Random Discrete Distributions

By J. F. C. KINGMAN

University of Oxford

[Read before the ROYAL STATISTICAL SOCIETY at a meeting organized by the RESEARCH SECTION, on Wednesday, October 16th, 1974, Professor R. L. PLACKETT in the Chair]

SUMMARY

It is impossible to choose at random a probability distribution on a countably infinite set in a manner invariant under permutations of that set. However, approximations to such a choice can be made by considering exchangeable probability measures on the class of probability distributions over a finite set, and letting the size of that set increase without limit. Under suitable conditions the resulting probabilities, when arranged in descending order, have non-degenerate limiting distributions. These apparently arcane considerations lead to rather concrete conclusions in certain problems in applied probability.

Keywords: DISCRETE DISTRIBUTION; DIRICHLET DISTRIBUTION; GAMMA PROCESS; SUBORDINATOR; ORDER STATISTICS; LIMIT THEOREM; RANDOM MEASURE

1. INTRODUCTION

IN recent years there has been a good deal of interest in the problem of describing the distributions of random elements which are themselves probability distributions on finite or infinite sets. Much of this has been motivated by problems of Bayesian inference and decision theory, in which the "unknown state of nature" takes the form of a probability distribution; this Society has heard papers by Good (1967) and Ericson (1969) in which just this problem arises. A good example of recent research in this area is the paper of Ferguson (1973), which deals with distributions on quite general sets. (Ferguson's paper contains a useful bibliography of earlier work. For some historical and technical comments on his analysis, see Appendix 2.)

In the case of a finite set (of N elements, say) the problem is in principle fairly simple. A probability distribution on such a set is just a collection of N numbers p_i $(i = 1, 2, \ldots, N)$ satisfying

$$p_i \geqslant 0, \quad \sum_{i=1}^{N} p_i = 1. \tag{1}$$

If the p_i are regarded as the co-ordinates of a point $\mathbf{p}$ in N-dimensional space R^N, then the conditions (1) delineate a subset Δ_N of R^N (an $(N-1)$-dimensional simplex). To choose a probability distribution at random is therefore tantamount to choosing a point $\mathbf{p}$ at random, according to some probability measure concentrated on Δ_N.

By far the most common such measure to be found in the literature is the Dirichlet distribution with probability element

$$\frac{\Gamma(\alpha_1 + \alpha_2 + \ldots + \alpha_N)}{\Gamma(\alpha_1)\,\Gamma(\alpha_2)\ldots\Gamma(\alpha_N)} p_1^{\alpha_1 - 1} p_2^{\alpha_2 - 1} \ldots p_N^{\alpha_N - 1}\, dp_1\, dp_2 \ldots dp_{N-1} \tag{2}$$

on Δ_N, where the parameters α_i are strictly positive, and Γ denotes the usual gamma function. It is often appropriate to consider exchangeable distributions for $\mathbf{p}$, in which case the α_i must be taken to be equal, and (2) reduces to

$$\frac{\Gamma(N\alpha)}{\Gamma(\alpha)^N}(p_1 p_2 \ldots p_N)^{\alpha-1}\, dp_1\, dp_2 \ldots dp_{N-1}. \tag{3}$$

The situation is much trickier if we wish to consider random distributions on countably infinite sets, for such distributions are described by points in the infinite-dimensional space Δ_∞ of sequences

$$\mathbf{p} = (p_1, p_2, p_3, \ldots) \tag{4}$$

satisfying

$$p_i \geqslant 0 \quad (i = 1, 2, 3, \ldots),\ \sum_{i=1}^{\infty} p_i = 1, \tag{5}$$

and probability measures on Δ_∞ are not easy to construct or to manipulate. To illustrate the difficulties which can arise, remark that there is no exchangeable distribution on Δ_∞, or indeed any probability measure for which the marginal distribution of p_i does not depend on i, for such a measure would have to satisfy the incompatible conditions†

$$\mathbf{E}(p_1) = \mathbf{E}(p_2) = \ldots = \mathbf{E}(p_i) = \ldots,$$

$$\sum_{i=1}^{\infty} \mathbf{E}(p_i) = 1.$$

My own interest in this problem has nothing to do with Bayesian inference, but arises from the conjunction of two facts, one practical and the other theoretical.

(i) In some applied situations for which stochastic models can be constructed, the performance of a system is a function of some underlying discrete distribution, and it is relevant to consider the "average" performance over "typical" distributions. The type of problem I have in mind is illustrated in Section 2.

(ii) If in (3) we let $N \to \infty$ and $\alpha \to 0$ in such a way that $N\alpha$ converges to a finite positive limit, then each p_i converges in probability to zero, but the distribution of the largest (and of the second largest, and so on) of the p_i converges to a non-degenerate limit. This is proved in Section 5.

In the next five sections I shall try to explain the significance of, and the connection between, (i) and (ii). In Section 7 and thereafter, a broader context will emerge in which to set the rather special considerations arising from the Dirichlet distribution.

2. HEAPS

The problem of "heaps" has been raised by Mr P. J. Burville, and he and I have discussed it in detail elsewhere (Burville and Kingman (1973); see also Hendricks (1972) and Burville (1974)). A number of different items $I_1, I_2, \ldots, I_N$ (papers on a desk, pieces of information in a computer, books on a shelf) are stored, literally or figuratively, in a heap. From time to time an item is demanded—it is I_i with probability $p_i > 0$ and successive demands are independent—and it is searched for through the heap, starting at the top. After being found and used, the item is

† In this paper **E** denotes expectation and **P** probability; there are no matrices to confuse the notation.

replaced on the top of the heap, and the system is ready for the next demand. Clearly, when the system has reached a steady state, there will be a tendency for the most popular items to be near the top of the heap, and thus to be quickly found when needed.

