POISSON PROCESSES REVISITED

J.F.C. Kingman

In Memoriam Kazimierz Urbanik

Summary The thesis of this paper is that a good basis for defining Poisson processes on a general state space is to assume that the mean measure satisfies a simple *bisection property*, that every set of finite measure can be divided into two disjoint subsets of equal measure. This assumption is weaker than those usually made, and leads to simple and concrete proofs of the basis results. As an illustration, a very general version of Rényi's characterisation theorem is proved. The paper also gives a straightforward account of the Poisson-Dirichlet distribution.

2000 Subject Classification 60D05, 28A12

Keywords Poisson process, bisection property, Rényi's characterisation, Poisson-Dirichlet distribution

1 Poisson processes on general spaces

In my 1993 book [6], I offered a treatment of the theory of Poisson processes in which they were regarded as random countable subsets of a state space about which only minimal assumptions were made. The point of considering very general state spaces was to avoid special considerations such as topology or ordering, which obscure the essential simplicity of the theory.

Experience in using the book for teaching postgraduate courses has shown that this general approach is sound. Students develop an intuition in which they think of the state space as the plane $\mathbb{R}^2$, but realise that the arguments apply much more generally. They then have no difficulty coping with Poisson processes on, for instance, complicated manifolds of the sort that arise in stochastic geometry [5].

I have however also come to realise that the particular assumptions made in [6] are clumsy and lack intuitive appeal. They also lead to unnecessarily complex proofs, involving quite subtle uses of Fubini's theorem. There is a better way, which it is the purpose of this paper to explain.

The state space of which the Poisson process is to be a random countable subset is a quite general measurable space S. That is to say, S is equipped with a non-empty family of subsets called measurable sets, and this family is closed

under the formation of complements, countable unions and intersections. Let μ be a (positive) measure on S. For simplicity, μ will be assumed σ-finite, although this is not strictly necessary. A Poisson process on S with mean measure μ is then defined to be a random countable subset $\Pi \subseteq S$ such that, if $N(A)$ is the number of points of Π in the measurable set $A \subseteq S$, then
(i) $N(A)$ is a random variable having the Poisson distribution with mean $\mu(A)$, and
(ii) for disjoint $A_1, A_2, \ldots, A_k$, the random variables $N(A_1), N(A_2), \ldots, N(A_k)$ are independent.

Such a random set needs to be defined on some probability space $(\Omega, \mathcal{F}, \mathbb{P})$ (and as usual the probability measure $\mathbb{P}$ is assumed to be complete), so that Π is a function from Ω into the set of all countable subsets of S, and $N(A)$ is an $\mathcal{F}$-measurable function from Ω into $\{0, 1, 2, \ldots, \infty\}$. Condition (i) means of course that, if $0 < \mu(A) < \infty$,

$$\mathbb{P}\{N(A) = n\} = \mu(A)^n e^{-\mu(A)}/n! \tag{1.1}$$

for $n = 0, 1, 2, \ldots$. If $\mu(A) = 0$ it means that

$$\mathbb{P}\{N(A) = 0\} = 1\,, \tag{1.2}$$

while if $\mu(A) = \infty$ it is to be read as

$$\mathbb{P}\{N(A) = \infty\} = 1\,. \tag{1.3}$$

Condition (ii) need only be verified when $0 < \mu(A_j) < \infty$ for $j = 1, 2, \ldots, k$.

In order to prove the existence of Π and to develop its properties, some mild condition must be imposed on S and μ. In [6] it is assumed that the diagonal

$$D = \{(x, x); x \in S\} \tag{1.4}$$

is a measurable subset of the product space $S \times S$. This implies that every singleton $\{x\}$ is measurable in S, and the fact that $N(\{x\}) \leqslant 1$ then requires us to assume that

$$\mu\{x\} = 0 \qquad (x \in S)\,, \tag{1.5}$$

that μ has no point atoms. By Fubini's theorem, this is equivalent to the statement that

$$(\mu \times \mu)(D) = 0\,, \tag{1.6}$$

where $(\mu \times \mu)$ is the uncompleted product measure on $S \times S$.

These conditions are quite easy to check in particular cases, but they cannot be said to be natural or transparent. A much better approach will be described in the next section.

2 The bisection property

In this alternative approach no assumptions at all are made about the measurable space S. The measure μ is said to have the *bisection property* if, for any measurable $A \subseteq S$ with $\mu(A) < \infty$, there exists a measurable $B \subseteq A$ with

$$\mu(B) = \tfrac{1}{2}\mu(A)\,. \tag{2.1}$$

This implies (1.5) if $\{x\}$ is measurable, but is in general stronger. In fact, Halmos ([4], sect. 41) has shown that the bisection property is equivalent to a more general version of non-atomicity of μ.

However, the Halmos result lies rather deep, and is not really relevant to the theory or application of Poisson processes. The bisection property has three great pedagogical advantages:
(i) it is concrete and easy to visualise,
(ii) it is easy to check in particular cases, and involves very little loss of generality, and
(iii) it leads to straightforward proofs.
These assertions will be justified below.

The easiest way to prove the bisection property for a measure μ is to construct a *cheesewire*. (The name comes from the device used to cut a measured portion of cheese in old-fashioned shops.) A cheesewire for μ is a measurable function $f : S \to \mathbb{R}$ with the property that, for any $\xi \in \mathbb{R}$, the measurable set

$$f^{-1}\{\xi\} = \{x \in S; f(x) = \xi\} \tag{2.2}$$

has

$$\mu\left(f^{-1}\{\xi\}\right) = 0\,. \tag{2.3}$$

If μ admits such a function, and $A \subseteq S$ has $\mu(A) < \infty$, the function $g : \mathbb{R} \to \mathbb{R}$ defined by

$$g(\xi) = \mu\{x \in A; f(x) \leqslant \xi\} \tag{2.4}$$

is monotone increasing, with

$$\lim_{\xi \to -\infty} g(\xi) = 0\,, \lim_{\xi \to \infty} g(\xi) = \mu(A)\,. \tag{2.5}$$

Because it is monotone, g has only jump discontinuities, and such a discontinuity at ξ would contradict (2.2). Thus g is continuous, and takes every value strictly between 0 and $\mu(A)$. In particular, there exists ξ with

$$g(\xi) = \tfrac{1}{2}\mu(A)\,, \tag{2.6}$$

and

$$B = \{x \in A\,; f(x) \leqslant \xi\} \tag{2.7}$$

then satisfies (2.1).

