Partitioning of a line segment

Hilde Grude Borgos
Department of Mathematical Sciences
Norwegian University of Science and Technology
Trondheim, Norway
hgb@stat.ntnu.no

Abstract

Different procedures for partitioning of a line segment is studied. An interval on the real line is partitioned according to a breakage procedure, resulting in a number of smaller segments. Conditioned on the number of resulting segments the different procedures give different length distributions of the segments. The length distributions are studied analytically.

1 Introduction

In a discussion on geological faults and fault patterns, a question turned up concerning the resulting length distribution of smaller segments, when a line segment is partitioned according to different partitioning processes. The aim of this report is to answer some of these questions, by studying the partitioning processes analytically.

In section 2 the partitioning of a line segment is expressed as a partitioning of an interval on the real line. In section 3 through 7 different partitioning procedures are studied. the resulting distributions of the lengths of the smaller segments are found, using the notation of section 2. In section 8 some of the different distributions are compared graphically.

2 Partitioning of a line segment

A line segment of length L can be represented by an interval of length L on the real line $\mathbb{R}$. The interval is partitioned into n intervals Y_i whose lengths sums up to L. Depending on the partitioning process, the resulting intervals Y_i will follow different distributions. If $n-1$ points X_l is distributed along the line, these points, together with the endpoints of the line, define the endpoints of the intervals Y_i. Placing the line on the real line $\mathbb{R}$, with one end at the origin and the other at L, the partitioning can be expressed as follows:

- Let the line be represented by the interval $[0, L]$.
- Distribute the points $X_1, X_2, \ldots, X_{n-1}$ in the interval.
- Order the points; $X_{(1)}, X_{(2)}, \ldots, X_{(n-1)}$.
- The ordered points, together with the endpoints of the line, define n intervals of lengths $Y_1, Y_2, \ldots, Y_n$, where $\sum_{i=1}^{n} Y_i = L$. These lengths are given by $Y_1 = X_{(1)}$, $Y_i = X_{(i)} - X_{(i-1)}$, $i = 2, 3, \ldots, n-1$ and $Y_n = L - X_{(n-1)}$.

3 Equidistant partitioning points

Let the points X_l be placed equidistant in the interval $[0, L]$, as in figure 1. The positions of the points are then

$$X_l = X_{(l)} = \frac{Ll}{n}, \quad l = 1, 2, \ldots, n-1,$$

where the ordered points $X_{(l)}$ are equal to the unordered points X_l. Denote the endpoints of the interval $X_{(0)} = 0$ and $X_{(n)} = L$. This yields intervals Y_i with lengths

$$Y_i = X_{(i)} - X_{(i-1)} = \frac{Li}{n} - \frac{L(i-1)}{n} = \frac{L}{n}, \quad i = 1, 2, \ldots, n.$$

All intervals have the same length, determined by the length of the line and the number of intervals, and the density function of an interval length Y is given by the *Dirac's delta function*,

$$f_Y(y) = \delta\left(y - \frac{L}{n}\right),$$

where $\delta(t) = 0$ for $t \neq 0$ and $\int_{-\infty}^{\infty} \delta(t)dt = 1$.

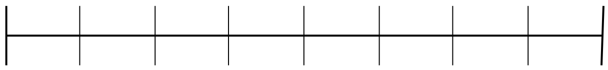

Figure 1: *The* $n-1$ *points* X_l *are placed equidistant in the interval* $[0, L]$.

4 Independent uniformly distributed partitioning points

Consider a stationary Poisson point process in $\mathbb{R}$, with intensity λ (Stoyan, Kendall and Mecke, 1995). The distance between two succeeding points is then exponentially distributed with the same parameter λ. Now, consider the finite interval $[0, L]$. Conditioned on the number of points falling in this interval, we want to find the distribution of the lengths between the points. This distribution will no longer be exponential, as all intervals have an upper bound of L. Given the number of points in the interval, the positions of the points are independent identically uniformly distributed on the interval, forming a binomial point process (Stoyan and Stoyan, 1994). Let the points $X_1, X_2, \ldots, X_{n-1}$ be i.i.d. with a uniform distribution $X_l \sim U[0, L]$. An example is shown in figure 2. The density function of an interval length Y is then given by

$$f_Y(y) = \frac{n-1}{L} \cdot \left(1 - \frac{y}{L}\right)^{n-2}, \quad 0 \leq y \leq L. \tag{1}$$

For a proof, see appendix A.1. The density function (1) corresponds to the scaled length $W = Y/L$ following a Beta distribution, $W \sim \text{Beta}(1, n-1)$. The mean and variance of an interval length Y are

$$E[Y] = \frac{L}{n}, \qquad \text{Var[Y]} = L^2 \frac{n-1}{n^2(n+1)}. \tag{2}$$

The mean is equivalent to the lengths of the intervals when the points X_l are equidistantly distributed. If we let $L \to \infty$ and $n \to \infty$ while $n/L = \lambda$ remains constant, the limiting distribution is the exponential distribution with parameter λ. For proof, see appendix B.1. The limiting distribution corresponds to the X's being a stationary Poisson process on the real line.