It is fairly straightforward to calculate probabilities and expectations which measure the performance of the system in statistical equilibrium;† for example, Burville and Kingman show that, if the search time for an item (in appropriate units) is equal to the number of items lying above it in the heap, then the expected search time for a typical item demanded is given by

$$\mu = \sum_{\substack{i,j=1 \\ i \neq j}}^{N} \frac{p_i p_j}{p_i + p_j}. \tag{6}$$

This quantity may be compared with others obtained by varying the conditions of the problem. For instance, if the item demanded is returned after use to a random position in the heap, then the mean search time becomes $\frac{1}{2}(N-1)$, and it is easy to show that

$$\mu \leqslant \tfrac{1}{2}(N-1), \tag{7}$$

with equality if and only if the p_i are all equal.

A more interesting variant is that in which the items are kept in the heap in order of popularity (supposing this to be known to the organizer of the system). For this arrangement the mean search time is

$$m = \sum_{j=1}^{N} (j-1) p_{(j)}, \tag{8}$$

where the numbers $p_{(j)}$ are the p_i arranged in descending order

$$p_{(1)} \geqslant p_{(2)} \geqslant \ldots \geqslant p_{(N)}. \tag{9}$$

It can be shown that μ and m are related by the inequalities

$$m \leqslant \mu \leqslant 2m. \tag{10}$$

The quantities μ and m depend of course on the underlying distribution $\mathbf{p} = (p_1, p_2, \ldots, p_N)$ of demand, and expressions like (6), though explicit, are scarcely transparent. It is natural to ask what sort of values μ and m take for typical distributions $\mathbf{p}$. This question can to some extent be answered by computing these quantities for simple distributions, but an alternative approach which is perhaps more convincing is to average μ and m over the set Δ_N of possible distributions.

But to average, one needs a probability measure on Δ_N, and following the crowd one is led to use the Dirichlet measure (3), if only for the possibility of explicit integration.

3. Expectations with Respect to the Dirichlet Distribution

If $\mathbf{p} = (p_1, p_2, \ldots, p_N)$ is distributed over Δ_N according to (3), it is well known that each p_i has the probability density

$$\frac{\Gamma(N\alpha)}{\Gamma(\alpha)\,\Gamma((N-1)\,\alpha)} x^{\alpha-1}(1-x)^{(N-1)\alpha-1} \tag{11}$$

† Formally, the successive orders of items in the heap form a Markov chain with $N!$ states, which is irreducible and aperiodic and hence has a unique equilibrium distribution.

in $0 \leqslant x \leqslant 1$, and that if $N \geqslant 3$ and $i \neq j$, (p_i, p_j) has the joint probability density

$$\frac{\Gamma(N\alpha)}{\Gamma(\alpha)^2\,\Gamma((N-2)\,\alpha)}(xy)^{\alpha-1}(1-x-y)^{(N-2)\alpha-1} \tag{12}$$

in $x, y \geqslant 0$, $x+y \leqslant 1$. These results enable the expectations of μ and m, regarded as functions of the random point $\mathbf{p}$ in Δ_N, to be evaluated:

$$\begin{aligned}
\mathbf{E}(\mu) &= \sum_{i \neq j} \mathbf{E}\left(\frac{p_i p_j}{p_i + p_j}\right) \\
&= N(N-1)\frac{\Gamma(N\alpha)}{\Gamma(\alpha)^2\,\Gamma((N-2)\,\alpha)} \iint\limits_{\substack{x,y \geqslant 0\\ x+y \leqslant 1}} \frac{xy}{x+y}(xy)^{\alpha-1}(1-x-y)^{(N-1)\alpha-1}\,dx\,dy, \\
\mathbf{E}(m) &= \mathbf{E}\left\{\sum_{j=1}^{N} p_j \times (\text{number of } i \text{ with } p_i > p_j)\right\} \\
&= N(N-1)\frac{\Gamma(N\alpha)}{\Gamma(\alpha)^2\,\Gamma((N-2)\,\alpha)} \iint\limits_{\substack{x > y \geqslant 0\\ x+y \leqslant 1}} y(xy)^{\alpha-1}(1-x-y)^{(N-1)\alpha-1}\,dx\,dy.
\end{aligned}$$

After some calculation these reduce to

$$\mathbf{E}(\mu) = \frac{(N-1)\,\alpha}{1+2\alpha} \tag{13}$$

and

$$\mathbf{E}(m) = (N-1)\left(\frac{1}{2} - \frac{\Gamma(1+2\alpha)}{\Gamma(1+\alpha)^2\,2^{1+2\alpha}}\right). \tag{14}$$

Other expectations may be calculated in a similar way. For instance, if ϕ is any integrable function, and

$$s_\phi = s_\phi(\mathbf{p}) = \sum_{j=1}^{N} \phi(p_j), \tag{15}$$

then

$$\mathbf{E}(s_\phi) = \sum_{j=1}^{N} \mathbf{E}\{\phi(p_j)\} = N\mathbf{E}\{\phi(p_1)\},$$

so that

$$\mathbf{E}(s_\phi) = \frac{N\Gamma(N\alpha)}{\Gamma(\alpha)\,\Gamma((N-1)\,\alpha)} \int_0^1 \phi(x)\,x^{\alpha-1}(1-x)^{(N-1)\alpha-1}\,dx. \tag{16}$$

One feature of all these formulae should be noticed. If $N \to \infty$ and $\alpha \to 0$ in such a way that

$$N\alpha \to \lambda \quad (0 < \lambda < \infty), \tag{17}$$

then each converges to a limit:

$$\mathbf{E}(\mu) \to \lambda, \tag{18}$$

$$\mathbf{E}(m) \to \lambda \log 2 = 0{\cdot}69\lambda, \tag{19}$$

and (so long as the function ϕ makes this integral absolutely convergent),

$$\mathbf{E}(s_\phi) \to \lambda \int_0^1 \phi(x)\, x^{-1}(1-x)^{\lambda-1}\, dx. \tag{20}$$

4. THE POISSON–DIRICHLET LIMIT

The form of the symmetric Dirichlet distribution (3), for given N, depends on the parameter α, which may take any strictly positive value. When α is large, the distribution is concentrated near the centre $(N^{-1}, N^{-1}, \ldots, N^{-1})$ of Δ_N, and as $\alpha \to \infty$ it converges to the degenerate distribution at this point. On the other hand, when α is small, the distribution assigns high probability to neighbourhoods of the faces, edges and vertices of Δ_N. Thus, if one expects the distribution $\mathbf{p}$ to be fairly uniform, then high values of α are appropriate, while small values of α correspond to a situation in which the different p_j are expected to be widely different.

