THE BRANCHING PROCESS METHOD IN LAGRANGE RANDOM VARIATE GENERATION

Luc Devroye

School of Computer Science
McGill University
Montreal, Canada H3A 2A7

KEYWORDS AND PHRASES. Lagrange distributions. Random variate generation. Branching processes. Probability inequalities. Rejection method. Borel-Tanner distribution. Haight's distribution. Expected time analysis.

ABSTRACT. The generalized Lagrange probability distributions include the Borel-Tanner distribution, Haight's distribution, the Poisson-Poisson distribution and Consul's distribution, to name a few. We introduce two universally applicable random variate generators for this family of distributions. In the branching process method, we produce the generation sizes in a Galton-Watson branching process. In the uniform bounding method, we employ the rejection method based upon a simple probability inequality that is valid for all members in a given subfamily.

1. Lagrange probability distributions.

The purpose of this note is to point out a few properties of the *generalized Lagrange probability distributions* (GLPD) that are useful for the purpose of generating random variates with these discrete distributions. First introduced by Consul and Shenton (1972), the probabilities defining the discrete distributions typically have complicated analytic formats that do not lend themselves to fast methods of random variate generation. Also, the vastness of the family makes any investment in efficient universal methods for the family a worthy one.

At the basis of the GLPD family is Lagrange's formula given below in a form that can be found e.g. in Dieudonné (1971).

PROPERTY 1. *Let f be analytic in a disk $D(0,r)$, $r = 0$, and let $f(0) \neq 0$. Then there exists an $R = 0$ and an analytic function g on $D(0,R)$ such that*

$$g(s) = sf(g(s)) \ , \text{ all } s \in (0,R) \ .$$

Furthermore, if y is analytic on $D(0,r)$, then for all $s \in D(0,R)$,

$$y(g(s)) = y(0) + \sum_{n=1}^{\infty} \frac{s^n}{n!} \left[(\frac{d}{dz})^{n-1} (y'(z))(f(z))^n \right]_{z=0} \ .$$

In particular, if f and y are generating functions for probability distributions, and $f(0) \neq 0$ and $f'(1) \leq 1$, then $y(g(s))$ is the generating function of a random variable N defined by

$$\mathbf{P}\{N = n\} = \frac{1}{n!} \left[(\frac{d}{dz})^{n-1} (y'(z))(f(z))^n \right]_{z=0} \ , \ n \geq 0 \ .$$

Lagrange's formula provides us with exact expressions for $\mathbf{P}\{N = n\}$. Consul and Shenton (1973, section 3; 1975, Theorem 2.1) have pointed out the relationship between GLPD's and the number of customers served in a busy period of a single-server queueing process started up with Y customers at time 0. This connection, when translated into an algorithm, leads to complexities of the order of $\mathbf{E}N$ since each customer is handled individually. A relationship between the above queueing process and branching processes was pointed out in Exercise 14 on p. 64 of Athreya and Ney (1972). We will introduce below the *branching process method*, in which N is generated based upon the partial recreation of a certain branching process. The advantage of this method is that no heavy numerical computations are required at all.

In the second part of the paper, we will show that certain GLPD subfamilies have the property that if θ is the collection of parameters,

$$\sup_{\theta} p_{\theta,i} \leq q_i$$

for all i, where $p_{\theta,i} \stackrel{\text{def}}{=} \mathbf{P}\{N = i\}$, and $\sum_i q_i < \infty$. Families with this property can be dealt with by employing the rejection method based upon rejection from a distribution with probabilities proportional to the q_i's. (Note: if $\{p_i\}$ is a probability distribution saisfying $p_i \leq q_i$ for all i, then a random variate with distribution p_i can be obtained by the rejection method by generating pairs of independent random variates (X, U), where X has distribution $\{cq_i\}$ for some constant c and U is uniform $[0,1]$, until $Uq_X < p_X$, and returning X.) More importantly, the design does not have to be altered in any way when θ changes, so good speed should be expected when θ changes frequently between calls. We will refer to this as the *uniform bounding method*. The method is extremely simple to implement and comprehend.

There are instances in which none of the two methods discussed above will satisfy the true perfectionist; indeed, for some GLPD families, the expected running time cannot be uniformly bounded over all members in the family. In those cases, one has to look at the limiting members in the family in order to design good rejection methods. Consul and Shenton (1973) have shown that for many families, one can obtain the normal law in the limit, as well as the entire family of inverse gaussian distributions. Based on this observation, Devroye (1989) worked out the details of a uniformly fast algorithm for the Poisson-Poisson distribution of Consul and Jain (1973). For other families, similar ad hoc algorithms can be derived as well.

