GENERALISED DIRICHLET-TYPE DISTRIBUTIONS

G.S. LINGAPPAIAH*

SUMMARY

Series of independent life tests based on exponential model are considered. The distribution of a linear combination of spacings (LCOS) as well as the distribution of the sum of LCOS are used to develop new distributions which are generalised Dirichlet-type distributions.

1. INTRODUCTION

Dirichlet distribution in its elementary form as well as in generalised forms is extensively discussed by many authors. For example, Aitcheson [1] uses gamma distribution to develop Dirichlet distribution. Yassaee [14] uses the same to develop inverted form of the Dirichlet distribution. Connor and Mosimann [2] develop the generalised Dirichlet distribution starting from the classical beta distribution of the first kind, and introduce the concept of "neutrality". Lochner [11] uses the same generalised Dirichlet distribution as an application to life tests using the idea of "force of mortality." Lingappaiah [8] introduces the inverted form of the generalised Dirichlet distribution, using beta distribution of the second kind. Darroch and James [3] introduce "F-independence" and suggest that Dirichlet distribution is by far the best example of F-independence. Darroch and Redcliff [5] discuss this F-independence as related to fingerprint analysis. Institute of Mathematical Statistics Tables [6] give the probability integrals of the Dirichlet variables and Yassaee [15]

* Department of Mathematics Sir George Williams Campus Concordia University Montreal, Canada.

gives the probability integral for the inverted Dirichlet distribution along with computer programmes and applications. In this paper, first, the idea of independence of spacings in the exponential case is made use of. Firstly, the distribution of a linear combination of spacings (LCOS) is considered. Ratio of two such LCOS in a life test is taken up and then considering a set of p independent life tests, a distribution dealing with the functions of these ratios, is obtained, which turns out to be a generalised Dirichlet distribution. Next, the distribution of the sum of LCOS within a life test is considered. Again, another distribution involving the ratios of theses sums, in a series of life tests is developed which also turns out to be another generalised Dirichlet-type distribution. In both the above developments, some special cases are considered where either only spacings are considered or the ristricted ranges (difference between the largest and the smallest of the available observations) are considered.

2. DISTRIBUTION OF THE RATIOS OF LCOS

Consider a life test based on the exponential model

$$f(x) = \theta \exp(-\theta x),\ \theta > 0,\ x > 0. \tag{1}$$

Let n items be put on test and $x_{(i)}$ denote the i-th life (i-th order statistics in a sample of size n). Then it is well known that the spacings $x_{(i)}-x_{(i-1)} = v_i$, $i=1,2,\ldots,n$ are independent and v_i has the distribution

$$f(v_i) = (n-i+1)\theta \exp\left[-(n-i+1)\theta\, v_i\right]. \tag{2}$$

Now, consider a linear combination of spacings (LCOS)

$$u_i = \sum_{j=r_i+2}^{n-s_i} b_j v_j,\ b_j > 0$$
$$1 \le r_i < s_i \le n. \tag{3}$$

Now by finding the characteristic function of u_i, using (2) and inverting, we get the probability density of u_i as

$$f(u_i) = \bar{E}_i \sum_{k_i=0}^{a_i} (-1)^{a_i} \theta [\exp(-\theta\, d_i u_i)]/C_k(i)Q(i) \quad (4)$$

where $m_i = n-r_i-1$, $a_i = n-r_i-s_i-2$,

$$\bar{E}_i = m_i!/s_i!, d_i = (m_i-k_i)/C_k(i) \quad (4a)$$

$$C_k(i) = b_{r_i+k_i+2}, \; C_j(i) = b_{r_i+j_i+2}$$

$$Q(i) = \prod_{\substack{j_i=0 \\ j_i \neq k_i}}^{a_i} [\, d_i C_j(i) = (m_i-j_i) \,]$$

and $C_j(i)$, $C_k(i)$ are such no term in the denominator of (4) vanishes. Now taking the joint pdf of u_i and u where u_i is in terms of r_i, s_i such that