In the practical problem which led to the formulation of the "heaps" model, the typical situation was one in which a few items were relatively popular, while there was a long tail of items more rarely demanded. Such a context therefore calls for large values of N combined with small values of α, and the limiting situation (17) describes an appropriate approximation. Then (18) and (19) are simple and useful conclusions, elaborating (10). Moreover, if assertion (ii) of Section 1 is valid, then the limiting process assigns a non-degenerate distribution for the frequency $p_{(1)}$ with which the most popular item is demanded.

The suggestion is therefore that, in a problem which seems to make the impossible demand for an exchangeable probability measure on Δ_∞, it may be sensible to proceed as follows:

(1) Replace Δ_∞ by Δ_N, for a large value of N.

(2) Compute the expectations of interesting quantities (and the probabilities of interesting events) with respect to the Dirichlet distribution (3) on Δ_N.

(3) Let $N \to \infty$ and $\alpha \to 0$ in such a way that (17) holds.

If this procedure is accepted as valid, then its application to the "heaps" model shows that the ratio $\mathbf{E}(m)/\mathbf{E}(\mu)$, measuring the efficiency of the self-regulating heap relative to the (often unpractical) ideal of filing in order of popularity, takes the limiting value 0·69, independent of λ.

Another simple consequence flows from (20) on taking

$$\phi(x) = -\beta x \log x, \tag{21}$$

where $\beta = (\log 2)^{-1}$. Then s_ϕ is just the entropy

$$h(\mathbf{p}) = -\sum_{j=1}^{N} p_j \log_2 p_j \tag{22}$$

of the distribution $\mathbf{p}$. In the limiting situation (17),

$$\mathbf{E}\{h(\mathbf{p})\} \to -\lambda\beta \int_0^1 \log x (1-x)^{\lambda-1}\, dx,$$

which simplifies to

$$\mathbf{E}\{h(\mathbf{p})\} \to \beta \sum_{n=1}^{\infty} \frac{\lambda}{n(n+\lambda)}. \tag{23}$$

Thus the limiting process yields, in a sense, a random distribution with finite expected entropy on the positive integers. Note that the right-hand side of (23) reduces to

$$\beta\left(1+\frac{1}{2}+\frac{1}{3}+\ldots+\frac{1}{\lambda}\right) \tag{24}$$

if λ is an integer.

5. THE LIMITING ORDER STATISTICS

In this section we prove a theorem which is the formal expression of assertion (ii) of Section 1, and which also exhibits the limiting joint distribution of the order statistics $p_{(1)} \geqslant p_{(2)} \geqslant p_{(3)} \geqslant \ldots$. To state it, let ∇_∞ be the subset of Δ_∞ consisting of infinite sequences (p_j) satisfying (5) and

$$p_1 \geqslant p_2 \geqslant p_3 \geqslant \ldots. \tag{25}$$

Then the mapping $(p_1, p_2, \ldots) \to (p_{(1)}, p_{(2)}, \ldots)$, which rearranges the components in descending order, is a function from Δ_∞ onto ∇_∞. If $\mathbf{P}$ is any probability measure on ∇_∞, and n is any positive integer, then the random n-vector $(p_1, p_2, \ldots, p_n)$ has a distribution depending on $\mathbf{P}$, which might be called the nth marginal distribution of $\mathbf{P}$. The measure $\mathbf{P}$ is uniquely determined by its marginal distributions.

Theorem. For each λ in $0 < \lambda < \infty$, there is a probability measure $\mathbf{P}_\lambda$ on ∇_∞ with the following property. If for each N the random vector $\mathbf{p}$ is distributed over Δ_N according to the distribution (3) with $\alpha = \alpha_N$, and if $N\alpha_N \to \lambda$ as $N \to \infty$, then for any n the distribution of the random vector $(p_{(1)}, p_{(2)}, \ldots, p_{(n)})$ converges to the nth marginal distribution of $\mathbf{P}_\lambda$.

The key to the proof of this theorem is a simple and ancient device, which is to note that if $y_1, y_2, \ldots, y_N$ are independent random variables, each having the gamma distribution with probability density

$$y^{\alpha-1} e^{-y}/\Gamma(\alpha) \quad (y > 0), \tag{26}$$

and if $s = y_1 + y_2 + \ldots + y_N$, then the vector

$$(y_1/s, y_2/s, \ldots, y_N/s) \tag{27}$$

has the distribution (3).

To exploit this fact, construct a gamma process ξ, a random process $(\xi(t);\, t \geqslant 0)$ with $\xi(0) = 0$ such that the increments of ξ on disjoint intervals are independent, and such that $\xi(t_2) - \xi(t_1)$ has the distribution (26) with α replaced by $(t_2 - t_1)$ for $0 \leqslant t_1 < t_2$. Such processes are commonplace for instance in the theory of dams, and it is known that, as a function of t, $\xi(t)$ increases only in jumps. The positions of these jumps form a random countable dense subset $J(\xi)$ of $(0, \infty)$, with

$$\mathbf{P}\{t \in J(\xi)\} = 0 \tag{28}$$

for all $t > 0$.

