Chapter Six
THE POISSON PROCESS

1. THE POISSON PROCESS.

1.1. Introduction.

One of the most important processes occurring in nature is the Poisson point process. It is therefore important to understand how such processes can be simulated. The methods of simulation vary with the type of Poisson point process, i.e. with the space in which the process occurs, and with the homogeneity or non-homogeneity of the process. We will not be concerned with the genesis of the Poisson point process, or with important applications in various areas. To make this material come alive, the reader is urged to read the relevant sections in Feller (1965) and Cinlar (1975) for the basic theory, and some sections in Trivedi (1982) for computer science applications.

In a first step, we will define the homogeneous Poisson process on $[0,\infty)$: the process is entirely determined by a collection of random events occurring at certain random times $0<T_1<T_2<\cdots$. These events can correspond to a variety of things, such as bank robberies, births of quintuplets and accidents involving Montreal taxi cabs. If $N(t_1,t_2)$ is the number of events occurring in the time interval (t_1,t_2), then the following two conditions are often satisfied:

(i) For disjoint intervals $(t_1,t_2),(t_3,t_4),\ldots,$ the random variables $N(t_1,t_2),N(t_3,t_4),\ldots$ are independent.

(ii) $N(t_1,t_2)$ is distributed as $N(0,t_2-t_1)$, i.e. the distribution of the number of events in a certain time interval just depends upon the length of the interval.

The amazing fact is that these two conditions imply that all random variables $N(t_1,t_2)$ are Poisson distributed, and that there exists a constant $\lambda\geq 0$ such that $N(t,t+a)$ is Poisson λa for all $t\geq 0$, $a>0$. See e.g. Feller (1965). Thus, the Poisson distribution occurs very naturally.

The previous concept can be generalized to R^d. Let A be a subset of R^d, and let N be a random variable taking only integer values. Let $X_1,\ldots,X_N$ be a sequence of random vectors taking values in A. Then we say that the X_i's

define a uniform (or: homogeneous) Poisson process on A if

(A) For any finite collection of finite-volume nonoverlapping subsets of A, say $A_1, \ldots, A_k$, the random variables $N(A_1), \ldots, N(A_k)$ are independent.

(B) For any Borel subset $B \subseteq A$, the distribution of $N(B)$ depends upon $Vol(B)$ only.

Again, these assumptions imply that all $N(B)$'s are Poisson distributed with parameter $\lambda\ Vol(B)$ for some $\lambda \geq 0$. λ will be called the rate, or rate parameter, of the homogeneous Poisson process on A. Examples of such processes in multidimensional Euclidean space include bacteria on a Petri plate and locations of murders in Houston.

Theorem 1.1.

Let $B \subseteq A$ be fixed sets from R^d, and let $0 < Vol(B) < \infty$. Then:

(I) If $X_1, X_2, \ldots$ determines a uniform Poisson process on A with parameter λ, then for any partition $B_1, \ldots, B_k$ of B, we have that $N(B_1), \ldots, N(B_k)$ are independent Poisson distributed with parameters $\lambda\ Vol(B_i)$.

(II) Let N be Poisson distributed with parameter $\lambda\ Vol(B)$, and let $X_1, \ldots, X_N$ be the first N random vectors from an iid sequence of random vectors uniformly distributed on B. For any partition $B_1, \ldots, B_k$ of B, the sequence $N(B_1), \ldots, N(B_k)$ is sequence of independent Poisson random variables with parameters $\lambda\ Vol(B_1), \ldots, \lambda\ Vol(B_k)$. In other words, $X_1, \ldots, X_N$ determines a uniform Poisson process on B with rate parameter λ.

Proof of Theorem 1.1.

We will only show part (II). Assume that $Vol(B)=1$ and that B is partitioned into two sets, A_1, A_2 with respective volumes p and $q=1-p$. For any two integers $i, j \geq 0$ with $i+j=k$, we have

$$
\begin{aligned}
&P(N(A_1)=i, N(A_2)=j) \\
&= P(N(B)=k) P(N(A_1)=i, N(A_2)=j \mid N(B)=k) \\
&= (e^{-\lambda} \frac{\lambda^k}{k!}) \binom{k}{i} p^i q^j \\
&= (e^{-\lambda p} \frac{\lambda p^i}{i!})(e^{-\lambda q} \frac{\lambda q^j}{j!}),
\end{aligned}
$$

and therefore, $N(A_1)$ and $N(A_2)$ are independent Poisson random variables as claimed. This argument can be extended towards all finite partitions and all positive values for $Vol(B)$. ■

1.2. Simulation of homogeneous Poisson processes.

If we have to simulate a uniform Poisson process on a set $A \subseteq R^d$, then we need to generate a number of random vectors $X_i \in A$. This can be done as follows (by Theorem 1.1):

Homogeneous Poisson process generator

Generate a Poisson random variate N with parameter $\lambda\, Vol(A)$.
Generate iid random vectors $X_1, \ldots, X_N$ uniformly distributed on A.
RETURN $X_1, \ldots, X_N$

To generate N it is virtually useless to use an $O(1)$ expected time algorithm because in the remainder of the algorithm, at least time $\Omega(N)$ is spent. Thus, it is recommended that if the algorithm is used, the Poisson random variate be generated by a very simple algorithm (with expected time typically growing as λ). For specific sets A, other methods can be used which do not require the explicit generation of a Poisson random variate. There are three cases that we will use to illustrate this:

(i) A is $[0,\infty)$.
(ii) A is a circle.
(iii) A is a rectangle.

To do so, we need an interesting connection between Poisson processes and the exponential distribution.

Theorem 1.2.

Let $0<T_1<T_2<\cdots$ be a uniform Poisson process on $[0,\infty)$ with rate parameter $\lambda>0$. Then

$$\lambda(T_1-0),\lambda(T_2-T_1),\lambda(T_3-T_2),\ldots$$

are distributed as iid exponential random variables.

Proof of Theorem 1.2.

For any $k \geq 0$ and any $x>0$,

$$P(T_{k+1}>T_k+x \mid T_k) = P(T_{k+1}\notin[T_k,T_k+x] \mid T_k)$$
$$= P(N_{[0,x]}=0)$$
$$= e^{-\lambda x}\frac{\lambda^0}{0!}$$
$$= e^{-\lambda x}.$$

Thus, given T_k, $T_{k+1}-T_k$ is exponential with parameter λ. Generalizing this argument to obtain the claimed independence as well, we see that for any finite k, and any sequence of nonnegative numbers $x_0, x_1, \ldots,$

$$\begin{aligned} & P(T_{k+1}-T_k > x_k, T_k - T_{k-1} > x_{k-1}, \ldots, T_2 - T_1 > x_1, T_1 - 0 > x_0) \\ &= P(N_{(T_k, T_k + x_k)} = 0, \ldots, N_{(0,x_0)} = 0) \\ &= P(N_{(0, x_0 + x_1 + \cdots + x_k)} = 0) \\ &= e^{-\lambda \sum_{i=0}^{k} x_i} \\ &= \prod_{i=0}^{k} e^{-\lambda x_i} . \end{aligned}$$

This concludes the proof of Theorem 1.2. ■

Theorem 1.2 suggests the following method for simulating a uniform Poisson process on $A = [0,\infty)$:

Uniform Poisson process generator on the real line: the exponential spacings method

$T \leftarrow 0$ (auxiliary variable used for updating the "time")
$k \leftarrow 0$ (initialize the event counter)
REPEAT
    Generate an exponential random variate E.
    $k \leftarrow k+1$
    $T \leftarrow T + \frac{E}{\lambda}$
    $T_k \leftarrow T$
UNTIL False (this is an infinite loop; a stopping rule can be added if desired).

This algorithm is easy to implement because no Poisson random variates are needed. For other simple sets A, there exist trivial generalizations of Theorem 1.2. For example, when A is $[0,t]\times[0,1]$ where possibly $t=\infty$, $0 < T_1 < T_2 < \cdots$ is a uniform Poisson process with rate λ on $[0,t]$, and $U_1, U_2, \ldots$ is a sequence of iid uniform [0,1] random variables, then

$$(T_1, U_1), (T_2, U_2), \ldots$$

determines a uniform Poisson process with rate λ on A.

Example 1.1. A uniform Poisson process on the unit circle.

If the set A is the circle with unit radius, then the various properties of uniform Poisson processes can be used to come up with several methods of generation (these can be extended to d dimensional spheres). Assume that λ is the desired rate. First, we could simply generate a Poisson $\lambda\pi$ random variate N, and then return a sequence of N iid random vectors uniformly distributed in the unit circle. If we apply the order statistics method suggested by Theorem 1.2, then the Poisson random variate is implicitly obtained. For example, by switching to polar coordinates (R,θ), we note that for a uniform Poisson process, R and θ are independent, and that a randomly chosen R has density $2r$ $(0\leq r\leq 1)$ and that a randomly chosen θ is uniformly distributed on $[0,2\pi]$. Thus, we could proceed as follows: generate a uniform Poisson process $0<\theta_1<\theta_2<\cdots<\theta_N$ with rate parameter $\frac{\lambda}{2\pi}$ on $[0,2\pi]$ by the exponential spacings method. Exit with

$$(\theta_1,R_1),\ldots,(\theta_N,R_N)$$

where the R_i's are iid random variates with density $2r$ $(0\leq r\leq 1)$ which can be generated individually as the maxima of two independent uniform $[0,1]$ random variates. There is no special reason for applying the exponential spacings method to the angles. We could have picked the radii as well. Unfortunately, the ordered radii do not form a one-dimensional uniform Poisson process on $[0,1]$. They do form a nonhomogeneous Poisson process however, and the generation of such processes will be clarified in the next subsection. ■

1.3. Nonhomogeneous Poisson processes.

There are situations in which events occur at "random times" but some times are more likely than others. This is the case for arrivals in intensive care units, for job submissions in a computer centre and for injuries to NFL players. A very good model for these cases is the nonhomogeneous Poisson process model, defined here for the sake of convenience on $[0,\infty)$. This is the most important case because "time" is usually the running variable.

