REFERENCE

[1] D. A. Darling, "On a class of problems related to the random division of an interval," *Ann. Math. Stat.*, Vol. 24 (1953), pp. 239–253.

GENERALIZED D_n^+ STATISTICS[1]

By A. P. Dempster

Harvard University

1. Introduction. The purpose here is to present simplified derivation methods which can be applied to generalizations of some distributions derived by Birnbaum and Tingey [1] and Birnbaum and Pyke [2]. In the case of [1] the generalization is explicitly written down as equation (5)′. Other authors have noticed this generalization; it appears implicitly in equation (31) of Chapman [3] and is given explicitly by Pyke [4]. However the derivation given in the following section differs from the methods of other authors and gives a probabilistic meaning to each term in the summation formula (5)′. In the case of [2] explicit formulas are given for a special case of our generalization different from that considered by Birnbaum and Pyke.

Consider a sample of n from the uniform distribution on (0, 1). Denote the sample c.d.f. by $F_n(x)$. The relevant part of the curve $y = F_n(x)$ is entirely contained by the closed unit square $0 \leqq x \leqq 1$ and $0 \leqq y \leqq 1$, and within this square the population c.d.f. is represented by the line $y = x$. For $0 \leqq \delta < 1$ and $0 < \epsilon < 1$ the line joining $(0, \delta)$ and $(1 - \epsilon, 1)$ will be referred to as barrier (δ, ϵ). A set of such barriers moving away from $y = x$ may be conceived of, and we are concerned with a set of probabilistic questions about which barriers are crossed and where by the curve $y = F_n(x)$ as it passes from (1, 1) to (0, 0)

2. The basic derivation. Denote by $f_j (0 \leqq j \leqq n - 1)$ the probability that $y = F_n(x)$ crosses the barrier (δ, ϵ) at level $y = (n - j) / n$ not having crossed it at any level $y = (n - i) / n$ for $i < j$. Denote the abscissa of the intersection of the barrier (δ, ϵ) and level $y = (n - j) / n$ by m_j. Then it is easily checked that

$$m_j = \frac{1 - \epsilon}{1 - \delta}\left(1 - \delta - \frac{j}{n}\right). \tag{1}$$

Finally, let us use $b(r, s, p)$ for the binomial probability $\binom{s}{r} p^r (1 - p)^{s-r}$

An expression for f_j may be derived as follows. Given that $y = F_n(x)$ passes

Received May 27, 1958; revised November 24, 1958.

[1] Work done in part at Princeton University while the author was supported by the National Research Council of Canada, and in part at Bell Telephone Laboratories while the author was a Member of Technical Staff.

through $(m_i, 1 - i/n)$ the conditional probabilities of passing through $(m_j, 1 - j/n)$ and $(m_j, 1 - (j-1)/n)$ are respectively

$$b(j - i, n - i, (m_i - m_j) / m_i)$$

and $b(j - i - 1, n - i, (m_i - m_j) / m_i)$, for $j > i$. The unconditional probabilities of arriving at these 2 points can therefore each be computed in 2 ways as

$$(2) \qquad b(j, n, 1 - m_j) = \sum_{i=0}^{j-1} f_i\, b\left(j - i, n - i, \frac{m_i - m_j}{m_i}\right) + f_j$$

and

$$(3) \qquad b(j - 1, n, 1 - m_j) = \sum_{i=0}^{j-1} f_i b\left(j - i - 1, n - i, \frac{m_i - m_j}{m_i}\right).$$

If equation (3) is multiplied by $[(n - j + 1) / (j - i)][(m_i - m_j) / m_j]$, which factor is independent of i from (1), and (3) is then subtracted from (2), then $f_0, \cdots, f_{j-1}$ are eliminated and so,

$$f_j = b(j, n, 1 - m_j) - \frac{n - j + 1}{j - i}\frac{m_i - m_j}{m_j} b(j - 1, n, 1 - m_j)$$

or after reduction

$$(4) \qquad f_j = \epsilon \binom{n}{j} (1 - m_j)^{j-1} m_j^{n-j}$$

or

$$(4') \qquad f_j = \epsilon \binom{n}{j} \left(\epsilon + \frac{1 - \epsilon}{1 - \delta}\frac{j}{n}\right)^{j-1} \left(1 - \epsilon - \frac{1 - \epsilon}{1 - \delta}\frac{j}{n}\right)^{n-j}.$$

If $P(n, \delta, \epsilon) = 1 - Q(n, \delta, \epsilon)$ is defined to be the probability that $y = F_n(x)$ nowhere crosses the barrier (δ, ϵ) then

$$(5) \qquad Q(n, \delta, \epsilon) = \sum_{j=0}^{k_\delta} f_j$$

where k_δ is the largest integer such that $k/n < 1 - \delta$. Thus in full form

$$(5') \qquad Q(n, \delta, \epsilon) = \sum_{j=0}^{k_\delta} \epsilon \binom{n}{j} \left(\epsilon + \frac{1 - \epsilon}{1 - \delta}\frac{j}{n}\right)^{j-1} \left(1 - \epsilon - \frac{1 - \epsilon}{1 - \delta}\frac{j}{n}\right)^{n-j}$$

This formula is a direct generalization of the formula in [1] for $P_n(\epsilon)$ where $P_n(\epsilon) = P(n, \epsilon, \epsilon)$.