For each value of N, write

$$q_j = q_j(N) = \frac{\xi(j\alpha_N) - \xi\{(j-1)\,\alpha_N\}}{\xi(N\alpha_N)}. \tag{29}$$

By the result cited above, the vector $\mathbf{q} = (q_1, q_2, \ldots, q_N)$ has the same distribution as $\mathbf{p}$, and it therefore suffices to prove the theorem with $\mathbf{p}$ replaced by $\mathbf{q}$. We shall in fact prove that

$$\lim_{N\to\infty} q_{(j)}(N) = \delta\xi_{(j)}/\xi(\lambda), \tag{30}$$

where $\delta\xi_{(j)}$ are the magnitudes of the jumps of ξ in $(0,\lambda)$, arranged in descending order. This will suffice to prove the theorem, with $\mathbf{P}_\lambda$ the distribution of the sequence

$$(\delta\xi_{(j)}/\xi(\lambda);\, j=1,2,\ldots), \tag{31}$$

for this sequence lies in ∇_∞ as a consequence of the equality

$$\xi(\lambda) = \sum_{j=1}^{\infty} \delta\xi_{(j)}. \tag{32}$$

Equation (30) is a result in elementary real analysis. Fix attention on any realization of the random process ξ. For any integer n, choose N_0 so large that, for any $N \geqslant N_0$, the discontinuities of height $\delta\xi_{(j)}$ $(j=1,2,\ldots,n)$ are contained in distinct intervals $((i-1)\alpha_N, i\alpha_N)$. Then (29) implies that

$$\xi(N\alpha_N)q_{(j)} \geqslant \delta\xi_{(j)} \quad (1 \leqslant j \leqslant n, N \geqslant N_0),$$

so that

$$\liminf_{N\to\infty} q_{(j)} \geqslant \delta\xi_{(j)}/\xi(\lambda) \tag{33}$$

for $j=1,2,\ldots,n$. Since n is arbitrary, (33) holds for all j, and Fatou's lemma and (32) give

$$\begin{aligned}\limsup_{N\to\infty} q_{(j)} &= \limsup_{N\to\infty}\Big\{1-\sum_{i\neq j} q_{(i)}\Big\}\\ &\leqslant 1-\sum_{i\neq j}\liminf_{N\to\infty} q_{(j)}\\ &\leqslant 1-\sum_{i\neq j}\{\delta\xi_{(i)}/\xi(\lambda)\}\\ &= \delta\xi_{(j)}/\xi(\lambda).\end{aligned}$$

Comparing this inequality with (33) we obtain (30), and the proof of the theorem is complete.

Moreover, the limiting distributions of the order statistics are identified as the distributions of the normalized jumps

$$\delta\xi_{(j)}/\xi(\lambda) \tag{34}$$

of the gamma process on the interval $(0,\lambda)$. Explicit formulae for these distributions are not easy to derive; some partial results are set out in Appendix 1.

It is important to note that this limit theorem is not robust or distribution-free in the sense in which, for instance, the central limit theorem is. The distribution $\mathbf{P}_\lambda$ contains unmistakable signs of its genesis from the Dirichlet distribution. There is in fact a connection with the work of Plackett (1969), in the discussion to whose paper I sought to draw this distinction.

6. THE UNSYMMETRICAL CASE

The proof of the theorem of the last section extends without change to the more general distribution (2). To see this, let ξ bear the same meaning as before, and write

$$q_j = \frac{\xi(\alpha_1+\alpha_2+\ldots+\alpha_j)-\xi(\alpha_1+\alpha_2+\ldots+\alpha_{j-1})}{\xi(\alpha_1+\alpha_2+\ldots+\alpha_N)}, \tag{35}$$

where the positive quantities α_j $(j = 1, 2, \ldots, N)$ depend on N. Then $\mathbf{q} = (q_1, q_2, \ldots, q_N)$ is a random point in Δ_N having the unsymmetrical Dirichlet distribution (2).

An examination of the proof of (30) shows that it continues to hold in this more general situation, so long as

$$\alpha_1 + \alpha_2 + \ldots + \alpha_N \to \lambda, \tag{36}$$

and the dissection defined by the points

$$\alpha_1 + \alpha_2 + \ldots + \alpha_j \quad (j = 1, 2, \ldots, N)$$

becomes arbitrarily fine in the sense that

$$\max(\alpha_1, \alpha_2, \ldots, \alpha_N) \to 0. \tag{37}$$

Thus if conditions (36) and (37) are satisfied, the descending order statistics from the distribution (2) have non-degenerate limiting distributions which derive from the measure $\mathbf{P}_\lambda$.

In some problems, particularly those arising from Bayesian theory, (37) may fail in that some of the α_j fail to converge to zero. In such cases (35) may still be used to derive limiting distributions, and the necessary changes in the argument will be obvious to the reader.

7. SUBORDINATORS

The gamma process, which, as has been seen, lies at the root of the Poisson–Dirichlet limit, is of course a very special case of a process with stationary independent increments. If such a process ξ is increasing, it is called a subordinator, and if it has no deterministic drift its distributions are determined by the Lévy formula

$$\mathbf{E}[\exp\{-\theta\xi(t)\}] = \exp\{-t\psi(\theta)\} \quad (\theta \geqslant 0, t \geqslant 0), \tag{38}$$

where

$$\psi(\theta) = \int_{(0,\infty)} \{1 - \exp(-\theta x)\}\, \mu(dx) \tag{39}$$

for some measure μ on $(0, \infty)$. If μ is totally finite, then ξ is a compound Poisson process, but the interesting case is that in which μ has infinite total mass and yet still makes the integral in (39) converge.