$$1 \leq r_i < s_i < r_\ell < s_\ell \leq n$$

and using the transofrmation $w = u_\ell/u_i$, we get the pdf of w from (4), as

$$f(w) = \bar{E}_i \bar{E}_\ell \sum_{k_i=0}^{a_i} \sum_{k_\ell=0}^{a_\ell} (-1)^{a_i+a_\ell}/C_k(i)C_i(\ell)Q(i)Q(\ell)[d_i+d_\ell w]^2 \quad (5)$$

where $C_k(\ell)$, $Q(\ell)$ are similar to $C_k(i)$ and $Q(i)$ in (4a). Now consider p independent life tests and take the joint density of $w_1, w_2, \ldots, w_p$, which can be written as

$$\prod_{i=1}^{p} f(w_i) = \prod_{i=1}^{p} \left[\prod_{t=1}^{2} (\bar{E}_{it}) \sum_{k_{it}=0}^{a_{it}} \frac{(-1)^{a_{it}}}{C_k(it)Q(it)[d_{i1}+d_{i2}w_i]^2} \right]. \quad (6)$$

The terms inside the braces in (6) are actually (5) in the new notations. $C_k(it)$, $Q(it)$ are exactly $C_k(i)$ and $Q_k(i)$ only with i replaced by (it) with $i = 1,2,..,p$; $t = 1,2$ (t runs within a life test while i runs among life tests). For example, $C_k(it) = b_{r_{it}+k_{it}+2}$. (6) can also be written as

$$f(z_1,z_2,....,z_p) = \prod_{i=1}^{p} \left[\prod_{i=1}^{2} (\bar{E}_{it}) \sum_{k_{it}} \frac{(-1)^{a_{it}}}{C_k(it)Q(it)d_{i1}d_{i2}(1+z_1)^2} \right]. \tag{6a}$$

Now make the transformations,

$$z_i = y_i/T_{i-1}, T_i = 1+y_1+...+y_i \tag{7}$$

$$i = 1,2,...,p.$$

From (7), we have

$$T_i = \prod_{j=1}^{i} (1 + z_j)$$

$$y_i = z_i T_{i-1} = z_i \prod_{j=1}^{i-1} (1 + z_j). \tag{8}$$

Noting the Jacobian of transformations as

$$|J| = 1/ \prod_{i=1}^{p} (1 + z_i)^{p-1} \tag{9}$$

and using it in (6a), we get the pdf of $y_1, y_2, ..., y_p$ as

$$f(y_1,...,y_p) = \prod_{i=1}^{p} \left[\prod_{t=1}^{2} \left(E_{it} \sum_{k_{it}} \frac{(-1)^{a_{it}}}{Q(it)C_k(it)d_{i1}d_{i2}} \right) \frac{1}{T_1T_2...T_{p-1}T_p^2} \right] \tag{10}$$

Terms d_{i1}, d_{i2} in (10) correspond to d_i, d_ℓ in (5).

The exponential (10) represents a generalised Dirichlet-type distribution since, if $a_{it} = 0$ and $C_k(it) = C_j(it) = 1$, $i=1,2,\ldots,p$, $t = 1,2$, then all the sums in (10) vanish and now we have

$$f(y_1,\ldots,y_p) = 1/(T_1 T_2 \ldots T_{p-1} T_p^2). \tag{11}$$

Note now $d_{i1} = m_{i1}$, $d_{i2} = m_{12}$ and $E_{it} = m_{it}$, $t = 1,2$. The pdf (11) is a special case of the generalised Dirichlet distribution of Lingappaiah [8] which is

$$f(y_1,\ldots,y_p) = C\, y_1^{a_1-1} \ldots y_p^{a_p-1} / T_1^{2a_1-a_2} \ldots T_{p-1}^{2a_{p-1}-a_p} T_p^{2a_p} \tag{12}$$

where $C = 1/\prod_{j=1}^{p} B(a_j,a_j)$. It is easily seen that (11) is (12) for $a_i = 1$, $i = 1,\ldots,p$. [a's in (12) are different from our a's here.] Actually by setting $a_{it}=0$, we are considering the ratios of two spacings in each of p tests. Two such spacings in each life test depend on r_{i1}, r_{i2}, $i = 1,\ldots,p$, assuming that we are putting the same number of items ($n_i=n$, $i = 1,\ldots,p$) in each of the test.