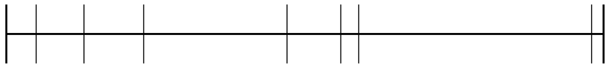

Figure 2: *The $n-1$ points X_l are i.i.d. random variables with a uniform distribution $U[0, L]$.*

5 Repulsion between the partitioning points

Repulsion between the points X_l can be created by thinning of another point process. Let $\mathbf{Z}$ be a set of points $\mathbf{Z} = \{Z_j\}_{j=1}^{s}$ distributed on the interval $[0, L]$, and let the set of points $\mathbf{X}$ be a subset of $\mathbf{Z}$, that is $\mathbf{X} \subset \mathbf{Z}$. Let $Z_{(j)}$, $j = 1, 2, \ldots, s$, be the ordered points. If the points X_l is given by $X_l = Z_{(pl)}$, $l = 1, 2, \ldots, n-1$, where p is a positive integer, there is located $p-1$ points $Z_{(pl+1)}, \ldots, Z_{(pl+p-1)}$ between every neighbour pair of ordered points $X_{(l)}$. These $p-1$ points "force" the two points away from each other, creating repulsion. An example is shown in figure 3. Let the endpoints of the interval be the points $X_{(0)} = 0$ and $X_{(n)} = L$. To get $p-1$ points between every $X_{(l)}$, $l = 0, 1, \ldots, n$, a total number of $s = pn - 1$ points Z_j has to be distributed along the line. If the s points Z_j are independent uniformly distributed on $[0, L]$, forming a binomial point process (Stoyan and Stoyan, 1994), the density function of an interval length $Y_i = X_{(i)} - X_{(i-1)} = Z_{(pi)} - Z_{(p(i-1))}$ is

$$f_Y(y) = \frac{p}{L} \cdot \binom{pn-1}{p} \left(\frac{y}{L}\right)^{p-1} \left(1 - \frac{y}{L}\right)^{p(n-1)-1}, \quad 0 \leq y \leq L. \tag{3}$$

This is shown by Mouland (1997), and an alternative proof is given in appendix A.2. The density function (3) corresponds to the scaled length $W = Y/L$ following a Beta distribution, $W \sim \mathrm{Beta}(p, p(n-1))$. Substituting $p = 1$ in equation (3) we return to equation (1), as expected, since $p = 1$ corresponds to the X_l's being independent uniformly distributed. The mean and variance of the interval lengths in this case are

$$E[Y] = \frac{L}{n}, \qquad \mathrm{Var}[Y] = L^2 \frac{n-1}{n^2(pn+1)}.$$

As in the previous cases, the mean equals the lengths of the intervals when the X_l's are distributed equidistantly. Substituting $p = 1$ in the expression for the variance, gives the variance obtained in equation (2). The limiting distribution of Y when $L \to \infty$, $n \to \infty$ and $n/L = \lambda$ remains constant, is the Gamma distribution with shape parameter p and scale parameter $1/p\lambda$. A proof is given in appendix B.2. The limiting distribution corresponds to the Z's being a stationary Poisson process on the real line, implying the distance between every p'th point follow a Gamma distribution.

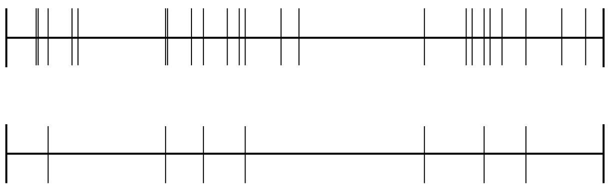

Figure 3: *On the top line the $3n-1$ points are i.i.d. random variables with a uniform distribution $U[0, L]$. On the bottom line, only each 3rd ordered point is kept, giving $n-1$ points with repulsion between them.*

6 Successive partitioning of all existing intervals

The partitioning described in this section consist of successive partitioning of all existing intervals. The partitioning of the line is done in a stepwise procedure, where in each step all existing intervals are partitioned once. Thus, after k steps the number of intervals is $n = 2^k$. The position of partitioning of an existing interval is drawn uniformly on the interval. A partitioning of the interval $[0, L]$ is illustrated in figure 4. The stepwise procedure is described in the following way:

Step 1 Let $X_1 \sim U[0, L]$, giving the interval lengths $Y_1^{(1)} = X_1$, $Y_2^{(1)} = L - X_1$, both having the distribution $Y_i^{(1)} \sim U[0, L]$.

Step 2 The two intervals with lengths $Y_1^{(1)}$ and $Y_2^{(1)}$ are partitioned by placing a point Z_i in each of them, where the Z_i's are uniformly distributed on the existing intervals:

- Let $Z_1|y_1^{(1)} \sim U[0, y_1^{(1)}]$, giving the interval lengths $Y_1^{(2)} = Z_1$, $Y_2^{(2)} = Y_1^{(1)} - Z_1$, both having the distribution $Y_i^{(2)}|y_1^{(1)} \sim U[0, y_1^{(1)}]$, $i = 1, 2$.
- Let $Z_2|y_2^{(1)} \sim U[0, y_2^{(1)}]$, giving the interval lengths $Y_3^{(2)} = Z_2$, $Y_4^{(2)} = Y_2^{(1)} - Z_2$, both having the distribution $Y_i^{(2)}|y_2^{(1)} \sim U[0, y_2^{(1)}]$, $i = 3, 4$.
- The points X_2 and X_3 are given by $X_2 = Z_1$ and $X_3 = X_1 + Z_2$.