The gamma process has

$$\psi(\theta) = \log(1 + \theta) \tag{40}$$

and the corresponding Lévy measure is given by

$$\mu(dx) = x^{-1} \exp(-x)\, dx. \tag{41}$$

Suppose, however, that μ is any measure on $(0, \infty)$ which has infinite total mass but which makes the integral (39) converge (for some and then for all $\theta > 0$). Then there is a subordinator ξ satisfying (38) and (39), and $\xi(t)$ increases with t only at its jumps, which form a random countable dense set in $(0, \infty)$. To simplify notation and to avoid slight difficulties with fixed discontinuities, it will be assumed throughout that μ has a density h.

The construction of Section 5 can be carried out with the gamma process replaced by the more general subordinator. Thus, taking $\lambda = 1$ without loss of generality, the vector $\mathbf{p} = (p_1, p_2, \ldots, p_N)$ defined by

$$p_j = \frac{\xi(jN^{-1}) - \xi\{(j-1)N^{-1}\}}{\xi(1)} \tag{42}$$

is randomly distributed over Δ_N with a distribution determined by μ. More precisely, the probability element of $\mathbf{p}$, generalizing (3), is

$$F(p_1, p_2, \ldots, p_N)\, dp_1\, dp_2 \ldots dp_{N-1}, \tag{43}$$

where, if $f(t, \cdot)$ denotes the probability density of $\xi(t)$,

$$F(p_1, p_2, \ldots, p_N) = \int_0^\infty f(N^{-1}, up_1) f(N^{-1}, up_2) \ldots f(N^{-1}, up_N)\, u^{N-1}\, du. \tag{44}$$

The argument of Section 5 carries through without change and shows that, if $\mathbf{p}$ has the distribution (43), and if $p_{(1)} \geqslant p_{(2)} \geqslant \ldots \geqslant p_{(N)}$ are the ordered values of the p_i, then the limiting joint distribution of the sequence $(p_{(j)})$ is the joint distribution of the normalized jumps

$$\delta\xi_{(j)}/\xi(1)$$

of the subordinator ξ on the interval (0, 1), and this is a probability measure $\mathbf{P}_\mu$ on ∇_∞ determined by the measure μ.

Of course, (44) is much more complicated than (3), especially since f must be determined by inverting the Laplace transform

$$\int_0^\infty f(t, x) \exp(-\theta x)\, dx = \exp\{-t\psi(\theta)\}. \tag{45}$$

However, some of the calculations generalizing those of Section 3 can be carried out. For instance, each p_i has the probability density

$$\int_0^\infty f(N^{-1}, ux) f\{1 - N^{-1}, u(1-x)\}\, u\, dx, \tag{46}$$

and the joint density of p_i and p_j $(i \neq j, N \geqslant 3)$ is

$$\int_0^\infty \int_0^\infty f(N^{-1}, ux) f(N^{-1}, uy) f\{1 - 2N^{-1}, u(1-x-y)\}\, u^2\, dx\, dy. \tag{47}$$

Moreover, in the notation of Section 2,

$$\begin{aligned} \mathbf{E}(\mu) &= N(N-1)\,\mathbf{E}\left(\frac{p_1 p_2}{p_1 + p_2}\right) \\ &= N(N-1)\,\mathbf{E}\left[\frac{\xi(N^{-1})\{\xi(2N^{-1}) - \xi(N^{-1})\}}{\xi(2N^{-1})\,\xi(1)}\right] \\ &= N(N-1) \int_0^\infty \int_0^\infty \int_0^\infty \frac{xy}{(x+y)(x+y+z)} f(N^{-1}, x) f(N^{-1}, y) f(1 - 2N^{-1}, z)\, dx\, dy\, dz, \end{aligned} \tag{48}$$

$$\lim_{N \to \infty} \mathbf{E}(\mu) = \int_0^\infty \int_0^\infty \int_0^\infty \frac{xy}{(x+y)(x+y+z)} h(x)\, h(y)\, f(1, z)\, dx\, dy\, dz, \tag{49}$$

and similarly

$$\mathbf{E}(m) = N(N-1)\int_0^\infty \underset{x<y}{\int_0^\infty}\int_0^\infty \frac{x}{x+y+z} f(N^{-1},x)f(N^{-1},y)f(1-2N^{-1},z)\,dx\,dy\,dz, \quad (50)$$

$$\lim_{N\to\infty}\mathbf{E}(m) = \int_0^\infty \underset{x<y}{\int_0^\infty}\int_0^\infty \frac{x}{x+y+z} h(x)\,h(y)f(1,z)\,dx\,dy\,dz. \quad (51)$$

The corresponding results for s_ϕ are slightly simpler:

$$\mathbf{E}(s_\phi) = N\int_0^\infty\int_0^\infty \phi\left(\frac{x}{x+y}\right) f(N^{-1},x)f(1-N^{-1},y)\,dx\,dy, \quad (52)$$

$$\lim_{N\to\infty}\mathbf{E}(s_\phi) = \int_0^\infty\int_0^\infty \phi\left(\frac{x}{x+y}\right) h(x)f(1,y)\,dx\,dy. \quad (53)$$

Two points should be noted about these forbidding formulae; firstly that quantities of interest have sensible limits as $N\to\infty$, and secondly that these limits depend in an essential way (through h and f) on the choice of the measure μ.

In the special case of the gamma process, the various integrals written down may be evaluated explicitly, to recover the results of Section 3. Another special case which is to some extent analytically tractable will be discussed in the next section.

8. The Stable Case

For any γ in $0<\gamma<1$ and any $a>0$, the measure μ with density

$$h(x) = ax^{-\gamma-1}\,dx \quad (54)$$

has infinite total mass but makes (39) converge; indeed

$$\psi(\theta) = b\theta^\gamma, \quad (55)$$

where

$$b = a\gamma^{-1}\Gamma(1-\gamma). \quad (56)$$

The corresponding subordinator ξ is *stable with exponent* γ in the sense that $\xi(t)$ has the same distribution as $t^{1/\gamma}\,\xi(1)$, so that

$$f(t,x) = t^{-1/\gamma}f(t^{-1/\gamma}x), \quad (57)$$

with $f(\cdot) = f(1,\cdot)$.