A nonhomogeneous Poisson process on $[0,\infty)$ is determined by a rate function $\lambda(t)\geq 0$ $(t\geq 0)$, which can be considered as a density of sorts, with the difference that $\int_0^\infty \lambda(t)\,dt$ is not necessarily 1 (usually, it is ∞). The process is defined by the following property: for all finite collections of disjoint intervals $A_1,\ldots,A_k$, the numbers of events happening in these intervals $(N_1,\ldots,N_k)$ are independent Poisson random variables with parameters

$$\int_{A_i} \lambda(t)\, dt \qquad (1 \leq i \leq k)\,.$$

Let us now review how such processes can be simulated. By simulation, we understand that the times of occurrences of events $0 < T_1 < T_2 < \cdots$ are to be given in increasing order. The major work on simulation of nonhomogeneous Poisson processes is Lewis and Shedler (1979). This entire section is a reworked version of their paper. It is interesting to observe that the general principles of continuous random variate generation can be extended: we will see that there are analogs of the inversion, rejection and composition methods.

The role of the distribution function will be taken over by the integrated rate function

$$\Lambda(t) = \int_0^t \lambda(u)\, du\,.$$

We begin by noting that given $T_n = t$, $T_{n+1} - T_n$ has distribution function

$$F(x) = 1 - e^{-(\Lambda(t+x)-\Lambda(t))} \qquad (x \geq 0)$$

provided that $\lim_{t\to\infty} \Lambda(t) = \infty$ (i.e, $\int_0^\infty \lambda(t)\, dt = \infty$). This follows from the fact that

$$\begin{aligned} F(x) &= P(T_{n+1} - T_n > x \mid T_n = t) \\ &= P(N(t, t+x) = 0 \mid T_n = t) \\ &= e^{-(\Lambda(t+x)-\Lambda(t))} \qquad (x \geq 0)\,. \end{aligned}$$

Thus, T_{n+1} is distributed as $T_n + F^{-1}(U)$ where U is a uniform [0,1] random variate. Interestingly, writing U as $1 - e^{-E}$ (where E denotes an exponential random variable), we see that T_{n+1} is also distributed as $\Lambda^{-1}(E + \Lambda(T_n))$. In other words, we need to invert Λ. Formally, we have (see also Cinlar (1975) or Bratley, Fox and Schrage, 1983):

Algorithm based on inversion of the integrated rate function

$T \leftarrow 0$ (T will be an auxiliary variable)
$k \leftarrow 0$ (k is a counter)
REPEAT
    Generate an exponential random variate E.
    $k \leftarrow k+1$
    $T \leftarrow T+\Lambda^{-1}(E+\Lambda(T))$
    $T_k \leftarrow T$
UNTIL False

Example 1.2. Homogeneous Poisson process.

For the special case $\lambda(t)=\lambda$, $\Lambda(t)=\lambda t$, it is easily seen that in the algorithm given above, the step $T \leftarrow T+\Lambda^{-1}(E+\Lambda(T))$ reduces to $T \leftarrow T+\frac{E}{\lambda}$. Thus, we obtain the exponential spacings method again. ■

Example 1.3.

To model morning pre-rush hour traffic, we can sometimes take $\lambda(t)=t$, which gives $\Lambda(t)=\frac{t^2}{2}$. The step $T \leftarrow T+\Lambda^{-1}(E+\Lambda(T))$ now needs to be replaced by

$$T \leftarrow \sqrt{T^2+2E} \quad .\blacksquare$$

If the rate function can be split into a sum of rate functions, as in

$$\lambda(t) = \sum_{i=1}^{n} \lambda_i(t)$$

and if $0<T_{i1}<T_{i2}< \cdots$, $1 \leq i \leq n$ are independent realizations of the individual nonhomogeneous Poisson processes, then the merged ordered sequences form

a realization of the nonhomogeneous Poisson process with rate function $\lambda(t)$. This corresponds to the **composition method**, but the difference now is that we need realizations of all component processes. The decomposition can be used when there is a natural decomposition dictated by the analytical form of $\lambda(t)$. Because the basic operation in merging the processes is to take the minimal value from the n processes, it could be advantageous for large n to store the times in a heap containing n elements. We summarize:

The composition method

Generate $T_{11}, \ldots, T_{n1}$ for the n Poisson processes, and store these values together with the indices of the corresponding processes in a table.
$T \leftarrow 0$ (T is the running time)
$k \leftarrow 0$ REPEAT
    Find the minimal element (say, T_{ij}) in the table and delete it.
    $k \leftarrow k+1$
    $T_k \leftarrow T_{ij}$
    Generate the value $T_{i\ j+1}$ and insert it into the table.
UNTIL False

The third general principle is that of **thinning** (Lewis and Shedler, 1979). Similar to what we did in the rejection method, we assume the existence of an easy dominating rate function $\mu(t)$:

$$\lambda(t) \leq \mu(t), \text{ all } t .$$

Then the idea is to generate a homogeneous Poisson process on the part of the positive halfplane between 0 and $\mu(t)$, then to consider the homogeneous Poisson process under λ, and finally to exit with the x-components of the events in this process. This requires a theorem similar to that preceding the rejection method.

Theorem 1.3.

Let $\lambda(t)\geq 0$ be a rate function on $[0,\infty)$, and let A be the set of all (x,y) with $x\geq 0, 0\leq y\leq \lambda(x)$. The following is true:

(I) If $(X_1,Y_1),\ldots$ (with ordered X_i's) is a homogeneous Poisson process with unit rate on A, then $0<X_1<X_2<\cdots$ is a nonhomogeneous Poisson process with rate function $\lambda(t)$.

(II) If $0<X_1<X_2<\cdots$ is a nonhomogeneous Poisson process with rate function $\lambda(t)$, and $U_1,U_2,\ldots$ are iid uniform $[0,1]$ random variables, then $(X_1,U_1\lambda(X_1)),(X_2,U_2\lambda(X_2)),\ldots$ is a homogeneous Poisson process with unit rate on A.

(III) If $B\subseteq A$, and $(X_1,Y_1),\ldots$ (with ordered X_i's) is a homogeneous Poisson process with unit rate on A, then the subset of points (X_i,Y_i) belonging to B forms a homogeneous Poisson process with unit rate function on B.

Proof of Theorem 1.3. We verify that for nonoverlapping intervals $A_1,\ldots,A_k$, the number of X_i's falling in the intervals (which we shall denote by $N(A_1),\ldots,N(A_k)$), satisfy:

$$P(N(A_1)=i_1,\ldots,N(A_k)=i_k)$$
$$=P(N(\bar{A}_1)=i_1,\ldots,N(\bar{A}_k)=i_k)$$
$$=\prod_{j=1}^{k}\frac{\left(\int_{A_j}\lambda(t)dt\right)^{i_j}}{i_j!}e^{-\int_{A_j}\lambda(t)dt},$$

where $\bar{A}_i$ refers to the intersection of the infinite slice with vertical projection A_i with A. This concludes the proof of part (I).

To show (II), we can use Theorem 1.1: it suffices to show that for all finite sets $\bar{A}_1$, the number of random vectors N falling in $\bar{A}_1$ is Poisson distributed with parameter $Vol(\bar{A}_1)$, and that every random vector in this set is uniformly distributed in it. The distribution of N is indeed Poisson with the given parameter because the X_i sequence determines a nonhomogeneous Poisson process with the correct rate function. Also, by Theorem II.3.1, a random vector $(X,U\lambda(X))$ is uniformly distributed in $\bar{A}_1$ if U is uniformly distributed on $[0,1]$ and X is a random vector with density proportional to $\lambda(x)$ restricted to A_1. Thus, it suffices to show that if an X is picked at random from among the X_i's in A_1, then X is a random vector with density proportional to $\lambda(x)$ restricted to A_1. Let B be a Borel set contained in A_1, and let us write λ_B and λ_{A_1} for the integrals of λ over B and A_1 respectively. Thus,

$$P(X\in B \mid X\in A_1)=P(X\in B \mid N(A_1)=1)$$
$$=\frac{P(N(B)=1,N(A_1-B)=0)}{P(N(A_1)=1)}$$

$$= \frac{\lambda_B e^{-\lambda_B} e^{-(\lambda_{A_1} - \lambda_B)}}{\lambda_{A_1} e^{-\lambda_{A_1}}}$$

$$= \frac{\lambda_B}{\lambda_{A_1}}$$

$$= \frac{1}{\lambda_{A_1}} \int_B \lambda(x)\, dx \ ,$$

which was to be shown.