A list of some GLPD distributions can be found in Consul and Shenton (1972). It is convenient to say that a distribution is an (f, y) GLPD when f and y are the generating functions of the probability distributions employed in the definition of the family. An important subclass is that of the (f, s^k) GLPD, or so-called delta GLPD's, in which $y(s) \equiv s^k$ puts mass one at k. The following property shows why the delta GLPD's play a key role for us — its proof is a simple corollary of the results on the branching process method.

Property 2. *If Y is a random variable with generating function $y(s)$, and given Y, X is a random variable with the (f, s^Y) GLPD, then X is (f, y) GLPD.*

In other words, we never need to look at the general (f, y) case as the delta GLPD's are the basic building blocks with which we shall work. Here are a few important delta GLPD's:

A. The binomial-delta GLPD. Consul's distribution. Here we take f binomial (m, p) (with $mp \le 1$) and $y(s) \equiv s^k$. This yields

$$\mathbf{P}\{N = i\} = \frac{k}{i}\binom{mi}{i-k} p^{i-k}(1-p)^{mi+k-i}\ ,\ i \ge k.$$

If X is binomial-delta with $m = 1$ (i.e., it is Bernoulli-delta), then X is distributed as $k + Y$, where Y is negative binomial (k, p). Thus, the geometric and negative binomial families are shifted Bernoulli-delta GLPD's!

B. Haight's distribution. The geometric-delta GLPD. Haight (1961) introduced a distribution which corresponds to a GLPD with $y \equiv s$ (hence $Y = 1$) and $f(s) = (1-p)/(1-ps)$, $p \in (0, 1/2)$ (the geometric distribution). This leads to the probabilities

$$\mathbf{P}\{N=i\} = \frac{(2i-2)!}{i!(i-1)!} p^{i-1}(1-p)^i \ , \ i \geq 1.$$

Note that in this simple case, the generating function g for N can be obtained without great difficulty as the solution of $g(s) = sf(g(s))$:

$$g(s) = \frac{s - sp}{1 - pg(s)} \ ,$$

or of $pg^2(s) - g(s) + s - ps = 0$. This yields

$$g(s) = \frac{1 - \sqrt{1 - 4p(1-p)s}}{2p} \ .$$

This gives us another interpretation for the Haight distribution: $g(s^2)/s$ is the generating function of the first visit to $+1$ in a simple asymmetric random walk on the integers in which a positive move is made with probability $1-p$.

C. THE NEGATIVE-BINOMIAL DELTA DISTRIBUTION. Consul and Shenton (1972) discussed the generalization of Haight's distribution, called the negative-binomial delta GLPD, in which $y(s) = s^k$, $f(s) = ((1-p)/(1-ps))^m$, $k \geq 1$, $m \geq 1$ and $mp/(1-p) \leq 1$. They obtain

$$\mathbf{P}\{N=i\} = \frac{k}{i} \frac{\Gamma(mi+i-k)}{(i-k)!\Gamma(mi)} p^{i-k}(1-p)^{mi} \ , \ i \geq k \ .$$

D. THE BOREL-TANNER DISTRIBUTION. THE POISSON-DELTA GLPD. The Borel-Tanner distribution (Tanner, 1951, 1953, 1961; Haight and Breuer, 1960) is the (f, s^k) GLPD, in which $f(s) = e^{\lambda(s-1)}$ is Poisson (λ) with $\lambda < 1$. It has probabilities

$$\mathbf{P}\{N=i\} = \frac{k}{i} \frac{e^{-\lambda i}(\lambda i)^{i-k}}{(i-k)!} \ , \ i \geq k \ ,$$

and was identified by Otter (1949) as the distribution of the size of a branching process with Poisson reproduction distribution, when the process is started up with one father ($k=1$). Tanner (1953, 1961; see also Takács, 1963) showed that N is the number of customers served in a busy period of a single server queue with k customers at the start, Poisson arrivals, and constant service times. For $k=1$, we obtain the Borel distribution:

$$\mathbf{P}\{N=i\} = \frac{(\lambda i)^{i-1} e^{-\lambda i}}{i!} \ , \ i \geq 1.$$

This is also the shifted Haight mixture distribution (Haight, 1961, entry 8.89), for which a uniformly fast generator is given in Devroye (1989).