Step 3 In the same way as in step 2, partition the existing intervals by drawing a point Z_i uniformly in each of the intervals. This yields $Y_i^{(3)}|y_j^{(2)} \sim U[0, y_j^{(2)}]$, where $i = 2j - 1, 2j$ and $j = 1, 2, 3, 4$. The points X_l, $l = 4, \ldots, 7$ can be found based on X_l, $l = 1, 2, 3$ and $Y_i^{(3)}$.

Step k After $k - 1$ steps there are 2^{k-1} intervals. Partition each of the intervals to get new intervals with length distributions $Y_i^{(k)}|y_j^{(k-1)} \sim U[0, y_j^{(k-1)}]$, where $i = 2j - 1, 2j$ and $j = 1, 2, \ldots, 2^{k-1}$. As above, the points X_l can be found.

Suppose the interval $[0, L]$ is partitioned into $n = 2^k$ intervals by the above procedure. The density function of an interval length $Y^{(k)}$ after k steps is then given by

$$f_{Y^{(k)}}(y) = \frac{1}{L(k-1)!} \cdot \left(-\ln \frac{y}{L}\right)^{k-1}, \quad 0 \leq y \leq L. \tag{4}$$

The proof is given in appendix A.3. The density function (4) corresponds to the variable $W^{(k)} = -\ln(Y^{(k)}/L)$ following a Gamma distribution, $W^{(k)} \sim \text{Gamma}(k, 1)$. The mean and variance of an interval length $Y^{(k)}$ are

$$E[Y^{(k)}] = \frac{L}{2^k} = \frac{L}{n}, \qquad \text{Var}[Y^{(k)}] = L^2 \left(\frac{1}{3^k} - \frac{1}{2^{2k}}\right) = L^2 \left(\frac{1}{n^{\ln 3/\ln 2}} - \frac{1}{n^2}\right).$$

Once again the mean equals the lengths of the intervals when the points X_l are equidistantly distributed.

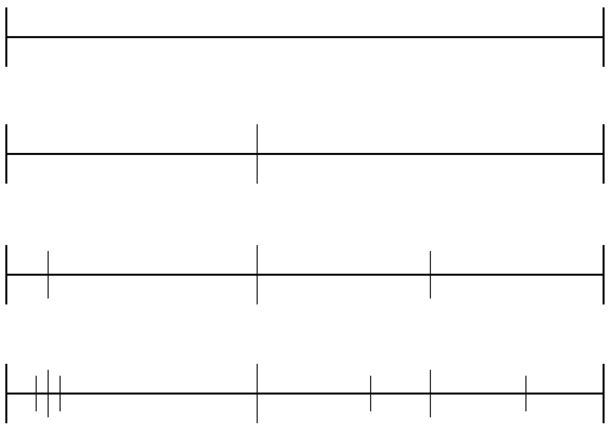

Figure 4: *The 2^{k-1} existing intervals are partitioned into 2^k intervals. The partition points are drawn uniformly on each of the preexisting intervals.*

7 Successive partitioning of one existing interval

Siegel and Sugihara (1983) describes a sequential breakage procedure where at each stage in the procedure one of the existing intervals is further partitioned into two intervals. This partitioning is called a *Kolmogorov breakage process*. Suppose there are $n-1$ intervals at one stage. A move to the next stage is done by choosing one of the n intervals at random, with equal probability of each interval, and partition the interval according to a breakage distribution F. The resulting number of intervals is then n. Figure 5 shows a line partitioned in this way. For a partitioning following this procedure, the distribution of the resulting interval lengths will differ from the distribution described in section 6, where all intervals are partitioned simultaneously.

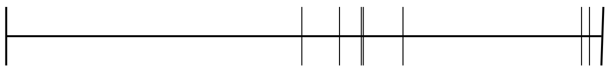

Figure 5: *At each stage one interval is chosen at random and partitioned, with the partition point drawn uniformly on the chosen interval. All existing intervals have equal probability of being chosen.*

Let $U^{(n)} = \log Y^{(n)}$ where $Y^{(n)}$ is the length of an interval when the total number of intervals is n. Siegel and Sugihara (1983) study the first two moments of the logarithm $U^{(n)}$. Using a uniform breakage distribution F, their results can be compared to the density function discussed in section 6, where all intervals were partitioned in each stage. When F is the uniform distribution, the mean and variance of $U^{(n)}$ are

$$E[U^{(n)}] = \ln L - 2\sum_{j=2}^{n} \frac{1}{j} \quad \text{and} \quad \mathrm{Var}[U^{(n)}] = 4\sum_{j=2}^{n} \frac{j-1}{j^2}. \tag{5}$$