Part 3 follows from Theorem 1.1 on homogeneous Poisson processes without further work. ■

Consider now the thinning algorithm of Lewis and Shedler (1979):

The thinning method (Lewis and Shedler)

$T \leftarrow 0$
$k \leftarrow 0$
REPEAT
    Generate Z, the first event in a nonhomogeneous Poisson process with rate function μ occurring after T. Set $T \leftarrow Z$.
    Generate a uniform [0,1] random variate U.
    IF $U \leq \frac{\lambda(Z)}{\mu(Z)}$
        THEN $k \leftarrow k+1$, $X_k \leftarrow T$
UNTIL False

The sequence of X_k's thus generated is claimed to determine a nonhomogeneous Poisson process with rate function λ. Notice that we have taken a nonhomogeneous Poisson process $0 < Y_1 < Y_2 < \cdots$ with rate function μ and eliminated some points. As we know, $(Y_1, U_1\mu(Y_1)), \ldots$ is a homogeneous Poisson process with unit rate under the curve of μ if $U_1, U_2, \ldots$ are iid uniform [0,1] random variates (Theorem 1.3). Thus, the subsequence falling under the curve of λ determines a homogeneous Poisson process with unit rate under that curve (part (iii) of the same theorem). Finally, taking the x-coordinates only of that subsequence gives a nonhomogeneous Poisson process with rate function λ.

The nonhomogeneous Poisson process with rate function μ is usually

obtained by the inversion method.

Example 1.4. Cyclic rate functions.

The following example is also due to Lewis and Shedler (1979): consider a cyclic rate function

$$\lambda(t) = \lambda(1+\cos(t))$$

with as obvious choice for dominating rate function $\mu(t)=2\lambda$. We have

$T \leftarrow 0$
$k \leftarrow 0$
REPEAT
  Generate an exponential random variate E.
  $T \leftarrow T + \frac{E}{2\lambda}$
  Generate a uniform [0,1] random variate U.
  IF $U \leq \frac{1+\cos(T)}{2}$
    THEN $k \leftarrow k+1$, $X_k \leftarrow T$
UNTIL False

It goes without saying that the squeeze principle can be used here to help avoiding the cosine computation most of the time. ■

A final word about the efficiency of the algorithm when used for generating a nonhomogeneous Poisson process on a set $[0,t]$. The expected number of events needed from the dominating process is $\int_0^t \mu(u)\,du$, whereas the expected number of random variates returned is $\int_0^t \lambda(u)\,du$. The ratio of the expected values can be considered as a fair measure of the efficiency, comparable in spirit to the rejection constant in the standard rejection method. Note that we cannot use the expected value of the ratio because that would in general be ∞ in view of the positive probability ($e^{-\int_0^t \lambda(u)\,du}$) of returning no variates.

1.4. Global methods for nonhomogeneous Poisson process simulation.

Nonhomogeneous Poisson processes on $[0,\infty)$ can always be obtained from homogeneous Poisson processes on $[0,\infty)$ by the following property (see e.g. Cinlar (1975, pp. 98-99)):

Theorem 1.4.

If $0<T_1<T_2<\cdots$ is a homogeneous Poisson process with unit rate on $[0,\infty)$, and if Λ is an integrated rate function, then

$$0<\Lambda^{-1}(T_1)<\Lambda^{-1}(T_2)<\cdots$$

determines a nonhomogeneous Poisson process with integrated rate function Λ.

Proof of Theorem 1.4.

We have implicitly shown this in the previous section. Let $i_1,\ldots$ be integers, let k be an integer, and let $N(A)$ be the number of points in a set $A\subseteq[0,\infty)$. Then, $N(A)$ is equal to $N^*(\Lambda(A))$ where N^* refers to the homogeneous Poisson process, and $\Lambda(A)$ is the set A transformed under Λ. Thus, if $A_1,\ldots,A_k$ are disjoint sets, it is easily seen that $N(A_1),\ldots,N(A_k)$ are distributed as $N^*(\Lambda(A_1)),\ldots,N^*(\Lambda(A_k))$, which is a sequence of independent Poisson random variables with parameters equal to the Lebesgue measures of the sets $\Lambda(A_i)$, i.e. $\int_{A_i}\lambda(t)\,dt$ where λ is the a.e. derivative of Λ. This shows that the transformed process is a nonhomogeneous Poisson process with integrated rate function Λ. ■

We observe that if $\int_0^\infty \lambda(t)\,dt<\infty$, then the function Λ^{-1} is not defined for very large arguments. In that case, the T_i's with values exceeding $\int_0^\infty \lambda(t)\,dt$ should be ignored. We conclude thus that only a finite number of events occur in such cases. No matter how large the finite value of the integral is, there is always a positive probability of not having any event at all.

Let us apply this theorem to the simulation restricted to a finite interval $[0,t_0]$. This is equivalent to the infinite interval case provided that $\lambda(t)$ is replaced by

$$\begin{cases}\lambda(t) & (0\le t\le t_0)\\ 0 & (t>t_0)\end{cases}$$

Thus, it suffices to use $\Lambda^{-1}(T_1),\ldots$ for all T_i's not exceeding $\Lambda(t_0)$. The inversion of Λ is sometimes not practical. The next property can be used to avoid it, provided that we have fast methods for generating order statistics with non-uniform

densities (see e.g. chapter V). The straightforward proof of its validity is left to the reader (see e.g. Cox and Lewis, 1966, chapter 2).

Theorem 1.5.

Let N be a Poisson random variate with parameter $\Lambda(t_0)$. Let $0<T_1<T_2<\cdots<T_N$ be order statistics corresponding to the distribution function

$$\frac{\Lambda(t)}{\Lambda(t_0)} \quad (0\le t\le t_0),$$

then this subsequence determines a nonhomogeneous Poisson process on $[0,t_0]$ with integrated rate function Λ.

Both Theorem 1.4 and Theorem 1.5 lead to global methods, i.e. methods in which a nonhomogeneous Poisson process can be obtained from another process, usually in a separate pass of the data. The methods of the previous section, in contrast, are sequential: the event times of the process are generated directly from left to right. Since the one-pass sequential approach allows optional stopping and restarting anywhere in the process, it is definitely of more practical value. In some applications, there is also a considerable savings in storage because no intermediate (or auxiliary) process needs to be stored. Finally, some global methods require the computation of Λ^{-1}, whereas the thinning method does not. This is an important consideration when Λ is difficult to compute.

For more examples, and additional details, we refer to the exercises and the other sections in this chapter. Readers who do not specialize in random process generation will probably not gain very much from reading the other sections in this chapter.

1.5. Exercises.

1. When $\int_0^\infty \lambda(t)\,dt<\infty$, the inversion and thinning methods for nonhomogeneous Poisson process generation need modifying. Show how.
2. Let N be the total number of events (points) in a nonhomogeneous Poisson process on the positive real line with rate function $\lambda(t)$. Show that there are only two possible situations:

$$P(N<\infty)=1 \quad (\int_0^\infty \lambda(t)\,dt<\infty)$$

$$P(N<\infty)=0 \quad (\int_0^\infty \lambda(t)\,dt=\infty).$$

3. The following rate function is given to you: $\lambda(t)$ is piecewise constant with breakpoints at $a, 2a, 3a, 4a, \ldots$, where for $t\in[ia,(i+1)a)$, $\lambda(t)=\lambda_i$, $i=0,1,2,\ldots$. Generalize the exponential spacings method for generating a nonhomogeneous Poisson process with this rate function. Hint: do not use transformations of exponential random variates when you cross breakpoints, but rely on the memoryless property of the exponential distribution.

4. We are interested in the generation of a nonhomogeneous Poisson process with log-linear rate function

$$\lambda(t) = c_0 e^{c_0+ct} \quad (t\geq 0).$$

where $c_0>0$, $c\in R$. There are two important situations: when $c<0$, the process dies out and only a finite number of events occurs. The process corresponds to an exponential population explosion however when $c>0$. Generate such a process by the inversion-of-Λ method.

5. This is a continuation of the previous exercise related to a method of Lewis and Shedler (1976) for simulating non-homogeneous Poisson processes with log-linear rate function. Show that if N is a Poisson $(-\frac{c_0}{c})$ random variable, and $E_1, E_2, \ldots$ is a sequence of iid exponential random variates, then, assuming $c<0$,

$$-\frac{E_i}{c(N-i+1)} \quad (1\leq i\leq N)$$

are distributed as the gaps between events in a nonhomogeneous Poisson process with rate function $\lambda(t) = c_0 e^{c_0+ct}$ $(t\geq 0)$ on $[0,\infty)$. Give the algorithm that exploits this property. Note that this implies that the expected number of events in such a process is $-\frac{c_0}{c}<\infty$. For the case $c>0$, show how by flipping the time axis around, you can reduce the problem to that of the case $c<0$ provided that one is only interested in simulation on a finite time interval.

6. Give an algorithm for generating random variates with a log-quadratic rate function. Hint: consider several cases as in the previous two exercises (Lewis and Shedler, 1979).

2. GENERATION OF RANDOM VARIATES WITH A GIVEN HAZARD RATE.

2.1. Hazard rate. Connection with Poisson processes.

In this section we consider the problem of the computer generation of random variables with a given hazard rate h on $[0,\infty)$. If X is a random variable with density f and distribution function F, then the **hazard rate** h and **cumulative hazard rate** H are inter-related as follows:

$$h(x) = \frac{f(x)}{1-F(x)} ;$$

$$H(x) = \int_0^x h(y)\,dy = -\log(1-F(x)) ;$$

$$F(x) = 1-e^{-H(x)} ;$$

$$f(x) = h(x)e^{-H(x)} .$$

The hazard rate plays a crucial role in reliability studies (Barlow and Proschan, 1965) and in all situations involving lifetime distributions. Note that $\int_0^\infty h(y)\,dy = \infty$ and thus $\lim_{x\to\infty} H(x) = \infty$. The key distribution now is the exponential: it has constant hazard rate of value 1. Roughly speaking, hazard rates tending to 0 correspond to densities with larger-than-exponential tails, and diverging hazard rates are for densities with smaller-than-exponential tails. For compact support distributions, we have $\lim_{x\uparrow c} H(x) = \infty$ for some finite c (corresponding to the rightmost point in the support). Sometimes, h or H is given, and not f or F. In particular, when only h is given, f cannot be computed exactly because we would first need to compute H by numerical integration. Thus, there is a need for methods which allow us to generate random variates with a given hazard rate h. Fortunately, such random variates are intimately connected to Poisson point processes.

