Probab. Theory Relat. Fields 92, 21–39 (1992)

Probability Theory and Related Fields

Size-biased sampling of Poisson point processes and excursions*

Mihael Perman,[1] Jim Pitman,[2] and Marc Yor[3]

[1] Department of Economics, University of Ljubljana, Kardeljeva ploscad 17, YU-61109 Ljubljana, Yugoslavia
[2] Department of Statistics, University of California, Berkeley, CA 94720, USA
[3] Laboratoire de Probabilités, Université Pierre et Marie Curie, 4, Place Jussieu-Tour 56, F-75252 Paris Cedex 05, France

Received July 7, 1991; in revised form September 11, 1991

Summary. Some general formulae are obtained for size-biased sampling from a Poisson point process in an abstract space where the size of a point is defined by an arbitrary strictly positive function. These formulae explain why in certain cases (gamma and stable) the size-biased permutation of the normalized jumps of a subordinator can be represented by a stickbreaking (residual allocation) scheme defined by independent beta random variables. An application is made to length biased sampling of excursions of a Markov process away from a recurrent point of its statespace, with emphasis on the Brownian and Bessel cases when the associated inverse local time is a stable subordinator. Results in this case are linked to generalizations of the arcsine law for the fraction of time spent positive by Brownian motion.

1 Introduction

Let $T > 0$ be an infinitely divisible random variable with no drift component:

(1.a) $$T = \sum_i \Delta_{(i)} \qquad \text{where}$$

(1.b) $$\Delta_{(1)} \geqq \Delta_{(2)} \geqq \cdots$$

are the ranked values of points in a Poisson point process on $(0, \infty)$ whose mean measure Λ is the Lévy measure associated with T, and

(1.c) $$E[e^{-yT}] = \exp\left(-\int_0^\infty (1 - e^{-yx})\Lambda(dx)\right), \quad y > 0 .$$

It is well known that the $\Delta_{(i)}$ may be interpreted as the size of the jump at time σ_i of an increasing process with independent increments, or *subordinator*

(1.d) $$T(s) = \sum_i \Delta_{(i)} 1(\sigma_i \leqq s), \quad 0 \leqq s \leqq 1, \quad \text{with } T(1) = T ,$$

---

* Research supported in part by NSF grant DMS88-01808 and DMS91-07351

where the σ_i are independent random times distributed uniformly on (0, 1), independent also of the $\Delta_{(i)}$.

Kingman (1975, Sect. 7), considered the distribution on the infinite simplex of

$$(P_{(1)}, P_{(2)}, \ldots) \overset{\text{def}}{=} (\Delta_{(1)}/T, \Delta_{(2)}/T, \ldots) . \tag{1.e}$$

In particular, Kingman showed that in case T has gamma distribution with shape parameter a, the distribution of $(P_{(1)}, P_{(2)}, \ldots)$ is the Poisson-Dirichlet distribution on the infinite simplex with parameter a. This distribution turns up in a number of different contexts, for example in the asymptotics for the ranked cycle lengths of a random permutation (Shepp and Lloyd 1966; Vershik and Schmidt 1977 and 1978); as the distribution of the ranked sizes of atoms in the Dirichlet process prior in Bayesian statistics (Blackwell and McQueen (1973); Ferguson (1973)), and in limiting models for the abundances of genes in population genetics and species in mathematical ecology (see e.g. Patil and Taillie (1977), Hoppe (1987).) Explicit but rather intractable formulae for features of the Poisson-Dirichlet distribution such as the joint density of the first n components can be found in Watterson (1976); Ewens (1988), and some of the above references. Perman (1990, 1991) gives extensions of these formulae for more general infinitely divisible laws in terms of the solution of an integral equation.

Another case of interest arises when T has a stable distribution of index α, for $0 < \alpha < 1$. The distribution of $(P_{(1)}, P_{(2)}, \ldots)$ in this case was encountered by Pitman and Yor (1990) (abbreviated PY throughout this paper), in the study of arcsine laws and interval partitions associated with a stable subordinator.

In settings for both the gamma and stable cases mentioned above, another random sequence which turns up naturally is the sequence $(P_1, P_2, \ldots)$ derived from $(P_{(1)}, P_{(2)}, \ldots)$ by *size-biased random permutation.* This means the following: conditionally given $(P_{(1)}, P_{(2)}, \ldots)$, the first term P_1 equals $P_{(i)}$ with probability $P_{(i)}$; given $(P_{(1)}, P_{(2)}, \ldots)$ and $P_1 = P_{(i)}$, the next term P_2 equals $P_{(j)}$ for $j \neq i$ with probability $P_{(j)}/(1 - P_{(i)})$, and so on. See Donnelly and Joyce (1989) for further motivation and background discussion. Let $U_1 = 1 - P_1, U_2 = (1 - P_1 - P_2)/(1 - P_1)$, and so on, so that

$$P_i = (1 - U_i) \prod_{j=1}^{i-1} U_j, \quad i \geqq 1 . \tag{1.f}$$

The U_i may be interpreted as residual fractions in a *stick-breaking* or residual allocation scheme leading to overall proportions P_i with sum 1. Here are two known results:

Theorem 1.1 (McCloskey 1965). *If T has gamma distribution with shape parameter $a > 0$ then the U_i are independent with identical beta$(a, 1)$ distribution.*

Theorem 1.2 (Perman 1990, Corollary 3.19) *If T has stable distribution with index α, $0 < \alpha < 1$, then the U_i are independent with beta$(i\alpha, 1 - \alpha)$ distributions.*

Donnelly and Joyce (1989) derived Theorem 1.1 by a method suggested by Patil and Taillie (1977), namely passing to the limit as $n \to \infty$ from an analogous result for size biased sampling from a Dirichlet distribution on the n-simplex. This approach does not work as smoothly for Theorem 1.2, because in the corresponding discrete setup there is not exact independence, only asymptotic independence as $n \to \infty$. Perman (1990) derived Theorem 1.2 indirectly, via results of PY.

One purpose of this paper is to derive Theorems 1.1 and 1.2 in a unified way. This is done in Sect. 2 by a general calculation of the joint density of $T, U_1, \ldots, U_n$ for any strictly positive infinitely divisible T with absolutely continuous Lévy measure Λ. This broadens the approach of McCloskey (1965), who noted that Theorem 1.1 appears as a special case characterized by independence of T and U_1. Theorem 1.2 is the special case characterized by independence of TU_1 and U_1.

Another purpose is to provide a simplified approach to an interesting collection of results concerning excursions of Brownian motion and Bessel processes, first obtained by PY and Perman (1990). This is done in Sect. 3 by lifting the discussion of Sect. 2 from the Poisson point process of jumps of a subordinator to Itô's Poisson point process of excursions. Aldous and Pitman (1991) show how these results in the Brownian case are related to Brownian bridge asymptotics for functionals of random mappings.

The key to the calculations in both Sects. 2 and 3 is a general formula for size-biased sampling from a Poisson point process that is tied to the simple characterization of a Poisson point process via its family of Palm distributions. Section 4 presents this formula and some of its consequences in a general setting, for an abstract space of points and any positive measurable function defining "size". Also appearing in Sect. 4 is a simple general representation of size-biased sampling. This has a number of interesting consequences, including the representation of the jumps of a gamma process due to Tavaré (1987). See also Pitman (1991) for further applications.

