Some simple method of estimation for the parameters of the discrete stable distribution with the probability generating function

Louis G. Doray*[1], Shu Mei Jiang* and Andrew Luong**

*Département de mathématiques et de statistique, Université de Montréal, C.P. 6128, Succursale Centre-ville, Montréal, Québec, Canada H3C 3J7

**École d'actuariat, Université Laval, Cité Universitaire, Ste-Foy, Québec, Canada G1K 7P4

Key Words and Phrases: probability generating function, quadratic distance, linear regression, asymptotic bias, normality, truncation.

ABSTRACT

In this paper, we develop a method to estimate the two parameters of the discrete stable distribution. By minimizing the quadratic distance between transforms of the empirical and theoretical probability generating functions, we obtain estimators simple to calculate, asymptotically unbiased and normally distributed. We also derive the expression for their variance-covariance matrix. We simulate several samples of discrete stable distributed datasets with different parameters, to analyze the effect of truncation on the right tail of the distribution.

1 Introduction

The discrete two-parameter stable distribution introduced by Steutel and van Harn in 1979 can be obtained as a certain mixture of Poisson distributions. It allows skewness and a heavy right tail and has many interesting mathematical properties. However, the lack of a closed form expression for the probability mass function makes it difficult to estimate its parameters, and to compute probabilities or quantiles and has been a major drawback to its use by practitioners.

When no explicit expression for the probability function exists, an estimation method like maximum likelihood can not be used. We will develop an alternative estimation procedure based on minimizing numerically the quadratic

[1]corresponding author

distance between the empirical and theoretical probability generating functions. Using the classical linear regression model, we study the properties of the estimators, such as their consistency and asymptotic normality. Numerical examples are provided, using Devroye's (1993) simulation method to generate observations from the discrete stable distribution. We observe that the bias of the estimators is little affected by truncation of observations in the right tail.

The paper is organized as follows. In section 2, we review some properties of the discrete stable distribution and in section 3, some statistical results used later. We define our model, give the algorithm and obtain the aymptotic properties of the estimators in section 4. Using simulations, we present numerical examples to illustrate the estimation method in section 5 and give some recommendations for the algorithm. Finally, we draw some conclusions.

2 Properties of the distribution

Steutel and van Harn (1979) introduced the discrete stable distribution for integer valued random variables (the discrete stable family), and analyzed some of the properties of this distribution (see also Steutel and van Harn (2004)); we review them briefly here.

Let X be a discrete random variable taking values on some subset of the non-negative integers $\{0, 1, ...\}$; its probability generating function (pgf) is defined as

$$P_X(z) = E(z^X) = \sum_{i=0}^{\infty} p_i z^i, \text{ where } p_i = P[X = i].$$

For a discrete stable random variable X with parameters $\alpha \in (0, 1]$ and $\lambda > 0$, its pgf is given by

$$P_X(z) = \exp[-\lambda(1 - z)^\alpha], \quad |z| \leq 1.$$

With $\alpha = 1$, we obtain the pgf of a Poisson(λ) distribution,

$$P_X(z) = \exp[\lambda(z - 1)], \quad |z| \leq 1, \quad \lambda > 0.$$

a) Stability of a random variable

A random variable X is stable if for X_1 and X_2, two independent random variables with the same distribution as X, and any positive constants a and b, the equality

$$aX_1 + bX_2 \stackrel{D}{=} cX + d$$

holds for some positive constant c and $d \in R$, where the symbol $\stackrel{D}{=}$ means equality in distribution. The random variable is strictly stable if this equation holds with $d = 0$ for all choices of a and b.

Stability of a random variable means its shape is unchanged under summations of the above type. Examples of stable distributions include the normal, Cauchy and Lévy distributions.

b) Compound Poisson distribution

The discrete stable random variable can be represented as

$$X \stackrel{D}{=} M_1 + M_2 + \ldots + M_Y,$$

where $Y \sim$ Poisson(λ), Y is independent of M_i, and the M_i's are i.i.d. random variables following a Sibuya (see Sibuya (1979)) distribution with parameter α and pgf

$$P_{M_i}(z) = 1 - (1 - z)^\alpha.$$

See also Christoph and Schreiber (2000) for properties of the Sibuya(α) distribution. Hence the discrete stable distribution is a compound Poisson distribution.