Notice that the variance of $U^{(n)}$ is independent of the length L. The density function of $Y^{(k)}$ in equation (4) corresponds to the variable $W^{(k)} = \ln L - \ln Y^{(k)}$ following a Gamma distribution, $W^{(k)} \sim \mathrm{Gamma}(k, 1)$. Let $U^{(k)} = \ln Y^{(k)} = \ln L - W^{(k)}$, such that $(\ln L - U^{(k)}) \sim \mathrm{Gamma}(k, 1)$. The mean and variance of $U^{(k)}$ in this case are

$$E[U^{(k)}] = \ln L - k \quad \text{and} \quad \mathrm{Var}[U^{(k)}] = k, \tag{6}$$

which can be compared to the mean and variance given in (5), where $n = 2^k$. Figure 6 shows plots of the different means and variances for $n \in [1, 100]$ and $L = 1$. The mean and variance given in (6) are both smaller in absolute value than the mean and variance given in (5).

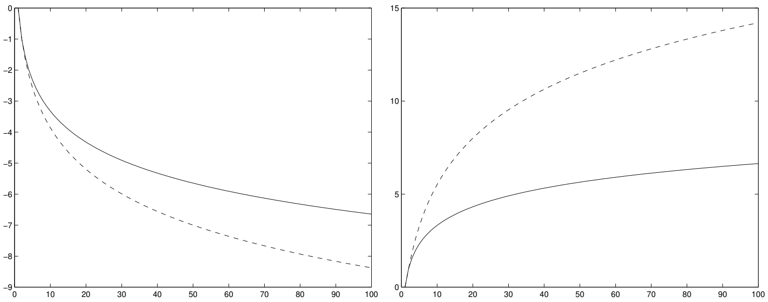

Figure 6: *To the left is the mean of U plotted as a function of n, and to the right the variance of U. The dotted lines are the mean and variance given in equation (5) and the solid lines are the mean and variance given in (6), where $k = \ln n / \ln 2$.*

Siegel and Sugihara (1983) study the two first moments of the logarithm $U = \ln Y$, but not the density functions of U or Y. Consider a uniform breakage distribution F. Theoretically, the density function of the interval lengths Y can be found, based on the results in section 6. When there are n intervals, each interval length Y will have a length distribution with the density function given in (4), but the number k will differ for each interval, taking values in the set $\{1, \ldots, n-1\}$. Thus, conditioned on k the distribution of Y, $f_{Y|K}(y|k)$, is known. Let k_i be the number of k corresponding to the interval Y_i. All possible combinations of $(k_1, \ldots, k_n)$ can be found, and from the combinations the density function of K, $f_K(k)$, can be found. The conditional distribution of Y is then given by $f_Y(y) = \sum_k f_{Y|K}(y|k) f_K(k)$, and thus the density function of an interval length Y can be found. This is a simple task when n is small, but can be tedious work for large n, since there are $(n-1)!$ possible combinations of $(k_1, \ldots, k_n)$[1]. For $n = 4$ there are $3! = 6$ possible combinations of $(k_1, \ldots, k_4)$, and the density function of Y, expressed by the density functions $f_{Y^{(k)}}(y)$ in equation (4), is

$$f_Y(y) = \frac{1}{6}\left(f_{Y^{(1)}}(y) + 3 f_{Y^{(2)}}(y) + 2 f_{Y^{(3)}}(y)\right) = \frac{1}{6}\left(1 - 3\ln y + (\ln y)^2\right), \quad 0 \le y \le 1.$$

For $n = 8$ there are $7! = 5040$ possible combinations of $(k_1, \ldots, k_8)$, and the density function of Y is a linear combination of $f_{Y^{(k)}}(y)$, $k = 1, \ldots, 7$.

8 Comparisons of length distributions

Figure 7, 9 and 8 shows the density functions for the interval lengths Y with different ways of partitioning. The density functions are shown for partitionings into 2, 4, 8, 16, 32 and 64 intervals, using original intervals of length $L = 1$. In figure 7 the points X_l are independent uniformly distributed on $[0, L]$, in figure 8 the intervals are created with repulsion between the interval endpoints, and in figure 9 the intervals are created by successive partitioning of all existing intervals. In the case of independent uniformly distributed points the density function is monotone decreasing, with $f_Y(0) = (n-1)/L$. The density function of the interval lengths produced by the successive partitioning of all existing intervals are also monotone decreasing, but with $f_Y(y) \to \infty$ as $y \to 0$. When there is repulsion between the points, the density function is a one peaked function starting in $f_Y(0) = 0$. The mode of the function is at the point $y_{top} = L(p-1)/(pn-2)$, which approaches zero as $n \to \infty$.

[1] Some of these combinations might be equal, but they result from different partitioning sequences.