3. RATIOS OF SUMS OF LCOS

Now consider the sum of LCOS within a test. That is,

$$z = \sum_{i=1}^{\ell} u_i \tag{13}$$

where u_i is the LCOS corresponding to the i-th segment of the life test. Similar to Section 2 by taking the characteristic function of z and inverting, we get the pdf of z as

$$f(z) = \left[\prod_{i=1}^{\ell} (E_i)(-1)^{\ell-1}\right] \cdot \left[\sum_{i=1}^{\ell} \sum_{k_i=0}^{a_i} \frac{\exp[-\theta d_i z]}{C_k(i)Q(i)Q(i,t)}\right] \tag{14}$$

where $E_i = m_i!(-1)^{a_i}/s_i!$, while m_i, s_i, a_i and $Q(i)$, $C_k(i)$ as in Section 2 with

$$Q(i,t) = \prod_{\substack{t=1\\t\neq i}}^{\ell} \prod_{j_t=0}^{a_t} \left[d_i C_j(t) - (m_t - j_t)\right] \tag{15}$$

$$C_j(t) = b_{r_t+j_t+2}, \quad m_t = n - r_t - 1.$$

It is to be noted that $Q(i)$ runs within a LCOS while $Q(i,t)$ runs among all other LCOS <u>except</u> i-th and further $Q(i,t)$ runs <u>all over</u> within the $\ell-1$ LCOS other then i-th while $Q(i)$ runs only on a_i-1 values except $j_i=k_i$. Now consider the joint density of such sums of LCOS, each from (p+1) independent life tests, which can be written as,

$$\prod_{i=1}^{p+1} g(z_i) = \prod_{q=1}^{p+1}\left[A(\ell_q) \sum_{i=1}^{\ell_q} \sum_{k_{qi}=0}^{a_{qi}} \frac{\left(\prod_{i=1}^{p+1} \theta_i\right) \exp\left[-\sum_{q=1}^{p+1} \theta_q z_q d_{qi}\right]}{C_k(qi)\,Q(qi)\,Q(qi,qt)}\right] \tag{16}$$

where $A(\ell_q) = \left[\prod_{i=1}^{\ell_q} (E_{qi})\right](-1)^{\ell_q-1}$, $E_{qi} = (m_{qi})!\,(-1)^{a_{qi}}/s_{qi}!$

$C_k(qi)$, $C_j(qi)$, $Q(ki)$, $Q(qi,qt)$ are all similar to $C_k(i)$, $C_j(i)$, $Q(i)$, and $Q(i,t)$ with i replaced by (it). For example,

$$Q(qi,qt) = \prod_{\substack{t=1\\t\neq i}}^{\ell_q} \prod_{j_{qt}=0}^{a_{qt}} \left[d_{qi} C_j(qt - (m_{qt} - j_{qt})\right]$$

$$d_{qi} = (m_{qi} - k_{qi})/C_k(qi), \qquad C_k(qi) = b_{r_{qi}+k_{qi}+2}, \tag{16a}$$

$$C_j(qt) = b_{r_{qt}+j_{qt}+2}, \qquad C_j(ki) = b_{r_{qi}+j_{qi}+2}.$$

Now make the transformations,

$$w_i = z_i/z_{p+1}, \quad i = 1,2,\ldots,p, \quad w_{p+1} = z_{p+1}, \tag{17}$$

(Note the difference in notations, z_i's, w_i's from the previous section.) Noting the Jacobian of transformation is w_{p+1}^{p} and using this along with the transformations (17) in (16) and integrating out w_{p+1}, we get

$$f(w_1,w_2,\ldots,w_p) = \prod_{q=1}^{p+1}\left[A(\ell_q) \sum_{i=1}^{\ell_q} \sum_{k_{qi}} \frac{\left(\prod_{i=1}^{p+1} \theta_i\right)\Gamma(p+1)}{\left[\sum_{q=1}^{p} \theta_q w_q d_{qi} + \theta_{p+1} d_{(p+1)i}\right]^{p+1}} \cdot \frac{1}{C_k(qi)Q(qi)Q(qi,qt)} \right]. \tag{18}$$