2 Size-biased sampling of the jumps of a subordinator

Throughout this section let $(\Delta_1, \Delta_2, \ldots)$ be a size-biased permutation of the ranked jumps of a subordinator $(\Delta_{(1)}, \Delta_{(2)}, \ldots)$, as in (1.a), with sum T. So the P_i in (1.f) are $P_i = \Delta_i/T$. Let $T_0 = T$, and for $n \geqq 1$ let

$$T_n = T - \Delta_1 - \cdots - \Delta_n .$$

Now the U_i in (1.f) are $U_i = T_i/T_{i-1}, i \geqq 1$. Blanket assumptions on the Lévy measure Λ, to ensure $0 < T < \infty$ a.s., are

$$\Lambda(0, \infty) = \infty , \quad \Lambda(1, \infty) < \infty, \quad \text{and} \quad \int_0^1 x\Lambda(\mathrm{d}x) < \infty .$$

Assume also that Λ has a density ρ:

$$\Lambda(\mathrm{d}x) = \rho(x)\mathrm{d}x, \quad x > 0 .$$

Then T has a probability density:

$$P(T \in \mathrm{d}t) = f(t)\,\mathrm{d}t, \quad t > 0 ,$$

for some f determined by its Laplace transform (1.c). See Brockett and Hudson (1980). Formula (2.a) in the following theorem appears under stronger regularity conditions as Lemma 1 of McCloskey (1965), attributed to Herman Rubin. See also Lemma 4.1 below for a still more general formulation.

Theorem 2.1 *The sequence $T, T_1, T_2, \ldots$ is a Markov chain with stationary transition probabilities*

$$\text{(2.a)} \qquad P(T_1 \in dt_1 | T = t) = \frac{\Theta(t-t_1)}{t} \frac{f(t_1)}{f(t)} dt_1, \quad \textit{where}$$

$$\text{(2.b)} \qquad \Theta(x) = x\rho(x) .$$

Consequently, the joint density of (T, T_1, T_2) is

$$\text{(2.c)} \qquad f(t, t_1, t_2) = \frac{\Theta(t-t_1)}{t} \frac{\Theta(t_1 - t_2)}{t_1} f(t_2) ;$$

the joint density of $(T, U_1, U_2) = (T, T_1/T, T_2/T_1)$ is

$$\text{(2.d)} \qquad g(t, u_1, u_2) = \Theta(\bar{u}_1 t)\Theta(u_1 \bar{u}_2 t) f(u_1 u_2 t), \quad \textit{where } \bar{u} = 1 - u ;$$

the joint density of (T_2, U_1, U_2) is

$$\text{(2.e)} \qquad h(t_2, u_1, u_2) = (u_1 u_2)^{-1} \Theta\left(\frac{\bar{u}_1 t_2}{u_1 u_2}\right) \Theta\left(\frac{\bar{u}_2 t_2}{u_2}\right) f(t_2) ;$$

and for every $n \geqq 1$ there are similar product formulae for the $n+1$ dimensional joint densities of $(T, T_1, \ldots, T_n)$, of $(T, U_1, \ldots, U_n)$, and of $(T_n, U_1, \ldots, U_n)$.

The proof of this theorem is based on the following lemma:

Lemma 2.2 *Let $\Delta^-_{(\cdot)}$ be the ranked sequence*

$$\Delta^-_{(\cdot)} = (\Delta^-_{(1)}, \Delta^-_{(2)}, \ldots)$$

obtained by deleting the point Δ_1 from the ranked sequence

$$\Delta_{(\cdot)} = (\Delta_{(1)}, \Delta_{(2)}, \ldots) .$$

Then for $x > 0$, $t_1 > 0$ and measurable subsets B of the sequence space,

$$\text{(2.f)} \qquad P(\Delta_1 \in dx, T_1 \in dt_1, \Delta^-_{(\cdot)} \in B) = [\rho(x)\,dx] P(T \in dt_1, \Delta_{(\cdot)} \in B) \frac{x}{x + t_1} .$$

Proof. The following argument is informal, but can be made entirely rigorous by using the concept of Palm distribution to justify the conditioning involved. See Sect. 4 of this paper, or Lemma 2.2 of PY.

Formula (2.f) is made evident by presenting the event $(\Delta_1 \in dx, T_1 \in dt_1, \Delta^-_{(\cdot)} \in B)$ as the following chain of events:

1) some point $\Delta_{(i)} \in dx$;
2) given 1), the sequence $\Delta^{x-}_{(\cdot)}$ derived from $\Delta_{(\cdot)}$ by deletion of the point in dx is such that $\sum_j \Delta^{x-}_{(j)} \in dt_1$ and $\Delta^{x-}_{(\cdot)} \in B$;
3) given 1) and 2) the value $\Delta_{(i)} \in dx$ is chosen by size-biased sampling from $\Delta_{(\cdot)}$.

The product of corresponding probabilities is the right side of (2.f). In particular, the conditional probability for 2) given 1) is $P(T \in dt_1, \Delta_{(\cdot)} \in B)$, because conditioning the Poisson point process defined by $\Delta_{(\cdot)}$ to have a point in dx leaves the remaining points $\Delta^{x-}_{(\cdot)}$ distributed like the original points. □

Proof of Theorem 2.1. Since $T = T_1 + \Delta_1$, it is immediate from (2.f) that the conditional distribution of $\Delta^-_{(\cdot)}$ given T and $T_1 = t_1$ is identical to the conditional distribution of $\Delta_{(\cdot)}$ given $T = t_1$. And given $\Delta_{(\cdot)}^{-}$, the sequence $T_2, T_3, \ldots$ is obtained by the same rule of size-biased deletion used to derive $T_1, T_2, \ldots$ from $\Delta_{(\cdot)}$. Thus the conditional distribution of $(T_2, T_3, \ldots)$ given T and $T_1 = t_1$ is identical to the conditional distribution of $(T_1, T_2, \ldots)$ given $T = t_1$. This is the time homogeneous Markov property for $(T, T_1, T_2, \ldots)$. Formula (2.a) for the transition probabilities is also immediate from (2.f). Now (2.c) follows at once, as do (2.d) and (2.e) by elementary changes of variables. □

Part (i) of the following Corollary was obtained by McCloskey (1965), assuming smoothness conditions on the densities. The "if" assertion of part (ii) is due to Perman (1990).

Corollary 2.3 *In the setting of Theorem 2.1, with* $U_1 = T_1/T$:
(i) *The random variables* T *and* U_1 *are independent if and only if* T *has the gamma distribution with density*

(2.g) $$f(t) = \frac{1}{\Gamma(a)} \lambda^a t^{a-1} e^{-\lambda t}$$

for some $a > 0, \lambda > 0$. *Equivalently,*

(2.h) $$\Theta(x) = a e^{-\lambda x} .$$

Then $T, U_1, U_2, \ldots, U_n$ *are independent for all* n, *and* U_n *has beta*$(a, 1)$ *distribution.*
(ii) *The random variables* T_1 *and* U_1 *are independent if and only if* T *has stable distribution wth index* α, *as specified by the Laplace transform*

(2.i) $$E e^{-yT} = \int_0^\infty e^{-yt} f(t)\, dt = \exp(-c y^\alpha) ,$$

for some $0 < \alpha < 1, c > 0$. *Equivalently,*

(2.j) $$\Theta(x) = K x^{-\alpha} \quad \text{where } K = c\alpha/\Gamma(1-\alpha) .$$

Then $T_n, U_1, U_2, \ldots, U_n$ *are independent for all* n, *and* U_n *has beta* $(n\alpha, 1-\alpha)$ *distribution.*

Proof of Corollary 2.3 (i). If T has gamma distribution, substituting the formulae for f and Θ into formula (2.d) for the joint density of (T, U_1, U_2) gives

$$g(t, u_1, u_2) = a e^{-\bar{u}_1 \lambda t} a e^{-u_1 \bar{u}_2 \lambda t} \frac{1}{\Gamma(a)} \lambda^a (u_1 u_2 t)^{a-1} e^{-u_1 u_2 \lambda t}$$

$$= \left[\frac{1}{\Gamma(a)} \lambda^a t^{a-1} e^{-\lambda t} \right] [a u_1^{a-1}] [a u_2^{a-1}] .$$

