STICHTING
MATHEMATISCH CENTRUM
2e BOERHAAVESTRAAT 49
AMSTERDAM

AFDELING MATHEMATISCHE STATISTIEK

REPORT S 314 (VP 21)

Random division of an interval

by

F.W. Steutel

Preliminary Report

May 1963

1 Introduction and summary

Distributions arising from the random division of an interval have been studied for a long time. The distribution function of e.g. the length of the largest interval was computed by Whitworth [10] in 1897 and by Girault [2] in 1962. These and most other computations are based on combinatorial methods. The relation between the random division of an interval and the Poisson process is well known, but has not been used very extensively, although it provides an often effective alternative approach.
It is used - not in a very efficient manner - by Moran [6] in connection with the distribution of the sum of the squared interval lengths. Dwass [1] uses an other form of this relation to obtain the distributions of linear combinations of interval lengths. These distributions have also been studied by Mauldon ([4] and [5]), who uses a special type of integral transform. In this paper, which is of an expository character, we use the relation between the random division of an interval and the Poisson process in a manner that unifies and generalizes the methods of Mauldon and Dwass. Some applications are given, most of which are known.

2. Definitions and notations

We consider the following situation:

the interval $(0,t]$ is divided into n sub-intervals by n-1 random points, i.e. points, which are drawn independently from a rectangular distribution on $(0,t]$. Denoting the coordinates of these points (in increasing order) by $\underline{y}_{(1)}(t),\ldots,\underline{y}_{(n-1)}(t)$ [1] and putting $\underline{y}_{(0)}(t) = 0$, $\underline{y}_{(n)}(t)=t$ for the lenghts $\underline{x}_1(t),\ldots,\underline{x}_n(t)$ of the sub-intervals we have

$$\underline{x}_j(t) = \underline{y}_j(t)-\underline{y}_{j-1}(t) \quad (j=1,2,\ldots,n).$$

---

1) Random variables are denoted by underlined symbols.

We will be concerned with the distributions of functions of the $\underline{x}_j(t)$.

$\underline{x}_{(1)}(t),\ldots,\underline{x}_{(n)}(t)$ will denote the $\underline{x}_j(t)$ in increasing order. As we will often take t=1 it will be convenient to omit the argument in that case. So we write e.g.

$$\underline{x}_j = \underline{x}_j(1) \; .$$

For the same reason we write $\underline{u}_j=\underline{u}_j(1)$ etc. for the functions of τ we will meet in the following sections.

3 Relation with Poisson process

The content of the following lemma is well known:

Lemma: if $\underline{u}_1(\tau),\ldots,\underline{u}_n(\tau)$ are independent random variables with the common distribution function $1-e^{-\tau u}$ ($u \geq 0$, τ a positive constant) and if $\underline{x}_1(t),\ldots,\underline{x}_n(t)$ are random variables as defined in section 2, then the conditional distribution of $(\underline{u}_1(\tau),\ \underline{u}_1(\tau)+\underline{u}_2(\tau),\ldots,\ \underline{u}_1(\tau)+\ldots+\underline{u}_{n-1}(\tau))$, given that $\underline{u}_1(\tau)+\ldots+\underline{u}_n(\tau)=t$, is the same as the distribution of $(\underline{x}_1(t),\ \underline{x}_1(t)+\underline{x}_2(t),\ldots,\underline{x}_1(t)+\ldots+\underline{x}_{n-1}(t)) = (\underline{y}_1(t),\ldots,\underline{y}_{n-1}(t))$.

proof: considering the density functions $f(z_1,\ldots,z_{n-1}|t)$, $g(z_1,\ldots,z_n)$ and $h(z_n)$ of $(\underline{u}_1(\tau),\ldots,\underline{u}_1(\tau)+\ldots+\underline{u}_{n-1}(\tau)\ |\ \underline{u}_1(\tau)+\ldots+\underline{u}_n(\tau)=t)$, $(\underline{u}_1(\tau),\ldots,\underline{u}_1(\tau)+\ldots+\underline{u}_n(\tau))$ and $\underline{u}_1(\tau)+\ldots+\underline{u}_n(\tau)$ respectively, we have

$$f(z_1,\ldots,z_{n-1}|t)=\frac{g(z_1,\ldots,z_{n-1},t)}{h(t)}=$$

$$=(n-1)!\,\frac{\exp\left[\tau\{z_1+(z_2-z_1)+\ldots+(t-z_{n-1})\}\right]\tau^n}{\tau^n\ t^{n-1}\ \exp\ (\tau t)}=\frac{(n-1)!}{t^{n-1}}\ ,$$ the density

function of $(\underline{y}_1(t),\dots,\underline{y}_{n-1}(t))$.