Theorem 2.1.

Let $0<T_1<T_2<\cdots$ be a nonhomogeneous Poisson process with rate function h (and thus integrated rate function H). Then T_1 is a random variable with hazard rate h.

Proof of Theorem 2.1.

Note that for $x>0$,

$$P(T_1 \le x) = 1-P(\text{no event times in } [0,x])$$

$$= 1-e^{-\int_0^x h(t)\,dt}$$

$$= 1-e^{-H(x)},$$

which was to be shown. ■

This connection helps us understand the algorithms of this section. We will discuss the inversion, composition and thinning methods. For special sub-classes of hazard rate functions, there are universally applicable (black box) methods that are worth reporting. In particular for **DHR distributions** (distributions with decreasing hazard rate), the method of dynamic thinning will be introduced and analyzed (Devroye, 1985). Other classes, such as the class of **IHR distributions** (distributions with increasing hazard rate), are dealt with indirectly in the text and exercises.

2.2. The inversion method.

For generating a random variate with cumulative hazard rate H, it suffices to invert an exponential random variate:

Inversion method

Generate an exponential random variate E.
RETURN $X \leftarrow H^{-1}(E)$

If H^{-1} is not explicitly known, then we are forced to solve $H(X)=E$ for X by some iterative method. Here the discussion of the standard inversion method for distribution functions applies again.

We can easily verify that the algorithm is valid, either by using the connection with Poisson processes given in Theorem 2.1, or directly: for $x>0$ observe that if H is strictly increasing, then

$$P(H^{-1}(E) \leq x) = P(E \leq H(x)) = 1-e^{-H(x)} = F(x).$$

When H is not strictly increasing, then the chain of inequalities remains valid for any consistent definition of H^{-1}.

This method is difficult to attribute to one person. It was mentioned in the works of Cinlar (1975), Kaminsky and Rumpf (1977), Lewis and Shedler (1979) and Gaver (1979). In the table below, a list of examples is given. Basically, this list contains distributions with an easily invertible distribution function because

$F(x)=1-e^{-H(x)}$.

$f(x)$	$h(x)$	$H(x)$	$H^{-1}(E)$
$ax^{a-1}e^{-x^a}$ $(a>0)$(Weibull)	ax^{a-1}	x^a	$E^{\frac{1}{a}}$
$\frac{a}{(1+x)^{a+1}}$ (Pareto)	$\frac{a}{1+x}$	$a\log(1+x)$	$e^{\frac{E}{a}}-1$
ax^{a-1} $(a>0, x\leq 1)$ (power function)	$\frac{ax^{a-1}}{1-x^a}$	$-\log(1-x^a)$	$(1-e^{-E})^{\frac{1}{a}}$

2.3. The composition method.

When $h=h_1+\cdots+h_n$ where the h_i's are in turn hazard rates, then we can use Theorem 2.1 directly and use the fact that it suffices to consider the minimum of n random variables $X_1, \ldots, X_n$ with the individual hazard rates h_i. When the individual cumulative hazard rates are H_i, then this can be shown directly: for $x>0$,

$$P(\min(X_1, \ldots, X_n)\geq x) = \prod_{i=1}^{n} e^{-H_i(x)} = e^{-H(x)}.$$

If the decomposition is such that for some h_i we have $\int_0^\infty h_i(t)\,dt<\infty$, then the method is still applicable if we switch to nonhomogeneous Poisson processes.

Composition method

$X\leftarrow\infty$
FOR $i=1$ TO n DO
    Generate Z distributed as the first event time in a nonhomogeneneous Poisson process with rate function h_i (ignore this if there are no events in the process; if $\int_0^\infty h_i=\infty$, then Z has hazard rate h_i).
    IF $Z<X$ THEN $X\leftarrow Z$
RETURN X

Usually, the composition method is slow because we have to deal with all the individual hazard rates. There are shortcuts to speed things up a bit. For example, after we have looked at the first component and set X equal to the random variate with hazard rate h_1, it suffices to consider the nonhomogeneous Poisson processes restricted to $[0,X]$. The point is that if X is small, then the probability of observing one or more event times in this interval is also small. Thus, often a

quick check suffices to avoid random variate generation for the remaining nonhomogeneous Poisson processes. To illustrate this, decompose h as follows:

$$h(x) = h_1(x) + h_2(x)$$

where h_1 is a hazard rate which puts its mass near the origin. The function h_2 is nonnegative, but does not have to be a hazard rate. It can be considered as a small adjustment, h_1 being the main (easy) component. Then the following algorithm can be used:

Composition method with quick acceptance

Generate a random variate X with hazard rate h_1.
Generate an exponential random variate E.
IF $E \leq H_2(X)$ (H_2 is the cumulative hazard rate for h_2)
THEN RETURN $X \leftarrow H_2^{-1}(E)$
ELSE RETURN X

Something can be gained if we replace $X \leftarrow H_2^{-1}(E)$ by a step in which we return a random variate X distributed with hazard rate

$$h_2(x)\frac{1-F_2(x)}{F_2(X)-F_2(x)}$$

which can be done by methods that do not involve inversion. The expected number of times that we need to use the second (time-consuming) step in the algorithm is the probability that $E \leq H_2(X)$ where X has hazard rate h_1:

$$P(E \leq H_2(X)) = \int_0^\infty h_1(y)e^{-H_1(y)}(1-e^{-H_2(y)})\,dy$$

$$= 1-\int_0^\infty h_1(y)e^{-H(y)}\,dy$$

$$= 1-\int_0^\infty (h(y)-h_2(y))e^{-H(y)}\,dy$$

$$= \int_0^\infty h_2(y)e^{-H(y)}\,dy$$

$$= \int_0^\infty \left(\frac{h_2(y)}{h(y)}\right) f(y)\,dy$$

where f is the density corresponding to f. From the last expression we conclude that it is important to keep $\frac{h_2}{h}$ small.

2.4. The thinning method.

Combining the theorem about thinning Poisson processes (Theorem 1.4) with Theorem 2.1 shows that the following algorithm produces a random variate with hazard rate h, provided that we can generate a nonhomogeneous Poisson point process with rate function g where

$$h(x) \leq g(x) \quad (\text{all } x) .$$

Thinning method (Lewis and Shedler, 1979)

$X \leftarrow 0$
REPEAT
    Generate a random variate Δ with hazard rate $g(X+x)$ $(x \geq 0)$ (equivalently, generate the first occurrence in a nonhomogeneous Poisson point process with the same rate function).
    Generate a uniform [0,1] random variate U.
    $X \leftarrow X+\Delta$
UNTIL $Ug(X) \leq h(X)$
RETURN X

This algorithm is most efficient when g is very simple. In particular, constant dominating rate functions $g = g_0$ are practical, because Δ can be obtained as $\frac{E}{g_0}$ where E is an exponential random variate. We will now see what the expected complexity is for this algorithm. It is annoying that the distribution of the number of iterations (which we shall call N) depends very heavily on h and g. Recall, in comparison, that for the rejection method, the distribution is always geometric. For the thinning method, we might even have $E(N)=\infty$, so that it is absolutely essential to clarify just how $E(N)$ depends upon h and g. The following theorem is due to Devroye (1985):

Theorem 2.2. (Analysis of the thinning method.)
Let f and F be the density and distribution function corresponding to a hazard rate h. Let $g \geq h$ be another hazard rate having cumulative hazard rate G. Then the expected number of iterations in the thinning algorithm given above is

$$E(N) = \int_0^\infty g(x)(1-F(x))\,dx = \int_0^\infty f(x)G(x)\,dx .$$

Proof of Theorem 2.2.

Let us call the X variates in subsequent iterations X_i, where $i=1,2,\ldots$. Similarly, the uniform [0,1] random variates used in the algorithm have also subscripts referring to the iteration, as in $U_1, U_2, \ldots$. In Theorem 1.4 we have shown that $(X_1, U_1 g(X_1)), (X_2, U_2 g(X_2)), \ldots$ if continued at infinitum form a homogeneous Poisson process with unit rate on the area bounded by the x-axis and the curve g. The only thing that we introduce in the thinning method is a stopping rule. We condition now on X, the random variate returned in the algorithm. Notice that N is 1 plus the number of event times in a nonhomogeneous Poisson process with rate function $g-h$ restricted to $[0,X)$. Thus, conditioned on X, $N-1$ is Poisson distributed with parameter $\int_0^X (g-h)$. This observation uses the properties of Theorem 1.4 connecting homogeneous Poisson processes in the plane with nonhomogeneous Poisson processes on the line.