Other examples are given in Consul and Shenton (1972). Among the important GLPD's that are not of the delta type, we cite the Poisson-Poisson distribution, the binomial-binomial GLPD, and the Poisson-geometric GLPD. All of these will be briefly commented upon in a further section.

Consider next the following property due to Consul and Shenton (1972) — also a simple corollary of the branching process interpretation of N (see further on).

PROPERTY 3. *If X and W are independent (f, s^k) and (f, s^l) GLPD random variables respectively, then $X + W$ is (f, s^{k+l}) GLPD.*

Thus, random variates for the (f, s^k) GLPD can be generated if we have generators available to us for the (f, s) GLPD's. These distributions are truly the atomic building blocks. Examples include the geometric-delta distribution and the Borel distribution. Unfortunately, summing k i.i.d. random variates with these distributions leads to methods whose running time grows linearly with k. As we will see below, the branching process and uniform bounding methods do much better, reducing the times roughly to $O(\log k)$ and $\sqrt{k}$ respectively. Nevertheless, we will give indications for good generators for these important atomic GLPD's.

2. The size of a branching process.

The *generating function* $f(s)$ for a nonnegative discrete random variable X is defined by

$$f(s) = \sum_{i=0}^{\infty} p_i s^i = \mathbf{E} s^X ,$$

where $p_i = \mathbf{P}\{X = i\}$. In a Galton-Watson branching process started with one element, every element produces children in an independent fashion according to the distribution defined by f. Such a process is known to be be finite with probability one (i.e. the population becomes extinct) if f has mean $m = \mathbf{E}X \leq 1$ (in the case $m = 1$, we assume that $p_1 \neq 1$). We define the size N of such a process to be the number of elements that ever live in such a finite population. The generating function of N is $g(s)$.

PROPERTY 4. *For $s \in [0, 1]$, the generating function $g(s)$ is equal to the unique solution u of the equation*

$$u = sf(u).$$

Thus, the total population size N is (f, s) GLPD.

PROOF. We trivially have $u \leq f(u)$. Also, $uf'(u) < f(u)$ by a simple geometric argument. Thus, $f(u)/u \downarrow 1$, since $(f'(u)/u) - (f(u)/u^2) = (uf'(u) - f(u))/u^2 < 0$. Hence, $u/f(u) \uparrow 1$ and the given equation has indeed a unique solution. If the father of the entire population has X kids, then $N = 1 + \sum_{i \leq X} N_i$, where the N_i's are i.i.d. and distributed as N. Thus,

$$g(s) = \mathbf{E}s^N = \mathbf{E}s^{1+\sum_{i \leq X} N_i} = s\mathbf{E}g(s)^X = sf(g(s)).\square$$

PROPERTY 5. *If we begin the branching process with Y fathers, where Y is a random variable having generating function $y(s)$, then the generating function for N is $\mathbf{E}g(s)^Y = y(g(s))$, where $g(s)$ is as in Property 4. The distribution of N is (f, y) GLPD.*

3. The branching process method.

The *branching process algorithm* for generating N merely mimicks the definition:

```
Branching process method.

generate Y with generating function y(s)
N ← Y
while Y > 0 do
      generate Z with generating function f^Y(s)
      (N, Y) ← (N + Z, Z)
return N
```

Let us look at the expected time per random variate. We note first that the sequence of Y's generated in the algorithm corresponds to the number of elements in subsequent generations of the branching process. The number of iterations through the `while` loop corresponds to the number of generations before the population becomes extinct. We must distinguish between several situations. If a random variate with generating function f^n can be generated in expected time $O(1)$, then the total expected time is bounded by

$$c_1\mathbf{E}T_y + c_2\mathbf{E}T_G,$$

where T_y is the time needed to generate a random variate with generating function $y(s)$, and T_G is the number of generations in the branching process. The c_i's are constants. If a random variate for f^n is generated as a sum of n i.i.d. random variates with generating function f, then the expected time is of the form $c_1\mathbf{E}T_y + c_2\mathbf{E}N$. This follows from the fact that every element ever alive is once and only once considered as a potential father.

But $\mathbf{E}N$ is simply the sum of the expected values of all elements that live in all generations, and thus

$$\mathbf{E}N = \mathbf{E}Y \sum_{i=0}^{\infty} m^i = \frac{\mathbf{E}Y}{1-m} .$$

If $m = 1$, $\mathbf{E}N = \infty$. In the inversion method (Devroye, 1986, p. 86), the expected complexity as measured by the expected number of comparisons is $\mathbf{E}(1+N)$. This is comparable with this version of the branching process algorithm. The difference is that in the inversion method, we need to have access to the probabilities $\mathbf{P}\{N=i\}$, whereas no steps in the branching process algorithm involve heavy numerical computations. The generator for f should of course be fast; since $m \le 1$, it too could efficiently be handled by the inversion method; depending upon the situation, a table method may prove to be convenient here.