Pakes (1998) and Bouzar (2002) presented some distributions derived from the discrete stable distribution, such as the discrete Linnik distribution.

c) Mixed Poisson

Devroye (1993) showed that a discrete stable random variable with parameters λ and α is a conditional Poisson random variable with parameter $\lambda^{1/\alpha} S_{\alpha,1}$, where $S_{\alpha,1}$ follows a continuous positive stable distribution with parameter α and Laplace transform equal to

$$E(e^{-sS_{\alpha,1}}) = e^{-s^\alpha}, \quad s > 0.$$

$S_{\alpha,1}$ can be generated by the method given by Kanter (1975),

$$S_{\alpha,1} \stackrel{D}{=} \left(\frac{\sin((1-\alpha)\pi U)}{E \times \sin(\alpha\pi U)} \right)^{(1-\alpha)/\alpha} \left(\frac{\sin(\alpha\pi U)}{\sin(\pi U)} \right)^{1/\alpha},$$

where $U \sim$ Uniform(0,1), $E \sim$ Exponential(1), and U and E are independent. Note that for $\alpha = 1$, $S_{\alpha,1}$ becomes the degenerate distribution with mass at $x = 1$ and Laplace transform equal to $e^{-s}, s > 0$.

This algorithm will be used in section 5 to generate observations from the discrete stable distribution.

d) Infinite divisibility

A discrete stable random variable X is infinitely divisible since its pgf $P_X(z)$ can be written as

$$P_X(z) = \exp[\lambda(G(z) - 1)],$$

where $\lambda > 0$ and $G(z) = 1 - (1-z)^\alpha$ is the pgf of a Sibuya(α) distribution with $G(0) = 0$ (see Steutel and van Harn (1979)).

e) Discrete self-decomposability

The discrete stable distribution is self-decomposable since its pgf satisfies

$$P_X(z) = P_X(1 - \alpha + \alpha z)P_\alpha(z), \quad |z| \le 1, \quad \alpha \in (0,1),$$

where $P_\alpha(z)$ is a probability generating function. Discete self-decomposable distributions are unimodal. See Steutel and van Harn (1979) for a necessary and sufficient condition to have discrete self-decomposability.

f) Probabilities

Expanding the pgf $P_X(z)$ of the discrete stable random variable X in a power series, Christoph and Schreiber (1998) obtain for p_k the series

$$p_k = (-1)^k e^{-\lambda} \sum_{m=0}^{k} \sum_{j=0}^{m} \binom{m}{j} \binom{\alpha j}{k} (-1)^j \frac{\lambda^m}{m!}.$$

However, this formula is difficult to use in practice, since it grows very fast in length as k increases.

In figure 1, we evaluated the expression for the probability function of the discrete stable distribution with various values of α and λ to see the effect of the parameters on the shape of the probability function.

g) Moments

For $\alpha \in (0,1)$, the moment $E(X^r)$ is finite only for $0 \le r < \alpha < 1$. For $r > \alpha, \alpha < 1$, the moments $E(X^r)$ do not exist (see Christoph and Schreiber (1998)). For $\alpha = 1$, the Poisson distribution, all moments exist.

3 Statistical Review

3.1 Moments of multinomial distribution

Johnson, Kotz and Balakrishnan (1997) give some properties of the multinomial distribution. Let $f_1, \ldots, f_k$ be the random variables denoting the numbers of occurrences of k mutually exclusive events in n independent trials, with corresponding probability of occurrence of an event in any trial equal to $p_1, \ldots, p_k$, with $p_1 + \ldots + p_k = 1$. The joint distribution of $f_1, \ldots, f_k$ is multinomial $(n, p_1, \ldots, p_k)$ and the marginal distribution of f_i is binomial (n, p_i); we have $E(f_i) = np_i, Var(f_i) = np_i(1-p_i)$ and $Cov(f_i, f_j) = -np_ip_j$.

