BULL. AUSTRAL. MATH. SOC. 60E05, 62G99
VOL. 7 (1972), 227-232.

Simple proofs of Steck's determinantal expressions for probabilities in the Kolmogorov and Smirnov tests

E.J.G. Pitman

This paper gives simple proofs of two theorems of Steck concerning the distribution of sample distribution functions.

Theorems I and II below were stated and proved by Steck in two notable papers [2], [3]. As Steck showed, Theorem I enables us to determine:

(i) the probability that the empirical distribution function lies between two other distribution functions;

(ii) very general confidence regions for an unknown distribution function;

(iii) the power of a test based on the empirical distribution function.

From Theorem II we can obtain the null distribution of the two-sample Smirnov statistics for arbitrary sample sizes. Steck's proofs were indirect, and somewhat complicated. Mohanty [1] gave a shorter proof of Theorem II. Here I give simple, direct proofs of both theorems.

LEMMA. *Let*

$$A_1, A_2, \ldots, A_{m-1},$$

$$B_1, B_2, \ldots, B_m$$

denote events such that for any integer k, *the sets*

Received 28 April 1972.

$$\{A_r, B_s \; ; \; r < k, \; s \le k\} \; , \quad \{A_r, B_s \; ; \; r > k, \; s > k\}$$

are independent, then

$$P\left(B_1 B_2 \ldots B_m A_1 A_2 \ldots A_{m-1}\right) = \Delta_m = \det\left(d_{ij}\right) \; , \quad 1 \le i, \; j \le m \; ,$$

where

$$\begin{aligned} d_{ij} &= 0 && \text{if } i > j+1 \, , \\ &= 1 && \text{if } i = j+1 \, , \\ &= P\left(B_i\right) && \text{if } i = j \, , \\ &= P\left(B_i B_{i+1} \ldots B_j A_i^* A_{i+1}^* \ldots A_{j-1}^*\right) && \text{if } i < j \, , \end{aligned}$$

and A_r^* *is the complement of* A_r .

Note that the conditions on the events make $B_1, B_2, \ldots, B_m$ independent. The events $A_1, A_2, \ldots$ are 1-dependent. Put

$$\overline{A}_r = A_1 A_2 \ldots A_r \; , \quad \overline{B}_r = B_1 B_2 \ldots B_r \; .$$

The lemma may be proved by use of the principle of inclusion and exclusion.

$$P\left(\overline{B}_m A_1 A_2 \ldots A_{m-1}\right) = P\left(\overline{B}_m\right) - \sum P\left(\overline{B}_m A_r^*\right) + \sum_{r<s} P\left(\overline{B}_m A_r^* A_s^*\right)$$
$$- \ldots + (-1)^{m-1} P\left(\overline{B}_m A_1^* A_2^* \ldots A_{m-1}^*\right) \; ,$$

which can be shown directly to be the expansion of Δ_m . A proof by induction is shorter, and easier to print.

Assume the lemma true for $m = n$.

$$\Delta_n = P\left(\overline{B}_n \overline{A}_{n-1}\right) \; .$$

Consider Δ_{n+1} . The elements of its last row are all zero except

$$d_{n+1,n} = 1 \; , \quad d_{n+1,n+1} = P\left(B_{n+1}\right) \; .$$

Therefore

$$\Delta_{n+1} = P\left(B_{n+1}\right)\Delta_n - \Delta_n' \; ,$$

where Δ_n' differs from Δ_n only in having B_n in the last column of Δ_n

https://doi.org/10.1017/S0004972700044981 Published online by Cambridge University Press

replaced by $B_nB_{n+1}A_n^*$, which satisfies the same conditions relative to the other events appearing in Δ_n' as does B_n . Therefore

$$\Delta_n' = P\left(\overline{B}_nB_{n+1}A_n^*\overline{A}_{n-1}\right) = P\left(\overline{B}_{n+1}\overline{A}_{n-1}A_n^*\right) ;$$

$$P\left(B_{n+1}\right)\Delta_n = P\left(B_{n+1}\right)P\left(\overline{B}_n\overline{A}_{n-1}\right) = P\left(\overline{B}_{n+1}\overline{A}_{n-1}\right) .$$

