting distribution, or practical methods of obtaining $P_n(\lambda)$ have been given. Such a method is given here.

It has been shown [2] that $P_n(\lambda)$ is independent of $F(x)$ provided only that $F(x)$ is continuous, and thus it is sufficient to consider only the case

$$F(x) = \begin{cases} 0 \text{ if } x < 0, \\ x \text{ if } 0 \leq x \leq 1, \\ 1 \text{ if } x \geq 1. \end{cases}$$

We will find the probability that $S_n(x)$ falls wholly in the band $F(x) \pm k/n$ (here $\lambda = k/\sqrt{n}$) where k is an integer or a rational number, and intermediate values may be obtained by interpolation. To illustrate the method we shall assume that k is an integer.

Divide the interval (0, 1) into n parts by the points $1/n, 2/n, \cdots, (n-1)/n$. The step function $S_n(x)$ rises by jumps of exactly $1/n$. Thus, in order to be inside the band at $x = i/n$, $S_n(x)$ would have to pass through exactly one of the lattice points whose ordinates are $(i-k+1)/n, (i-k+2)/n, \cdots, (i+k-1)/n$.

Suppose that the step function stays inside the band by means of α_i of the observations falling in the interval $\left(\frac{i-1}{n}, \frac{i}{n}\right)$ $i = 1, 2, \cdots, n$. The a priori probability of this happening is given by the multinomial law as

$$P_r(\alpha_1 \cdots \alpha_n) = \frac{n!}{\alpha_1!\, \alpha_2! \cdots \alpha_n!} \left(\frac{1}{n}\right)^{\alpha_1} \left(\frac{1}{n}\right)^{\alpha_2} \cdots \left(\frac{1}{n}\right)^{\alpha_n}$$

$$= \frac{1}{\alpha_1! \cdots \alpha_n!} \frac{n!}{n^n}$$

since $\sum_1^n \alpha_i = n$.

Thus the probability of the step function staying in the band is given by

$$P_n(\lambda) = \sum \frac{n!}{n^n} \frac{1}{\alpha_1!\, \alpha_2! \cdots \alpha_n!} = \frac{n!}{n^n} \sum \frac{1}{\alpha_1! \cdots \alpha_n!}$$

where the summation is over all possible combinations of $\alpha_1, \cdots, \alpha_n$ such that $\max_x | S_n(x) - x | < \frac{\lambda}{\sqrt{n}}$ and $\sum_{i=1}^{n} \alpha_i = n$.

Let $U_i(m) = \sum_i \frac{1}{\alpha_1! \cdots \alpha_m!}$, $i = 1, 2, \cdots, 2k-1$ be the sum of all the terms indicated such that $S_n(x)$ arrives at the lattice point $\left(\frac{m}{n}, \frac{m-k+i}{n}\right)$ by a route that stays inside the band. Since the $S_n(x)$ is non-decreasing it can only pass through a point

$$\left(\frac{m+1}{n}, \frac{m-k+1+j}{n}\right), \quad m = 0, 1, \cdots, n-1; \qquad j = 1, 2, \cdots, 2k-1,$$

if it previously passed through one of the points

$$\left(\frac{m}{n}, \frac{m-k+1}{n}\right) \cdots, \left(\frac{m}{n}, \frac{m-k+2}{n}\right) \cdots, \left(\frac{m}{n}, \frac{m-k+j+1}{n}\right).$$

If it passed through $\left(\frac{m}{n}, \frac{m-k+h}{n}\right)$ the value of α_{m+1} would have to be $(j+1-h)$ and the product $U_h(m) \frac{1}{(j+1-h)!}$ would be part of $U_j(m+1)$. This is true for all $h = 1, 2, \cdots, j+1$ and all of these terms would give different paths for $S_n(x)$ so we have

$$U_j(m+1) = \sum_{h=1}^{j+1} \frac{1}{(j+1-h)!} U_h(m), \qquad j = 1, 2, \cdots, 2k-1,$$

where it is understood $U_h(m) = 0$ if $h \geq m+k$.