And the joint density of $(T, U_1, \ldots, U_n)$ factors similarly.
Conversely, if T and U_1 are independent, then from (2.d) we must have

$$g(t, u_1) = \Theta(\bar{u}_1 t) f(u_1 t) = \phi(u_1) f(t)$$

where ϕ is the density of U_1. But then from (2.d) again

$$\begin{aligned} g(t, u_1, u_2) &= \Theta(\bar{u}_1 t)\Theta(u_1 \bar{u}_2 t) f(u_1 u_2 t) \\ &= \Theta(\bar{u}_1 t)\phi(u_2) f(u_1 t) \\ &= \phi(u_1)\phi(u_2) f(t) , \end{aligned}$$

and so on, implying that $U_1, U_2, \ldots$ are i.i.d. and independent of T. That this is characteristic of the gamma case follows from the result of Lukacs (1955) that if X and Y are independent, positive, non-constant, and $X + Y$ is independent of $X/(X + Y)$, then X and Y have gamma distributions with a common scale parameter. For let $X = T(1/2)$, $Y = T - T(1/2)$, where $(T(s), s \geqq 0)$ is a subordinator with $T(1) = T$ as in (1.d). If I_n is the indicator of the event that the nth jump of size $T_n - T_{n-1}$ occurs before time 1/2, then $I_1, I_2, \ldots$ are independent Bernoulli (1/2) variables independent of $T_0, T_1, T_2, \ldots$, and

$$X = T(1/2) = \sum_n I_n(T_n - T_{n-1}) = \sum_n I_n T U_1 U_2 \cdots U_{n-1} \bar{U}_n$$

so that $X/(X + Y) = X/T$ is a function of $(I_1, I_2, \ldots)$ and $(U_1, U_2, \ldots)$, hence independent of $T = X + Y$. □

Proof of Corollary 2.3 (ii). In the stable case (2.e) gives the following joint density for (T_2, U_1, U_2):

$$h(t_2, u_1, u_2) = K^2[u_1^{\alpha-1}\bar{u}_1^{-\alpha}][u_2^{2\alpha-1}\bar{u}_2^{-\alpha}]t_2^{-2\alpha} f(t_2) , \tag{2.k}$$

so T_2, U_1, and U_2 are independent. The joint density of $(T_n, U_1, \ldots, U_n)$ factors similarly for every n. Conversely, if T_1 and U_1 are independent, then by (2.e) the joint density of T_1 and U_1 is of the form

$$h(t_1, u_1) = u_1^{-1}\,\Theta\left(\frac{\bar{u}_1 t_1}{u_1}\right) f(t_1) = f_1(t_1)\phi(u_1) \tag{2.l}$$

for some densities f_1 and ϕ. This forces $\Theta(tx) = \eta(t)\psi(x)$ for some positive measurable functions η and ψ, hence $\Theta(x) = Kx^{-\alpha}$ for some K and α. Finally, $0 < \alpha < 1$ by the constraints on the Lévy measure of a subordinator. □

Remarks on the stable case. In the stable case the density of T_n appears easily from the above calculation: for every $n \geqq 0$.

$$P(T_n \in dt)/dt = K_n^{-1} t^{-n\alpha} f(t) , \tag{2.m}$$

where the constant of integration K_n is easily calculated as

$$K_n = E(T^{-n\alpha}) = \frac{\Gamma(n+1)}{\Gamma(n\alpha + 1)} c^{-n} . \tag{2.n}$$

It will be seen below that formula (2.n) is valid for an arbitrary real number $r > -1$ in place of the non-negative integer n. For such r let μ_r denote the distribution on $(0, \infty)$ whose density at t is $K_r^{-1} t^{-r\alpha} f(t)$.

Consider a Markov chain $T_0, T_1, T_2, \ldots$, with stationary transition probabilities as in (2.a) derived from the setup (2.i) and (2.j) for the stable case. But now assign T_0 an arbitrary initial distribution μ. Let $U_i = T_i/T_{i-1}$ as before. It follows from Corollary 2.3 (ii) and (2.m), simply by shifting everything r steps along, that the following assertion holds for every non-negative integer r:

(2.o) *If T_0 has law μ_r, then for all $n = 1, 2, \ldots$ the law of T_n is μ_{r+n}, the law of U_n is beta $((r+n)\alpha, 1-\alpha)$, and $T_n, U_1, U_2, \ldots, U_n$ are independent.*

We now claim further that

(2.p) *the above assertion* (2.o) *holds for every real $r > -1$, and*

(2.q) *the laws μ_r are the only laws for T_0 which make U_1 and T_1 independent.*

Proof of (2.p) and (2.q). Note first from (2.a) that if T_0 is given a law with density $d_0(t)f(t)$, then instead of (2.1) the joint density of U_1 and T_1 becomes

(2.r) $$u_1^{-1}\,\Theta\left(\frac{\bar{u}_1 t_1}{u_1}\right) f(t_1)\, d_0\left(\frac{t_1}{u_1}\right).$$

In the case at hand, $d_0(t) = K_r^{-1} t^{-r\alpha}$, so the above joint density equals

$$K K_r^{-1} u_1^{(r+1)\alpha-1} \bar{u}_1^{-\alpha} t_1^{-(r+1)\alpha} f(t_1)\,.$$

Now (2.o) is clear for any real r and $n = 1$ provided the moment K_r is finite, which is seen below to be so iff $r > -1$. The assertion (2.o) for such r and $n = 1, 2, \ldots$ follows easily by induction. And (2.q) follows by a variation of the argument below (2.1)

Proof of (2.n). The right side of (2.r) defines a joint density iff

$$K K_r^{-1} B((r+1)\alpha, 1-\alpha) K_{r+1} = 1\,,$$

where $B(a, b) = \Gamma(a)\Gamma(b)/\Gamma(a+b)$, whence using also (2.j)

(2.s) $$K_{r+1} = \frac{\Gamma(r\alpha+1)K_r}{c\alpha\Gamma((r+1)\alpha)}, \quad r > -1\,.$$

Now (2.n) follows by induction for positive integer n using the recursion for the gamma function.

Formula (2.n) can also be checked for any real $r > 0$ instead of n using the formula

(2.t) $$E(T^{-p}) = \frac{1}{\Gamma(p)} \int_0^\infty \lambda^{p-1}\phi(\lambda)\,\mathrm{d}\lambda$$

valid for any positive r.v. T with Laplace transform $\phi(\lambda) = Ee^{-\lambda T}$, and any $p > 0$. This follows by Fubini's theorem from the identity

$$t^{-p} = \frac{1}{\Gamma(p)} \int_0^\infty \lambda^{p-1} \mathrm{e}^{-t\lambda}\,\mathrm{d}\lambda\,.$$

Finally, (2.n) for $-1 < r < 0$ instead of n follows from the case $0 < r < 1$ by the recursion (2.s).

3 Bridges and excursions

This section shows how, by an extension of ideas of the preceding section, we can derive quite easily an interesting collection of results concerning Brownian motion and Bessel processes. These results were first obtained in a more roundabout way by PY and Perman (1990). Here the idea is to lift the whole discussion of the previous section from the Poisson point process of jumps of a subordinator to the level of Itô's Poisson point process of excursions. We start with a result in a general Markovian setting.

Theorem 3.1 *Let* $(S_t, t \geqq 0)$ *be a continuous local time process associated with a recurrent point* 0 *in the statespace of a strong Markov process* $X = (X_t, t \geqq 0)$ *with cadlag paths starting from* $X_0 = 0$. *Let* $T(s) = \inf\{t: S_t > s\}$, *and suppose the subordinator* $(T(s), s \geqq 0)$ *has no drift component, or equivalently that* $P(X_t = 0) = 0$ *for all* $t > 0$. *Fix* $s > 0$ *and let* $T = T(s)$. *Let* (G_{UT}, D_{UT}) *denote the excursion interval of* X *away from* 0 *that contains the random time* UT, *where* U *is uniformly distributed on* $(0, 1)$ *independent of* X. *Let* $\Delta_1 = G_{UT} - D_{UT}$, *let* $\varepsilon^{\#} = (\varepsilon^{\#}_u, 0 \leqq u < \Delta_1)$ *be the excursion of* X *straddling time* UT:

$$\varepsilon^{\#}_u = X_{G_{UT}+u} \quad \textit{if} \quad 0 \leqq u < \Delta_1 \,,$$

let $T_1 = T - \Delta_1$, *and let* $X^{\#} = (X^{\#}_u, 0 \leqq u \leqq T_1)$ *be the process with lifetime* T_1 *obtained from* X *on* $[0, T]$ *by deleting the excursion* $\varepsilon^{\#}$ *and closing up the gap*:

$$\begin{aligned} X^{\#}_u &= X_u \textit{ if } 0 \leqq u < G_{UT} \\ &= X_{\Delta_1+u} \textit{ if } G_{UT} \leqq u \leqq T_1 \,. \end{aligned}$$

The $\varepsilon^{\#}$ *and* $X^{\#}$ *are conditionally independent given* T_1, *and the following conditional laws are identical for* $t \geqq 0$:

(3.a) *the conditional law of* $X^{\#}$ *on* $[0, T_1]$ *given* $T_1 = t$;

(3.b) *the conditional law of* X *on* $[0, T]$ *given* $T = t$.