3.2 Empirical probability generating function

Nakamura and Pérez-Abreu (1993) define the empirical probability generating function following the idea of Press (1972) who defines empirical characteristic functions.

Let $X_1, \ldots, X_n$ be a random sample from a discrete distribution over $0, 1, 2, ...$, with probabilities p_k, $k = 0, 1, 2, ...$ The empirical pgf, equal to

$$P_n(z) = \frac{1}{n}\sum_{i=1}^{n} z^{X_i}, \text{ is estimated by } \sum_{j=0}^{h} \frac{n_j}{n} z^j$$

for $|z| \leq 1$, where n_j is the number of observations in the sample equal to j (observed frequency), and h is the largest observation. $P_n(z)$ is an estimator of the theoretical pgf

$$P_X(z) = E(z^X) = \sum_{k=0}^{\infty} p_k z^k, \quad |z| \leq 1.$$

Rémillard and Theodorescu (1991) have proved that, as $n \to \infty$, $sup_{z\in(0,1]}|P_n(z) - P_X(z)|$ converges to zero almost surely, i.e.

$$P\left(\lim_{n\to\infty} sup_{z\in(0,1]}\,|P_n(z) - P_X(z)| = 0\right) = 1.$$

For a discrete random variable X, with a fixed z_0 where $|z_0| \leq 1$, we have

$$E(P_n(z_0)) = \sum_{j=0}^{\infty} E(f_j/n) z_0^j = \sum_{j=0}^{\infty} p_j z_0^j = P_X(z_0).$$

Evaluated at z_0, the empirical pgf is an unbiased estimator of the theoretical pgf. By the central limit theorem, the standardized empirical pgf evaluated at z_0 will converge to a standard normal distribution $N(0, 1)$.

In section 4, to estimate the two parameters λ and α of the discrete stable distribution, we minimize the quadratic distance between its empirical and theoretical pgf.

4 Estimation of the parameters

4.1 The model

In order to define the linear regression model, we define the function $g(x) = \ln(-\ln(x))$; we obtain

$$\begin{aligned} g(P_X(z)) &= \ln\left[-\ln\left(P_X(z)\right)\right] \\ &= \ln\left[\lambda(1-z)^{\alpha}\right] \\ &= \ln\lambda + \alpha\ln\left(1-z\right) \\ &= \beta + \alpha\ln\left(1-z\right), \end{aligned}$$

where $\beta = \ln \lambda$. It is a linear function of the parameters β and α. We can define a linear model in terms of parameters β, α, and an error term ϵ, with the empirical probability generating function.

The model is the following:

$$\begin{aligned} g(P_n(z_s)) &= g(P_X(z_s)) + \epsilon_s, \quad s = 1, 2, ..., k \\ &= \beta + \alpha \ln(1 - z_s) + \epsilon_s, \end{aligned}$$

where $z_1, z_2, ..., z_k$ are k selected points in the interval $(-1, 1)$.

From the model, we get

$$\epsilon_s = \ln[-\ln P_n(z_s)] - \beta - \alpha \ln(1 - z_s), \quad s = 1, 2, ..., k.$$

Using the method of statistical differentials, we have, asymptotically,

$$\begin{aligned} E(\epsilon_s) &\simeq \ln[-\ln E(P_n(z_s))] - \ln[-\ln P_X(z_s)] \\ &= \ln[-\ln P_X(z_s)] - \beta - \alpha \ln(1 - z_s) \\ &= 0. \end{aligned}$$

With vector and matrix notation, the model becomes

$$Y = X\theta + \epsilon,$$

where

$$Y_{k\times 1} = \begin{pmatrix} \ln(-\ln P_n(z_1)) & \ln(-\ln P_n(z_2)) & ... & \ln(-\ln P_n(z_k)) \end{pmatrix}'$$

$$X_{k\times 2} = \begin{pmatrix} 1 & \ln(1 - z_1) \\ 1 & \ln(1 - z_2) \\ ... & ... \\ 1 & \ln(1 - z_k) \end{pmatrix}$$

$$\theta_{2\times 1} = \begin{pmatrix} \beta & \alpha \end{pmatrix}'$$

$$\epsilon_{k\times 1} = \begin{pmatrix} \epsilon_1 & \epsilon_2 & ... & \epsilon_k \end{pmatrix}'.$$