Another interesting special case is $\delta = 0$. Here it should be made definite that we are considering only intersections of $y = F_n(x)$ and barriers $(0, \epsilon)$ occurring at points other than $(0, 0)$. A special derivation for $Q(n, 0, \epsilon)$ is given as follows. Considering once more the movement of $y = F_n(x)$ downward from $(1, 1)$ in the general case of barrier (δ, ϵ) we see that f_j is the probability that

$y = F_n(x)$ passes through $(m_j, 1 - j/n)$ times the conditional probability that it did not cross the barrier between (1, 1) and $(m_j, 1 - j/n)$, i.e.

$$f_j = b(j, n, 1 - m_j)P\left(j, 0, \frac{\epsilon}{1 - m_j}\right) \tag{6}$$

from which using (4) we find,

$$P\left(j, 0, \frac{\epsilon}{1 - m_j}\right) = \epsilon \binom{n}{j} (1 - m_j)^{j-1} m_j^{n-j} \left[\binom{n}{j} (1 - m_j)^j m_j^{n-j}\right]^{-1} = \frac{\epsilon}{1 - m_j}$$

whence by reparametrizing we may deduce that

$$P(n, 0, \epsilon) = \epsilon \tag{7}$$

independently of n. Finally from (5)′ and (7) we have

$$Q(n, 0, \epsilon) = 1 - \epsilon = \epsilon \sum_{j=0}^{n-1} \binom{n}{j}\left(\frac{j}{n} + \epsilon\left[1 - \frac{j}{n}\right]\right)^{j-1} \cdot \left(1 - \frac{j}{n} - \epsilon\left[1 - \frac{j}{n}\right]\right)^{n-j} \tag{8}$$

the algebraic identity here having been circuitously derived.

It is worth remarking that if $G(x)$ is any continuous c.d.f. and $G_n(x)$ the sample c.d.f. for a sample of n from $G(x)$ we have by transformation from

$$\begin{aligned} P(n, \delta, \epsilon) &= \Pr\left(F_n(x) \leqq \delta + \frac{1 - \delta}{1 - \epsilon} x \quad \text{for} \quad 0 < x < 1\right) \quad \text{that} \\ P(n, \delta, \epsilon) &= \Pr\left(G_n(x) \leqq \delta + \frac{1 - \delta}{1 - \epsilon} G(x) \quad \text{for} \quad -\infty < x < +\infty\right) \end{aligned} \tag{9}$$

or

$$P(n, \delta, \epsilon) = \Pr([1 - \epsilon]G_n(x) - [1 - \delta]G(x) \leqq \delta[1 - \epsilon] \quad \text{for} \quad -\infty < x < +\infty) \tag{10}$$

for $0 \leqq \delta < 1$ and $0 < \epsilon < 1$. Therefore given any real numbers a, b and c we can express $\Pr(aG_n(x) + bG(x) \leqq c$ for $-\infty < x < \infty)$ as 0 or 1 or $P(n, \delta, \epsilon)$ for correctly chosen δ and ϵ depending only on a, b and c. In particular

$$\Pr(G_n(x) \leqq \alpha G(x) \quad \text{for} \quad -\infty < x < \infty) = \begin{cases} \dfrac{\alpha - 1}{\alpha} & \text{if} \quad \alpha \geqq 1 \\ 0 & \text{if} \quad \alpha \leqq 1 \end{cases} \tag{11}$$

3. The statistics D_n^+, U^* and i^*. Suppose we consider the class of barriers (d, d) moving away from line $y = x$ as d moves from 0 to 1. The d corresponding to the furthest barrier reached defines random variable D_n^+ and the point $(U^*, i^*/n)$ where this furthest barrier is touched defines random variables U^* and i^*. Since

$$U^* = \frac{i^*}{n} - D_n^+ \tag{12}$$

the joint distribution of any 2 of the 3 random variables determines the joint distribution of all 3. In [1] the distribution of D_n^+ is presented: $\Pr(D_n^+ \leqq d) = P_n(d)$. In [2] the joint and marginal distributions of U^* and i^* are derived.