In the important GLPD families, f is Poisson, binomial or negative binomial, so that f^n is again Poisson, binomial or negative binomial. If we use a uniformly fast generator for these families, then the complexity is largely reduced, and we only have to look at $\mathbf{E}T_G$, the expected number of generations. To analyze this, we will first look at the finiteness of $\mathbf{E}T_G$, and then obtain upper bounds for this quantity.

THE FINITENESS OF $\mathbf{E}T_G$. We note that if $y(s) \equiv s$ (i.e., we have one initial element), and if $N_0 = 1, N_1, N_2, \ldots$ denote the sizes of the generations, then N_i has generating function $f_n(s)$, where $f_n(s) = f(f(f(\ldots f(s))$ is the n times iterated version of f. See e.g. Athreya and Ney (1972, pp. 1-6). Thus,

$$\mathbf{E}T_G = \sum_{n=0}^{\infty} \mathbf{P}\{T_G = n\} = \sum_{n=0}^{\infty} \mathbf{P}\{N_n = 0\} = \sum_{n=0}^{\infty} (1 - f_n(0)).$$

By definition, $f_0(s) = s$. If we begin with k fathers, then $\mathbf{P}\{N_n = 0\} = (f_n(0))^k$, so that, in general, starting with a random number Y of fathers, we have

$$\mathbf{E}T_G = \sum_{n=0}^{\infty} \mathbf{E}(1 - (f_n(0))^Y) = \sum_{n=0}^{\infty} (1 - y(f_n(0))).$$

As $n \to \infty$, we have $f_n(0) \to 1$, hence $1 - y(f_n(0)) \sim y'(1)(1 - f_n(0))$ if $y'(1) = \mathbf{E}Y < \infty$. In such cases, the finiteness of $\mathbf{E}T_G$ is solely determined by the finiteness of $\mathbf{E}T_G$ in the single father case $y(s) \equiv s$. Assume thus $y(s) \equiv s$ (and $Y \equiv 1$).

- If $m < 1$, we have $1 - f_n(0) = P(N_n = 0) \leq \mathbf{E}N_n = m^n$, so that $\mathbf{E}T_G \leq 1/(1-m) < \infty$.
- If $m = 1$, then the situation depends upon the behavior of f near one. If $\sigma^2 < \infty$, where σ^2 is the variance for f, then $1 - f_n(0) \sim 2/(n\sigma^2)$ (Kolmogorov, 1938; Kesten, Ney and Spitzer, 1966; see Athreya and Ney, 1972, pp. 19-22), so that $\mathbf{E}T_G = \infty$.
- Finally, if $m = 1$ but $\sigma^2 = \infty$, both $\mathbf{E}T_G < \infty$ and $\mathbf{E}T_G = \infty$ can occur. □

PROPERTY 6. *If $y(s) = s$, then*

$$\mathbf{E}T_G \leq 1 + \frac{1 - f(0)}{1 - m} \leq \frac{1}{1 - m} \ .$$

If $y(s) = s^k$ for $k \geq 1$ (i.e. $Y \equiv k$), then

$$\mathbf{E}T_G \leq 1 + \frac{1 + \log k}{1 - m} \ .$$

Finally, for general Y,

$$\mathbf{E}T_G \leq 1 + \frac{1 + \mathbf{E}\log_+ Y}{1 - m} \leq 1 + \frac{1 + \log_+ \mathbf{E}Y}{1 - m} \ ,$$

where by definition $\log_+ x = \log x$ when $x \geq 1$ and $\log_+ x = x - 1$ when $0 < x < 1$.