Let us represent the asymptotic variance-covariance matrix of ϵ by $\Sigma = Var(\epsilon) = E(\epsilon\epsilon')$, since $E(\epsilon) \to 0$ as $n \to \infty$. In subsection 4.2, we derive the expression for Σ, which will be a function of the unknown parameters β and α.

4.2 The variance-covariance matrix

By the method of statistical differentials, we have, as $n \to \infty$,

$$Var(\epsilon_s) = Var[\ln(-\ln P_n(z_s)] = (1/\ln P_X(z_s))^2 \times Var[-\ln P_n(z_s)]$$

where $Var[-\ln P_n(z_s)] = (1/P_X(z_s))^2 \times [P_X(z_s \cdot z_s) - P_X(z_s) \cdot P_X(z_s)]$, so that the asymptotic variance of ϵ_s is given by

$$Var(\epsilon_s) = \frac{[P_X(z_s \cdot z_s) - P_X(z_s) \cdot P_X(z_s)]}{[P_X(z_s) \ln P_X(z_s)]^2}$$

where $s = 1, 2, ..., k$.

Similarly, we can find the asymptotic covariances of the error terms,

$$Cov(\epsilon_r, \epsilon_s) = Cov[\ln(-\ln P_n(z_r)), \ln(-\ln P_n(z_s))], \quad \text{for } r \neq s$$

$$= \frac{[P_X(z_r \cdot z_s) - P_X(z_r) \cdot P_X(z_s)]}{[P_X(z_r) \ln P_X(z_r)][P_X(z_s) \ln P_X(z_s)]}.$$

These terms can be estimated with $P_n(\cdot)$ replacing $P_X(\cdot)$.

We have the expressions to evaluate all the elements of the asymptotic variance-covariance matrix Σ. Let us define the matrix $\Sigma^* = n\Sigma$. These variance and covariance elements are also given by the matrix W in Theorem 3.5 of Rémillard and Theodorescu (2001).

4.3 Initial values of the parameters

To estimate the parameter vector θ, we first need to calculate initial values of the parameters, denoted $\hat{\theta}_0 = (\hat{\beta}_0, \hat{\alpha}_0)'$, where $\hat{\beta}_0 = \ln \hat{\lambda}_0$, using either of the following two methods.

Method 1. By replacing Σ by the identity matrix, we obtain a consistent estimator of the parameter vector θ,

$$\hat{\theta}_0 = (X'X)^{-1}X'Y.$$

However, it is a less efficient estimator of θ than the weighted version one given in subsection 4.4.

Method 2. See Rémillard and Theodorescu (2001).

Since $P_n(z_0) \xrightarrow{P} P_X(z_0)$, as $n \to \infty$, and $g(x) = \ln x$ is bounded and continuous for $|z_0| \leq 1$, then $\ln(P_n(z_0)) \xrightarrow{P} \ln(P_X(z_0))$ (see Roussas (1997)).

We can therefore estimate $\ln(P_X(z_0))$ by $\ln(P_n(z_0))$. Using two distinct points z_1 and z_2, we obtain two equations in two unknowns:

$$\ln P_n(z_1) = -\lambda(1 - z_1)^\alpha$$

$$\ln P_n(z_2) = -\lambda(1 - z_2)^{\alpha}.$$

Solving this system of equations, we obtain the estimators

$$\hat{\alpha}_0 = \frac{\ln\left[\ln(P_n(z_1))/\ln(P_n(z_2))\right]}{\ln\left(\frac{1-z_1}{1-z_2}\right)},$$

and

$$\hat{\lambda}_0 = -\frac{\ln(P_n(z_1))}{(1 - z_1)^{\alpha_0}}.$$

In order to get good initial values for the two parameters, we should select the two values of z far apart. A more general version based on q points is also given in their paper. The estimator can be viewed as a moments-type estimator which was introduced in Press (1972).