Now it is possible to generalize this situation by defining a class of linear barriers moving away from the line $y = x$ restricted only by the requirement that no 2 members of the class intersect within the unit square. These barriers may be indexed by real variable d with at most one barrier corresponding to a given d, and described as barriers $(\delta(d), \epsilon(d))$ where $\delta(d)$ and $\epsilon(d)$ are monotone non-decreasing. We can allow d to take values in a discrete or continuous set. Random variables D_n^+, U^* and i^* may be defined as in the previous paragraph where relation (12) becomes

$$U^* = \left[\frac{i^*}{n} - \delta(D_n^+)\right] \frac{1 - \epsilon(D_n^+)}{1 - \delta(D_n^+)} . \tag{13}$$

Clearly now $\Pr(D_n^+ < d) = P(n, \delta(d), \epsilon(d))$ and this gives the distribution of the generalized D_n^+. A method of writing down the joint distribution of U^* and i^* which applies in general will be demonstrated in the following section applied to a special case where the formulas become fairly simple.

4. The class of barriers $(0, d)$. The class of barriers $(0, d)$ formed by rotating a line through the origin may be of some statistical interest. It has been seen in this case that

$$\Pr(D_n^+ \leqq d) = d. \tag{14}$$

It is proposed now to investigate the joint distribution of U^* and i^*. Define $H(i, u) = \Pr(i^* = i \text{ and } U^* \leqq u)$ and $h(i, u) = dH(i, u) / du$, so that $h(i, u)$ is the density of the joint distribution along the line $y = i/n$. It is evident that $h(i, u) = P_1P_2P_3$, where P_1 is the density function at u of the ith order statistic, P_2 is the conditional probability given that $y = F_n(x)$ passes through $(u, i/n)$ that it does not touch the barrier through $(u, i/n)$ above level i/n, and P_3 is as P_2 but replacing above by below. Thus

$$P_1 = \frac{n!}{(i-1)!(n-i)!} u^{i-1}(1-u)^{n-i}$$

$$P_2 = P\left(n - i, 0, \frac{1 - (n/i)u}{1 - u}\right) = \frac{1 - (n/i)u}{1 - u}$$

and

$$P_3 = P(i - 1, 0, 1/i) = 1/i$$

so that

$$h(i, u) = \binom{n}{i} u^{i-1}(1 - u)^{n-i-1}(1 - (n/i)u) \tag{15}$$

for $0 \leqq u \leqq i/n$ which integrates to give

$$H(i, u) = \frac{1}{i}\binom{n}{i} u^{i}(1 - u)^{n-i}. \tag{16}$$

The marginal distributions are given by

$$p_i = \Pr(i^* = i) = H\left(i, \frac{i}{n}\right) = \frac{1}{i}\binom{n}{i}\left(\frac{i}{n}\right)^{i}\left(1 - \frac{i}{n}\right)^{n-i} \tag{17}$$

and

$$\begin{aligned} K(u) = \Pr(U^* \leqq u) &= \sum_{i=1}^{n} H\left(i, \min\left[\frac{i}{n}, u\right]\right) \\ &= \sum_{i=1}^{n} \frac{1}{i}\binom{n}{i}\left(\min\left[\frac{i}{n}, u\right]\right)^{i-1}\left(1 - \min\left[\frac{i}{n}, u\right]\right)^{n-i} \end{aligned} \tag{18}$$

The algebraic identity implied by the relation $\sum_{i=1}^{n} p_i = 1$, like that in (8), has been indirectly derived. Both identities may be algebraically proved using the formula quoted as (5) in [2].

We repeat that the method used here to obtain $H(i, u)$ is applicable to the general class of barriers $(\delta(d), \epsilon(d))$.

REFERENCES

[1] Z. W. Birnbaum and F. H. Tingey, "One-sided confidence contours for probability distribution functions," *Ann. Math. Stat.*, Vol. 22 (1951), pp. 592–596.

[2] Z. W. Birnbaum and R. Pyke, "On some distributions related to the statistic D_n^+," *Ann. Math. Stat.*, Vol. 29 (1958), pp. 179–187.

[3] D. G. Chapman, "A Comparative study of several one-sided goodness-of-fit tests," *Ann. Math. Stat.*, Vol. 29 (1958), pp. 655–674.

[4] R. Pyke, "The supremum and infimum of the Poisson process," *Technical Report No.* 39, Applied Mathematics and Statistics Laboratory, Stanford University, March 1958 (Abstract in *Ann. Math. Stat.*, Vol. 29 (1958), p. 327).

APPLICATIONS OF A CERTAIN REPRESENTATION OF THE WISHART MATRIX

By Robert A. Wijsman

University of Illinois

0. Summary. Apart from pre- and post-multiplication by a fixed matrix and its transpose, the Wishart matrix **A** can be written as the product of a triangular matrix and its transpose, whose elements are independent normal and chi variables. Various applications of this representation are indicated. Examples

Received July 17, 1958.