Thus the existence of a cheesewire implies the bisection property (and it is an interesting exercise to prove the converse for σ-finite μ). It is usually possible to write down a cheesewire by inspection of μ. For instance, if $S = \mathbb{R}^d$ (with the usual measurable structure) and μ has a density with respect to Lebesgue measure, any coordinate function is a cheesewire. If S is a manifold embedded in $\mathbb{R}^d$, it may be necessary to choose f more carefully to cut across S, but I know of no significant case in which the construction of f presents any real difficulty.

Nevertheless, it is important to understand just how strong is the restriction imposed by the bisection property, and this is an issue to which we shall return in Section 5.

3 Uses of the bisection property

Let us now test assertion (iii), that the bisection property leads to straightforward proofs of the basic theorems about Poisson processes. The first difficult proof encountered by the reader of [6] is that of the Disjointness Lemma. This states that if Π_1 and Π_2 are independent Poisson processes on the same space S, and if their mean measures μ_1 and μ_2 are both finite, then they are disjoint with probability 1:

$$\mathbb{P}\{\Pi_1 \cap \Pi_2 = \emptyset\} = 1\,. \tag{3.1}$$

If μ_1 and μ_2 both have the bisection property, proceed as follows. Let $n = 2^\nu$ be any power of 2. Use (2.1) ν times to express S as a disjoint union of measurable sets $S_1, S_2, \ldots, S_n$ with

$$\mu_1(S_i) = n^{-1}\mu_1(S) \qquad (i = 1, 2, \ldots, n)\,. \tag{3.2}$$

Then apply (2.1) to μ_2 to express each S_i as a disjoint union of measurable S_{ij} with

$$\mu_2(S_{ij}) = n^{-1}\mu_2(S_i) \quad (j = 1, 2, \ldots, n)\,. \tag{3.3}$$

Observe that

$$\{\omega; \Pi_1 \cap \Pi_2 \neq \emptyset\} \subseteq E_n\,,$$

where

$$E_n = \bigcup_{i,j=1}^{n} \{\omega; N_1(S_{ij}) \geqslant 1\,,\ N_2(S_{ij}) \geqslant 1\}$$

belongs to $\mathcal{F}$ and has probability

$$\begin{aligned}
\mathbb{P}(E_n) &\leqslant \sum_{i,j=1}^{n} \mathbb{P}\{N_1(S_{ij}) \geqslant 1\,,\ N_2(S_{ij}) \geqslant 1\} \\
&= \sum_{i,j=1}^{n} \mathbb{P}\{N_1(S_{ij}) \geqslant 1\}\,\mathbb{P}\{N_2(S_{ij}) \geqslant 1\} \\
&\leqslant \sum_{i,j=1}^{n} \mu_1(S_{ij})\mu_2(S_{ij}) \\
&= n^{-1} \sum_{i,j=1}^{n} \mu_1(S_{ij})\mu_2(S_i) \\
&= n^{-1} \sum_{i=1}^{n} \mu_1(S_i)\mu_2(S_i) \\
&= n^{-2} \sum_{i=1}^{n} \mu_1(S)\mu_2(S_i) \\
&= n^{-2}\mu_1(S)\mu_2(S)\,.
\end{aligned}$$

Letting $n \to \infty$ shows that

$$\mathbb{P}\left\{\bigcap_{n=1}^{\infty} E_n\right\} = 0\,,$$

which implies (3.1) since $\mathbb{P}$ is complete.

A very similar argument proves the Mapping Theorem, whose proof in [6] again involves a subtle Fubini argument. This theorem concerns a function $f : S \to S^*$, where S^* is another measurable space and f is measurable. It gives conditions to ensure that, if Π is a Poisson process on S with σ-finite mean measure μ, then

$$f(\Pi) = \{f(x); x \in \Pi\} \tag{3.4}$$

is a Poisson process on S^* whose mean measure μ^* is given by

$$\mu^*(A) = \mu(f^{-1}(A)) \quad (A \subseteq S^*)\,. \tag{3.5}$$

This will be true if, with probability 1, no two points of Π map under f into the same point of S^*.

This will be the case if μ^* has the bisection property. To see this, suppose that μ^* has that property, and let $A \subseteq S^*$ be measurable with $\mu^*(A) < \infty$. For $n = 2^\nu$, dissect A into disjoint $A_1, A_2, \ldots, A_n$ with

$$\mu^*(A_i) = n^{-1}\mu^*(A) \quad (i = 1, 2, \ldots, n)\,. \tag{3.6}$$

If two points of Π map into the same point of A then there is a value of i with two points of $f(\Pi)$ in A_i, and so

$$N[f^{-1}(A_i)] \geqslant 2\,. \tag{3.7}$$

Now the probability that a Poisson random variable with mean μ is 2 or more is at most $\frac{1}{2}\mu^2$, so that the probability that (3.6) holds for some i is at most

$$\begin{aligned}
&\sum_{i=1}^{n} \mathbb{P}\left(\left\{N\left[f^{-1}(A_i)\right] \geqslant 2\right\}\right) \\
&\leqslant \tfrac{1}{2}\textstyle\sum_{i=1}^{n} \mu\left[f^{-1}(A_i)\right]^2 \\
&= \tfrac{1}{2}\textstyle\sum_{i=1}^{n} n^{-2}\mu^*(A)^2 \\
&= \mu^*(A)^2/2n\,.
\end{aligned}$$

Letting $n \to \infty$ gives the required result.