It is a simple matter to compute $E(N)$:

$$\begin{aligned} E(N) &= 1+E\left(\int_0^X (g-h)\right) = 1+E\left(\int_0^X g\right)-E(H(X)) \\ &= E\left(\int_0^X g\right) \\ &= \int_0^\infty \int G \\ &= \int_0^\infty g(1-F)\,. \end{aligned}$$

Here we used the fact that $H(X)$ is exponentially distributed, and, in the last step, partial integration. ■

Theorem 2.2 establishes a connection between $E(N)$ and the size of the tail of X. For example, when $g=c$ is a constant, then

$$E(N) = cE(X)\,.$$

Not unexpectedly, the value of $E(N)$ is scale-invariant: it depends only upon the shapes of h and g. When g increases, as for example in

$$g(x) = \sum_{i=0}^{n} c_i x^i\,,$$

then $E(N)$ depends upon more than just the first moment:

$$E(N) = \sum_0^n \frac{c_i}{i+1} E(X^{i+1})\,.$$

There are plenty of examples for which $E(N)=\infty$ even when $g(x)=1$ for all x. Consider for example $h(x)=\frac{1}{x+1}$, which corresponds to the long-tailed density $f(x)=\frac{1}{(x+1)^2}$. Generally speaking, $E(N)$ is small when g and h are close. For example, we have the following helpful inequalities:

Theorem 2.3.

The thinning algorithm satisfies

$$E(N) \leq \sup_{x>0} \frac{g(x)}{h(x)} ;$$

$$E(N) \leq \sup_{x>0} \frac{1-F(x)}{1-F^*(x)} ,$$

where F, F^* are the distribution functions for h and g respectively.

Proof of Theorem 2.3.

The first inequality follows from

$$E(N) = \int_0^\infty \frac{g(x)}{h(x)} f(x)\, dx ,$$

and the second inequality is a consequence of

$$E(N) = \int_0^\infty f^*(x) \frac{1-F(x)}{1-F^*(x)}\, dx ,$$

where f^* is the density corresponding to g. ■

There are examples in which g and h appear to be far apart ($\lim_{x\uparrow\infty} \frac{g(x)}{h(x)} = \infty$), yet $E(N)<\infty$: consider for example $h(x)=\frac{1}{x+1}$, $g(x)=\frac{1}{(x+1)^a}$, $0<a\leq 1$. The explanation is that g and h should be close to each other near the origin and that the difference does not matter too much in low density regions such as the tails.

The expression for $E(N)$ can be manipulated to choose the best dominating hazard rate g from a parametrized class of hazard rates. This will not be explored any further.

2.5. DHR distributions. Dynamic thinning.

In this section we will try to obtain a black box generator for DHR distributions, i.e. a generator which does not require a priori explicit knowledge of the form of h. The method that will be given in this section is the method of **dynamic thinning.** This principle in itself is also useful for other distributions and for the nonhomogeneous Poisson process on the real line. The algorithm resembles the thinning algorithm, but the dominating hazard rate is dynamic, i.e. it varies during the execution of the algorithm.

The DHR distributions form a sub-class of the monotone densities because $f = he^{-H}$, $h \downarrow$ and $H \uparrow$. It contains the Pareto distribution with parameter $a > 0$:

$$h(x) = \frac{a}{x+1},$$

the Weibull distribution with parameter $a \leq 1$ and the gamma distribution with parameter $a \leq 1$. The peak of the density is at 0, with value $f(0) = h(0)$. This value can of course be ∞ as for the gamma (a) density with $0 < a < 1$. The class has some desirable properties, for example, it is closed under convex combinations (see exercises), which means that mixtures of DHR distributions are again DHR.

The inversion method is based upon the fact that the solution X of $H(X) = E$ where E is exponentially distributed, has cumulative hazard rate H. But for DHR distributions, H is concave (its derivative h is nonincreasing). Thus, Newton-Raphson iterations started at 0 converge whenever $h(0) < \infty$:

Inversion method for DHR distributions

$X \leftarrow 0$
REPEAT
$$X \leftarrow X + \frac{E - H(X)}{h(X)}$$
UNTIL False

In practical applications, an appropriate stopping rule must be added. An exact solution usually requires infinite time (this is not the case if h is piecewise constant !). The thinning method, if it is to be used in black box mode, can only use the constant dominating hazard rate $g = h(0)$, in which case the expected number of iterations becomes

$$h(0)E(X).$$

We recall however that DHR distributions have heavier-than-exponential tails. Thus, the fact that $E(N)$, the expected number of iterations, is proportional to $E(X)$ could be a serious drawback. The two prototype examples that we will consider throughout this section are the exponential density $(E(N) = h(0)E(X) = 1)$ and the Pareto (a) density

$$f(x) = \frac{a}{(1+x)^{a+1}},$$

for which $h(x) = \frac{a}{1+x}$, $H(x) = a \log(1+x)$, $h(0) = a$, and, if $a > 1$, $E(X) = \frac{1}{a-1}$. Thus,

$$E(N) = \begin{cases} \infty & 0 < a \leq 1 \\ \frac{a}{a-1} & a > 1 \end{cases}.$$

We are now ready to present the dynamic thinning algorithm:

Dynamic thinning algorithm for DHR distributions

$X \leftarrow 0$
REPEAT
  $\lambda \leftarrow h(X)$
  Generate an exponential random variate E and a uniform [0,1] random variate U.
  $X \leftarrow X + \frac{E}{\lambda}$
UNTIL $\lambda U \leq h(X)$
RETURN X

The method uses thinning with a constant but continuously adjusted dominating hazard rate λ. When h decreases as X grows, so will λ. This forces the probability of acceptance up. The complexity can again be measured in terms of the number of iterations before halting, N. Note that the number of evaluations of h is $1+N$ (and not $2N$ as one might conclude from the algorithm shown above, because some values can be recuperated by introducing auxiliary variables). If $\lambda \leftarrow h(X)$ is taken out of the loop, and replaced at the top by $\lambda \leftarrow h(0)$, we obtain the standard thinning algorithm. While both algorithms do not require any knowledge about h except that h is DHR, a reduction in N is hoped for when dynamic thinning is used. In Devroye (1985), various useful upper bounds for $E(N)$ are obtained. Some of these are given in the next subsection and in the exercise section. The value of $E(N)$ is always less than or equal that of the thinning method. For example, for the Pareto (a) distribution, we obtain

$$E(N) = \frac{1}{\int_0^\infty e^{-z} \left(1+\frac{z}{a}\right)^{-1} dz},$$

which is finite for all $a > 0$. In fact, we have the following chain of inequalities showing the improvement over standard thinning:

$$E(N) = \frac{1}{\int_0^\infty e^{-z}(1+\frac{z}{a})^{-1}\,dz}$$

$$\leq \frac{a+1}{a} \quad \text{(use Jensen's inequality; note: } \frac{1}{(1+\frac{z}{a})} \text{ is convex in } z\text{)}$$

$$< \frac{a}{a-1} \quad \text{(for all } a>1\text{)}$$

$$= \mu = h(0)E(X)\,.$$

For example, at $a=1$, we have $E(N)\leq 2$ whereas $h(0)E(X)=\infty$.

2.6. Analysis of the dynamic thinning algorithm.

Throughout this section, we will use the following notation:

$$\mu = h(0)E(X)\,,$$

$$\beta = \sup_{x\geq 0}\int_0^\infty e^{-yh(x)}(h(x)-h(x+y))\,dy\,,$$

$$\gamma = \sup_{x\geq 0}\frac{h(x)}{h(x+\frac{1}{h(x)})}\,,$$

$$\xi = E(\log_+(h(0)X))$$

where μ,β,γ and ξ are various quantities that will appear in the upper bounds for $E(N)$ given in this subsection. Note that ξ is the logarithmic moment of $h(0)X$, for which we have, by Jensen's inequality,

$$\xi \leq E(\log(h(0)X+1)) \leq \log(\mu+1) \leq \mu\,,$$

so that ξ is always finite when μ is finite. Obtaining an upper bound of the form $O(\xi)$ is, needless to say, strong proof that dynamic thinning is a drastic improvement over standard thinning. This is the goal of this subsection. Before we proceed with the technicalities, it is perhaps helpful to collect all the results.

Bounds C and D are never better than bound B, but often γ is easier to compute than β. For the Pareto family, we obtain via D,

$$E(N) \leq \gamma = \frac{a+1}{a},$$

a result that can be obtained from B via Jensen's inequality too. Inequalities E-H relate the size of the tail of X to $E(N)$, and give us more insight into the behavior of the algorithm. Of these, inequality H is perhaps the easiest to understand: $E(N)$ cannot grow faster than the logarithm of μ. Unfortunately, when $\mu=\infty$, it is of little help. In those cases, the logarithmic moment ξ is often finite. For example, this is the case for all members of the Pareto family. We will now prove Theorem 2.4. It requires a few Lemmas and other technical facts. Yet the proofs are instructive to those wishing to learn how to apply embedding techniques and well-known inequalities in the analysis of algorithms. Non-technical readers should most certainly not read beyond this point.

Proof of Theorem 2.4.

Part A. This part uses embedding. Consider the sequence of random vectors $(Y_1,h(0)U_1),(Y_2,h(0)U_2),\ldots$ where the U_i's are iid uniform [0,1] random variables, and $0=Y_0<Y_1<Y_2<\cdots$ are defined by the relations:

$$Y_{i+1} = Y_i + \frac{E_{i+1}}{h(0)}$$

where $E_1,E_2,\ldots$ are iid exponential random variates. This is the sequence considered in standard thinning, where we stop when for the first time $h(0)U_i \leq h(Y_i)$. We recall from Theorem 2.3 that in that case $E(N)=\mu=E(h(0)X)$. Let us use starred random variables for the subsequence satisfying $h(0)U_i \leq h(Y_{i-1})$. Observe first that this sequence is distributed as the sequence of random vectors used in dynamic thinning. Then, part A follows without work because we still stop when the first random vector falling below the curve of h is encountered.