PROOF. Note the relationship $f(s) \geq 1 - m(1 - s)$, i.e. $1 - f(s) \leq m(1 - s)$. Thus,

$$1 - f_n + 1(s) \leq m(1 - f_n(s)).$$

By induction, $1 - f_n(0) \leq m^{n-1}(1 - f(0))$. Hence, if $y(s) = s$,

$$\mathbf{E}T_G \leq 1 + \sum_{n=1}^{\infty} m^{n-1}(1 - f(0)) = 1 + \frac{1 - f(0)}{1 - m} \ .$$

For the second part, we use $1 - f_n(0) \leq m^n$, so that

$$\begin{aligned}
\mathbf{E}T_G &\leq \sum_{n=0}^{\infty}(1 - (1 - m^n)^k) \\
&\leq \inf_{l \text{ integer}} \left(l + \sum_{n=l}^{\infty} k m^n \right) \\
&= \inf_{l \text{ integer}} \left(l + \frac{k m^l}{1 - m} \right) \\
&\leq 1 + \frac{\log k}{\log(1/m)} + \frac{1}{1 - m} \quad (\text{take } l = \lceil \log k / \log(1/m) \rceil) \\
&\leq 1 + \frac{1 + \log k}{1 - m} \ .
\end{aligned}$$

The third part of Property 6 follows if we replace k by Y and take expected values. Note, in particular, that we used $\mathbf{E}\log_+ Y \leq \log_+ \mathbf{E}Y$ since $\log_+$ is concave. □

The cases of interest are when one or both of $\mathbf{E}Y$, $1/(1-m)$ are large. Note that if f^Y is handled by generating sums $\sum_{i=1}^{Y} N_i$, where the N_i's are i.i.d. with generating function f, then the expected time complexity has as main term $\mathbf{E}Y/(1-m)$. If however, we replace it by a method with uniformly bounded generation times for f^Y, then the main term in the expected time complexity is reduced to something smaller than a constant times $\log_+(\mathbf{E}Y)/(1-m)$. This results in spectacular savings. The savings are even more outspoken in the case $m=1$. Since we have to compare infinities, it is perhaps better to look at the expected work done up to n generations. Assume that $Y=1$ and that f has a finite variance σ^2. In the "sum" method, we have work that grows as the expected number of elements in those n generations, which in turn grows linearly with n. In the "shortcut" method, we have work growing as $\sum_{i=0}^{n}(1-f_i(0)) \sim 2\sigma^{-2}\log n$.

Uniformly bounded times for standard distributions. For the Poisson, binomial and negative binomial distributions, we refer to the general treatment in Devroye (1986). Recently developed fast methods include Stadlober (1989), Kachitvichyanukul and Schmeiser (1988, 1989), Ahrens and Dieter (1987), Stadlober (1988) and Pokhodzei (1985). $\square$

The Poisson-delta GLPD. The expected running time is $O(\log k/(1-\lambda))$. $\square$

The binomial-binomial GLPD. This distribution is obtained by taking $f(s)=(1-p+ps)^m$ binomial (m,p) and $y(s)=(1-p+ps)^k$ binomial (k,p). Here $mp \le 1$ (Consul and Jain, 1973; Consul and Shenton, 1972). We have

$$\mathbf{P}\{N=i\} = \frac{k}{k+mi}\binom{k+mi}{i}p^i(1-p)^{k+mi-i}\ ,\ i \ge 0.$$

The expected time is bounded by $c_1 + c_2\log(kp)/(1-mp)$. $\square$

The Poisson-Poisson distribution. If $f(s)=e^{\theta(s-1)}$ $(\theta \le 1)$ and $g(s)=e^{\lambda(s-1)}$, we obtain the so-called Poisson-Poisson GLPD of Consul and Jain (1973), given by

$$\mathbf{P}\{N=i\} = \lambda(\lambda+\theta i)^{i-1}e^{-(\lambda+\theta i)}/i!\ ,\ i \ge 0\ .$$

The branching process algorithm leads to a very simple method of expected complexity roughly bounded by $c_1 + c_2\log(\lambda)/(1-\theta)$ when $\theta < 1$, where c_1 and c_2 are positive constants. Devroye (1989) has developed uniformly fast generators for this family for all $\lambda = 0$ and $\theta \in (0,1]$, but the code becomes more involved, and the overhead per variate is more substantial. Occasional users are probably better off with the branching process algorithm. For $\theta = 1$, we obtain the Abel distribution, and for $\theta = 0$, the Poisson distribution. Although applicable, the branching process algorithm is not recommended for the Abel distribution. The branching process method and the algorithms of Devroye (1989) are not applicable when $\theta > 1$. For an in–depth study, see Consul (1989), and for yet another view of generalized Poisson distributions, see Consul and Shoukri (1988). □

THE BINOMIAL-DELTA GLPD. The expected time is bounded by $c_1 + c_2 \log k/(1-mp)$. □

THE NEGATIVE BINOMIAL-DELTA GLPD. The expected complexity of the branching process method grows as $\log k/(1 - mp(1 - p))$. □