The expression (18) denotes another generalised Dirichlet-type distribution. Because, if all $C_k(qi)=C_j(qi)=1=C_j(qt)$ and $a_{qi}=0, q=1,\ldots,p+1,\ i=1,\ldots,\ell_q$, then (18) reduces to

$$f(\bar{w}_1,\ldots,\bar{w}_p) = \prod_{q=1}^{p+1}\left[\prod_{i=1}^{\ell_q}(m_{qi}) \sum_{i=1}^{q} \frac{\left(\prod_{i=1}^{p+1} \theta_i\right)\Gamma(p+1)}{\left[\sum_{q=1}^{p+1} \theta_q m_{qi}\bar{w}_q + \theta_{p+1} m_{(p+1)i}\right]^{p+1}} \cdot \frac{1}{\prod_{\substack{t=1\\t\neq i}}^{\ell_q} (m_{qi}-m_{qt})} \right]. \tag{19}$$

From (19), we have

$$\int_0^\infty \cdots \int_0^\infty f(w_1,\ldots,w_p)dw_1,\ldots,dw_p = \prod_{q=1}^{p+1}\left[\prod_{i=1}^{\ell_q}(m_{qi}) \sum_{i=1}^{q} \frac{1}{m_{qi}\prod_{\substack{t=1\\t\neq i}}^{q}(m_{qi}-m_{qt})}\right] \quad (20)$$

Now each of the sums $\sum_{i=1}^{\ell_q}$ in (20) is $(-1)^{\ell_q-1} \Big/ \prod_{i=1}^{\ell_q}(m_{qi})$ and hence (20) reduces to unity. Again if $C_k(qi)=C_j(qi)=1$ and $q_{qi}\neq 0$, then (18) reduce to

$$f(w_1^o,\ldots,w_p^o) = \prod_{q=1}^{p+1}\left[A(\ell_q)\sum_{i=1}^{\ell_q}\sum_{k_{qi}} \frac{\binom{a_{qi}}{k_{qi}}(-1)^{k_{qi}}\left(\prod_{i=1}^{p+1}\theta_i\right)\Gamma(p+1)}{\left[\sum_{i=1}^{p}\bar{d}_{qi}w_q^o\theta_q+\bar{d}_{(p+1)i}\theta_{p+1}\right]^{p+1}\bar{Q}(qi,qt)}\right] \quad (21)$$

where $\bar{d}_{qi} = (m_{qi} - k_{qi})$ and

$$\bar{Q}(qi,qt) = \prod_{\substack{t=1\\t\neq i}}^{q}\prod_{j_{qt}=0}^{a_{qt}}\left[\bar{d}_{qi}-(m_{qt}-j_{qt})\right]. \quad (21a)$$

For example, if $i=2$, $p=2$, $\ell_q=2$, $m_{q1}=7$, $m_{q2}=3$, $r_{q1}=2$, $r_{q2}=6$, $s_{q1}=5$, $s_{q2}=1$, $q=1,2,3$. Then, we have $a_{qi}=1$, $q=1,2,3$, $i=1,2$ and (20) now reduces to

$$(-1)\left(\frac{7!\;3!}{5!\;1!}\right)^3\prod_{q=1}^{3}\left[\sum_{k_{q1}=0}^{1}\binom{1}{k_{q1}}(-1)^{k_{q1}}\Big/(7-k_1)\prod_{j_{q2}=0}^{1}(4-k_{q1}+j_{q2})\right.$$

$$\left.+\sum_{k_{q2}=0}^{1}\binom{1}{k_{q2}}(-1)^{k_{q2}}\Big/(3-k_2)\prod_{j_{q1}=0}^{1}(-4-k_{q2}+j_{q1})\right] \quad (22)$$

and (22) is equal to unity.