Part B. Let the E_i's be as before. The sequence $Y_0<Y_1<\cdots$ used in dynamic thinning satisfies: $Y_0=0$, and

$$Y_{i+1} = Y_i + \frac{E_{i+1}}{h(Y_i)}.$$

Note that this is the sequence of possible candidates for the returned random variate X in the algorithm. The index i refers to the iteration. Taking the stopping rule into account, we have for $i \geq 1$,

$$P(N>i \mid Y_0, \ldots, Y_i) = \prod_{j=1}^{i} \left(1-\frac{h(Y_j)}{h(Y_{j-1})}\right).$$

Theorem 2.4.

The expected number of iterations in the dynamic thinning algorithm applied to a DHR distribution with bounded h does not exceed any of the following quantities:

A. μ;

B. $\frac{1}{1-\beta}$;

C. $\frac{e}{e-1}\gamma$;

D. γ (when h is also convex);

E. $(8\mu)^{\frac{1}{2}}+4(8\mu)^{\frac{1}{4}}$;

F. $\inf_{p>2}(4p+2+\frac{2\xi}{\log(p-1)})$;

G. $O(\frac{\xi}{\log(\xi)})$ as $\xi\to\infty$;

H. $O(\frac{\log(\mu)}{\log\log(\mu)})$ as $\mu\to\infty$.

Part A states that we have an improvement over standard thinning. Inequalities B and D are sharp: for example, for the exponential distribution, we have $\beta=0$, $\gamma=1$, which leads to $E(N)\leq 1$. Inequality B is also sharp for the Pareto family defined above. One can easily verify that

$$\beta = \int_0^\infty e^{-\frac{ya}{1+x}}\left(\frac{a}{1+x}-\frac{a}{1+x+y}\right)dy$$

$$= \int_0^\infty e^{-z}\left[1-\frac{1}{1+\frac{z}{a}}\right]dz$$

where we used the transformation $z=\frac{ay}{1+x}$. By carefully checking the induction argument used in the proof of Theorem 2.4, we see that for any $i\geq 0$, $P(N>i)=\beta^i$ and thus that

$$E(N)=\frac{1}{\int_0^\infty e^{-z}(1+\frac{z}{a})^{-1}\,dz}$$

$$=\frac{1}{1-\beta}.$$

Thus, for $i \geq 2$,

$$P(N>i \mid Y_0, \ldots, Y_{i-1})$$
$$= \prod_{j=1}^{i-1} (1-\frac{h(Y_j)}{h(Y_{j-1})}) \int_0^\infty e^{-yh(Y_{i-1})}(h(Y_{i-1})-h(Y_{i-1}+y))\, dy$$
$$\leq \beta \prod_{j=1}^{i-1} (1-\frac{h(Y_j)}{h(Y_{j-1})}),$$

and we obtain, by a simple induction argument on i, that

$$P(N>i) \leq \beta^i \quad (i \geq 0).$$

Thus,

$$E(N) = \sum_{i=0}^{\infty} P(N>i) \leq \frac{1}{1-\beta}.$$

Part C. Part C is obtained from B by bounding β from above. Fix x and $c>0$. Then

$$\int_0^\infty e^{-yh(x)}(h(x)-h(x+y))\, dy$$
$$\leq \int_{y>\frac{c}{h(x)}} e^{-yh(x)}(h(x)-h(x+y))\, dy + \int_0^{\frac{c}{h(x)}} e^{-yh(x)}(h(x)-h(x+y))\, dy$$
$$\leq \int_c^\infty e^{-z}\, dz + \int_0^{\frac{c}{h(x)}} e^{-yh(x)}(h(x)-h(x+\frac{c}{h(x)}))\, dy$$
$$= e^{-c} + (1-e^{-c})\left(\frac{h(x+\frac{c}{h(x)})}{h(x)}\right)$$
$$= 1-(1-e^{-c})\frac{h(x+\frac{c}{h(x)})}{h(x)}.$$

Inequality C follows after taking $c=1$.

Part D. Inequality D follows by applying Jensen's inequality to an intermediate expression in the preceding chain of inequalities:

$$\int_0^\infty e^{-yh(x)}(h(x)-h(x+y))\, dy$$
$$= \int_0^\infty e^{-yh(x)}h(x)(1-\frac{h(x+y)}{h(x)})\, dy$$
$$\leq 1-\frac{1}{h(x)}h\left(x+\int_0^\infty e^{-yh(x)}h(x)y\, dy\right)$$

$$= 1-\frac{1}{h(x)}h(x+\frac{1}{h(x)}) .$$

Lemma 2.1, needed for parts E-H. We will show that for $x \geq 0$, $p > 2$, and integer m in $\{0,1, \ldots, n\}$,

$$P(N>n) \leq P(X>x)+\frac{h(0)x}{p^{n-m}}+(1-\frac{1}{p})^m \qquad (n>0) .$$

Define the E_i and Y_i sequences as in the proof of part B, and let $U_1, U_2, \ldots$ be a sequence of iid uniform [0,1] random variables. Note that the random variate X returned by the algorithm is Y_N where N is the first index i for which $U_i h(Y_{i-1}) \leq h(Y_i)$. Define N_1, N_2 by:

$$N_1 = \sum_{i=1}^{n} I_{[h(Y_i) \leq \frac{1}{p}h(Y_{i-1})]} ,$$

$$N_2 = \sum_{i=1}^{n} I_{[h(Y_i) > \frac{1}{p}h(Y_{i-1})]} .$$

Then we can write the following:

$$[N>n] \subseteq [X>x] \cup [X \leq x, N_1 \geq n-m, N>n] \cup [N_2 \geq m, N>n] .$$

Now,

$$P(X \leq x, N_1 \geq n-m, N>n) \leq P(E_1 \leq \frac{xh(0)}{p^{n-m}}) \leq \frac{xh(0)}{p^{n-m}}$$

and

$$P(N_2 \geq m, N>n) \leq P(N>n \mid N_2 \geq m) \leq (1-\frac{1}{p})^m .$$

This concludes the proof of the Lemma.

Part E. Consider Lemma 2.1, and take $x = x_n$ random, independent of X and uniformly distributed on $[\frac{n}{Ch(0)}, \frac{n+1}{Ch(0)}]$ where $C>0$ is a constant to be chosen further on. Take $m = m_n = \left\lceil \frac{n}{2} \right\rceil$, and take p constant and independent of n. We will apply the formula

$$E(N) = \sum_{n=0}^{\infty} P(N>n)$$

and use Lemma 2.1, averaged over x_n. This yields an upper bound consisting of three terms:

(1)

$$\sum_{n=0}^{\infty} P(X>x_n) = \sum_{n=0}^{\infty} \int_{n}^{n+1} P(Ch(0)X>t)\, dt$$

$$= \int_0^\infty P(Ch(0)X > t)\, dt = E(Ch(0)X) = C\mu .$$

(ii)

$$\sum_{n=0}^{\infty} (1-\frac{1}{p})^{m_n} = 1+2\sum_{j=1}^{\infty}(1-\frac{1}{p})^{j} = 1+\frac{2(1-\frac{1}{p})}{\frac{1}{p}} = 2p-1 .$$

(iii)

$$\frac{1}{C}\sum_{n=0}^{\infty}(\int_n^{n+1} t\ dt)p^{-(n-m_n)} = \frac{1}{C}\sum_{n=0}^{\infty}\frac{2n+1}{2}p^{-(n-m_n)}$$

$$= \frac{2}{C}\sum_{n=0}^{\infty}(2n+1)p^{-n}$$

$$= \frac{2}{C}(\frac{1}{1-\frac{1}{p}}+2\sum_{n=0}^{\infty}np^{-n})$$

$$= \frac{2}{C}(\frac{1}{1-\frac{1}{p}}+\frac{2}{p}\frac{1}{(1-\frac{1}{p})^2})$$

$$= \frac{2}{C}\frac{1+\frac{1}{p}}{(1-\frac{1}{p})^2} .$$

These estimates are substituted in

$$E(N) \leq 1+\sum_{n=1}^{\infty}(P(X>x_n)+\frac{h(0)x_n}{p^{n-m_n}}+(1-\frac{1}{p})^{m_n} .$$

This gives the upper bound

$$E(N) \leq 1+C\mu-P(X>x_0)+2(p-1)+\frac{2}{C}(\frac{p(p+1)}{(p-1)^2}-\frac{1}{4}) .$$

Since $h(0)X$ is stochastically greater than an exponential random variate, we have

$$P(X>x_0) = \int_0^1 P(Ch(0)X>t)\, dt \geq \int_0^1 e^{-\frac{t}{C}}\, dt$$

$$= C\int_0^{\frac{1}{C}} e^{-z}\, dz = C(1-e^{-\frac{1}{C}}) \geq 1-\frac{1}{2C} .$$

Thus,

$$E(N) \leq C\mu+2(p-1)+\frac{2}{C}\frac{p(p+1)}{(p-1)^2} .$$

The optimal choice for C is

$$C = \sqrt{\frac{2p(p+1)}{\mu(p-1)^2}} ,$$

which, after substitution, gives

$$\begin{aligned} E(N) &\leq 2(p-1)+\sqrt{8\mu}\sqrt{\frac{p(p+1)}{(p-1)^2}} \\ &< 2(p-1)+\sqrt{8\mu}\frac{p+1}{p-1} \\ &= 2(p-1)+\frac{2\sqrt{8\mu}}{p-1}+\sqrt{8\mu} . \end{aligned}$$

The right-hand-side is minimal for $p-1=(8\mu)^{\frac{1}{4}}$, and this choice gives inequality E.