Remark 3.2. Assume that X admits sensible bridge distributions $P(\cdot \mid X_t = 0)$ defined either by a smooth conditional distribution given $X_t = x$ as x varies, or as Palm distributions associated with the local time random measure with distribution function $(S_t, t \geqq 0)$, as in Kallenberg (1981). Then we can make sense of the intuitive identity

$$P(\cdot \mid T(s) = t) = P(\cdot \mid X_t = 0, S_t = s) \,.$$

Then the common conditional law in (3.a) and (3.b) is that of a bridge of length t conditioned to accumulate local time s at 0 up to time t.

Remark 3.3. The partial description of the joint law of $X^{\#}$ and $\varepsilon^{\#}$ in Theorem 3.1 is completed using ideas of excursion theory, applied also in the following proof. Let Q^z_{ex} be the probability law for excursions of X away from 0 of a given length z obtained by disintegration of Itô's excursion law Q_{ex} with respect to the lifetime ζ of an excursion, meaning that for $t > 0$ there is the following identity of measures on the path space for excursions:

$$Q_{ex}(\cdot \cap \zeta > t) = \int_t^{\infty} \Lambda(\mathrm{d}z) Q^z_{ex}(\cdot)$$

where Λ is the Q_{ex} distribution of ζ, which is the Lévy measure associated with the inverse local time subordinator. Since the excursion straddling UT picks a jump of the inverse local time subordinator by length biased sampling,

(3.c) *the joint law of* $\Delta_1 = \zeta(\varepsilon^{\#})$ *and* T_1 *is as specified by Theorem* 2.1.

Or see (4.f) below in case Λ does not have a density. And the following argument shows that

(3.d) *the law of* $\varepsilon^{\#}$ *given* $\zeta(\varepsilon^{\#}) = z$ *is* Q_{ex}^{s} .

Theorem 3.1 combined with (3.c) and (3.d) fully specifies the joint law of $X^{\#}$ and $\varepsilon^{\#}$.

Remark 3.4. To obtain a fully fledged path decomposition in the spirit of Williams (1974), just one more ingredient is required to recreate the whole path of X on $[0, T]$ from $X^{\#}$ and $\varepsilon^{\#}$. This is $\sigma^{\#} = S_{UT}$, the value of the local time during the excised excursion. The argument below shows that $\sigma^{\#}$ is distributed uniformly on $[0, s]$ independently of both $X^{\#}$ and $\varepsilon^{\#}$. And given $\varepsilon^{\#}$, $X^{\#}$ and $\sigma^{\#}$, the original path of X on $[0, T]$ is recovered by inserting $\varepsilon^{\#}$ into $X^{\#}$ at time $\inf\{t: S_t^{\#} = \sigma^{\#}\}$, where $S_t^{\#}$ is the local time of $X^{\#}$ at 0 up to time t.

Proof of Theorem 3.1. Without loss of generality we take $s = 1$. Let $(\Omega_{ex}, \mathbf{F}_{ex})$ be a suitable path space for the description of excursions of X away from 0. Let N stand for the point process on $(0, 1) \times \Omega_{ex}$ which puts a point at $(\sigma_i, \varepsilon_i)$ for each i, where σ_i is the value of local time during the ith longest excursion ε_i of X away from 0 up to time T. According to Itô (1970), the point process N is Poisson with mean measure Lebesgue $\times Q_{ex}$, where the σ-finite measure Q_{ex} on $(\Omega_{ex}, \mathbf{F}_{ex})$ is Itô's excursion law. And the path $(X_t, 0 \leqq t \leqq T)$ can be reconstructed in a measurable manner from the point process N by stringing together the excursions ε_i in the order specified by the local times σ_i. See Salisbury (1986) for details. Write this as

(3.e) $$(X_t, 0 \leqq t \leqq T) = \Psi(N)$$

where Ψ is the reconstruction operation. Now let $N^{\#}$ be the point process obtained from N by deletion of the point corresponding to the excursion of X on (G_{UT}, D_{UT}). Then by definition of the process $X^{\#}$,

(3.f) $$(X_t^{\#}, 0 \leqq t \leqq T_1) = \Psi(N^{\#})\,.$$

Since $N^{\#}$ is derived from N by removal of a point $(\sigma^{\#}, \varepsilon^{\#})$ by length-biased sampling, where the length of the point $(\sigma_i, \varepsilon_i)$ is the duration of excursion ε_i, a slight variation of Lemma 2.2 shows that for $0 < u < 1$,

(3.g) $$P(\sigma^{\#} \in du, \varepsilon^{\#} \in de, T_1 \in dt, N^{\#} \in \cdot) = du\, Q_{ex}(de) P(T \in dt, N \in \cdot) \frac{\zeta(e)}{\zeta(e) + t}\,,$$

where $\zeta(e)$ is the lifetime of excursion e. Now the conclusions of the Theorem follow (3.e), (3.f) and (3.g), as also do the claims of Remarks 3.3 and 3.4. See Sect. 4 for details of this argument in a general PPP setting. □

Corollary 3.5 *Let* $X^{n\#}$ *be the process with lifetime* T_n *derived from* X *on the interval* $[0, T]$ *by first deleting* n *excursions* $\varepsilon^{1\#}, \ldots, \varepsilon^{n\#}$, *by length biased sampling, then* *closing up the gaps. The sequence* (T_n) *is Markov with stationary transition probabilities, as in Theorem 2.1. And conditionally given* $T = t_0, T_1 = t_1, \ldots, T_n = t_n$ *the excursions* $\varepsilon^{1\#}, \ldots, \varepsilon^{n\#}$ *and the remaining process* $X^{n\#}$ *are mutually independent, with the* $\varepsilon^{i\#}$ *distributed according to the laws* $Q_{ex}^{z(i)}$ *for excursions with prescribed lengths* $z(i) = t_i - t_{i-1}$, *and with* $X^{n\#}$ *distributed like* X *on* $[0, T]$ *given* $T = t_n$.

Proof. Repeated application of Theorem 3.1, (3.c) and (3.d). □

For the rest of this section we make the following assumption:

Assumption 3.6 *The process* $B = (B_t, t \geqq 0)$ *is a Brownian motion on* $\mathbb{R}$, *or a Bessel process of dimension* δ *on* $\mathbb{R}^+$, *for* $0 < \delta < 2$, *started at* $B_0 = 0$.

The special feature of B which connects to the results of Sect. 2 is this: the inverse local time at zero is a stable subordinator with index α for $\alpha = 1 - \delta/2$. See Molchanov and Ostrowski (1969); Barlow et al. (1989), PY. For the rest of the section δ and α are fixed and largely suppressed in the notation. The simplest definition of the local time process (S_t) in this setting is

$$S_t = c_\alpha \lim_{\varepsilon \to 0} \varepsilon^\alpha \# \{\text{excursions longer than } \varepsilon \text{ started by time } t\}$$

where c_α is a normalization constant whose value has no effect on the following results. The same formula serves to define a local time process at zero for processes derived from B in the following discussion by various conditioning and scaling operations, in particular for the corresponding Brownian or Bessel bridges. We denote by BB the standard bridge obtained by conditioning B to return to zero at time 1.

The following lemma is well known. (See Lévy (1939); Dynkin (1961); Kallenberg (1991); Barlow-Pitman-Yor (1989)).