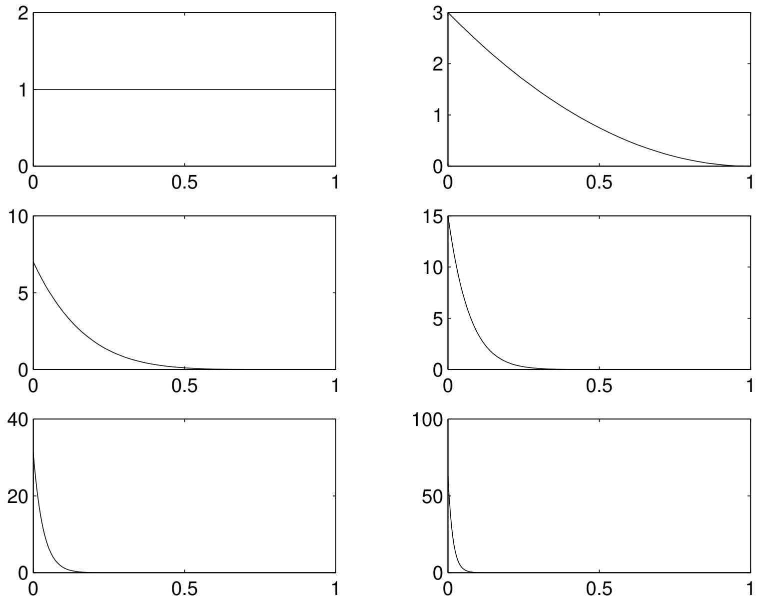

Figure 7: *Density function of length distributions when the X_l's are independent uniformly distributed. The number of intervals is* $2, 4$ *(top),* $8, 16$ *(middle) and* 32 *and* 64 *(bottom).*

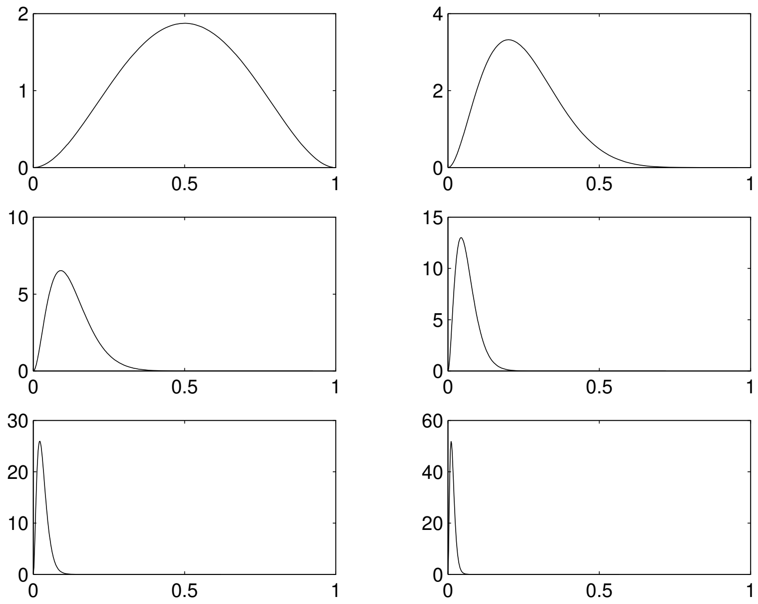

Figure 8: *Density function of length distributions when the X_l's are given by $X_l = Z_{(3l)}$, where the Z_j's are independent uniformly distributed. The number of intervals is* $2, 4$ *(top),* $8, 16$ *(middle) and* 32 *and* 64 *(bottom).*

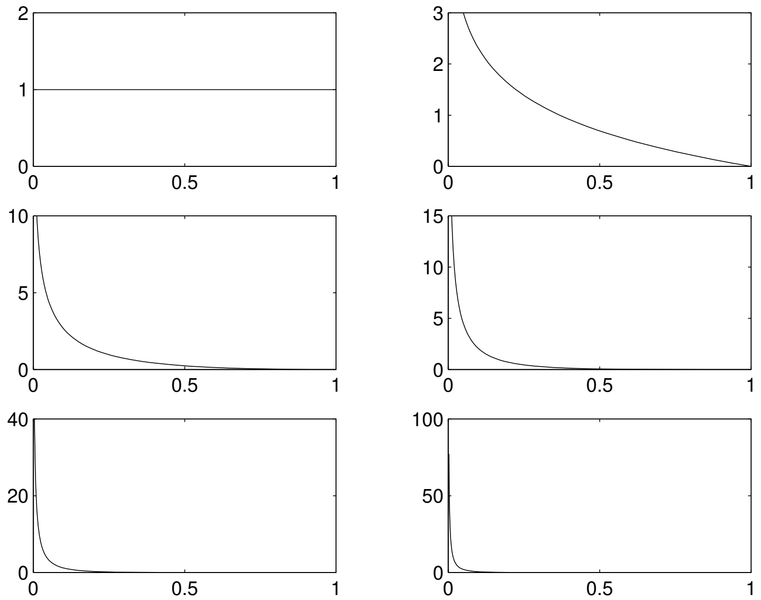

Figure 9: *Density function of length distributions when the intervals are created by successive partitioning of all existing intervals. The number of intervals is* $2, 4$ *(top),* $8, 16$ *(middle) and* 32 *and* 64 *(bottom), which corresponds to* $k = 1, \ldots, 6$.