These two examples should be enough to show the power of the bisection property, but there is a third which is glossed over in [6]. In the proof of the Existence Theorem, it is necessary to know that a number of independent random variables take distinct values with probability one. An obvious application of the bisection property deals with this too.

4 Poisson random measures

It might be argued that the complications of the last section could be avoided by working with the integer-valued random measure $N(\cdot)$ rather than the random set Π. Thus $N(\cdot)$ is a random measure on S whose values $N(A)$ have Poisson distributions, and are independent on disjoint sets.

In this approach the mean measure

$$\mu(A) = \mathbb{E}\{N(A)\} \tag{4.1}$$

may have point atoms. If $\{x\}$ is measurable and $\mu\{x\} > 0$, the variable $N(\{x\})$ is greater than 1 with positive probability. The proofs of results like the Mapping Theorem are almost trivial.

There comes a point, however, when one needs to know whether or not $N(\cdot)$ does have multiple points. Under what conditions, in other words, is it true that

$$\mathbb{P}\left\{N(\{x\}) \leqslant 1 \text{ for all } x \in S\right\} = 1\,? \tag{4.2}$$

Arguments just like those of Section 3 show easily that a sufficient condition is that μ be σ-finite and have the bisection property.

The two approaches are mathematically (but not pedagogically) equivalent, and a choice between them is a matter of taste. It is true that random sets with multiple points arise in applied probability (think of queues with batch arrivals) but the multiplicities are not usually Poisson distributed. It is better to handle multiple points by means of the theory of marked Poisson processes, allowing more general distributions.

5 The force of the bisection property

It is natural to ask how the bisection property compares with the conditions assumed in [6], measurability of the diagonal and absence of point atoms. If the bisection property were more restrictive, this might outweigh its greater transparency. However, the opposite is the case, as the next theorem shows.

Theorem 1 *Let μ be a σ-finite measure on the measurable space S. Then μ has the bisection property if and only if the diagonal $D \subseteq S \times S$ has zero outer measure for the product measure $\mu \times \mu$. In particular, if D is a measurable subset of $S \times S$ and $\mu\{x\} = 0$ for all x, then μ has the bisection property.*

Proof Suppose first that μ is σ-finite and has the bisection property. Dissect S into disjoint $S_1, S_2, \ldots$ with

$$\mu(S_i) = m_i < \infty \quad (i = 1, 2, \ldots)\,. \tag{5.1}$$

For any $\epsilon > 0$, let n_i be a power of 2 with

$$n_i > \epsilon^{-1} m_i^2 2^i\,, \tag{5.2}$$

and use the bisection property to dissect S_i into disjoint $S_{ij}(j = 1, 2, \ldots, n_i)$ with

$$\mu(S_{ij}) = m_i/n_i \ (j = 1, 2, \ldots, n_i)\,. \tag{5.3}$$

Then, since

$$D \subseteq \bigcup_{i=1}^{\infty} \bigcup_{j=1}^{n_i} (S_{ij} \times S_{ij})\,, \tag{5.4}$$

the $(\mu \times \mu)$ outer measure of D is at most

$$\begin{aligned}
&\sum_{i=1}^{\infty}\sum_{j=1}^{n_i}(\mu \times \mu)(S_{ij} \times S_{ij}) \\
&= \sum_{i=1}^{\infty}\sum_{j=1}^{n_i}\mu(S_{ij})^2 \\
&= \sum_{i=1}^{\infty}\sum_{j=1}^{n_i}(m_i/n_i)^2 \\
&= \sum_{i=1}^{\infty} m_i^2/n_i \\
&< \sum_{i=1}^{\infty} \epsilon 2^{-i} = \epsilon\,.
\end{aligned}$$

Thus the outer measure of D is 0.

To prove the converse, suppose that μ does not have the bisection property. The result of Halmos cited above shows that there is a measurable $A \subseteq S$ with

$$0 < \mu(A) = m < \infty\,, \tag{5.5}$$

and that every measurable $B \subseteq A$ has either $\mu(B) = 0$ or $\mu(B) = \mu(A) = m$.

Consider these families of subsets of $A \times A$:
$\mathcal{A}$ consists of all sets of the form

$$(N_1 \times A)_\cup (A \times N_2) \tag{5.6}$$

with N_1 and N_2 measurable and

$$\mu(N_1) = \mu(N_2) = 0\,. \tag{5.7}$$

Clearly $\mathcal{A}$ is closed under countable unions, and every set in $\mathcal{A}$ has $(\mu \times \mu)$ measure 0.
$\mathcal{B}$ consists of all subsets of A (measurable or not) which are contained in some member of $\mathcal{A}$; it is closed under countable unions and intersections, and every member of $\mathcal{B}$ has $(\mu \times \mu)$ outer measure 0.
$\mathcal{C}$ consists of all sets in $\mathcal{B}$ and all sets whose complements in $A \times A$ are in $\mathcal{B}$. It is easy, but not quite trivial, to check that $\mathcal{C}$ is a σ-algebra, and of course

$$\mathcal{A} \subseteq \mathcal{B} \subseteq \mathcal{C}\,. \tag{5.8}$$

If B and C are measurable subsets of A, then

$$B \times C \subseteq (B \times A)_\cap (A \times C) \tag{5.9}$$

is in $\mathcal{B}$ unless $\mu(B)$ and $\mu(C)$ are both non-zero. If this is so, $\mu(B) = \mu(C) = m$, and the complements N_1 and N_2 of B and C in A satisfy (5.7). The complement of $B \times C$ in $A \times A$ is then given by (5.6), so that this complement is in $\mathcal{B}$. Hence $\mathcal{C}$ contains $B \times C$ for all measurable B, C and since it is a σ-algebra it contains every measurable subset of $A \times A$.

Now suppose that E is any measurable subset of $A \times A$ which contains

$$D_A = \{(x, x); x \in A\}\ . \tag{5.10}$$

Then $E \in \mathcal{C}$, and therefore either E or its complement belongs to $\mathcal{B}$. If $E \in \mathcal{B}$, there are sets N_1 and N_2 satisfying (5.4) with

$$D_A \subseteq E \subseteq (N_1 \times A)_\cap (A \times N_2)\,. \tag{5.11}$$

This implies that $\mu(A) = 0$, contradicting (5.5).