Thus we have a set of $2k - 1$ linear homogeneous difference equations. They may be reduced to a single difference equation by eliminating $2k - 2$ of the variables by substitution. This results in the following difference equation.

$$\sum_{h=1}^{2k-1} (-1)^h \frac{(2k-h)^h}{h!} U_k(2k - 1 - h + m) = 0.$$

TABLE 1

k	$n = 5$	10	20	25	30	35	40	45
1.0	.0384	.0004						
1.5	.3276	.0449						
2.0	.6521	.2513	.0238					
2.5	.8880	.5139						
3.0	.9699	.7331	.2955					
3.5	.9947	.8522						
4.0	.99935	.9410	.6473					
5.0		.9922	.8624	.7637	.6629	.5674	.4808	.4042
6.0		.9994	.9569	.9057	.8420	.7725	.7016	.6322
7.0			.9892	.9683	.9359	.8945	.8471	.7962
8.0			.9979	.9911	.9774	.9566	.9295	.8974
9.0			.9997	.9979	.9931	.9842	.9708	.9529

k	$n = 50$	55	60	65	70	75	80
5.0	.3377	.2807	.2324	.1918	.1577	.1294	.1060
6.0	.5662	.5046	.4478	.3954	.3492	.3072	.2696
7.0	.7439	.6916	.6403	.5908	.5435	.4987	.4566
8.0	.8616	.8234	.7837	.7434	.7031	.6633	.6244
9.0	.9312	.9063	.8789	.8496	.8189	.7874	.7554

Initial conditions on either the simultaneous equations or on the single equation are

$$U_i(0) = 0 \text{ for } i \neq k,$$

$$U_k(0) = 1 \text{ for } i = k.$$

After values of $U_k(n)$ have been found the value of $P_n\left(\frac{k}{\sqrt{n}}\right)$ can be found by multiplying $U_k(n)$ by $\frac{n!}{n^n}$.

The values of $U_k(n)$ can be obtained numerically either from the simultaneous equations or from the single equation. Table 1 was computed partly by numerical solution of the simultaneous equations above and partly by setting up similar equations connecting $U_i(x + 5)$ to $U_t(x)$, $t = 1, 2, \cdots, i + 5$. Either method could be set up on punch cards if an extensive table was desired. Notice that to get $U_k(n)$ all $U_k(t)$, $t = 1, 2, \cdots, n - 1$ are also found. Table 1 gives some computed values of $P_n(k)$. Table 2 gives results interpolated from Table 1, showing the approach of $P_n(\lambda)$ to its limiting distribution.

If the width of the band is $2\left(\frac{k}{l}\right)$ when k and l are integers a similar procedure to that above can be used. However instead of dividing the interval (0, 1) into n parts it is necessary to divide it into $l \cdot n$ parts.

TABLE 2

n	$\lambda = .9$	1.0	1.10	1.20	1.30	1.40
10	.66	.78	.85	.91	.95	.97
20	.65	.77	.85	.91	.94	.97
30	.65	.76	.85	.90	.94	.96
40	.64	.76	.84	.90	.94	.96
50	.64	.75	.84			
60	.63	.75	.84			
70	.63	.75	.83			
80	.63	.74				
∞	.607	.730	.822	.888	.932	.960

It has been suggested (2) that instead of a band bounded by $y = x \pm c$ it might be convenient to use a band bounded by the lines $y = px + q$ and $y = p'x + q'$. If $p = p'$ and if p, q, q' are rational the probability of $S_n(x)$ staying inside the band can be evaluated by the method presented above. If $p \neq p'$ and if p, p', q, q' are all rational a similar procedure could be used but it would be very tedious.

[1] N. Smirnov, "Sur les écarts de la courbe de distribution empirique," *Recueil Math. de Moscou*, Vol. 6 (1939), pp. 3–26.

[2] A. Wald and J. Wolfowitz, "Confidence limits for continuous distribution functions," *Annals of Math. Stat.*, Vol. 10 (1939), pp. 105–118.

[3] N. Smirnov, "On the estimation of the discrepancy between empirical curves of distribution for two independent samples," *Bulletin Mathematique de l'Universite de Moscou*, Vol. 2 (1939), fasc. 2.

[4] A. Kolmogorov, "Sulla determinazione empirica di una legge di distribuzione," *Ist. Ital. att.* Giorn. Vol. 4 (1933), p. 1–11.

[5] W. Feller, "On the Kolmogorov-Smirnov limit theorems for empirical distribution" *Annals of Math. Stat.*, Vol. 19 (1948), pp. 177–190.