THE POISSON-GEOMETRIC GLPD. In 1973, Consul and Shenton introduced the GLPD with f Poisson (θ), $0 < \theta < 1$, and $y(s) = y_1(s)y_2(s)$, where $y_1(s) = e^{\lambda(s-1)}$ is Poisson (λ), and $y_2(s) = (1 - \theta)/(1 - \theta s)$ is geometric with parameter θ. The corresponding probabilities are

$$\mathbf{P}\{N = i\} = (1 - \theta)(\lambda + \theta i)^i e^{-(\lambda+\theta i)}/i! \ , \ i \geq 0 \ .$$

This is the distribution of the busy period of certain single server queues. The expected time complexity of the branching process method grows as $O((1-\theta)^{-1} \log_+(\lambda+\theta/(1-\theta)))$. This is not an atomic GLPD family, as random variables can be obtained as sums of random variables that are (Poisson (θ), Poisson (λ)) GLPD and (Poisson (θ), geometric (θ)) GLPD respectively. The probabilities in the latter case are given by

$$\mathbf{P}\{N = i\} = (1 - \theta)(\theta i)^i e^{-\theta i}/i! \ , \ i \geq 0 \ . \ \square$$

4. The uniform bounding method.

In the uniform bounding method, we find the supremum with respect to one or more parameters of $\mathbf{P}\{N = i\}$, and apply the rejection method based on these suprema. This principle cannot be applied to just any family. For example, for a Poisson (λ) distribution,

$$\sup_\lambda \frac{e^{-\lambda}\lambda^i}{i!} = \left(\frac{i}{e}\right)^i \frac{1}{i!} \sim \frac{1}{\sqrt{2\pi i}}$$

as $i \to \infty$. The upper bound in this case is not summable, and is therefore useless in the rejection method. Interestingly, for many GLPD subfamilies, the uniform bounding method *does* work. We consider in each case an (f, s^k) GLPD with fixed k — the parameters are thus all borrowed from f. We will call the parameter or parameters θ, and we will write $p_{\theta,i}$ for the probabilities in question. In the examples considered below, we will obtain inequalities of the form given below.

PROPERTY 7. *Consider the binomial (m,p)-delta (with $m \geq 2, mp < 1$), Poisson (λ)-delta ($\lambda > 0$), and negative binomial (m,p)-delta ($mp < 1$, $m > 0$ real-valued) GLPD's, with k fixed. Then*

$$\sup_\theta p_{\theta,i} \leq \begin{cases} \frac{ck}{i\sqrt{(i-k)}} \ , & \text{if } i > k; \\ 1 \ , & \text{if } i = k, \end{cases}$$

where c is a constant depending upon k only. We can take $c = 1/\sqrt{2\pi}$ for the Poisson-delta GLPD, $c = e^{1/24k}/\sqrt{\pi}$ for the binomial-delta GLPD, and $c = e^{1/12}/\sqrt{2\pi}$ for the negative binomial-delta GLPD.

PROOF. Take the Borel-Tanner distribution (Poisson-delta GLPD) with parameters $k \geq 1$ and $\lambda \in (0,1)$:

$$p_{\lambda,i} \stackrel{\text{def}}{=} \frac{k}{i} \frac{e^{-\lambda i}(\lambda i)^{i-k}}{(i-k)!} \ , \ i \geq k \ ,$$

Note in particular that for $a, b > 0$, $\sup_{u>0} u^a e^{-bu} = (a/be)^a$. Using this, we see that

$$\sup_{0<\lambda\leq 1} p_{\lambda,i} \leq \frac{k}{i} \frac{((i-k)/e)^{i-k}}{(i-k)!} \ , \ i \geq k \ ,$$

Apply Stirling's approximation ($i! \geq (i/e)^i \sqrt{2\pi i}$) to the last expression.