From this lemma we immediately have the well known

<u>Theorem 1:</u> if $\underline{u}_1(\tau),\dots,\underline{u}_n(\tau)$ are independent random variables with distribution function $1-e^{-\tau u}$ $(u \geq 0,\ \tau$ a positive constant) and if $\underline{x}_1(t),\dots,x_n(t)$ are random variables as defined in section 2, then the conditional distribution of $(\underline{u}_1(\tau),\dots,\underline{u}_n(\tau))$ given that $\underline{u}_1(\tau)+\dots+\underline{u}_n(\tau)= t$ is the same as the distribution of $(\underline{x}_1(t),\dots,\underline{x}_n(t))$, i.e. we have

$$(1)\qquad P\{\underline{x}_1(t)\leq x_1,\dots,\underline{x}_n(t)\leq x_n\} = P\{\underline{u}_1(\tau)\leq x_1,\dots,\underline{u}_n(\tau)\leq x_n \mid \underline{u}_1(\tau)+\dots+u_n(\tau)=t\}.$$

From (1) we get, multiplying both sides by $\frac{\tau^n}{(n-1)!}\, t^{n-1}e^{-\tau t}$ and integrating,

$$(2)\qquad \frac{\tau^n}{(n-1)!}\int_0^\infty P\{\underline{x}_1(t)\leq x_1,\dots,\underline{x}_n(t)\leq x_n\}\, t^{n-1}e^{-\tau t}dt = P\{\underline{u}_1(\tau)\leq x_1,\dots,\underline{u}_n(\tau)\leq x_n\} = \prod_{j=1}^{n}(1-e^{-\tau x_j}).$$

If now $f(x_1,\dots,x_n)$ is a Borel-measurable function and if by $\mathcal{E}\underline{a}$ we denote the expectation of $\underline{a}$, then in the same way we prove

$$(3)\qquad \frac{\tau^n}{(n-1)!}\int_0^\infty \mathcal{E} f(\underline{x}_1(t),\dots,x_n(t))\ t^{n-1}e^{-\tau t}dt= \mathcal{E} f(\underline{u}_1(\tau),\dots,\underline{u}_n(\tau)),$$

if $\int_0^\infty \mathcal{E}|f(\underline{x}_1(t),\dots,\underline{x}_n(t))|\ t^{n-1}e^{-\tau t}dt < \infty$. Hence

$$(4)\qquad \int_0^\infty \mathcal{E}\ f(\underline{x}_1(t),\dots,\underline{x}_n(t))t^{n-1}e^{-\tau t}dt = \frac{(n-1)!}{\tau^n}\ \mathcal{E}\ f(\underline{u}_1(\tau),\dots,\underline{u}_n(\tau))$$

and therefore the left-hand side of (4) is in fact the Laplace-transform with respect to τ [2] of $\mathcal{E}f(\underline{x}_1(t),\dots,\underline{x}_n(t))t^{n-1}$. The $\underline{u}_j(\tau)$ are independent, so $\mathcal{E}f(\underline{u}_1(\tau),\dots,\underline{u}_n(\tau))$ is often easy to compute; $\mathcal{E}f(\underline{x}_1(t),\dots,\underline{x}_n(t))$ may then be obtained by an (often quite simple) inversion. Substituting in (4)

$$(5) \qquad f(x_1,\dots,x_n)=\begin{cases}1 & \text{if } g(x_1,\dots,x_n)\leq z\\ 0 & \text{other wise}\end{cases}$$

we obtain the distribution function of $g(\underline{x}_1(t),\dots,\underline{x}_n(t)$.