Thus

$$\begin{aligned}\Delta_{n+1} &= P\left(\overline{B}_{n+1}\overline{A}_{n-1}\right) - P\left(\overline{B}_{n+1}\overline{A}_{n-1}A_n^*\right) \\ &= P\left(\overline{B}_{n+1}\overline{A}_{n-1}A_n\right) = P\left(\overline{B}_{n+1}\overline{A}_n\right) ,\end{aligned}$$

and so the lemma is true for $m = n + 1$. It is easy to show that it is true for $m = 2$, and so it is true for all m .

COROLLARY. *Taking the case where every* $P\left(B_i\right) = 1$ *, we obtain the following result for the* 1*-dependent sequence of events* $A_1, A_2, \ldots$ *,*

$$P\left(A_1A_2 \ldots A_{m-1}\right) = \det\left(d_{ij}\right) ,$$

where

$$\begin{aligned}d_{ij} &= 0 && \text{if } i > j + 1 , \\ &= 1 && \text{if } i = j \text{ or } j + 1 , \\ &= P\left(A_i^*A_{i+1}^* \ldots A_{j-1}^*\right) && \text{if } i < j .\end{aligned}$$

THEOREM I. *Let*

$$0 \le u_1 \le u_2 \le \ldots \le u_m \le 1 ,$$

$$0 \le v_1 \le v_2 \le \ldots \le v_m \le 1 ,$$

be given constants such that

$$u_i < v_i , \quad i = 1, 2, \ldots, m .$$

If $U_1, U_2, \ldots, U_m$ *are the order statistics (in ascending order) from a sample of* m *independent uniform random variables with range* 0 *to* 1 *,*

$$P\left(u_i \le U_i \le v_i, 1 \le i \le m\right) = m!\det\left[\left(v_i-u_j\right)_+^{j-i+1}/(j-i+1)!\right] ,$$

where $(x)_+ = \max(x, 0)$ *, and it is understood that determinant elements* *for which* $i > j + 1$ *are all zero.*

Proof. Let $Y_1, Y_2, \ldots, Y_m$ be independent random variables, each with a uniform distribution from 0 to 1 . The required probability is equal to

$$(1) \qquad m!P\left(u_i \le Y_i \le v_i,\ 1 \le i \le m;\ Y_1 \le Y_2 \le \ldots \le Y_m\right) .$$

Denote by B_i the event $u_i \le Y_i \le v_i$. Denote by A_i the event $Y_i \le Y_{i+1}$, and by A_i^* the complement of A_i , that is, the event $Y_i > Y_{i+1}$. The events A_i, B_i satisfy the conditions of the lemma. Hence

$$(2) \quad P\left(u_i \le Y_i \le v_i,\ 1 \le i \le m;\ Y_1 \le Y_2 \le \ldots \le Y_m\right)$$
$$= P\left(B_1B_2 \ldots B_mA_1A_2 \ldots A_{m-1}\right) = \det\left(d_{ij}\right) .$$

$$d_{ii} = P\left(B_i\right) = v_i - u_i .$$

If $i < j$, $d_{ij} = P\left(B_iB_{i+1} \ldots B_jA_i^*A_{i+1}^* \ldots A_{j-1}^*\right)$. The event $B_iB_{i+1} \ldots B_jA_i^*A_{i+1}^* \ldots A_{j-1}^*$ is

$$u_r \le Y_r \le v_r , \quad i \le r \le j ,$$

$$Y_i > Y_{i+1} > \ldots > Y_j .$$

This is equivalent to

$$v_i \ge Y_i > Y_{i+1} > \ldots > Y_j \ge u_j ,$$

the probability of which is $\left(v_i-u_j\right)_+^{j-i+1}/(j-i+1)!$. The theorem then follows from (1) and (2).