Next, consider the binomial-delta GLPD in which the binomial has parameters $m \geq 2$ and $p \in (0, 1/m)$:

$$\mathbf{P}\{N = i\} = \frac{k}{i} \binom{mi}{i-k} p^{i-k}(1-p)^{mi+k-i} \ , \ i \geq k \ .$$

This is maximal with respect to p for $p = (i-k)/mi$. Resubstitution shows that for $i > k$,

$$\begin{aligned} \sup_p \mathbf{P}\{N = i\} &\leq \frac{k}{i} \binom{mi}{i-k} \left(\frac{i-k}{mi}\right)^{i-k} \left(\frac{mi-i+k}{mi}\right)^{mi+k-i} \\ &\leq \frac{k}{i} e^{1/12mi} \sqrt{\frac{mi}{2\pi(i-k)(mi-i+k)}} \ , \end{aligned}$$

where we used Stirling's approximation. But $mi/(mi - i + k) \leq 2i/(i + k)$ when $m \geq 2$, so that we have

$$\sup_{p,m\geq 2} \mathbf{P}\{N=i\} \leq \frac{k}{i} e^{1/24i} \sqrt{\frac{2i}{2\pi(i-k)(i+k)}}$$
$$\leq \frac{ke^{1/24k}}{\sqrt{\pi}\, i\sqrt{i-k}} \ .$$

Consider finally the negative binomial (m,p) distribution as a generator for the negative binomial-delta GLPD:

$$\mathbf{P}\{N=i\} = \frac{k}{i} \frac{\Gamma(mi+i-k)}{(i-k)!\Gamma(mi)} p^{i-k}(1-p)^{mi} \ , \ i \geq k \ .$$

Considered as function of p, this probability is maximal for $p = (i-k)/(mi+i-k)$. Resubstitution of this value yields, for $i > k$,

$$\sup_p \mathbf{P}\{N=i\} \leq \frac{mik}{i(mi+i-k)} \frac{\Gamma(mi+i-k+1)}{(i-k)!\Gamma(mi+1)} \frac{(i-k)^{i-k}(mi)^{mi}}{(mi+i-k)^{mi+i-k}}$$
$$\leq \frac{k}{i} \sqrt{\frac{mi}{2\pi(mi+i-k)(i-k)}}\, e^{1/(12(mi+i-k))}$$
$$\leq \frac{k}{i\sqrt{2\pi(i-k)}}\, e^{1/12} \ . \ \square$$

In view of property 7, we need only be concerned with one rejection algorithm. We take the standard method of defining a continuous density $p(x)$ that is related to the probabilities $p_i = \mathbf{P}\{N=i\}$ in the following manner: $f(x) = p_i$ for $i \leq x < i+1$. With this understanding, we have $p(x) \leq h(x)$, where the dominating function $h(x)$ is given by

$$h(x) = \begin{cases} 1 \ , & k \leq x < k+1; \\ c/\sqrt{x-1-k} \ , & k+1 \leq x < 2k+1; \\ kc/(x-1-k)^{3/2} \ , & x \geq 2k+1. \end{cases}$$

If $h(x)$ is used in the rejection method, then the rejection constant, or the expected number of iterations, is $\int_k^\infty h(x)\,dx = 1 + 4c\sqrt{k}$. Also, h is well-suited for the inversion method because

$$\int_{-\infty}^x h(y)\,dy = \begin{cases} 0 \ , & x < k; \\ x-k \ , & k \leq x < k+1; \\ 1+2c\sqrt{x-1-k} \ , & k+1 \leq x < 2k+1; \\ 1+4c\sqrt{k} - 2kc/\sqrt{x-1-k} \ , & 2k+1 \leq x \ . \end{cases}$$

```
The uniform bounding method based on Property 7.

define c
  Poisson-delta:  c = 1/√(2π)
  binomial-delta GLPD: c = e^{1/24k}/√π
  negative binomial-delta GLPD: c = e^{1/12}/√(2π)
repeat
  generate U uniformly on [0,1]
  V ← (1 + 4c√k)U
  case
      V ≤ 1:  return X ← k
      1 < V ≤ 1 + 2c√k:  Y ← k + 1 + (V − 1)^2/4c^2
                          T ← c/√(Y − 1 − k)  (= 2c^2/(V − 1))
      V > 1 + 2c√k:  Y ← k + 1 + (2kc/(1+4c√k−V))^2
                      T ← kc/(Y − 1 − k)^{3/2}
  generate W uniform [0,1]
until WT < p_⌊Y⌋
return X ← ⌊Y⌋
```

We observe that the parameters enter in the algorithm only via $p_{\lfloor Y \rfloor}$. The expected number of iterations before halting is $1+4c\sqrt{k}$. In all cases considered here, this is $O(\sqrt{k})$.

5. Acknowledgment.

The author's research was sponsored by NSERC Grant A3456 and FCAR Grant EQ-1678. Part of this research was carried out while visiting the Division of Statistics, University of California at Davis.