A formula equivalent to the Laplace-inversion of (4) occurs in Pollaczek [8] [3], where it is derived as a formal indentity without interpretation. A special case of (4) is used in Dwass [1].

4 Linear combinations

Mauldon [4] , [5] and Dwass [1] consider linear combinations of $\underline{x}_j(t)$ and $\underline{x}_{(j)}(t)$. As for the ordered variables $\underline{u}_{(j)}(\tau)$ we may write

$$\underline{u}_{(j)}(\tau) = \frac{\underline{u}_1(\tau)}{n} + \frac{\underline{u}_2(\tau)}{n-1} + \dots + \frac{\underline{u}_{(j)}(\tau)}{n-j+1}$$

(see e.g. Renyi [9]) we have by (2) the well known representation

$$(6) \qquad \underline{x}_{(j)}(t)= \frac{\underline{x}_1(t)}{n} + \frac{\underline{x}_2(t)}{n-1} + \dots + \frac{\underline{x}_j(t)}{n-j+1} .$$

----

2) Although relation (4) is derived for positive τ it may be extended (without interpretation) by analytic continuation to all complex values of τ , for which the left-hand side of (4) is absolutely convergent.

3) This was pointed out to me by Prof. Runnenburg, to whom I am indebted for some useful suggestions.

Hence we need only consider linear combinations of the $\underline{x}_j(t)$. If

$$\underline{a}_n(t)=\sum_{j=1}^{n}\alpha_j\underline{x}_j(t),$$

where the α_j are real constants, we have by (4)

$$(7)\quad \int_0^\infty \mathcal{E}e^{-s\underline{a}_n(t)}t^{n-1}e^{-\tau t}dt=\frac{(n-1)!}{\tau^n}\mathcal{E}e^{-s\sum\alpha_j\underline{u}_j(\tau)}=$$

$$=(n-1)!\prod_{j=1}^{n}\frac{1}{s\alpha_j+\tau}.$$

Dwass [1] uses

$$\frac{\tau^n}{(n-1)!}\int_0^\infty e^{-a(\underline{x}_1+\ldots+\underline{x}_k)t}\,t^{n-1}e^{-\tau t}dt=\left(\frac{\tau}{a+\tau}\right)^k$$

and the fact (not proved in [1]) that the density function $g_{a,k}(z)$ of $a(\underline{x}_1+\ldots+\underline{x}_n)$ is given by

$$g_{a,k}(z)=\frac{1}{a}(n-1)\binom{n-2}{k-1}\left(\frac{z}{a}\right)^{k-1}\left(1-\frac{z}{a}\right)^{n-k-1}\iota(a-z)\quad(a>0),$$

as is easily seen from (4) and (5).
His results may be obtained from (7) by partial fraction expansion of the right-hand side followed by Laplace-Stieltjes-inversion with respect to s and Laplace-inversion with respect to τ . The general formula is quite complicated and will not be given.

As, on the other hand, $\underline{a}_n(t)$ and $\underline{a}_n t$ have the same distribution, the first member of (7) may be written

$$\int_0^\infty\int_0^\infty e^{-sat}dF_n(a)t^{n-1}e^{-\tau t}dt=(n-1)!\int_0^\infty(sa+\tau)^{-n}dF_n(a),$$

$F_n(a)$ de noting the distribution function of $\underline{a}_n$, and so by (7)

$$(8)\quad \mathcal{E}(s\underline{a}_n+\tau)^{-n}=\prod_{j=1}^{n}(s\alpha_j+\tau)^{-1},$$

which is a generalized Stieltjes transform.

Relation (8) was obtained for special cases in Mauldon [4] in a much more complicated way. The simplification is, of course, partly due to the fact that, be cause of (6), we have to consider only the independent $\underline{x}_j(t)$.

In Mauldon [6] the inversion of $\mathcal{E}(s\underline{a} + \tau)^{-n}$ is considered more generally. Analogous to the relation between the Stieltjes and Laplace transforms (see e.g. Widder [11]) we have in the same way as above: if

$$\varphi(s) = \mathcal{E} e^{is\underline{x}}$$

and

$$v(s) = \mathcal{E}(s - i\underline{x})^{-n},$$

then

$$v(s) = \int_0^\infty \frac{u^{n-1}}{(n-1)!} \varphi(u) e^{-su} du.$$

Hence the properties of v(s) should follow from the properties of the Fourier-Stieltjes and Laplace transforms.