Thus the complement of E must belong to $\mathcal{B}$, so that

$$(\mu \times \mu)(E) = (\mu \times \mu)(A \times A) = m^2\,. \tag{5.12}$$

Since this holds for every measurable cover of D_A, the $(\mu \times \mu)$ outer measure of D_A is $m^2 > 0$, which shows that $D \subseteq D_A$ has non-zero outer measure, as required.

To complete the proof, note that if D is measurable and μ is σ-finite, then the bisection property is equivalent to (1.6), which is equivalent to (1.5) by Fubini's theorem.

6 Rényi's theorem

In [9] Rényi proved a very surprising result about Poisson processes in $\mathbb{R}$. Let Ξ be a random locally finite subset of $\mathbb{R}$ with the property that the number of points of Ξ in any finite union A of bounded intervals has a Poisson distribution with mean $\mu(A)$, μ being a non-atomic measure finite on bounded intervals. No independence assumption is made, but Rényi proves that Ξ is a Poisson process. The force of this theorem is emphasised by a counterexample due to Moran [8], which shows that 'finite union of bounded intervals' cannot be replaced by 'bounded interval'.

Section 3.4 of [6] generalises Rényi's result to $\mathbb{R}^d$, but the result is in fact much more general, depending only on the bisection property of μ.

Theorem 2 *Let μ be a σ-finite measure on S with the bisection property. Let Ξ be a random subset of S, and denote by $N(A) \leqslant \infty$ the number of points of Ξ in the measurable subset A of S. If*

$$\mathbb{E}\{N(A)\} = \mu(A) \ , \ \mathbb{P}\{N(A) = 0\} = e^{-\mu(A)} \tag{6.1}$$

for every A with $\mu(A)$ finite, then Ξ is a Poisson process with mean measure μ.

Proof Let $\mathcal{A}$ be any family of subsets of S with finite measure, such that any two sets in $\mathcal{A}$ are disjoint. Then, by (6.1), if $A_1, A_2, \ldots, A_n$ belong to $\mathcal{A}$,

$$\begin{aligned} \mathbb{P}\{N(A_1) &= N(A_2) = \ldots = N(A_n) = 0\} \\ &= \mathbb{P}\left\{N\left(\cup_{r=1}^n A_r\right) = 0\right\} \\ &= \exp\left[-\mu\left(\cup_{r=1}^n A_r\right)\right] = \exp\left[-\sum_{r=1}^n \mu(A_r)\right] \\ &= \prod_{r=1}^n \mathbb{P}\left\{N(A_r) = 0\right\} . \end{aligned}$$

Hence the events

$$E(A) = \{N(A) = 0\} = \{\Xi \cap A = \emptyset\} \ , \tag{6.2}$$

as A runs over $\mathcal{A}$, are independent.

Now fix a set A with $\mu(A) = m < \infty$. Use the bisection property to divide A into disjoint sets A_0, A_1 with

$$\mu(A_0) = \mu(A_1) = \tfrac{1}{2}m \, , \tag{6.3}$$

and think of A_0 and A_1 as the 'children' of A. Divide each of A_0 and A_1 into disjoint 'grandchildren' of measure $\frac{1}{4}m$, so that

$$A_0 = A_{00} \cup A_{01} \ , \ A_1 = A_{10} \cup A_{11} \, , \tag{6.4}$$

$$\mu(A_{00}) = \mu(A_{01}) = \mu(A_{10}) = \mu(A_{11}) = \tfrac{1}{4}m \, . \tag{6.5}$$

Continue in this way, so that the sets of the kth generation, labelled by strings of k binary digits, each have measure $2^{-k}m$.

The 2^k sets $A_{\dots}$ of the kth generation are pairwise disjoint, and so the events

$$E(A_{\dots}) \tag{6.6}$$

are independent, with equal probabilities

$$\exp(-2^{-k}m)\,. \tag{6.7}$$

Hence $N_k(A)$, defined as the number of the kth generation sets that contain points of Ξ, has a binomial distribution with

$$\mathbb{E}\{N_k(A)\} = 2^k\left[1 - \exp\left(-2^{-k}m\right)\right]\,. \tag{6.8}$$

Clearly

$$N_k(A) \leqslant N_{k+1}(A) \leqslant N(A)\,, \tag{6.9}$$

so that

$$N_\infty(A) = \lim_{k\to\infty} N_k(A) \tag{6.10}$$

exists, and

$$\mathbb{E}\{N_\infty(A)\} = \lim_{k\to\infty} \mathbb{E}\{N_k(A)\} = m \tag{6.11}$$

by (6.8). Since $N_\infty(A)$ is the limit of binomial variables $N_k(A)$, it has the Poisson distribution with mean $m = \mu(A)$. The random variable

$$N(A) - N_\infty(A) \tag{6.12}$$

is non-negative by (6.9) and has zero expectation by (6.1), so that

$$\mathbb{P}\{N(A) = N_\infty(A)\} = 1\,, \tag{6.13}$$

and therefore $N(A)$ has the Poisson distribution with mean $\mu(A)$.

Now let $A_1, A_2, \dots, A_n$ be disjoint sets of finite measure. Carry out the repeated bisection for each A_r, and for a fixed value of k let $\mathcal{A}_k$ consist of all the kth generation subsets of all the A_r. The sets $A_{\dots}$ in $\mathcal{A}_k$ are pairwise disjoint, and so that events $E(A_{\dots})$ are independent. Thus, for any fixed value of k, the random variables $N_k(A_r)(r = 1, 2, \dots, n)$ are independent. It follows from (6.10) that the $N_\infty(A_r)$ are independent, and (6.13) shows that the $N(A_r)$ are independent. This completes the proof.