THEOREM II. *Let* $b_1 \le b_2 \le \ldots \le b_m$ *and* $c_1 \le c_2 \le \ldots \le c_m$ *be sequences of integers such that* $b_i < c_i$. *The number of sets of integers* $\left(R_1, R_2, \ldots, R_m\right)$ *such that*

$$R_1 < R_2 < \ldots < R_m ,$$

$$b_i < R_i < c_i , \quad 1 \le i \le m$$

is the m-th order determinant $\det(d_{ij})$ *, where*

$$d_{ij} = 0 \qquad \text{if } i > j+1 \text{ or if } c_i - b_j \le 1 ,$$
$$= \binom{c_i-b_j+j-i-1}{j-i+1} \qquad \text{otherwise.}$$

Proof. Put $Y_i = R_i - i$, $u_i = b_i - i + 1$, $v_i = c_i - i - 1$. The conditions on the R_i are equivalent to

$$Y_i \text{ an integer, } u_i \le Y_i \le v_i \; ; \; 1 \le i \le m ,$$

$$Y_1 \le Y_2 \le \ldots \le Y_m .$$

As before, denote by A_i the event $Y_i \le Y_{i+1}$, and by A_i^* its complement, the event $Y_i > Y_{i+1}$. Put $N_i = v_i - u_i + 1$.

The required number is equal to

(3) $$N_1N_2 \ldots N_mP(A_1A_2 \ldots A_{m-1})$$

when the Y_i are independent random variables, and Y_i has a uniform distribution over the integers from u_i to v_i . By the corollary to the lemma, this is

$$N_1N_2 \ldots N_m\det(d'_{ij}) ,$$

where

$$d'_{ij} = 0 \qquad \text{if } i > j+1 ,$$
$$= 1 \qquad \text{if } i = j \text{ or } j+1 ,$$
$$= P(A_i^*A_{i+1}^* \ldots A_{j-1}^*) \qquad \text{if } i < j ;$$

$$P(A_i^*A_{i+1}^* \ldots A_{j-1}^*) = P(v_i \ge Y_i > Y_{i+1} > \ldots > Y_j \ge u_j)$$

as before. This is zero if $v_i - u_j < j - i$, that is if $c_i - b_j \le 1$. Otherwise it is equal to

$$\binom{v_i-u_j+1}{j-i+1} \Big/ N_iN_{i+1} \ldots N_j = \frac{d_{ij}}{N_iN_{i+1}\ldots N_j} .$$

https://doi.org/10.1017/S0004972700044981 Published online by Cambridge University Press

The numerator is the number of vectors of integers $(y_i, y_{i+1}, \ldots, y_j)$ satisfying $v_i \geq y_i > y_{i+1} > \ldots > y_j \geq u_j$, and the denominator is the number of vectors satisfying $u_r \leq y_r \leq v_r$, $i \leq r \leq j$.

Put $M_0 = 1$, $M_r = N_1N_2 \ldots N_r$; then in all cases

$$d'_{ij} = d_{ij}M_{i-1}/M_j .$$

The required number (3) is

$$M_m\det(d'_{ij}) = M_m\det(d_{ij}M_{i-1}/M_j) = \det(d_{ij}) ,$$

as may be obtained by taking factors out of rows and out of columns. This proves the theorem.

References

[1] S.G. Mohanty, "A short proof of Steck's result on two-sample Smirnov statistics", *Ann. Math. Statist.* 42 (1971), 413-414.

[2] G.P. Steck, "The Smirnov two sample tests as rank tests", *Ann. Math. Statist.* 40 (1969), 1449-1466.

[3] G.P. Steck, "Rectangle probabilities for uniform order statistics and the probability that the empirical distribution function lies between two distribution functions", *Ann. Math. Statist.* 42 (1971), 1-11.

301 Davey Street,
Hobart,
Tasmania 7000.