To obtain explicit results however it is often more practical to start from (4) than from (7) or (8), especially in the simpler cases. If, for instance, we wish to know the distribution function of $\underline{x}_{(n-k+1)}(t)$, we have by (4)

$$\int_0^\infty P\{\underline{x}_{(n-k+1)}(t) \leq z\} t^{n-1} e^{-\tau t} dt = \frac{(n-1)!}{\tau^n} \sum_{j=0}^{k-1} \binom{n}{j} e^{-\tau jz} (1 - e^{-\tau z})^{n-j} =$$

$$= (n-1)! \sum_{j=0}^{k-1} \sum_{l=0}^{n-j} \binom{n}{j} \binom{n-j}{l} (-)^l \frac{e^{-\tau z(j+l)}}{\tau^n}.$$

Using

$$\int_0^\infty \frac{(t-\alpha)^{p-1}}{\Gamma(p)} \iota(t-\alpha) e^{-\tau t} dt = \frac{e^{-\alpha\tau}}{\tau^p} \qquad (\alpha > 0, p > 0),$$

where

$$\iota(x) = \begin{cases} 0 & x < 0 \\ 1 & x \geq 0 \end{cases},$$

we obtain

$$P\{\underline{x}_{(n-k+1)}(t) \leq z\} = \sum_{j=0}^{k-1} \sum_{l=0}^{n-j} \binom{n}{j} \binom{n-j}{l} (-)^l \left\{1 - \frac{(j+l)z}{t}\right\}^{n-1} \iota(t-(j+l)z),$$

as may be found in [10].

With little more difficulty the result in Mauldon [4] may be obtained by inverting directly

$$\frac{(n-1)!}{\tau^n} P\{ \underline{u}_{(n)}(\tau)+ \underline{u}_{(n-1)}(\tau)+\ldots+\underline{u}_{(n-k+1)}(\tau)\leq z\} ,$$

for which we may write

$$\binom{n}{k} \frac{(n-1)!}{(k-2)!} \int_0^z e^{-\tau u} u^{k-2} du \int_0^{z-u} \frac{e^{-\tau x}(1-e^{-\tau \frac{x}{k}})^{n-k}}{\tau^{n-k}} dx.$$

5 More general functions

The method used in section 4 is particularly suited to obtain the distributions of linear combinations of the $\underline{x}_j(t)$, for which the right-hand side of (4) is simple. We may use (4) to calculate the moments of more general functions. From (4) we have at once, substituting $\tau = 1$

Theorem 2: if $f(x_1,\ldots,x_n)$ is homogeneous of the order p, i.e. if $f(\lambda x_1,\ldots,\lambda x_n)= \lambda^p f(x_1,\ldots,x_n)$ and if $\mathcal{E}|f(\underline{x}_1,\ldots,\underline{x}_n)|$ is finite then

$$\mathcal{E} f(\underline{x}_1,\ldots,\underline{x}_n) = \frac{(n-1)!}{\Gamma(n+p)} \mathcal{E} f(\underline{u}_1,\ldots,\underline{u}_n). \tag{9}$$

From (9) we get for instance

$$\mathcal{E}\, \underline{x}_1^{\alpha_1-1} \ldots \underline{x}_n^{\alpha_n-1} = (n-1)!\, \frac{\Gamma(\alpha_1)\ldots\Gamma(\alpha_n)}{\Gamma(\alpha_1+\ldots+\alpha_n)},$$

as given in Kendall and Moran [3], where it is derived by a geometrical method.

Finally we consider the distribution of

$$\underline{q}_n(t) = \sum_{j=1}^{n} \underline{x}_j^2(t)$$

as studied by Moran [6], [7] and others.