9 Conclusions

Different models for partitioning of line segments have been discussed. In section 3 the partitioning points are regularly spaced, resulting in equal lengths of all smaller intervals.

In section 4 the partitioning points are independent identically uniformly distributed on the line. This is a special case of the model presented in section 5, where points are independent identically uniformly distributed, but only every p'th point are kept as partitioning points, where p is a positive integer. The model in section 5 enforce repulsion between the partitioning points when $p > 1$. The limiting distributions are studied, as the length of the initial interval and the number of partitionings approach infinity, while the relation between the length and number remains constant. The limiting distributions corresponds to the initial points being a stationary Poisson process on the real line (Stoyan *et al.*, 1995).

In section 6 and 7 sequential breakage processes are discussed. In section 6 all existing intervals are partitioned simultaneously, while in section 7 only one existing interval is partitioned in each step. The two sequential procedures leads to different distributions of the resulting intervals. The density function of the partitioning procedure described in section 7 can, under certain conditions, be found based on the density functions found in section 6.

References

Mouland, L. (1997). Gjennkjenning av multiple hendelser i tidsserier, Report in subject 75800 Prosjekt 2, Norwegian University of Science and Technology.

Siegel, A. F. and Sugihara, G. (1983). Moments of particle size distributions under sequential breakage with applications to species abundance, *Journal of Applied Probability* **20**: 158–164.

Stoyan, D. and Stoyan, H. (1994). *Fractals, Random Shapes and Point Fields*, Wiley.

Stoyan, D., Kendall, W. S. and Mecke, J. (1995). *Stochastic Geometry and its Applications*, 2 edn, Wiley.

A Computation of length distributions

A.1 Independent uniformly distributed partitioning points

This proof is a special case of the proof given in appendix A.2. Let $X_{(0)} = 0$ and $X_{(n)} = L$. The interval lengths are then $Y_i = X_{(i)} - X_{(i-1)}$, $i = 1, 2, \ldots, n$. When the X_l's are independent uniformly distributed on the interval $[0, L]$, the density function of the l'th ordered point $X_{(l)}$ is

$$f_{X_{(l)}}(x_{(l)}) = \frac{1}{L^{n-1}} \cdot \frac{(n-1)!}{(l-1)!(n-l-1)!} \cdot x_{(l)}^{l-1}(L - x_{(l)})^{n-l-1}, \tag{7}$$

and the joint density of the neighbour points $X_{(l-1)}$ and $X_{(l)}$ is

$$f_{X_{(l-1)},X_{(l)}}(x_{(l-1)}, x_{(l)}) = \frac{1}{L^{n-1}} \cdot \frac{(n-1)!}{(l-2)!(n-1-l)!} \cdot x_{(l-1)}^{l-2}(L - x_{(l)})^{n-1-l}, \quad l = 2, 3, \ldots, n-1.$$

The distribution of Y_i can be found by a bivariate transformation. Let

$$\begin{array}{ll} Y_i = X_{(i)} - X_{(i-1)}, & 0 \leq Y_i \leq L, \\ W = X_{(i)}, & Y_i \leq W \leq L, \end{array}$$

implying

$$\begin{array}{ll} X_{(i-1)} = W - Y_i, & 0 \leq X_{(i-1)} \leq L, \\ X_{(i)} = W, & X_{(i-1)} \leq X_{(i)} \leq L, \end{array}$$

with the Jacobian

$$J = \begin{vmatrix} \partial x_{(i-1)}/\partial y_i & \partial x_{(i-1)}/\partial w \\ \partial x_{(i)}/\partial y_i & \partial x_{(i)}/\partial w \end{vmatrix} = \begin{vmatrix} -1 & 1 \\ 0 & 1 \end{vmatrix} = -1.$$

The joint density of Y_i and W are then

$$\begin{aligned} f_{Y_i,W}(y, w) &= f_{X_{(i-1)},X_{(i)}}(w - y, w)|J| \\ &= \frac{1}{L^{n-1}} \cdot \frac{(n-1)!}{(i-2)!(n-1-i)!} \cdot (w-y)^{i-2}(L-w)^{n-1-i}, \quad i = 2, 3, \ldots, n-1, \end{aligned}$$

and the marginal density of Y_i, $i = 2, 3, \ldots, n-1$, is found by

$$\begin{aligned} f_{Y_i}(y) &= \int_y^L f_{Y_i,W}(y, w)dw \\ &= \int_y^L \frac{1}{L^{n-1}} \cdot \frac{(n-1)!}{(i-2)!(n-1-i)!} \cdot (w-y)^{i-2}(L-w)^{n-1-i}dw \\ &= \frac{n-1}{L} \cdot \left(1 - \frac{y}{L}\right)^{n-2}. \end{aligned} \tag{8}$$

Since $Y_1 = X_{(1)}$ and $Y_n = L - X_{(n-1)}$, the density functions of Y_1 and Y_n are found from equation (7), and are easily shown to be the density function (8).