6. Bibliography.

J. H. Ahrens and U. Dieter, "A convenient sampling method with bounded computation times for Poisson distributions," in: *First International Conference on Statistical Computation, Izmir, Turkey*, pp. 4–17, 1987.

K. B. Athreya and P. E. Ney, *Branching Processes*, Springer Verlag, Berlin, 1972.

P. C. Consul, *Generalized Poisson Distributions*, Marcel Dekker, New York, 1989.

P. C. Consul and G. C. Jain, "A generalization of the Poisson distribution," *Technometrics*, vol. 15, pp. 791–799, 1973.

P. C. Consul and L. R. Shenton, "Use of Lagrange expansion for generating discrete generalized probability distributions," *SIAM Journal of Applied Mathematics*, vol. 23, pp. 239–248, 1972.

P. C. Consul and L. R. Shenton, "Some interesting properties of Lagrange distributions," *Communications in Statistics*, vol. 2, pp. 263–271, 1973.

P. C. Consul and L. R. Shenton, "On the probabilistic structure and properties of discrete Lagrangian distribution," in: *Statistical Distributions in Scientific Work, Vol. 1*, ed. G. P. Patil, S. Kotz and J. K. Ord, pp. 41–57, D. Reidel Publishing Company, Dordrecht, The Netherlands, 1975.

P. C. Consul and M. M. Shoukri, "Some chance mechanisms related to a generalized Poisson probability model," *American Journal of Mathematical and Management Sciences*, vol. 8, pp. 181–202, 1988.

L. Devroye, *Non-Uniform Random Variate Generation*, Springer-Verlag, New York, 1986.

L. Devroye, "Random variate generators for the Poisson-Poisson and related distributions," *Computational Statistics and Data Analysis*, vol. 8, pp. 247–278, 1989.

J. Dieudonné, *Infinitesimal Calculus*, Houghton Mills, Boston, 1971.

M. Dwass, "Poisson recurrence times," *Journal of Applied Probability*, vol. 4, pp. 605–608, 1967.

F. A. Haight, "A distribution analogous to the Borel-Tanner," *Biometrika*, vol. 48, pp. 167–173, 1961.

F. A. Haight and M. A. Breuer, "The Borel-Tanner distribution," *Biometrika*, vol. 47, pp. 143–150, 1960.

G. C. Jain, "On power series distributions associated with Lagrange expansion," *Biometrische Zeitschrift*, vol. 17, pp. 85–97, 1974.

V. Kachitvichyanukul and B. W. Schmeiser, "Binomial random variate generation," *Communications of the ACM*, vol. 31, pp. 216–222, 1988.

V. Kachitvichyanukul and B. W. Schmeiser, “Algorithm 678: BTPEC: sampling from the binomial distribution,” *ACM Transactions on Mathematical Software*, vol. 15, pp. 394–397, 1989.

H. Kesten, P. Ney, and F. Spitzer, *The Galton-Watson process with mean one and finite variance*, vol. 11, pp. 513–540, Theory of Probability and its Applications, 1966.

A. N. Kolmogorov, *Zur Lösung einer biologischen Aufgabe*, vol. 2, pp. 1–6, Izvestiya nauchno-issledovatelskogo instituta matematiki i mechaniki pri Tomskom Gosudarstvennom Universitete, 1938.

R. Otter, “The multiplicative process,” *Annals of Mathematical Statistics*, vol. 50, pp. 206–224, 1949.

B. B. Pokhodzei, “Modelling of a negative binomial distribution using the sampling method,” *USSR Computational Mathematics and Mathematical Physics*, vol. 25(3), pp. 85–87, 1985.

E. Stadlober, “Sampling from Poisson, binomial and hypergeometric distributions: ratio of uniforms as a simple fast alternative,” Habilitationsschrift, Institute of Statistics, Technical University of Graz, Austria, 1988.

E. Stadlober, “Binomial random variate generation: a method based on ratio of uniforms,” *American Journal of Mathematical and Management Sciences*, vol. 9, pp. 1–20, 1989.

L. Takács, “The stochastic law of the busy period for a single server queue with Poisson input,” *Journal of Mathematical Analysis and Applications*, vol. 6, pp. 33–42, 1963.

J. C. Tanner, “The delay to pedestrians crossing a road,” *Biometrika*, vol. 38, pp. 383–392, 1951.

J. C. Tanner, “A problem of interference between two queues,” *Biometrika*, vol. 40, pp. 58–69, 1953.

J. C. Tanner, “A derivation of the Borel distribution,” *Biometrika*, vol. 48, pp. 222–224, 1961.