<u>Comments</u>: As in the previous section 2, here in (19) we have the ratios of the sum of spacings from <u>each</u> of p life tests to the sum of spacings in (p+1)-th life test, while in (21) we have the ratio of the sum of the ristricted ranges fróm <u>each</u> life test. That is, z in (13) represents the sum of spacings if $a_{qt}=0$, $i=1,\ldots,\ell_q$; $q=1,2,\ldots,p+1$ and similarly z represents the sum of ristricted ranges, when $a_{qi}\neq 0$ and $C_i(k)=C_j(k)=1$, where the ristricted range for the segment i of test q is given by $(x_{n-s_{qi}})-(x_{r_{qi}+1})$ $q=1,2,\ldots,p+1$. Further it is to be noted that we have quite a large number of parameters in addition to $\theta_1,\theta_2,\ldots,\theta_{p+1}$, such as r_{qi},s_{qi}, $q=1,\ldots,p+1$; $i=1,2,\ldots,\ell_q$ and ℓ_q themselves. In addition to these, we can set different n_q's, $q=1,2,\ldots,p+1$, while we have taken here for simplicity, $n_q=n$. Of course, a_{qi}, m_{qi} depend in turn on n_q, s_{qi} and r_{qi}'s. It may be of interest to discuss the effect of the distances between r_{qi} and r_{qj}, $i\neq j$, $q=1,2,\ldots,p+1$, $i,j=1,2,\ldots,\ell_q$ and in turn the effect of a_{qi}'s on $T_1,\ldots,T_p$'s$(\ell_q=2)$ and also on z's in Section 3. Also, it is to be observed, tha t our transformation $z_i=y_i/T_{i-1}$, and in turn $y_i = z_i \prod_{j=1}^{i-1} (1+z_j)$ denotes the function of z's similar to the concept of "force of mortality" mentioned in Lochner [11] .

4. REFERENCES

[1] Aitchison, J. (1963) Inverse distributions and independent gamma distributed products of random variables. Biometrika, Vol. 50, pp. 505-507.

[2] Connor, Robert J. and Mosimann, James E. (1969) Concept of independence for proportions with a generalisation of the Dirichlet's distribution. Journal of American Statistical Association, Vol. 66, pp. 641-643.

[3] Darroch, J.N. and James, I.R. (1974) F-independence and null correlations of continuous bounded sums and positive variables. Journal of the Royal Statistical Scoiety, Series-B, Vol. 36.

[4] Darroch, J.N. and Ratcliff, D. (1971) Characterization of the Dirichlet's distribution. Journal of American Statistical Association, Vol. 66, pp. 642-643.

[5] Darroch, J.N. and Ratcliff, D. (1973) Tests of F-independence with reference to quasi-independence and Waite's finger print data. Biometrika, Vol. 60, pp. 395-401.

[6] Institute of Mathematical Statistics. Selected Tables in Mathematical Statistics, Vol. 4, (1977).

[7] James, I.R. (1972) Products of independent beta variables with application to Connor and Mosimann's generalised Dirichlet's distribution. Journal of American Statistical Association, Vol. 67, pp. 910-912.

[8] Lingappaiah, G.S. (1976) On the generalised inverted Dirichlet's distribution. Demonstratio Mathematica, Vol. 9, No. 3, pp. 423-433.

[9] Lingappaiah, G.S. (1977) On the Dirichlet's variables. Journal of Statistics Research, Vol. 11, No.2.

[10] Lingappaiah, G.S. (1978) On the extension of the generalised Inverted Dirichlet's distribution, Rev. Univ. Nac. Tucuman, Vol. 28.

[11] Lochner, R.H. (1975) A generalised Dirichlet's distribution in Bayesian life testing. Journal of Royal Statistical Society Series-B, Vol. 37, pp. 103-113.

[12] Mosimann, G.E. (1963) On the compound multinomial distribution, the multi-variate Beta distribution and correlation among proportions. Biometrika, Vol. 49, pp. 65-82.

[13] Mosimann, G.E. (1963) On the compound negative multinomial distribution and correlations among inversely sampled pollen counts, Biometrika, Vol. 49, pp. 47-54.

[14] Yassaee, H. (1974) Inverted Dirichlet's Distribution and Multivariate Logistic Distribution. Canadian Journal of Statistics, Vol. 2, pp. 99-105.

[15] Yassaee, H. (1976) Probability Integral of Inverted Dirichlet Distribution and its applications. Proceedings of COMSTAT, Second Symposium Berlin (West).