A.2 Repulsion between the partitioning points

This proof is similar to the proof in A.1, where the special case $p = 1$ is considered. We now find the distribution for general p. Let $Z_{(0)} = 0$ and $Z_{(pn)} = L$. The interval lengths are then $Y_i = X_{(i)} - X_{(i-1)} = Z_{(pi)} - Z_{(p(i-1))}$, $i = 1, 2, \ldots, n$. When the Z_l's are independent uniformly distributed on the interval $[0, L]$, the density function of the pl'th ordered point $Z_{(pl)}$ is

$$f_{Z_{(pl)}}(z_{(pl)}) = \frac{1}{L^{pn-1}} \cdot \frac{(pn-1)!}{(pl-1)!(p(n-l)-1)!} \cdot z_{(pl)}^{pl-1}(L - z_{(pl)})^{p(n-l)-1}, \tag{9}$$

and the joint density of the neighbour points $Z_{(p(l-1))}$ and $Z_{(pl)}$ is

$$\begin{aligned} & f_{Z_{(p(l-1))},Z_{(pl)}}(z_{(p(l-1))}, z_{(pl)}) \\ = \; & C(l) \cdot z_{(p(l-1))}^{p(l-1)-1}(z_{(pl)} - z_{(p(l-1))})^{p-1}(L - z_{(pl)})^{p(n-l)-1}, \quad l = 2, 3, \ldots, n-1. \end{aligned}$$

where

$$C(l) = \frac{1}{L^{pn-1}} \cdot \frac{(pn-1)!}{(p(l-1)-1)!(p-1)!(p(n-l)-1)!}.$$

The distribution of Y_i can be found by a bivariate transformation. Let

$$\begin{array}{ll} Y_i = Z_{(pi)} - Z_{(p(i-1))}, & 0 \le Y_i \le L, \\ W = Z_{(pi)}, & Y_i \le W \le L, \end{array}$$

implying

$$\begin{array}{ll} Z_{(p(i-1))} = W - Y_i, & 0 \le Z_{(i-1)} \le L, \\ Z_{(pi)} = W, & Z_{(i-1)} \le Z_{(i)} \le L, \end{array}$$

with the Jacobian

$$J = -1.$$

The joint density of Y_i and W are then

$$\begin{aligned} f_{Y_i,W}(y, w) &= f_{Z_{(p(i-1))},Z_{(pi)}}(w - y, w)|J| \\ &= C(i) \cdot (w-y)^{p(i-1)-1} y^{p-1}(L - w)^{p(n-i)-1}, \quad i = 2, 3, \ldots, n-1, \end{aligned}$$

and the marginal density of Y_i, $i = 2, 3, \ldots, n-1$, is found by

$$\begin{aligned} f_{Y_i}(y) &= \int_y^L f_{Y_i,W}(y, w)dw \\ &= \int_y^L C(i) \cdot (w-y)^{p(i-1)-1} y^{p-1}(L-w)^{p(n-i)-1}dw \\ &= \frac{p}{L} \cdot \binom{pn-1}{p} \left(\frac{y}{L}\right)^{p-1} \left(1 - \frac{y}{L}\right)^{p(n-1)-1}. \end{aligned} \tag{10}$$

Since $Y_1 = Z_{(p)}$ and $Y_n = L - Z_{(p(n-1))}$, the density functions of Y_1 and Y_n are found from equation (9), and are easily shown to be the density function (10).

A.3 Successive partitioning of all existing intervals

A partitioning of the line into $n = 2^k$ intervals is done by successive partitioning of all existing intervals. After the first partitioning the number of intervals is 2, after the second the number is $2^2 = 4$, and after $k-1$ steps the number of intervals is 2^{k-1}. The length distributions of the intervals created in step k, conditioned on the intervals after $k-1$ steps are

$$Y_i^{(k)} | y_j^{(k-1)} \sim U[0, y_j^{(k-1)}], \quad i = 2j-1, 2j, \; j = 1, 2, \dots, 2^{k-1}.$$

The density function of the length of the interval $Y_{2j}^{(k)}$ in the final partitioning is then given by the integral

$$f_{Y_{2j}^{(k)}}(y) = \int_y^L f_{Y_{2j}^{(k)} | Y_j^{(k-1)}}(y|y') f_{Y_j^{(k-1)}}(y') dy' = \int_y^L \frac{f_{Y_j^{(k-1)}}(y')}{y'} dy'. \qquad (11)$$

The density function of $Y_{2j-1}^{(k)}$ equals that of $Y_{2j}^{(k)}$. Since $Y_j^{(1)} \sim U[0, L]$, $j = 1, 2$, $f_{Y_j^{(1)}}(y)$ is known, and $f_{Y_j^{(k)}}(y)$, $k = 2, 3, \dots$, $j = 1, 2, \dots, 2^k$ can be found. The density function is identical for all $j = 1, 2, \dots, 2^k$, and is given by

$$f_{Y^{(k)}}(y) = \frac{1}{L(k-1)!} \cdot \left(-\ln \frac{y}{L} \right)^{k-1}, \quad 0 \le y \le L. \qquad (12)$$