7 Separating bisectors

Theorem 2 contains Rényi's theorem as a special case. It is stronger not only because the state space S is quite general, but also because it only uses the two properties (6.1), rather than the full panoply of the Poisson distribution (1.1). However, the theorem proved in Section 3.4 of [6] is still stronger (when $S = \mathbb{R}^d$), because it assumes only that

$$\mathbb{P}\{N(A) = 0\} = e^{-\mu(A)}\,. \tag{7.1}$$

An examination of the proof of Theorem 2 shows that the only point where we need the other condition of (6.1) is in proving that $N_\infty(A) = N(A)$ with probability 1, and this is achieved in a different way in [6]. Before explaining this for general S, consider what can be said if $N_\infty(A)$ can be strictly less than $N(A)$ with positive probability.

If A is a set of finite measure m, the repeated bisection of A described in the proof of Theorem 2 defines a function

$$\psi : A \to \Omega = \{0,1\}^\infty \tag{7.2}$$

from A into the space Ω of infinite binary sequences

$$\omega = (\omega_1, \omega_2, \ldots, \omega_k, \ldots) \ , \tag{7.3}$$

such that the kth generation set containing x has label

$$\psi_1(x)\,, \psi_2(x)\,, \ldots\,, \psi_k(x)\,. \tag{7.4}$$

Because each kth generation set has measure $2^{-k}m$, the restriction μ_A of μ to A induces under ψ the measure

$$\mu_A \psi^{-1} = m\beta\,, \tag{7.5}$$

where β is the Bernoulli (coin tossing) probability measure on Ω which makes the ω_k independent random variables taking the values 0 and 1 with probability $\frac{1}{2}$.

The random subset $\Xi_\cap A$ maps into a random subset

$$\Pi = \{\psi(x); x \in \Xi_\cap A\} \tag{7.6}$$

of Ω, and it is easy to see that, for $B \subseteq \Omega$, $N_\infty(\psi^{-1}B)$ is the number of points of Π in B. The argument of Theorem 2 shows that, if (7.1) holds, then Π is a Poisson process on Ω with mean measure $m\beta$. If $N_\infty(A) < N(A)$, it is because distinct points of $\Xi_\cap A$ map into the same point of Ω.

The reason why this possibility can be excluded when $S = \mathbb{R}^d$ is that, at least if μ has a density, it is always possible to carry out the bisection of A so that the function ψ is an injection. One has only to use as cheesewires the coordinate functions in rotation, the kth bisection using the cheesewire

$$(x_1, x_2, \ldots, x_d) \longmapsto x_{r(k)} \ , \tag{7.7}$$

where $r(k)$ is the remainder of k when divided by d (or d if d divides k).

More generally, Theorem 2 holds under the weaker assumption (6.1) if there is an injection (7.2) satisfying (7.5). If this holds for all A of finite measure, we say that (S, μ) admits *separating bisectors*. Thus $(\mathbb{R}^d, \mu)$ admits separating bisectors if μ is finite on bounded sets and has a density with respect to Lebesque measure.

The concept has other applications. For instance, Section 8.3 of [6] contains a version of an argument of Blackwell [3] for proving that certain random measures are purely atomic with probability 1. Blackwell makes topological assumptions which imply the existence of an injection (7.2), and then uses the sequential structure of Ω to bring to bear a technique from the theory of games. The proof is considerably simplified if one can control the induced measure under ψ, and if this is a multiple of β the simplification is optimal. Thus the concept of separating bisectors is a powerful one.

Although the proofs are similar, the contents of the theorems in [3] and [6] are distinct. Blackwell is concerned with a Dirichlet random measure on S, which is a random probability measure P such that, for disjoint $A_1, A_2, \ldots, A_n$ with union S, the vector

$$(P(A_1)\,,\ P(A_2)\,, \ldots, P(A_n)) \tag{7.8}$$

has the Dirichlet distribution with density

$$\frac{\Gamma(\alpha_1+\alpha_2+\ldots+\alpha_n)}{\Gamma(\alpha_1)\Gamma(\alpha_2)\ldots\Gamma(\alpha_n)} x_1^{\alpha_1-1} x_2^{\alpha_2-1} \ldots x_n^{\alpha_n-1} \tag{7.9}$$

on the simplex

$$x_1, x_2, \ldots, x_n \geqslant 0\,,\ x_1+x_2+\ldots+x_n = 1\,. \tag{7.10}$$

The parameters α_r are given by

$$\alpha_r = \mu(A_r)\,, \tag{7.11}$$

where μ is a finite measure on S.

On the other hand, Section 8.3 of [6] deals with completely random measures, for which the values on disjoint sets are independent. However, Blackwell's theorem can be brought within this framework although Dirichlet measures are not completely random. An argument used for a different purpose in [7] shows that, if Z is independent of $P(\cdot)$ and has a gamma distribution with parameter $\mu(S)$, then

$$ZP(\cdot) \tag{7.12}$$

is a completely random measure.

8 The Poisson-Dirichlet distribution

Chapter 9 of [6] is an introduction to the theory of the Poisson-Dirichlet distribution, a theory which has developed significantly since 1993. Arratia, Barbour and Tavaré give in [1] an exhaustive account of the theory and its applications. A weakness of [6] is that the fundamental properties of the distribution are proved using deep results from the Lévy theory of subordinators. In fact, the elementary theory of Poisson processes is all that is needed.

Theorem 3 *Let* Π_θ *be a Poisson process on* $(0,\infty)$ *with density*

$$\theta y^{-1} e^{-y}\,, \tag{8.1}$$

where θ is a positive constant. Then the points of Π_θ may be written in descending order as

$$Y_1 > Y_2 > Y_3 > \ldots > 0 \tag{8.2}$$

and $Y_r \to 0$ with probability one. The random variable

$$Z = \sum_{r=1}^{\infty} Y_r \tag{8.3}$$

is almost certainly finite, and has the gamma distribution with probability density

$$\zeta^{\theta-1} e^{-\zeta}/\Gamma(\theta) \qquad (\zeta > 0)\,. \tag{8.4}$$

The random variables

$$X_r = Y_r/Z \tag{8.5}$$

are independent of Z and satisfy

$$X_1 > X_2 > \ldots > 0\,, \; \sum_{r=1}^{\infty} X_r = 1\,. \tag{8.6}$$

The joint distribution of the infinite sequence

$$(X_1, X_2, \ldots) \tag{8.7}$$

is the *Poisson-Dirichlet distribution* $\mathcal{PD}(\theta)$. It first arose [10] as a limiting form of the Dirichlet distribution (7.9), when n is large and α_j small. The only difficult aspect of Theorem 3 is the independence property, but the advantage of the proof below is that it also establishes the limit theorem in full generality.