Lemma 3.7 *The law of* $G_1 = \sup\{t : t \leqq 1, B_t = 0\}$ *is beta*$(\alpha, 1-\alpha)$. *Given* G_1, *the process* $(B_t, 0 \leqq t \leqq G_1)$ *is a bridge of length* G_1, *and the process* $(B(uG_1)/\sqrt{G_1}, 0 \leqq u \leqq 1)$, *is a* BB *independent of* G_1.

We will show that Theorem 3.1 and Lemma 3.7 imply the following central result of PY:

Theorem 3.8. (PY, Theorem 1.3.) *Let* $T = \inf\{t : S_t > 1\}$. *Let*

$$X_u = B(uT)/\sqrt{T}, \quad 0 \leqq u \leqq 1 \,. \tag{3.h}$$

Let U *be distributed uniformly on* $[0, 1]$, *independent of* B. *Let* (G, D) *be the maximal interval free of zeros of* X *which contains* U, *so* $G < U < D$ *and* $X_G = X_D = 0$. *Let* $U_1 = 1 - (D - G)$, *and let* Y *be the process with lifetime* U_1 *obtained by excising the excursion of* X *away from zero on the interval* (G, D), *and closing up the gap*:

$$Y_t = \begin{cases} X_t, & 0 \leqq t \leqq G \\ X_{D+t-G}, & G \leqq t \leqq U_1 \,. \end{cases}$$

Let $G_1 = \sup\{t : t \leqq 1, B_t = 0\}$. *Then there is the identity in distribution of processes*

$$(Y_t, 0 \leqq t \leqq U_1) \stackrel{d}{=} (B_t, 0 \leqq t \leqq G_1) \,. \tag{3.i}$$

Remark 3.9. The process $(X_u, 0 \leqq u \leqq 1)$ we call a *pseudo-bridge*. According to Biane et al. (1987), and PY Theorem 5.2, no matter what the dimension d, the law of the pseudo-bridge has density proportional to $1/S_{\mathrm{BB}}$ relative to that of the standard Brownian or Bessel bridge BB of the same dimension, where S_{BB} is the BB local time at zero up to time 1. This appears also as a byproduct of the following proof of Theorem 3.8. See Remark 3.13 below.

Remark 3.10. Lemma 3.7 and Corollary 2.3 (ii) show that $U_1 \overset{d}{=} G_1$, and that Theorem 3.8 amounts to

(3.j) *the process* $(Y(uU_1)/\sqrt{U_1}, 0 \leqq u \leqq 1)$, *is a* BB *independent of* U_1 .

In words, the process in (3.j) is derived as follows from the original Brownian or Bessel motion B run up to inverse local time T: delete a single excursion by length biased sampling, then rescale the remaining process of length $T_1 = U_1 T$ to have unit length, with Brownian scaling. Note that while the BB in (3.j) turns out to be independent of U_1, it is not independent of T or of T_1, because $T_1^{-\alpha}$ is the local time at 0 up to time 1 of the BB in (3.j), by scaling properties of local time.

Proof of Theorem 3.8. Let P_s^t denote the conditional distribution of $(B_u, 0 \leqq u \leqq T(s))$ given $T(s) = t$. By scaling,

(3.k) *the local time at zero up to time* 1 *of the process* $(X_u, 0 \leqq u \leqq 1)$ *is* $T^{-\alpha}$, *and*

(3.l) *conditionally given* $T^{-\alpha} = s$, *the process* $(X_u, 0 \leqq u \leqq 1)$ *has law* P_s^1 .

By application of Theorem 3.1,

(3.m) *conditionally given* $T^{-\alpha} = s$ *and* $U_1 = u$, *the process* $(Y_t, 0 \leqq t \leqq U_1)$ *has law* P_s^u .

On the other hand, by Lemma 3.7 above,

(3.n) *conditionally given* $S_1 = s$ *and* $G_1 = u$, *the process* $(B_t, 0 \leqq t \leqq G_1)$ *has law* P_s^u .

Now (3.i) follows from (3.m), (3.n) and the following equality in distribution (3.o). □

Lemma 3.11 *In the setup of Theorem* 3.8,

$$(T^{-\alpha}, U_1) \overset{d}{=} (S_1, G_1) \tag{3.o}$$

where G_1 *has beta*$(\alpha, 1-\alpha)$ *law, and*

$$S_1 = G_1^{\alpha} S_{\mathrm{BB}} \tag{3.p}$$

where S_{BB}, *independent of* G_1, *is the local time at* 0 *up to time* 1 *of the* BB *in Lemma* 3.7

Proof. Note first that (3.p) follows from Lemma 3.7 and the scaling property of local time. To derive (3.o), note that by definition of U_1

$$T^{-\alpha} = U_1^{\alpha} T_1^{-\alpha} , \tag{3.q}$$

where U_1 and T_1 are independent by Corollary 2.3. Moreover

(3.r) $$T^{-\alpha} \stackrel{d}{=} S_1, \quad \text{and} \quad U_1 \stackrel{d}{=} G_1 .$$

The first identity in (3.r) is an elementary application of scaling and the inverse relation between the processes $(T(s), s \geqq 0)$ and $(S_t, t \geqq 0)$, as shown in Sect. 2 of PY. The second is due to Corollary 2.3 and Lemma 3.7 above. Now (3.o) follows from (3.p), (3.q) and (3.r) as soon as it is verified that

(3.s) $$T_1^{-\alpha} \stackrel{d}{=} S_{\mathrm{BB}} .$$

But (3.r), (3.q) and (3.p) imply that $T_1^{-\alpha}$ and S_{BB} have identical positive integer moments. From (3.q), the common nth moment is the $n\alpha$th moment of T displayed in (2.n) divided by the $n\alpha$th moment of the beta$(\alpha, 1-\alpha)$ distribution of U_1. It is easily checked that the moment generating function converges in a neighbourhood of zero. So these moments determine a unique distribution. □

Remark 3.12. From (3.r) the density of S_1 is

(3.t) $$P(S_1 \in ds)/ds = \alpha^{-1} s^{-1-1/\alpha} f(s^{-1/\alpha}) ,$$

where f is the stable (α) density of T. And from (3.s) and (2.m),

(3.u) $$P(S_{\mathrm{BB}} \in ds) = sP(S_1 \in ds)/E(S_1) .$$

The assertion of Remark 3.9 follows from (3.k), (3.l) and (3.r) and (3.t).

Remark 3.13. Suppose, as in Corollary 3.5, that $B^{n^{\#}}$ is the process with lifetime T_n derived from B on the interval $[0, T]$ by length biased deletion of n excursions. According to Corollary 2.3, $T_n = TU_1U_2 \ldots U_n$ where the $U_i, 1 \leqq 1 \leqq n$ are independent beta $(i\alpha, 1-\alpha)$ variables, independent also of T_n. Let B^{n^*} be $B^{n^{\#}}$ reparameterized by $[0, 1]$ with Brownian scaling. The local time of B^{n^*}at 0 up to time 1 is $T_n^{-\alpha}$, and by (3.t) and a change of variables from (2.m)

(3.v) $$P(T_n^{-\alpha} \in ds) = K_n^{-1} s^n P(S_1 \in ds)$$

where $K_n = E(S_1^n) = E(T^{-n\alpha})$ as in (2.n). So by repeated application of Theorem 3.1, the process B^{n^*} is independent of $U_1, U_2, \ldots, U_n$ and B^{n^*} has law with density $K_1 K_n^{-1} S_{\mathrm{BB}}^{n-1}$ relative to the law of a BB. In particular B^{0^*} on $[0, 1]$ is the pseudobridge of Remark 3.9 with density $K_1 S_{\mathrm{BB}}^{-1}$ relative to the BB. And B^{1^*} is just a plain BB. But exactly as in the discussion of (2.o) and (2.p), with no new calculations, the whole set up can be shifted along to yield the following Corollary. Note from (3.s), (2.m) and (3.u) that the pth power S_{BB}^p of the BB local time is integrable iff $p > -2$:

Definition 3.14 For a real number $p > -2$, say a process with continuous paths $B_p = (B_p(t), 0 \leqq t \leqq 1)$ is a BB(p) if the law of B is absolutely continuous with respect to the law of a BB, with density proportional to the pth power of S_{BB}, the BB local time at 0 up to time 1.