Part F. In Lemma 2.1, replace n by $2j$, and sum over j. Set $m_{2j}=j$, $p_{2j}=p>2$, and $h(0)x_{2j}=(p-1)^j$. Since for any random variable Z,

$$\sum_{j=0}^{\infty} P(Z>j) \leq 1+\int_0^{\infty} P(Z>t)\,dt = 1+E(Z_+) ,$$

we see that

$$\begin{aligned} E(N) &\leq 2\sum_{j=0}^{\infty} P(N>2j) \\ &\leq 2\sum_{j=0}^{\infty}(P(h(0)X>(p-1)^j)+2(1-\frac{1}{p})^j) \\ &= 2\sum_{j=0}^{\infty} P(\frac{\log_+(h(0)X)}{\log(p-1)}>j)+4p \\ &\leq 2E(\frac{\log_+(h(0)X)}{\log(p-1)})+4p+2 . \end{aligned}$$

Part G. Inequality G follows from inequality F for the following choice of p:

$$p = 2+\frac{\xi}{2\log^2(1+\xi)} .$$

This value was obtained as follows: inequality F is sharpest when p is picked as the solution of $(p-1)\log^2(p-1)=\frac{\xi}{2}$. But because we want $p>2$, and because we want a good p for large values of ξ, it is good to obtain a rough solution by functional iteration, and then adding 2 to this to make sure that the restrictions on p are satisfied. Resubstitution yields:

$$E(N) \leq 10+\frac{2\xi}{\log^2(1+\xi)}+\frac{2\xi}{\log(1+\frac{\xi}{2\log^2(1+\xi)})} ,$$

which is $O(\frac{\xi}{\log(\xi)})$ as $\xi \to \infty$.

Part H. Use the bound of part G, and the fact that $\xi \leq \log(1+\mu)$. In fact, we have shown that

$$E(N) \leq (2+o(1))\frac{\log(\mu)}{\log\log(\mu)}$$

as $\mu \to \infty$. ■

2.7. Exercises.

1. Sketch the hazard rate for the halfnormal density for $x > 0$. Determine whether it is monotone, and show that $\lim_{x \uparrow \infty} \frac{h(x)}{x} = 1$.
2. Give an efficient algorithm for the generation of random variates from the left tail of the extreme value distribution truncated at $c < 0$ (the extreme value distribution function before truncation is $e^{-e^{-x}}$). Hint: when E is exponentially distributed, then $\frac{1}{b}\log(1+bEe^{-a})$ has hazard rate $h(x) = e^{a+bx}$ for $x > 0$, $b > 0$.
3. Show that when H is a cumulative hazard rate on $[0,\infty)$, then $\frac{H(x)}{x}$ is a hazard rate on $[0,\infty)$. Assume now that random variates with cumulative hazard rate H are easy to generate. How would you generate random variates with hazard rate $\frac{H(x)}{x}$?
4. Prove that $\frac{1}{x}$ cannot be a hazard rate on $[0,\infty)$.
5. Construct a hazard rate on $[0,\infty)$, continuous at all points except at $c > 0$, having the additional properties that $h(x) > 0$ for all $x > 0$, and that $\lim_{x \uparrow c} h(x) = \lim_{x \downarrow c} h(x) = \infty$.
6. In this exercise, we consider a tight fit for the thinning method: $M = \int (g-h) < \infty$. Show first that

$$E(N) \leq 1 + \int_0^\infty (g-h).$$

Prove also that the probability that N is larger than Me decreases very rapidly to 0, by establishing the inequality

$$P(N \geq i) \leq e^{-M}\left(\frac{eM}{i}\right)^i \qquad (i \geq M).$$

To do this, start with $P(N \geq i) \leq e^{-ti}E(e^{tN})$ where $t \geq 0$ is arbitrary (this is Jensen's inequality). Evaluate the expected value, bound this value by introducing M, and optimize with respect to t.

7. Consider the family of hazard rates $h_b(x)=\frac{x}{1+bx}$ $(x>0)$, where $b>0$ is a parameter. Discuss random variate generation for this family. The average time needed per random variate should remain uniformly bounded over b.

8. Give an algorithm for the generation of random variates with hazard rate $h_b(x)=b+x$ $(x>0)$ where $b\geq 0$ is a parameter. Inversion of an exponential random variate requires the evaluation of a square root, which is considered a slow operation. Can you think of a potentially faster method ?

9. Develop a thinning algorithm for the family of gamma densities with parameter $a\geq 1$ which takes expected time uniformly bounded over a.

10. The hazard rate has infinite peaks at all locations at which the density has infinite peaks, plus possibly an extra infinite peak at ∞. Construct a monotone density f which is such that it oscillates infinitely often in the following extreme sense:

$$\limsup_{x\uparrow\infty} h(x) = \infty ;$$

$$\liminf_{x\uparrow\infty} h(x) = 0 .$$

Notice that h is neither DHR nor IHR.

11. If X is a random variate with hazard rate h, and ψ is a suitable smooth monotone transformation, give a formula for the hazard rate of $\psi(X)$ and conditions under which your formula is valid. See Gaver (1979) for several examples of such transformations.

12. Show that a mixture of DHR distributions is again a DHR distribution (Barlow, Marshall and Proschan, 1963).

13. Show that for any DHR random variable X, $\mu=E(h(0)X)\geq 1$.

14. Construct a DHR distribution for which the logarithmic moment $\xi=E(\log_+(h(0)X))=\infty$.

15. For the Pareto family (density $f(x)=\frac{a}{(1+x)^{a+1}}$, $x>0$), find the rate of increase of ξ, the logarithmic moment, as $a\downarrow 0$ (the answer should be of the form: $\xi\sim$ simple expression involving a).

16. Develop a black box method for DHR distributions with $h(0)=\infty$.

17. Let the hazard rate h be piecewise constant with breakpoints at $0=x_0<x_1<x_2<\cdots$ and values h_i on $(x_{i-1},x_i]$, $i\geq 1$. Assume that these numbers are given in an infinite table. Describe the inversion algorithm. Determine the expected number of iterations as a function of the x_i's and the h_i's.

18. Show that for the dynamic thinning method for DHR distributions, $E(N)\leq 4+\sqrt{24\mu}$, where $\mu=E(h(0)X)$ (Devroye, 1985).

19. This exercise is concerned with an improvement over inequalities $F-H$ in Theorem 2.4. Define the random variable $Y=\log_+(h(0)X)$, and the quantity

$$\chi = E\left(\frac{Y}{\log(1+Y)}\right) .$$

A. Show that $\chi<\infty$ implies $\xi<\infty$ (try to do this by establishing an inequality).

B. Show by example that there exists a density f for which $\chi<\infty$, yet $\xi=\infty$.

C. Find positive constants $a>0, b>0$ such that for the dynamic thinning method, $E(N)\le a+b\chi$. Hint: in Lemma 2.1, choose $p=p_n=\dfrac{n}{\log^3(n+1)}$, $m=m_n=\left\lfloor \dfrac{n}{\log(n+1)} \right\rfloor$, $x=x_n=\dfrac{1}{h(0)}e^{n\log n-4n\,\mathrm{loglog}(n+1)}$; and use $E(N)\le n_0+\sum_{n=n_0}^{\infty} P(N>n)$ for an appropriate n_0 (Devroye, 1985).

3. GENERATING RANDOM VARIATES WITH A GIVEN DISCRETE HAZARD RATE.

3.1. Introduction.

Assume that we wish to generate a random variate with a given probability vector $p_1, p_2, \ldots$, and that the **discrete hazard rate function** $h_n, n=1,2,\ldots$ is given, where

$$h_n = \frac{p_n}{q_n},$$

$$q_n = \sum_{i=n}^{\infty} p_i .$$

One verifies quickly that

$$p_n = h_n \prod_{i<n} (1-h_i) .$$

In some applications, the original probability vector of p_n's has a more complicated form than the discrete hazard rate function.

The general methods for random variate generation in the continuous case have natural extensions here. As we will see, the role of the exponential distribution is inherited by the geometric distribution. In different sections, we will briefly touch upon various techniques, while examples will be drawn from the classes of logarithmic series distributions and negative binomial distributions. In general, if we have finite-valued random variables that remain fixed throughout the simulation, table methods should be used. Thus, it seems appropriate to draw all the examples from classes of distributions with unbounded support.

3.2. The sequential test method.

The following method will be called the **sequential test method.** Although it is conceptually very simple, it seems to have been formally proposed for the first time by Shanthikumar (1983,1985).

Sequential test method

$X \leftarrow 0$
REPEAT
    Generate a uniform [0,1] random variate U.
    $X \leftarrow X+1$
UNTIL $U \leq h_X$
RETURN X

The validity of this method follows directly from the fact that all h_n's are numbers in [0,1], and that

$$p_n = h_n \prod_{i<n} (1-h_i) .$$

It is obvious that the number of iterations needed here is equal to X. The strength of this method is that it is universally applicable, and that it can be used in the black box mode. When it is compared with the inversion method for discrete random variates, one should observe that in both cases the expected number of iterations is $E(X)$, but that in the inversion method, only one uniform random variate is needed, versus one uniform random variate per iteration in the sequential test method. If h_n is computed in $O(1)$ time and p_n is computed as the product of n factors involving $h_1, \ldots, h_n$, then the expected time of the inversion method grows as $E(X^2)$. Fortunately, there is a simple recursive formula for p_n:

$$p_{n+1} = p_n (\frac{h_{n+1}}{h_n})(1-h_n) .$$

Thus, if the p_n's are computed recursively in this manner, the inversion method takes expected time proportional to $E(X)$, and the performance should be comparable to that of the sequential test method.