Theorem 4 *Suppose that, for any $n \geqslant 1$, the random variables $\xi_{n1}, \xi_{n2}, \ldots, \xi_{nn}$ have joint distribution of the Dirichlet form (7.9), with parameters $\alpha_{n1}, \alpha_{n2}, \ldots, \alpha_{nn}$. Suppose that, as $n \to \infty$,*

$$\sum_{r=1}^{n} \alpha_{nr} \to \infty \in (0, \infty) \tag{8.8}$$

and that

$$\max\,(\alpha_{nr}, r = 1, 2, \ldots, n) \to 0\,. \tag{8.9}$$

Denote by X_{nr} the rth largest of the ξ_{nr}. Then the joint distribution of the random sequence

$$(X_{n1}, X_{n2}, \ldots, X_{nn}, 0, 0, \ldots) \tag{8.10}$$

converge as $n \to \infty$ to those of $\mathcal{PD}(\theta)$.

To prove these two theorems, we first construct a Poisson process Π on the positive quadrant

$$S = \{(u, v); u, v > 0\} \tag{8.11}$$

of $\mathbb{R}^2$ whose density

$$u^{-1} e^{-u} \tag{8.12}$$

depends only on u. For any $\theta > 0$, the Mapping Theorem [6] shows that

$$\Pi_\theta = \{U; (U,V) \in \Pi\,,\ V < \theta\} \tag{8.13}$$

is a Poisson process on $(0,\infty)$ whose density is given by (8.1). This density is, for any $\epsilon > 0$, integrable on (ϵ,∞) but not on $(0,\epsilon)$, so that Π_θ has a limit point at 0 but not at ∞. Its points can therefore be written in the form (8.2), and $Y_r \to 0$ with probability 1. We can and shall take the process Π_θ of Theorem 3 to have been constructed in this way, so that Y_r is the rth largest of the points U, where (U,V) runs over those points of Π with $V < \theta$.

For any $0 \leqslant a < b$, consider the random variable

$$W(a,b) = \sum_{(U,V)\in\Pi a<V<b} U\,. \tag{8.14}$$

Its distribution can be calculated by Campbell's Theorem [6]: for $t \geqslant 0$,

$$\begin{aligned}\mathbb{E}\left\{e^{-tW(a,b)}\right\} &= \exp\left\{\int_0^\infty \int_a^b \left(e^{-tu}-1\right) u^{-1} e^{-u} du dv\right\} \\ &= (1+t)^{-(b-a)}\,.\end{aligned}$$

This shows that $W(a,b)$ is finite with probability one, and has the gamma distribution (8.4) with parameter $(b-a)$. Moreover, the independence property of Π means that the $W(a,b)$ for disjoint intervals (a,b) are independent random variables.

In particular, for fixed $\theta > 0$ and any $n \geqslant 2$, the variables

$$\eta_{nr} = W[(r-1)\theta n^{-1}, r\theta n^{-1}], (r = 1,2,\ldots,n) \tag{8.15}$$

are independent with the same gamma distribution with parameter θn^{-1}, and

$$\sum_{r=1}^{a} \eta_{nr} = W(\theta)\,. \tag{8.16}$$

A well known fact, easily proved by change of variables, is that the variables

$$\xi_{nr} = \eta_{nr}/W(\theta) \qquad (r = 1,2,\ldots,n) \tag{8.17}$$

are independent of $W(\theta)$ and have joint distribution of Dirichlet form (7.9) in which all the parameters α_r are equal to θn^{-1}.

Let Y_{nr} be the rth largest of the $\eta_{ns}(s = 1,2,\ldots,n)$, so that

$$X_{nr} = Y_{nr}/W(\theta) \tag{8.18}$$

is the rth largest of the ξ_{ns}. Then it is a matter of elementary, non-stochastic, analysis to show that

$$\lim_{n\to\infty} Y_{nr} = Y_r\,. \tag{8.19}$$

To see this, first note that with probability 1, the points V for $(U, V) \in \Pi$ are distinct (proof as in Section 3). Hence for sufficiently large n, the points V corresponding to the u-values $Y_1, Y_2, \ldots, Y_r$ fall in different intervals $((s-1)\theta n^{-1}, s\theta n^{-1})$. Hence, for any $r \geqslant 1$,

$$Y_{nr} \geqslant Y_r \qquad (n \geqslant N(r)) \tag{8.20}$$

for some $N(r)$. On the other hand,

$$\sum_{r=1}^{n} Y_{nr} = \sum_{s=1}^{n} \eta_{ns} = W(\theta) = \sum_{r=1}^{\infty} Y_r \, . \tag{8.21}$$

If (8.19) is false, there exist $r \geqslant 1$ and $\epsilon > 0$ such that

$$Y_{nr} > Y_r + \epsilon \tag{8.22}$$

for infinitely many n. Choose $k > r$ so large that

$$\sum_{s=1}^{k} Y_s > W(\theta) - \epsilon \, , \tag{8.23}$$

and then choose $n \geqslant N(1), N(2), \ldots, N(k)$ to satisfy (8.22). Then

$$\sum_{s=1}^{k} Y_{ns} > \sum_{s=1}^{k} Y_s + \epsilon > W(\theta) \, , \tag{8.24}$$

which contradicts (8.21).