Corollary 3.15 *Suppose B_p is a* BB(p), *where $p > -2$, and let S denote the local time of B_p at zero up to time 1. Let B_p^{n*} be the process derived from B_p by deletion of n excursions of B_p away from* 0 *selected by length-biased sampling and closing up the gaps. Then let $\bar{B}_p^{n*}$ be B_p^{n*} scaled onto* [0, 1] *by Brownian scaling. Then B_p^{n*} has lifetime $U_1 U_2 \ldots U_n$ where the $U_i, 1 \leqq 1 \leqq n$ are independent beta $((p+1+i)\alpha, 1-\alpha)$ variables, independent also of $\bar{B}_p^{n*}$, which is a* BB($p+n$) *whose local time at zero up to time* 1 *is $S/(U_1 U_2 \ldots U_n)^\alpha$.*

4 Sampling from a Poisson process

The purpose of this section is to present some general formulae for size-biased sampling from a Poisson point process is an abstract space S, where the "size" of a point at $x \in S$ is defined by an arbitrary strictly positive function $h(x)$. As examples, we have in mind

(i) the *subordinator case* of Sect. 2 with $S = (0, \infty)$ and $h(x) = x$;

(ii) the *excursion case* as in the proof of Theorem 3.1, with S the product of an interval $[0, s]$ on the local time scale and a path space describing excursions of a Markov process, and $h(x) = h(\sigma, \varepsilon)$ the duration of an excursion ε during which the local time is σ.

Assume throughout this section that

$$N(\cdot) = \sum_i 1(X_{(i)} \in \cdot) \tag{4.a}$$

is the counting process associated with some sequence of random variables $X_{(1)}, X_{(2)}, \ldots$ defined on a basic probability space $(\Omega, \mathbf{F}, P)$ with values in a measurable space $(S, \mathbf{S})$. We view N as a random counting measure with values in the space of all counting measures v on S, equipped with the σ-field generated by evaluations on $\mathbf{S}$ measurable sets. It is assumed that each $X_{(i)}$ is a measurable function of N. Typically the $X_{(i)}$ would be obtained from N by ordering the points of N in some deterministic way, so we use the common notation for order statistics. For a positive measurable function h defined on S, write vh for the v integral of h. For the rest of the section, we fix a *strictly positive* function h, and assume that

$$T \stackrel{\text{def}}{=} Nh = \sum_i h(X_{(i)}) < \infty \quad \text{a.s.} \tag{4.b}$$

Say that X_1 *is picked from N by h-biased sampling,* if

$$P(X_1 = X_{(i)} | X_{(1)}, X_{(2)}, \ldots) = h(X_{(i)})/T\,. \tag{4.c}$$

Now suppose N is a PPP(μ), that is a Poisson point process with mean measure μ on S, assumed to be σ-finite. Note that we assume the random variables $X_{(i)}$ in (4.a) are defined everywhere on Ω, so $N(S) = \infty$ and hence $\mu(S) = \infty$. But we note that the following results are easily modified to allow the case $\mu(S) < \infty$, corresponding to $X_{(i)}$ defined only for a Poisson random number of terms.

For N a PPP(μ), it is well known that our assumption $T < \infty$ a.s. amounts to finiteness of

$$F(y) \stackrel{\text{def}}{=} \int (1 - e^{-h(x)y}) \mu(dx) = -\log(E e^{-yT}) \tag{4.d}$$

for some $y > 0$, hence for all $y > 0$. Here T has an infinitely divisible law, just as in (1.c) and Sect. 2. Each $\Delta_{(i)} = h(X_{(i)})$ represents a jump in the sum for T. These $\Delta_{(i)}$ are the points of a PPP(Λ) on $(0, \infty)$, where Λ is the μ distribution of h. By suitable definition of $X_{(i)}$ it can even be arranged that the $\Delta_{(i)}$ are ranked as in (1.a).

Lemma 4.1 *Let N_1 be the point process derived from N, a* PPP(μ), *by deletion of a point X_1 picked from N by h-biased sampling. Then*

$$\frac{P(X_1 \in \mathrm{d}x, N_1 \in \mathrm{d}v)}{\mu(\mathrm{d}x)P(N \in \mathrm{d}v)} = \frac{h(x)}{h(x) + vh}. \tag{4.e}$$

Proof. Since it is assumed that the $X_{(i)}$ are N-measurable, (4.c) can be rewritten as

$$P(X_1 \in \mathrm{d}x \mid N) = h(x)N(\mathrm{d}x)/Nh.$$

Now (4.e) follows easily from the characterization of a PPP(μ) via Palm distributions. See for instance Daley and Vere-Jones (1988, Sect. 12.1), or Lemma 2.2 of PY. □

To match notation with previous sections, set $\Delta_1 = h(X_1)$, $T_1 = T - \Delta_1 = N_1 h$. Formula (4.e) can be developed in a number of different ways to display the joint laws of various subcollections of N, N_1, X_1, T, T_1, and Δ_1. Formulae (2.f) and (3.g) are examples in point. The following formulae (4.f) to (4.i) present the most important features in the present setting.

$$\frac{P(\Delta_1 \in \mathrm{d}y, T_1 \in \mathrm{d}t)}{\Lambda(\mathrm{d}y)P(T \in \mathrm{d}t)} = \frac{y}{y+t}. \tag{4.f}$$

This formula, valid without any assumption of densities for the distribution of T or its Lévy measure Λ, implies (2.a) in case Λ has a density, and in any case there is a formula for the Radon-Nikodym derivative

$$d_1(t) \stackrel{\text{def}}{=} \frac{P(T_1 \in \mathrm{d}t)}{P(T \in \mathrm{d}t)} = \int \frac{y\Lambda(\mathrm{d}y)}{y+t} = \int \frac{h(x)\mu(\mathrm{d}x)}{h(x)+t}. \tag{4.g}$$

Inspection of (4.e) shows that N_1 and X_1 are conditionally independent given T_1, with

$$P(X_1 \in \mathrm{d}x \mid T_1 = t, N_1) = d_1(t)^{-1}\,\mu(\mathrm{d}x)\frac{h(x)}{h(x)+t}. \tag{4.h}$$

Integrating out x in (4.e) shows that the distribution of N_1 has a density at v relative to the PPP(μ) law of N which depends on v only through vh. By the usual factorization criterion for sufficient statistics, the conditional independence (4.h) gives also the following result, which amounts to Theorem 3.1 in the excursion case:

$$P(N_1 \in \cdot \mid T_1 = t, X_1, T) = P(N \in \cdot \mid T = t) \stackrel{\text{def}}{=} Q(t, \cdot), \tag{4.i}$$

an equality which holds for a.e. t relative to the law of either T or T_1, by (4.g). Here we suppose S is a nice space, say Polish, so $Q(t, \cdot)$ is a well defined regular conditional distribution for N given $T = t$.

Returning to the general setup of (4.a) and (4.b), without the PPP assumption for a moment, suppose that $X_1, X_2, \ldots$ is obtained by repeated h-biased sampling without replacement from the points of N. Put more carefully, X_1 is picked from N by h-biased sampling. Assuming that $X_1, \ldots, X_n$ have been defined, let N_n be the process of points left behind after removal of $X_1, \ldots, X_n$ from N. Then X_{n+1} is picked from N_n by h-biased sampling, in such a way that X_{n-1} and $X_1, \ldots, X_n$ are conditionally independent given N_n. Notice that

$$T_n \stackrel{\text{def}}{=} N_n h = T - \sum_{i=1}^{n} h(X_i) = \sum_{i=n+1}^{\infty} h(X_i) \quad \text{a.s.} \,. \tag{4.j}$$

Here the last equality holds almost surely because we assume $h > 0$ and $T < \infty$ a.s., from which it follows easily that the whole sequence $X_1, X_2, \ldots$ obtained by repeated h-biased sampling is almost surely an exhaustive sample of the points $X_{(1)}, X_{(2)}, \ldots$ of N. Put another way:

$$N(\cdot) = \sum_i 1(X_{(i)} \in \cdot) = \sum_i 1(X_i \in \cdot) \quad \text{a.s.} \,.$$

To keep this in mind we may call $X_1, X_2, \ldots$ an h-biased random permutation of $X_{(1)}, X_{(2)}, \ldots$.