Except when $\gamma=\frac{1}{2}$, there is no closed expression for f, but integrals like (49) and (51) may be evaluated by noting that, for $w>0$,

$$\int_0^\infty \frac{f(z)}{w+z}\,dz = \int_0^\infty\int_0^\infty \exp\{-\theta(w+z)\}f(z)\,d\theta\,dz = \int_0^\infty \exp(-\theta w - b\theta^\gamma)\,d\theta.$$

Hence the right-hand side of (49) is equal to

$$\int_0^\infty\int_0^\infty\int_0^\infty \frac{xy}{x+y}\,ax^{-\gamma-1}\,ay^{-\gamma-1}\exp\{-\theta(x+y)-b\theta^\gamma\}\,dx\,dy\,d\theta,$$

and that of (51) to

$$\int_0^\infty \underset{x<y}{\int_0^\infty}\int_0^\infty x\,ax^{-\gamma-1}\,ay^{-\gamma-1}\exp\{-\theta(x+y)-b\theta^\gamma\}\,dx\,dy\,d\theta.$$

Each of these integrals succumbs to the transformation $x = u(1-v)$, $y = uv$, to give

$$\lim_{N\to\infty} \mathbf{E}(\mu) = \frac{\gamma}{1-2\gamma} \tag{58}$$

and

$$\lim_{N\to\infty} \mathbf{E}(m) = \frac{\Gamma(1-2\gamma)}{\Gamma(1-\gamma)^2\, 2^{1-2\gamma}} - \frac{1}{2} \tag{59}$$

when $\gamma < \frac{1}{2}$. When $\gamma \geqslant \frac{1}{2}$, both integrals diverge.

It is interesting to note that the expressions in (58) and (59) depend only on γ, and not on a or b. The ratio $\mathbf{E}(m)/\mathbf{E}(\mu)$, which takes the value $\log 2$ in the Dirichlet case, here varies between

$$\log 2 = 0{\cdot}69 \quad \text{and} \quad 2/\pi = 0{\cdot}64, \tag{60}$$

as γ varies between 0 and $\frac{1}{2}$. Thus, although no universality can be ascribed to the Dirichlet figure of 0·69, the stable subordinator never gives a figure very far removed for the efficiency of the heap. It will be seen in the next section that the qualitative behaviour of the p_j is very different in the two cases, so that the small discrepancy in (60) is an encouraging sign of robustness.

9. The Tail of the Distribution

For any subordinator ξ satisfying (38) and (39), the jumps in $(0, 1)$ have heights which form a non-homogeneous Poisson process governed by the measure μ. That is to say, the number of jumps whose heights $\delta\xi$ satisfy $a < \delta\xi < b$ has a Poisson distribution with mean $\mu(a, b)$, and the numbers in disjoint intervals are independent. Thus, if we write

$$m(t) = \mu(t, \infty) = \int_t^\infty h(x)\, dx, \tag{61}$$

the random variables

$$m(\delta\xi_{(j)}) \quad (j = 1, 2, \ldots) \tag{62}$$

describe the points of a Poisson process of unit rate on the positive half-line.

The strong law of large numbers therefore implies that, with probability one,

$$\lim_{j\to\infty} j^{-1} m(\delta\xi_{(j)}) = 1. \tag{63}$$

Thus the behaviour of the tail of the random distribution

$$p_{(j)} = \delta\xi_{(j)}/\xi(1)$$

is governed by the way that $m(t) \to \infty$ as $t \to 0$.

Consider first the Dirichlet case for which (bearing in mind the change of scale involved in replacing $(0, \lambda)$ by $(0, 1)$)

$$h(x) = \lambda x^{-1} \exp(-x). \tag{64}$$

For this measure,

$$m(t) = \lambda \int_t^\infty x^{-1} \exp(-x)\, dx \sim -\lambda \log t$$

as $t \to 0$, so that (63) gives

$$\lim_{j\to\infty} j^{-1}\lambda \log \delta\xi_{(j)} = -1$$

or

$$\lim_{j\to\infty} j^{-1}\log p_{(j)} = -\lambda^{-1}. \tag{65}$$

Thus in the Dirichlet case the random distribution **p** *has a tail of approximately exponential form, resembling a geometric distribution with common ratio* $\exp(-1/\lambda)$.

A word of caution is in order here. If **p** is exactly a geometric distribution with common ratio ρ, then it is easily checked that

$$\mu = 2\sum_{j=1}^{\infty} \frac{\rho^j}{1+\rho^j}, \quad m = \frac{\rho}{1-\rho}. \tag{66}$$

In particular, when ρ is near 1,

$$\frac{m}{\mu} \sim (2\log 2)^{-1} = 0{\cdot}72, \tag{67}$$

a figure slightly different from the ratio

$$\mathbf{E}(m)/\mathbf{E}(\mu) = 0{\cdot}69$$

which we have seen to be characteristic of the Dirichlet case. In other words, the "heap" is slightly more efficient for a rigidly geometric distribution than for a random distribution of approximately geometric type.

The situation is quite different if (64) is replaced by the density (54), for which

$$m(t) = a\gamma^{-1} t^{-\gamma},$$

so that (63) implies that, with probability one,

$$\lim_{j\to\infty} j^{-1}(\delta\xi_{(j)})^{-\gamma} = a^{-1}\gamma$$

or

$$\lim_{j\to\infty} j^{1/\gamma} p_{(j)} = (a\gamma^{-1})^{1/\gamma}\, \xi(1)^{-1}. \tag{68}$$

Thus in the stable case $p_{(j)}$ *decays as* $j^{-1/\gamma}$*, resembling Zipf's law.*

In fact, there is a distant relationship between the present argument and recent discussions of the theoretical basis of Zipf's law, notably those of Hill (1970, 1974a). Notice also that (68) explains why $\gamma < \frac{1}{2}$ is necessary for the finiteness of $\mathbf{E}(m)$ in (59).