The contradiction proves (8.19), so that, for $r \geqslant 1$,

$$\lim_{n \to \infty} X_{nr} = Y_r / W(\theta) = X_r \, , \tag{8.25}$$

in the notation of Theorem 3. Since the X_{nr} are independent of $W(\theta) = Z$, their limits X_r are independent of Z, and Theorem 3 is proved.

Notice that we have also proved the special case of Theorem 4 in which

$$\alpha_{nr} = \theta / n \, . \tag{8.26}$$

However, the only use we have made of the dissection of $(0, \theta)$ into equal subintervals has been to ensure that distinct points are, for large enough n, in different subintervals. For general α_{nr} satisfying (8.8) and (8.9), the same argument works with (8.15) replaced by

$$\eta_{nr} = W\left(\alpha_{n1} + \alpha_{n2} + \ldots + \alpha_{n,r-1}, \alpha_{n1} + \alpha_{n2} + \ldots + \alpha_{nr}\right) \, . \tag{8.27}$$

This proves Theorem 4 in full generality.

The independence property asserted in Theorem 3 is characteristic of the density (8.1) and its mean relatives. To see this, suppose that Π is a Poisson process on $(0, \infty)$ with mean measure μ. Suppose that

$$\int y \mu(dy) < \infty \, , \tag{8.28}$$

so that

$$Z = \sum_{Y \in \Pi} Y \tag{8.29}$$

is finite with probability one. Suppose finally that the (non-Poisson) random set

$$\{Y/Z \ ; \ Y \in \Pi\} \tag{8.30}$$

is independent of Z. Colour the points of Π red or green with equal probabilities, distinct points coloured independently. Then the red points and the green points form independent Poisson processes.

Let Z_1 be the sum (8.28) taken over the red points, and Z_2 the same over the green points. Then Z_1 and Z_2 are independent and

$$Z = Z_1 + Z_2 \,.$$

Moreover, Z_1/Z_2 can be expressed in terms of a colouring of (8.29) which is independent of Z. Thus Z_1 and Z_2 are independent positive random variables with the property that Z_1/Z_2 and $Z_1 + Z_2$ are independent. It is easy to show that this can only happen if, for some $c > 0$, cZ_1 and cZ_2 have gamma distributions (8.4), so that cZ has that distribution for some $\theta > 0$.

Campbell's Theorem then shows that, for $t \geqslant 0$,

$$\begin{aligned}(1+t)^{-\theta} &= \mathbb{E}\{e^{-tcZ}\} \\ &= \exp\left\{\int \left(e^{-tcy} - 1\right) \mu(dy)\right\} ,\end{aligned}$$

so that

$$\int \left(1 - e^{-tcy}\right) \mu(dy) = \theta \log(1+t) \,.$$

and this inverts to give

$$\mu(dy) = \theta y^{-1} e^{-cy} dy \,. \tag{8.31}$$

9 The marginals of $\mathcal{PD}(\theta)$

For many purposes, Theorem 3 is an adequate description of $\mathcal{PD}(\theta)$, but what it does not give is explicit formulae for the marginal distributions of the infinite random sequence $(X_1, X_2, X_3, \ldots)$. These were first calculated by Billingsley [2] in the special case $\theta = 1$ which arises in number theory (see [7] for the tangled history). The general case is due to Watterson [10] in the context of population genetics. None of the derivations in the literature [1] is entirely transparent, and it seems worth giving a self-contained calculation in the spirit of [6].

Consider first a Poisson process Π on $(0,1)$ with density

$$\theta u^{-1} (0 < u < 1) \tag{9.1}$$

and let

$$T = \sum_{U \epsilon \Pi} U \,. \tag{9.2}$$

Campbell's Theorem shows that T is finite with probability 1, and for $t \geqslant 0$,

$$\mathbb{E}\left(e^{-tT}\right) = \exp\left(-\theta \int_0^1 \left(1 - e^{-tu}\right) u^{-1} du\right) .$$

By a well-known identity,

$$\int_0^1 (1 - e^{-tu}) u^{-1} du = \int_0^t (1 - e^{-v}) v^{-1} dv$$
$$= \int_t^\infty e^{-v} v^{-1} dv + \log t + \gamma ,$$

where γ is Euler's constant, so that

$$\mathbb{E}\left(e^{-tT}\right) = e^{-\gamma\theta} t^{-\theta} \exp\left(-\theta \int_t^\infty e^{-v} v^{-1} dv\right) . \tag{9.3}$$

Expanding the exponential shows that the right hand side is the Laplace transform of an integrable function, which must be the probability density of T. The process Π has now served its purpose and is discarded, retaining only the fact that, for any $\theta > 0$, there is a probability density p_θ on $(0, \infty)$, defined by its Laplace transform

$$\int_0^\infty p_\theta(u) e^{-tu} du = \exp\left(-\theta \int_0^t (1 - e^{-v}) v^{-1} dv\right) \tag{9.4}$$

$$= e^{-\gamma\theta} t^{-\theta} \exp\left(-\theta \int_t^\infty e^{-v} v^{-1} dv\right) . \tag{9.5}$$

A great deal of information about p_θ can be found in [1].

Returning to the Poisson process

$$\Pi_\theta = \{Y_1, Y_2, \ldots\}$$

of Theorem 3, note first that, conditional on the values of $Y_1, Y_2, \ldots, Y_n$, the points $Y_r (r > n)$ form a Poisson process on $(0, Y_n)$ with density

$$\theta y^{-1} e^{-y} \quad (0 < y < Y_n) .$$

Campbell's Theorem can then be used to find the conditional distribution of

$$Z_n = Y_{n+1} + Y_{n+2} + \ldots \qquad ;$$

for $t \geqslant 0$,

$$\mathbb{E}\left\{e^{-tZ_n} | Y_1, Y_2, \ldots, Y_n\right\} = \exp\left(-\theta \int_\theta^{Y_n} (1 - e^{-ty}) y^{-1} e^{-y} dy\right) . \tag{9.6}$$