Theorem 4.2 *Let $X_1, X_2, \ldots$ be an h-biased random permutation of the points of a* PPP(μ). *Let N_n and T_n and $Q(t, \cdot)$ be as defined above. Let $\mu(t, \cdot)$ be a μ regular conditional probability law for x given $h(x) = t$. Then*
(i) *The sequence $T, T_1, T_2, \ldots$ is a Markov chain with stationary transition probabilities, which are as in* (2.a) *in case the Lévy measure Λ has a density.*
(ii) *Conditionally given $T = t_0, T_1 = t_1, \ldots, T_n = t_n$, the remaining point process N_n has law $Q(t_n, \cdot)$, independently of the X_i for $1 \leqq i \leqq n$, which are independent, with laws $\mu(t_i - t_{i-1}, \cdot)$.*

Proof. This follows easily by iteration of formula (4.i). □

To illustrate, in the excursion case $Q(t, \cdot)$ governs a process of excursions over a fixed interval $[0, s]$ on the local time scale with given total duration t, while $\mu(t, \cdot)$ is the product of the uniform law on $[0, s]$ and the law for excursions of given length t obtained by a disintegration of Itô's excursion law. In the Bessel case the law for an excursion of length t would be described by Brownian scaling of a standard Bessel excursion of length 1.

Corollary 4.3 *The law of the point process N_n remaining after n points have been removed by h-biased sampling from N, a* PPP(μ), *is mutually absolutely continuous with respect to the original law of N, with density*

$$\frac{P(N_n \in \mathrm{d}v)}{P(N \in \mathrm{d}v)} = d_n(vh), \quad \text{where} \quad d_n(t) = \frac{P(T_n \in \mathrm{d}t)}{P(T \in \mathrm{d}t)} = \frac{t}{n!} \int_0^\infty F(y) y^n \mathrm{e}^{-yt} \mathrm{d}y \,. \tag{4.k}$$

Proof. Theorem 4.2 implies $P(N_n \in \cdot) = \int Q(t, \cdot) P(T_n \in \mathrm{d}t)$, which gives all of (4.k) except the final formula. This is obtained recursively as follows. Start from $d_0(t) = 1$, use the Markov property of $T, T_1, T_2, \ldots$ to extend the result (4.g) for $n = 0$ to $n = 1, 2, \ldots$ as follows:

$$d_{n+1}(t) = \int_0^\infty \frac{y \Lambda(\mathrm{d}y)}{y + t} \mathrm{d}_n(y + t) \mathrm{d}y \,.$$

This yields the formula for $d_n(t)$ by induction. □

While it is easy enough to prove the formula for $d_n(t)$ this way, it is not clear from this approach how one would ever guess that there would be such a formula. In fact we discovered this formula, and the following companion formulae for $n = 1, 2, \ldots$

$$\text{(4.l)} \qquad \frac{P(X_n \in \mathrm{d}x)}{\mu(\mathrm{d}x)} = \frac{h(x)}{(n-1)!} \int_0^\infty \mathrm{e}^{-h(x)y} \mathrm{e}^{-F(y)} F(y)^{n-1} \, \mathrm{d}y \,,$$

$$\text{(4.m)} \qquad \frac{P(X_n \in \mathrm{d}x, T_n \in \mathrm{d}t)}{\mu(\mathrm{d}x) P(T \in \mathrm{d}t)} = \frac{h(x)}{(n-1)!} \int_0^\infty \mathrm{e}^{-h(x)y - ty} \mathrm{e}^{-F(y)} F(y)^{n-1} \, \mathrm{d}y \,,$$

as consequences of Theorem 4.5 below. This theorem brings out the fact that N_n is a Cox process (i.e. a mixture of Poisson point processes) with a simple random intensity measure. This theorem is a development of Theorem 3.1 of PY, and a result of Tavaré (1987) in the case of jumps of a gamma process, stated as Corollary 4.7 below. The basic idea is that in a setup with suitable additional randomization we can find something to condition on, which makes X_n and N_n mutually independent, and N_n conditionally Poisson. Things like the left side of (4.m) are then easily computed by conditioning this way and then integrating out. But any such formula concerning the joint law of $X_1, \ldots, X_n$ and N_n must of course hold for any scheme of h-biased sampling. See also Pitman (1991) for analogous formulae in a setting of size-biased sampling without replacement from a sequence of i.i.d. random variables.

Lemma 4.4 (General representation of h-biased sampling)
Let $X_{(i)}$ be an arbitrary sequence of random variables, h a strictly positive function such that $\sum_i h(X_{(i)}) < \infty$ a.s. If

(4.n) $Y_i = \varepsilon_i / h(X_{(i)})$ *where the ε_i are i.i.d. standard exponential variables, and*

(4.o) $Y_{(1)} < Y_{(2)} < \cdots$ *are the order statistics of the Y_i's, and $X_1, X_2, \ldots$ are the corresponding X values, then*

(4.p) $X_1, X_2, \ldots$ *is an h-biased random permutation of the $X_{(i)}$.*

Proof. Immediate from standard properties of independent exponential variables.

Theorem 4.5 (Representation of h-biased sampling of a PPP(μ)
Suppose that $h > 0$ and μ are such that $F(y)$ in (4.d) *is finite. If*

$$\text{(4.q)} \qquad N(\cdot) = \sum_i 1(X_{(i)} \in \cdot) \quad \textit{is a } \mathrm{PPP}(\mu) \,,$$

and $X_1, X_2, \ldots$ is an h-biased radom permutation of the $X_{(i)}$ constructed as in (4.n) *and* (4.o), *then the following conditions hold:*

(4.r) $\sum_i 1[(X_{(i)}, Y_i) \in \cdot] = \sum_i 1[(X_i, y_{(i)}) \in \cdot]$ *is a* $\mathrm{PPP}[\mu(\mathrm{d}x) h(x) \mathrm{e}^{-h(x)y} \mathrm{d}y]$;

(4.s) $Y_{(1)} < Y_{(2)} < \cdots$ *are the points of a* $\mathrm{PPP}[F'(y)\mathrm{d}y]$ *on* $(0, \infty)$, *where*

$$F'(y) = \frac{\mathrm{d}}{\mathrm{d}y} F(y) = \int h(x) \mathrm{e}^{-h(x)y} \mu(\mathrm{d}x) \,.$$

(4.t) *given all the* $Y_{(i)}$, *the* X_i *are independent, with distributions* $P(Y_{(i)}, \cdot)$, *where*

$$P(y, \mathrm{d}x) = F'(y)^{-1} h(x) \mathrm{e}^{-h(x)y} \mu(\mathrm{d}x) .$$

Conversely, starting from a bivariate sequence of random variables $(X_i, Y_{(i)})$ *satisfying* (4.s) *and* (4.t), *let* $N(\cdot) = \sum_i 1(X_i \in \cdot)$, *let* $X_{(i)}$ *be defined as some N-measurable ordering of the* X_i, *and let* Y_i *correspond to* $X_{(i)}$. *Then* (4.n), (4.o), (4.p), (4.q), *and* (4.r) *all hold.*

Proof. It is well known that
(i) if the points of a PPP(μ) on S are assigned marks independently according to a transition probability function $Q(x, \mathrm{d}y)$ from S to some other space, the result is a PPP[$\mu(\mathrm{d}x)Q(x, \mathrm{d}y)$] on the product space, and
(ii) any PPP[$\mu(\mathrm{d}x)Q(x, \mathrm{d}y)$] on a product space can be decomposed as in (i).
Here the measure on the product space is

$$\mu(\mathrm{d}x) h(x) \mathrm{e}^{-h(x)y} \mathrm{d}y = F'(y) \mathrm{d}y \, P(y, \mathrm{d}x) . \tag{4.u}$$