This is shown by induction on k. If $k = 1$, the interval length $Y^{(1)}$ is uniformly distributed on the interval $[0, L]$, giving

$$f_{Y^{(1)}}(y) = \frac{1}{L},$$

which equals equation (12) with $k = 1$. If $k = 2$ the integral in (11) can be solved to give

$$f_{Y^{(2)}}(y) = \int_y^L \frac{f_{Y^{(1)}}(y')}{y'} dy' = \int_y^L \frac{1}{y' L} dy' = \frac{1}{L} \cdot \left(-\ln \frac{y}{L} \right),$$

which equals equation (12) with $k = 2$. Thus equation (12) is valid for $k = 1, 2$. Suppose (12) holds for $k-1$. Using the integral in (11) the density function $f_{Y^{(k)}}(y)$ is found to be

$$f_{Y^{(k)}}(y) = \int_y^L \frac{f_{Y^{(k-1)}}(y')}{y'} dy' = \int_y^L \frac{1}{y'} \cdot \frac{1}{L(k-2)!} \cdot \left(-\ln \frac{y'}{L} \right)^{k-2} dy' = \frac{1}{L(k-1)!} \cdot \left(-\ln \frac{y}{L} \right)^{k-1}.$$

B Limiting distributions

B.1 Independent uniformly distributed partitioning points

The density function of Y, given in equation (1), corresponds to $W = Y/L$ following a Beta distribution, $W \sim \mathrm{Beta}(1, n-1)$. The moment generating function of Y can be found through the Moment generating function of W:

$$M_Y(t) = E[\exp(tY)] = E[\exp(tLW)] = M_W(tL).$$

Using the Moment generating function of the Beta distribution, we get

$$\begin{aligned} M_Y(t) &= 1 + \sum_{k=1}^{\infty} \left\{ \prod_{r=0}^{k-1} \frac{1+r}{n+r} \right\} \cdot \frac{(tL)^k}{k!} = 1 + \sum_{k=1}^{\infty} \frac{k!}{n \cdot (n+1) \cdots (n+k-1)} \cdot \frac{(tL)^k}{k!} \\ &= 1 + \sum_{k=1}^{\infty} \frac{(tL)^k}{n \cdot (n+1) \cdots (n+k-1)} = 1 + \sum_{k=1}^{\infty} \frac{t^k}{\frac{n}{L} \cdot (\frac{n}{L} + \frac{1}{L}) \cdots (\frac{n}{L} + \frac{k-1}{L})} \\ &= 1 + \sum_{k=1}^{\infty} \frac{t^k}{\lambda \cdot (\lambda + \frac{1}{L}) \cdots (\lambda + \frac{k-1}{L})}. \end{aligned} \tag{13}$$

In the last step we let $n/L = \lambda$. For $t < \lambda$, taking limits of the sum in (13) gives

$$\lim_{L \to \infty} M_Y(t) = 1 + \sum_{k=1}^{\infty} \left(\frac{t}{\lambda} \right)^k = \sum_{k=0}^{\infty} \left(\frac{t}{\lambda} \right)^k = \frac{1}{1 - \frac{t}{\lambda}} = \frac{\lambda}{\lambda - t},$$

which is the Moment generating function of the exponential distribution with parameter λ.

B.2 Repulsion between the partitioning points

The density function of Y, given in equation (3), corresponds to $W = Y/L$ following a Beta distribution, $W \sim \mathrm{Beta}(p, p(n-1))$. As in appendix B.1, the Moment generating function of the Beta distribution can be used to find the limiting distribution of Y as $L \to \infty$, $n \to \infty$ and $n/L = \lambda$. The proof is similar to the proof in appendix B.1, where the special case $p = 1$ is considered. For general p we find

$$\begin{aligned} M_Y(t) &= 1 + \sum_{k=1}^{\infty} \left\{ \prod_{r=0}^{k-1} \frac{p+r}{pn+r} \right\} \cdot \frac{(tL)^k}{k!} \\ &= 1 + \sum_{k=1}^{\infty} \frac{(p+k-1)!}{(p-1)!k!} \cdot \frac{t^k}{\frac{pn}{L} \cdot (\frac{pn}{L} + \frac{1}{L}) \cdots (\frac{pn}{L} + \frac{k-1}{L})} \\ &= 1 + \sum_{k=1}^{\infty} \binom{p+k-1}{k} \cdot \frac{t^k}{p\lambda \cdot (p\lambda + \frac{1}{L}) \cdots (p\lambda + \frac{k-1}{L})}. \end{aligned} \tag{14}$$

For $t < \lambda$, taking limits of the sum in (14) gives

$$\begin{aligned} \lim_{L \to \infty} M_Y(t) &= 1 + \sum_{k=1}^{\infty} \binom{p+k-1}{k} \left(\frac{t}{p\lambda} \right)^k = \sum_{k=0}^{\infty} \binom{p+k-1}{k} \left(\frac{t}{p\lambda} \right)^k \\ &= \left(\frac{1}{1 - \frac{t}{p\lambda}} \right)^p = \left(\frac{p\lambda}{p\lambda - t} \right)^p, \end{aligned}$$

which is the Moment generating function of the Gamma distribution with shape parameter p and scale parameter $1/p\lambda$.