3.3. Hazard rates bounded away from 1.

Consider the class of discrete hazard rates h_n with supremum $\rho<1$. This class will be called the class $H(\rho)$. For such hazard rates, the sequential test method can be accelerated by observing that we can jump ahead more than 1 in each iteration. To see this, assume that X is geometrically distributed with parameter p:

$$P(X=n) = p(1-p)^{n-1} \quad (n \geq 1).$$

Then X has hazard rate $h_n = p$. But in that case the sequential test method counts the number of iid uniform [0,1] random variates generated until for the first time a number smaller than p is obtained. This is of course known to be geometrically distributed with parameter p. In this special case, the individual uniform random variates can be avoided, because we can generate X directly by inversion of a uniform random variate U as

$$X \leftarrow \left\lceil \frac{-\log U}{-\log(1-p)} \right\rceil$$

or as $X \leftarrow \left\lceil \frac{E}{-\log(1-p)} \right\rceil$, where E is an exponential random variate. For the limit case $p=1$, we have $X=1$ with probability one. The smaller p, the more dramatic the improvement. For non-geometric distributions, it is possible to give an algorithm which parallels to some extent the thinning algorithm.

Thinning method for discrete distributions

NOTE: This algorithm is valid for hazard rates in $H(\rho)$ where $\rho \in (0,1]$ is a given number.
$X \leftarrow 0$
REPEAT
  Generate iid uniform [0,1] random variates U, V.
  $X \leftarrow X + \left\lceil \frac{\log U}{\log(1-\rho)} \right\rceil$
UNTIL $V \leq \frac{h_X}{\rho}$
RETURN X

This algorithm is due to Shanthikumar (1983,1985). We have to show that it is valid, and verify what the expected time complexity is.

Theorem 3.1. (Shanthikumar, 1983,1985)

The discrete thinning method generates a random variate with discrete hazard rate h_n.

Proof of Theorem 3.1.

Let $G_1, G_2, \ldots$ be the sequence of iid geometric (p) random variates used in the discrete thinning method. Let X be the returned random variate. Thus, $X = G_1 + \cdots + G_N$ where N is the number of iterations. Let us define the partial sums $S_n = \sum_{i=1}^{n} G_i$. Thus, $X = S_N$. We compute the probability $P(S_N = n)$ from the following formula:

$$P(X=n, N=k+1, S_1=n_1, \ldots, S_r=n_k)$$
$$= h_n \rho^k (1-\rho)^{n-1-k} \prod_{i=1}^{k} \left(1-\frac{h_{n_i}}{\rho}\right) \quad (k \leq n-1).$$

This can be seen by just computing individual probabilities of independent events. To obtain $P(X=n)$, it suffices to sum over all possible values of k and n_i. We note now that the following multinomial expansion is valid:

$$\prod_{i=1}^{n-1} \left(\rho\left(1-\frac{h_i}{\rho}\right)+1-\rho\right)$$
$$= \sum_{k=0}^{n-1} \rho^k (1-\rho)^{n-1-k} \left(\sum_{1 \leq n_1 < n_2 < \cdots \leq n_k \leq n-1} \prod_{i=1}^{k} \left(1-\frac{h_{n_i}}{\rho}\right)\right).$$

Thus,

$$P(X=n) = \rho \frac{h_n}{\rho} \prod_{i=1}^{n-1} \left(\rho\left(1-\frac{h_i}{\rho}\right)+1-\rho\right)$$
$$= h_n \prod_{i=1}^{n-1} (1-h_i) \quad (n=1,2,\ldots),$$

which was to be shown. ■

If we use the algorithm with $\rho=1$ (which is always allowed), then the sequential test algorithm is obtained. For some distributions, we are forced into this situation. For example, when X has compact support, with $p_n > 0, p_{n+i} = 0$ for some n and all $i \geq 1$, then $h_n = 1$. In any case, we have

Theorem 3.2.

For the discrete thinning algorithm, the expected number of iterations $E(N)$ can be computed as follows:

$$E(N) = \rho E(X) .$$

Proof of Theorem 3.2.

We observe that in the notation of the proof of the previous theorem,

$$X = \sum_{i=1}^{N} G_i ,$$

so that by Wald's equation,

$$E(X) = E(N)E(G_1) = E(N)\rho ,$$

which was to be shown. ■

Example 3.1. The logarithmic series distribution.

For the logarithmic series distribution defined by

$$P(X=n) = \frac{1}{-\log(1-\theta)} \frac{\theta^n}{n} \qquad (n \geq 1) ,$$

where $\theta \in (0,1)$ is a parameter, we observe that h_n is not easy to compute (thus, some preprocessing seems necessary for this distribution). However, several key properties of h_n can be obtained with little difficulty:

(i) $\frac{1}{h_n} = 1 + \frac{n\theta}{n+1} + \frac{n\theta^2}{n+2} + \cdots ;$

(ii) $h_n \downarrow 1-\theta$ as $n \to \infty$;

(iii) $\rho = \sup_n h_n = h_1 = \frac{\theta}{-\log(1-\theta)} .$

Thus, while the sequential method has $E(N) = E(X) = \frac{\rho}{1-\theta}$, the discrete thinning method satisfies

$$E(N) = \rho E(X) = \frac{\rho^2}{1-\theta} .$$

Since $\rho \to 0$ as $\theta \to 1$, we see that the improvement in performance can be dramatic. Unfortunately, even with the thinning method, we do not obtain an algorithm that is uniformly fast over all values of θ:

$$\sup_{\theta\in(0,1)} E(N) = \infty \,. \blacksquare$$

3.4. Discrete dynamic thinning.

Shanthikumar (1983,1985) has also observed that for distributions with decreasing discrete hazard rate (also referred to below as DHR distributions) that the value of ρ can be dynamically modified to increase the jumps for the geometric random variates, and thus increase the performance. The formal algorithm is given below.

Dynamic thinning method for discrete DHR distributions

$X \leftarrow 0$
REPEAT
  Generate iid uniform [0,1] random variates U, V.
  $\rho \leftarrow h_{X+1}$
  $X \leftarrow X + \left\lceil \dfrac{\log U}{\log(1-\rho)} \right\rceil$
UNTIL $V \leq \dfrac{h_X}{\rho}$
RETURN X

The validity of this algorithm follows by a short recursive argument:

$$\begin{aligned} &P(X > n \mid X > n-1, \text{ the last partial sum of geometric variates} \\ &\qquad\qquad \text{less than } n \text{ takes the value } k) \\ &= (1-h_k) + h_k\left(1-\frac{h_n}{h_k}\right) \\ &= 1-h_n \,. \end{aligned}$$

Thus, because this does not depend upon k,

$$\begin{aligned} &P(X > n) = (1-h_n)P(X > n-1) \\ &= \prod_{i=1}^{n} (1-h_i) \,. \end{aligned}$$

3.5. Exercises.

1. Prove the following for the logarithmic series distribution with parameter $\theta \in (0,1)$:

   (i) $\frac{1}{h_n} = 1 + \frac{n\theta}{n+1} + \frac{n\theta^2}{n+2} + \cdots$,

   (ii) $h_n \downarrow 1-\theta$ as $n \to \infty$,

   (iii) $E(X) = \frac{\theta}{-\log(1-\theta)} \frac{1}{1-\theta}$.

2. Assume that discrete dynamic thinning is used for a DHR distribution. Obtain good upper bounds for $E(N)$ in terms of the size of the tail of the distribution. Show also that for the logarithmic series distribution the value of $E(N)$ is not uniformly bounded in $\theta \in (0,1)$, the parameter of the distribution.

3. Show that in the discrete thinning algorithm, quick acceptance and rejection steps can be introduced that would effectively reduce the expected number of evaluations of h_n. Compute the expected number of such evaluations for two squeezing sequences.

4. A continuation of exercise 3. For the logarithmic series distribution with parameter θ, show that

$$1-\theta \le h_n \le \frac{n(1-\theta)+1}{n+1} \quad (n \ge 1).$$

   Show that if these bounds are used for squeeze steps in the discrete dynamic thinning method, then the expected number of evaluations of h_n is $o(1)$ as $\theta \uparrow 1$. (The inequalities are due to Shanthikumar (1983,1985).)

5. **The negative binomial distribution.** A random variable Y has the negative binomial distribution with parameters (k,p) where $k \ge 1, p \in (0,1)$ if

$$P(Y=n) = \binom{n-1}{k-1} p^k (1-p)^{n-k} \quad (n \ge k).$$

   Then, the normalized random variable $X = Y-k+1$ has a distribution on all positive integers. For this random variable X, show that $h_n \uparrow p$ as $n \uparrow \infty$. (Hint: the relationship

$$h_{n+1} = \left(\frac{n+k-1}{n}\right)\left(\frac{h_n}{1-h_n}\right)(1-p) \quad (n \ge 1)$$

   is helpful.) Show that in the sequential test algorithm, $E(N) = \frac{k}{p}$, while in the discrete thinning algorithm (with $\rho = p$), we have $E(N) = k$. Compare this algorithm with the algorithm based upon the observation that Y is distributed as the sum of k iid geometric (p) random variates. Finally, show the squeeze type inequalities

$$1 - \frac{n+k-1}{n}(1-p) \le h_n \le p \quad (n \ge 1).$$

6. Example 3.1 for the logarithmic series distribution and the previous exercise for the negative binomial distribution require the computation of h_n. This can be done by setting up a table up to some large value. If the parameters of the distributions change very often, this is not feasible. Show that we can compute the sequence of values recursively during the generation process by

$$h_1 = p_1 ;$$

$$h_{n+1} = \frac{p_{n+1}}{p_n} \frac{h_n}{1-h_n} .$$