The two special cases of the general construction are together sufficient to model situations in which the tail of the random distribution decays either exponentially or as an inverse power. Moreover, the key parameters (λ and γ respectively) are determined by the tail behaviour of the distribution.

10. CONDITIONED SUBORDINATORS

It is possible (though not necessarily useful) to generalize the construction of Section 7 in a number of ways, of which for the sake of illustration just one will be described. Instead of normalizing the increments as in (42) by dividing by $\xi(1)$, one can condition the subordinator so that $\xi(1) = 1$. The effect is slightly to simplify the formulae; (44) is replaced by

$$F(p_1, p_2, \ldots, p_N) = f(N^{-1}, p_1) f(N^{-1}, p_2) \ldots f(N^{-1}, p_N) / f(1, 1), \tag{69}$$

(46) by

$$f(N^{-1}, x) f(1 - N^{-1}, 1 - x) / f(1, 1), \tag{70}$$

and (47) by

$$f(N^{-1}, x) f(N^{-1}, y) f(1 - 2N^{-1}, 1 - x - y) / f(1, 1). \tag{71}$$

However, the problem of calculating expectations with respect to these quantities is of comparable difficulty with that for the original construction.

The limiting behaviour of the order statistics as $N \to \infty$ follows the now familiar pattern, but $\mathbf{P}_\mu$ must be replaced by the joint distribution $\mathbf{P}_\mu^c$ of the sequence $(\delta\xi_{(j)})$ conditional on $\xi(1) = 1$. To put it another way, $\mathbf{P}_\mu^c$ is the distribution of the sequence (δ_j) conditional on $\sum \delta_j = 1$, where

$$\delta_1 \geqslant \delta_2 \geqslant \delta_3 \geqslant \ldots \tag{72}$$

is a non-homogeneous Poisson process governed by the measure μ.

For the special case of the gamma process, (69) and (44) coincide. Hence there is no point in using this conditioning for the Dirichlet distribution, for nothing new will result.

APPENDIX 1

The Limiting Distribution of $p_{(1)}$

In this appendix we illustrate the problems of the explicit calculation of $\mathbf{P}_\mu$ (or of its special case $\mathbf{P}_\lambda$) by considering the limiting distribution of the maximum probability $p_{(1)}$. The representation of Section 9 shows that we can write

$$p_{(1)} = \delta_1 \left(\sum_{j=1}^{\infty} \delta_j \right)^{-1}, \tag{73}$$

where, under $\mathbf{P}_\mu$, the δ_j are the points (72) in a non-homogeneous Poisson process with measure μ.

Assuming as before that μ has a density h, we have

$$\mathbf{P}(\delta_1 < x) = \exp\{-m(x)\},$$

so that δ_1 has probability density

$$g(x) = h(x) \exp\{-m(x)\}. \tag{74}$$

Conditional on δ_1, the point process $(\delta_j; j \geqslant 2)$ is a Poisson process on $(0, \delta_1)$ governed by the restriction of μ to that interval. Moreover,

$$\sigma = \sum_{j=2}^{\infty} \delta_j = \lim_{\varepsilon \to 0} \sigma(\varepsilon),$$

where $\sigma(\varepsilon)$ is the sum of those δ_j with $j \geqslant 2$ and $\delta_j > \varepsilon$. The number of such δ_j has a Poisson distribution with mean $\mu(\varepsilon, \delta_1)$, and they are distributed as the order statistics of a sample from the distribution $\mu(\varepsilon, \delta_1)^{-1} \mu(\cdot)$ on (ε, δ_1). From this it is easy to compute that for $\theta > 0$,

$$\mathbf{E}[\exp\{-\theta\sigma(\varepsilon)\} \mid \delta_1] = \exp\left[-\int_\varepsilon^{\delta_1} \{1 - \exp(-\theta x)\} h(x)\, dx\right],$$

so that

$$\mathbf{E}\{\exp(-\theta\sigma) \mid \delta_1\} = \exp\left[-\int_0^{\delta_1}\{1-\exp(-\theta x)\}\,h(x)\,dx\right].$$

From (73),

$$p_{(1)} = (1+\eta)^{-1}, \tag{75}$$

where $\eta = \sigma/\delta_1$, and

$$\begin{aligned}\mathbf{E}\{\exp(-\theta\eta)\} &= \mathbf{E}[\mathbf{E}(\exp\{-(\theta/\delta_1)\,\sigma\} \mid \delta_1)]\\ &= \int_0^\infty \exp\left[-\int_0^y \{1-\exp(-\theta x/y)\}\,h(x)\,dx\right] g(y)\,dy.\end{aligned}$$

Hence, for $\theta > 0$,

$$\mathbf{E}\{\exp(-\theta\eta)\} = \int_0^\infty \exp\left[-\int_0^y \{1-\exp(-\theta x/y)\}\,h(x)\,dx - \int_y^\infty h(x)\,dx\right] h(y)\,dy. \tag{76}$$

In principle, (75) and (76) determine the distribution of $p_{(1)}$ under $\mathbf{P}_\mu$, and it will be clear how the argument can be extended to deal with the joint distribution of $(p_{(1)}, p_{(2)}, \ldots, p_{(n)})$ for any n. In practice, it is not obvious how to extract useful information from (76). However, this expression does simplify in particular cases. For example, if μ has the density (54) corresponding to the stable subordinator with exponent γ, then (76) reduces to

$$\mathbf{E}\{\exp(-\theta\eta)\} = \left[1+\gamma\int_0^1 \{1-\exp(-\theta x)\}\,x^{-\gamma-1}\,dx\right]^{-1}. \tag{77}$$