The theorem follows by repeated application of (i) and (ii). □

Corollary 4.6 *Under the assumptions of Theorem* 4.5, *the following also obtain:*

$$P(Y_{(n)} \in \mathrm{d}y)/\mathrm{d}y = \Gamma(n)^{-1} F'(y) \mathrm{e}^{-F(y)} F(y)^{n-1}, \quad y > 0; \tag{4.v}$$

(4.w) *given* $Y_{(n)} = y$, *the counting process* N_n *associated with* $X_{n+1}, X_{n+2}, \ldots$ *is a* PPP[$\mathrm{e}^{-h(x)y} \mu(\mathrm{d}x)$], *independent of* $X_1, \ldots X_n$, *so*

$$P(N_n \in \mathrm{d}v \mid Y_{(n)} = y, X_1, \ldots X_n) = P(N \in \mathrm{d}v) \exp[F(y) - (vh)y] .$$

(4.x) *given* $Y_{(n)} = y$, *the counting process associated with* $X_1, \ldots, X_{n-1}$ *is distributed like the empirical distribution of* $n - 1$ i.i.d. *variables with law* $F(y)^{-1}[1 - \mathrm{e}^{-h(x)y}]\mu(\mathrm{d}x)$.

Proof. Immediate from Theorem 4.5 by standard properties of Poisson point processes. □

Remark 4.7. Formulae (4.k), (4.l) and (4.m) now follow easily from (4.v), (4.w) and (4.t) by integration, with an integration by parts for (4.k). As a check on (4.l), notice that the sum on n of the right side of (4.l) is identically equal to 1. This is obvious a priori, because $N(\cdot) = \sum_i 1(X_i \in \cdot)$, so the sum is the density of the mean measure of N with respect to μ. As a check on (4.k), in the stable case $F(y) = cy^\alpha$, the simple formulae (2.m) and (2.n) are recovered immediately. The gamma case does not simplify so pleasantly. For T as in (2.g) with $\lambda = 1$ and $a > 0$, $F(y) = a \log(1 + y)$, the density at t in (4.k) becomes

$$\frac{a^n}{\Gamma(n)} \int_0^\infty \frac{[\log(1 + y)]^{n-1}}{1 + y} \mathrm{e}^{-yt} \mathrm{d}y = \frac{a^n}{\Gamma(n)} \mathrm{e}^t \int_1^\infty v^{-1} \log(v)^{n-1} \mathrm{e}^{-vt} \mathrm{d}v . \tag{4.y}$$

As a final remark in the gamma case, we deduce the following result which we learned from Simon Tavaré. This result appears in Tavaré (1987) without the refinement that the random permutation is size-biased.

Corollary 4.8 (Representation of the size-biased jumps of a gamma process) *Let* V_i *and* W_i *be two independent sequences of i.i.d standard exponentials. Fix* $a > 0$ *and* *let* $S_n = V_1 + \cdots + V_n, X_n = W_n \exp(-S_n/a)$. *Then* $X_1, X_2, \ldots$ *is a size-biased random permutation of the points of a* $\mathrm{PPP}[ax^{-1}e^{-x}dx]$, *whose sum has gamma distribution with shape parameter a.*

Proof. This is the converse part of Theorem 4.5 in the case $S = (0, \infty)$, $\mu(dx) = ax^{-1}e^{-x}dx$, $h(x) = x$, $F(y) = a\log(1 + y)$. Set $Y_{(n)} = \exp(S_n/a) - 1$. Then conditions (4.s) and (4.t) are immediate from the definitions.

References

Aldous, D., Pitman, J.: Brownian bridge asymptotics for random mappings. (Preprint, 1991)

Barlow, M., Pitman, J., Yor, M.: Une extension multidimensionnelle de la loi de l'arc sinus. Séminaire de Probabilités XXIII. (Lect. Notes Math., vol. 1372, pp. 294–314) Berlin Heidelberg New York: Springer 1989

Biane, P., Le Gall, J.F., Yor, M.: Un processus qui ressemble au pont Brownien. Séminaire de Probabilités XXI. (Lect. Notes Math., vol. 1247, pp. 270–275) Berlin Heidelberg New York: Springer 1987

Blackwell, D., MacQueen, J.B.: Ferguson distributions via Pòlya urn schemes. Ann. Stat. **1**, 353–355 (1973)

Brockett, P.L., Hudson, W.N.: Zeros of the densities of infinitely divisible measures. Ann. Probab. **8**, 400–403 (1980)

Daley, D.J., Vere-Jones, D.: An introduction to the theory of point processes. Berlin Heidelberg New York: Springer 1988

Donnelly, P., Joyce, P.: Continuity and weak convergence of ranked and size-biased permutations on the infinite simplex. Stochastic Processes Appl. **31**, 89–103 (1989)

Dynkin, E.B.: Some limit theorems for sums of independent random variables with infinite mathematical expectations. Trans. Math. Stat. Probab. **1**, 171–189 (1961)

Ewens, W.J.: Population genetics theory — the past and the future. In: Lessard, S. (ed.) Mathematical and statistical problems in evolution. University of Montreal Press. Montreal, 1988

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

Hoppe, F.M.: The sampling theory of neutral alleles and an urn model in population genetics. J. Math. Biol. **25**, 123–159 (1987)

Itô, K.: Poisson point processes attached to Markov processes. Proc. Sixth Berkeley Symp. Math. Statist. Prob. University of California Press, Berkeley, pp. 225–239, 1970

Kallenberg, O.: Splitting at backward times in regenerative sets. Ann. Probab. **9**, 781–799 (1981)

Kingman, J.F.C.: Random discrete distributions. J.R. Stat. Soc. **37**, 1–22 (1975)

Lukacs, E.: A characterization of the gamma distribution. Ann. Math. Stat. **26**, 319–324 (1955)

Lévy, P.: Sur certains processus stochastiques homogènes. Compos. Math. **7**, 283–339 (1939)

Molchanov, S.A., Ostrovski, E.: Symmetric stable processes as traces of degenerate diffusion processes. Theory of Probab. Appl. **14**, (1) 128–131 (1969)

McCloskey, J.W.: A model for the distribution of individuals by species in an environment. Unpublished Ph.D. thesis, Michigan State University, 1965

Patil, G.P., Taillie, C.: Diversity as a concept and its implications for random communities, Proc. 41st. I.S.I. New Delhi, 497–515 (1977)

Perman, M.: Random discrete distributions derived from subordinators. Ph.D. thesis, Dept Statistics, U.C. Berkeley, 1990

Perman, M.: Order statistics for normalised jumps of subordinators. (Preprint, 1991)

Pitman, J.: Size-biased sampling of independent sequences. (Preprint, 1991)

Pitman, J., Yor, M.: Arcsine laws and interval partitions derived from a stable subordinator. Proc. Lond. Math. Soc. (to appear)

Salisbury, T.S.: Construction of right processes from excursions. Z. Wahrscheinlichkeitstheor. Verw. Geb. **73**, 351–367 (1986)

Shepp, L.A., Lloyd, S.P.: Ordered cycle lengths in a random permutation. Trans. Am. Math Soc. **121**, 340–357 (1966)

Tavaré, S.: The birth process with immigration, and the genealogical structure of large populations. J. Math Biol. **25**, 161–168 (1987)
Vershik, A.M., Schmidt, A.A.: Limit measures arising in the asymptotic theory of symmetric groups I. Theory Probab. Appl. **22**, 70–85 (1977)
Vershik, A.M., Schmidt, A.A.: Limit measures arising in the asymptotic theory of symmetric groups II. Theory Probab. Appl. **23**, 36–49 (1978)
Watterson, G.A.: The stationary distribution of the infinitely many neutral alleles diffusion model. J. Appl. Probab. **13**, 639–651 (1976)
Williams, D.: Path decomposition and continuity of local time for one-dimensional diffusions. Proc. London Math. Soc. **28**, 738–768 (1974)