Now

$$\int_0^{Y_n} \left(1-e^{-ty}\right) y^{-1} e^{-y} dy \quad = \quad \int_0^{(1+t)Y_n} (1-e^{-v}) v^{-1} dv$$
$$- \int_{Y_n}^{\infty} e^{-v} v^{-1} dv - \log Y_n - \gamma$$

and

$$S_n = Z - Z_n = Y_1 + Y_2 + \ldots + Y_n \,, \tag{9.7}$$

so that

$$\mathbb{E}\{e^{-tZ} | Y_1, Y_2, \ldots, Y_n\}$$
$$= e^{-tS_n} \int_0^\infty p_\theta(u) e^{-(1+t)Y_n u} du \exp\left(\theta \int_{Y_n}^\infty e^{-v} v^{-1} dv\right) Y_n^\theta e^{\gamma\theta} \,.$$

Regarded as a function of t, this equation is a Laplace transform identity, which can be inverted to show that, given $Y_1, Y_2, \ldots, Y_n$, Z has a conditional probability density

$$g(\zeta\,; Y_1\,, Y_2\,, \ldots\,, Y_n) = p_\theta\left(\frac{\zeta - S_n}{Y_n}\right) e^{-(\zeta - S_n)} Y_n^{\theta-1} e^{\gamma\theta} \exp\left(\theta \int_{Y_n}^\infty e^{-v} v^{-1} dv\right)$$

in $\zeta > S_n$. The joint distribution of $Y_1, Y_2, \ldots, Y_n$ has density

$$\prod_{r=1}^n \left(\theta y_r^{-1} e^{-y_r}\right) \exp\left(-\theta \int_{y_n}^\infty v^{-1} e^{-v} dv\right) \,.$$

and multiplying this by $g(\zeta; y_1, y_2\,, \ldots, y_n)$ gives the joint probability density of $Z, Y_1, Y_2, \ldots, Y_n$ in the form

$$p_\theta\left(\frac{\zeta - s_n}{y_n}\right) (y_1 y_2 \ldots y_n)^{-1} y_n^{\theta-1} e^{-\zeta} e^{\gamma\theta} \theta^n \tag{9.8}$$

in $\zeta > s_n = y_1 + y_2 + \ldots + y_n$.

Now make the change of variable

$$X_r = Y_r / Z \,, \tag{9.9}$$

to show that the joint distribution of $Z, X_1, X_2, \ldots, X_n$ has density

$$p_\theta\left(\frac{1 - x_1 - x_2 - \ldots x_n}{x_n}\right) (x_1 x_2 \ldots x_n)^{-1} \zeta^{\theta-1} x_n^{\theta-1} e^{-\zeta} e^{\gamma\theta} \theta^n$$
$$= f_n(x_1, x_2, \ldots, x_n) \frac{\zeta^{\theta-1} e^{-\zeta}}{\Gamma(\theta)} \,,$$

where

$$f_n(x_1, x_2, \ldots, x_n) = \frac{e^{\gamma\theta} \theta^n \Gamma(\theta) x_n^{\theta-1}}{x_1 x_2 \ldots x_n} p_\theta\left(\frac{1 - x_1 - x_2 - \ldots - x_n}{x_n}\right) \tag{9.10}$$

in

$$x_1 + x_2 + \ldots + x_n < 1\,. \tag{9.11}$$

This shows that, as we already know, $(X_1, X_2, \ldots, X_n)$ and Z are independent, and that Z has a gamma distribution. But it also shows that the joint distribution of $X_1, X_2, \ldots, X_n$ is given by (9.10), consistently with the calculations of Billingsley and Watterson.

The densities f_n must of course satisfy the consistency conditions

$$\int_0^{x_{n-1}} f_n(x_1, x_2, \ldots, x_{n-1}, \xi)d\xi = f_{n-1}(x_1, x_2, \ldots, x_{n-1}) \tag{9.12}$$

for $n \geqslant 2$, and

$$\int_0^1 f_1(\xi)d\xi = 1\,. \tag{9.13}$$

Substituting (9.10) into these equations gives the equations

$$\theta u^{\theta-1} \int_{\max(u,1)}^{\infty} v^{-\theta} p_\theta(v-1)dv = p_\theta(u) \tag{9.14}$$

and

$$\theta e^{\gamma\theta}\Gamma(\theta) \int_1^{\infty} v^{-\theta} p_\theta(v-1)dv = 1\,. \tag{9.15}$$

Taking Laplace transforms we easily recover (9.4), so that p_θ is uniquely determined by (9.14) and (9.15).

References

[1] R. Arratia, A.D. Barbour & S. Tavaré *Logarithmic Combinational Structures: a Probabilistic Approach* (European Mathematical Society, Zurich, 2003).

[2] P. Billingsley, On the distribution of large prime divisors, *Period. Math. Hungar.* **2** (1972) 283–289.

[3] D. Blackwell, Discreteness of Ferguson selections, *Ann. Statist.* **1** (1973) 356–358.

[4] P.R. Halmos, *Measure Theory* (van Nostrand, Princeton, 1950).

[5] E.F. Harding & D.G. Kendall (ed.) *Stochastic Geometry* (Wiley, New York, 1974).

[6] J.F.C. Kingman, *Poisson Processes* (Oxford, 1993).

[7] J.F.C. Kingman, The Poisson-Dirichlet distribution and the frequency of large prime divisors
(2004, http://www.newton.cam.ac.uk/preprints/NI04019.pdf).

[8] P.A.P. Moran, A non-Markovian quasi-Poisson process, *Stud. Sci. Math. Hungar.* **2** (1967) 425–429.

[9] A. Rényi, Remarks on the Poisson process, *Stud. Sci. Math. Hungar.* **2** (1967) 119–123.

[10] G.A. Watterson, The sampling theory of selectively neutral alleles, *Adv. Appl. Prob.* **6** (1974) 463–488.

J.F.C. Kingman
Isaac Newton Institute for Mathematical Sciences,
20 Clarkson Road,
Cambridge CB3 0EH,
UK

director@newton.cam.ac.uk