APPENDIX 2

Remarks on the Ferguson–Dirichlet distribution

Ferguson (1973) makes a far-reaching generalization of the Dirichlet distribution (2), which gives a random probability distribution on an arbitrary measurable space $\mathfrak{X}$. Let α be any finite measure on $\mathfrak{X}$. Then he describes a random probability measure $\mathfrak{p}$ on $\mathfrak{X}$ as a *Dirichlet process* on $(\mathfrak{X}, \alpha)$ if, whenever $A_1, A_2, \ldots, A_N$ form a dissection of $\mathfrak{X}$, the random variables $\mathfrak{p}(A_1), \mathfrak{p}(A_2), \ldots, \mathfrak{p}(A_N)$ have the joint distribution (2), with $\alpha_j = \alpha(A_j)$.

Suppose now that $\mathfrak{m}$ is a random measure on $\mathfrak{X}$, with the property that

(i) the values of $\mathfrak{m}$ on disjoint sets are independent, and
(ii) the distribution of $\mathfrak{m}(A)$ is given by (26), with $\alpha = \alpha(A)$.

Then, as in Section 5, it is clear that

$$\mathfrak{p}(A) = \mathfrak{m}(A)/\mathfrak{m}(\mathfrak{X}) \quad (A \subseteq \mathfrak{X}) \tag{78}$$

defines a Dirichlet process on $(\mathfrak{X}, \alpha)$.

Random measures satisfying condition (i) have been studied by Kingman (1967), and rather more rigorously by Tortrat (1968) and Morando (1969). The main result is that (excluding deterministic components) every such measure is equivalent to one constructed as follows. First construct a Poisson process Π on the product space $\mathfrak{X} \times (0, \infty)$ such that the mean number of points of Π in a subset B of the product space is $\gamma(B)$. Here γ is a measure, and if it is σ-finite there is no difficulty in constructing Π (Kingman, 1967). Define $\mathfrak{m}(A)$ as the sum of the values of y for all points (x, y) of Π for which x lies in A:

$$\mathfrak{m}(A) = \sum_{\substack{(x,y)\in\Pi \\ x\in A}} y. \tag{79}$$

Then $\mathfrak{m}$ is a random measure satisfying (i), and the distribution of $\mathfrak{m}(A)$ is determined from γ by the formula

$$\mathbf{E}[\exp\{-\theta\mathfrak{m}(A)\}] = \exp\left[-\int_{(0,\infty)}\{1-\exp(-\theta y)\}\gamma(A\times dy)\right]. \tag{80}$$

In particular, if we choose γ to be the product measure

$$\gamma = \alpha\times\beta, \tag{81}$$

where β is defined by

$$\beta(dy) = y^{-1}\exp(-y)\,dy, \tag{82}$$

then $\mathfrak{m}(A)$ has the gamma distribution (26), with $\alpha = \alpha(A)$, and hence (ii) is also satisfied. Notice that the measure (79) is purely atomic, its atoms being the points of the random set

$$\{x\in X;\ (x,y)\in\Pi \text{ for some } y\}. \tag{83}$$

Thus the Dirichlet measure (78) is also purely atomic, a result proved by Ferguson (whose proof is in effect the specialization of the present argument to his particular case).

What either Ferguson's analysis or mine establishes is that, given α (or more generally γ) there is a corresponding Dirichlet process (or random measure) which is purely atomic. Neither proves the stronger assertion that any such process (or measure) is purely atomic with probability one. But in a sequel to Ferguson's paper, Blackwell (1973) establishes the stronger assertion under the (weak) assumption that the σ-algebra of measurable subsets of $\mathfrak{X}$ is separable. His argument generalizes without difficulty from Dirichlet processes (and therefore random measures satisfying (i) and (ii)) to quite general random measures satisfying (i), subject always to the separability condition.

REFERENCES

BLACKWELL, D. (1973). Discreteness of Ferguson selections. *Ann. Statist.*, **1**, 356–358.

BURVILLE, P. J. (1974). HEAPS: a concept in optimization. *J. Inst. Math. Appl.* **13**, 263–278.

BURVILLE, P. J. and KINGMAN, J. F. C. (1973). On a model for storage and search. *J. Appl. Prob.*, **10**, 697–701.

ERICSON, W. A. (1969). Subjective Bayesian models in sampling finite populations. *J. R. Statist. Soc.* B, **31**, 195–233.

FERGUSON, T. S. (1973). A Bayesian analysis of some nonparametric problems. *Ann. Statist.*, **1**, 209–230.

GOOD, I. J. (1967). A Bayesian significance test for multinomial distributions. *J. R. Statist. Soc.* B, **29**, 399–431.

HENDRICKS, W. J. (1972). The stationary distribution of an interesting Markov chain. *J. Appl. Prob.*, **9**, 231–233.

HILL, B. M. (1970). Zipf's law and prior distributions for the composition of a population. *J. Amer. Statist. Ass.*, **65**, 1220–1232.

—— (1974a). The rank-frequency form of Zipf's law. *J. Amer. Statist. Ass.*, **69** (in the press).

KINGMAN, J. F. C. (1967). Completely random measures. *Pacific J. Math.* **21**, 59–78.

MORANDO, P. (1969). Mesures aléatoires. *Sém. Prob. Univ. Strasbourg*, III, 190–229.

PLACKETT, R. L. (1969). Stochastic models of capital investment. *J. R. Statist. Soc.* B, **31**, 1–28.

TORTRAT, A. (1968). Processus indéfiniment divisibles et mesures aléatoires. (Lecture notes, University of Paris.)