Again the moments are easily obtained. Here (9) yields

$$\mathcal{E}\underline{q}_n^l = \frac{(n-1)!}{(n+2l-1)!}\mathcal{E}(\underline{u}_1^2+\ldots+\underline{u}_n^2)^l = \frac{(n-1)!}{(n+2l-1)!}\sum_{j_1+\ldots+j_n=l}\frac{l!}{j_1!\ldots j_n!}\times$$

$$\times(2j_1)!\ldots(2j_n)!,$$

as given in Moran [6].

The use of (4) yields for the distribution function

$$G_n(z;t) = G_n\left(\frac{z}{t^2}\right) \text{ of } \underline{q}_n(t)$$

$$(10) \qquad \int_0^\infty G_n(z;t)\, t^{n-1}e^{-\tau t}dt = (n-1)! \underset{\substack{u_1^2+\ldots+u_n^2\le z \\ u_1>0,\ldots u_n>0}}{\int\ldots\int} e^{-\tau(u_1+\ldots+u_n)}du_1\ldots du_n.$$

Using

$$\int_0^\infty \delta(t-c)e^{-\tau t}dt = e^{-\tau c},$$

where $\delta(t)$ denotes Dirac's δ-function, by (10) we evidently have

$$(11) \qquad G_n(z;t)\, t^{n-1} = (n-1)! \underset{\substack{u_1^2+\ldots+u_n^2\le z \\ u_1+\ldots+u_n=t \\ u_1>0,\ldots,u_n>0}}{\int\ldots\int} du_1\ldots du_{n-1},$$

which may be interpreted geometrically as e.g. in Moran [6]. From (11) it follows that.

$$(12) \qquad G_n(z^2) = (n-1)\int_0^z G_{n-1}\left(\frac{z^2-u^2}{(1-u)^2}\right)(1-u)^{n-2}\,du.$$

From (11) one may calculate $G_n(z^2)$ for small values of n starting from $G_1(z) = \iota(z-1)$. The expressions become rapidly awkward with increasing n.

Finally, it is not difficult to prove by induction from (11) that

$$G_n(z)=\frac{\iota(z-\frac{1}{n})(n-1)!\,\pi^{\frac{n-1}{2}}(z-\frac{1}{n})^{\frac{n-1}{2}}}{\sqrt{n}\,\Gamma(\frac{n+1}{2})} \quad \text{for } z\leq\frac{1}{n-1}\ .$$

An expression extenting to values of $z\leq\frac{1}{n-2}$ is obtained in Moran [7] by geometrical methods. On the whole the distribution function of $\underline{q}_n$ remains rather untractable.

Concluding one may say that the method treated here yields most known results with little difficulty, while some of them, e.g. the distribution function of $\underline{x}_{(j)}$, are obtained with surprisingly little effort. In those cases, where the right-hand side of (4) is simple, the method will be successfully applicable. In other cases, as in our final example, little seems to be gained. Perhaps the method may serve to give some more insight in the problems connected with the random division of an interval.

References

[1] Dwass, M., The distribution of linear combinations of random divisions of an interval, Trabajos de Estadistica XII- I and II, (1961).

[2] Girault, M., Loi de probabilité du plus grand intervalle dans une parlition "au hasard". Revue de statistique appliquée, Vol X-2 (1962).

[3] Kendall. M.G and Moran, P.A.P., Geometrical Probability, Charles Griffin, London, (1963).

[4] Mauldon, J.G., Random division of an interval, Proc Cambridge Phil. Soc. 47 (1951)

[5] Mauldon, J.G., An inversion formula for a generalized transform, Mathematika, Vol.4, (1957).

[6] Moran, P.A.P., The random division of an interval I, J. Roy.Stat.Soc. suppl. 9 (1947).

[7] Moran, P.A.P., The random division of an interval III, J.Roy.Stat.Soc. XV (1953).

[8] Pollaczek, F., Problèmes stochastiques posés par le phénomène de formation d'une queue d'attente à un guichet et par des phénomènes apparentés, Gauthier-Villars, Paris, (1957).

[9] Renyi, A., On the theory of order statistics, Acta Math. Acad. Sci. Hung. IV (1953).

[10] Whitworth, W.A., Exercises in Choise and Chance, Deighton Bell & Co Cambridge (1897). (Republished by Hafner, New York, 1959).

[11] Widder, D.V., The Laplace Transform, Princeton University Press (1946).