4.4 The algorithm

The quadratic distance estimator (QDE) of the parameter vector $\theta = (\beta, \alpha)'$, denoted by $\hat{\theta}$, is obtained by minimizing the quadratic form

$$(Y - X\theta)'\Sigma^{-1}(Y - X\theta).$$

with respect to θ. Explicily, $\hat{\theta}$ can be expressed as

$$\hat{\theta} = (X'\Sigma^{-1}X)^{-1}X'\Sigma^{-1}Y. \qquad (4.1)$$

However, since Σ is a function of p_i and therefore of the parameters, the following iterative algorithm must be used to evaluate $\hat{\theta}$.

1. Calculate the initial value of $\hat{\theta}_0 = (\hat{\beta}_0, \hat{\alpha}_0)$, using one of the methods of subsection (4.3).

2. With the estimated values of the parameters, estimate the probabilities p_i using the series expansion of the pgf in terms of z

$$P_X(z) = \exp[-\lambda(1 - z)^{\alpha}] = \sum_{i=0}^{\infty} p_i z^i.$$

3. Estimate the variance-covariance matrix Σ using the estimated values of p_i from step 2 and the formulas of section (4.2) for $Var(\epsilon_s)$ and $Cov(\epsilon_r, \epsilon_s)$.

4. Reestimate $\hat{\lambda}$ and $\hat{\alpha}$ from equation (4.1).

5. Redo steps 2, 3 and 4 to estimate new values for p_i, Σ and $\hat{\theta}$, to the level of accuracy desired.

The estimator of λ will be $\hat{\lambda} = e^{\hat{\beta}}$, with variance equal to $e^{2\beta}Var(\hat{\beta})$. Note that in the formulas for $Var(\epsilon_s)$ and $Cov(\epsilon_r, \epsilon_s)$, we could also estimate p_i and p_j by f_i/n and f_j/n. Only one iteration is then required.

4.5 Properties of the estimator

Luong and Doray (2009) studied the asymptotic properties of the quadratic distance estimator $\hat{\theta}$. They showed that:
1. $\hat{\theta}$ is a consistent estimator and is aymptotically unbiased.
2. $\sqrt{n}(\hat{\theta}-\theta)$ has an asymptotic normal distribution with mean 0 and variance-covariance matrix $(X'\Sigma^{*-1}X)^{-1}$.
3. For certain parametric families, $\hat{\theta}$ has high efficiency if the score function can be approximated by linear combinations of $z_1^X, \ldots, z_k^X$ using the mean square error criterion.
4. A simple goodness-of-fit statistic for testing the model is

$$n(Y - X\hat{\theta})'\hat{\Sigma}^{-1}(Y - X\hat{\theta}),$$

which converges to a χ^2_{k-2} distribution, where $\hat{\theta}$ is the QDE given by (4.1).

Comparing our quadratic distance method to the weighted L2-distance method of Marcheselli, Baccini and Barabesi (2008), it can be viewed as based on a form which is a discretized version. The quadratic form of our discretized version with appropriate weights leads to test statistics with a chi-square distribution across the composite hypothesis; however, no numerical integrations are required, which make the method easy to apply in practice. These weights are based on the variances and covariances of the empirical probability generating function at various points. The weights of the continuous version of the L2 method do not take into account the covariance structure.

5 Numerical Examples

Using the algorithm mentioned in section 2c), we generated samples of discrete stable random variables and used the estimation method developed in section 4 to analyze the effects of the choice of points z_s and of truncating the sample, on the bias of the estimators.

5.1 Selection of the points z

We first generated 5000 observations from the discrete stable distribution with parameters $\lambda = 1$ and $\alpha = 0.9$. With those parameter values, the distribution is not too heavy-tailed. We wanted to determine the number of points k and the values of the points z_s, $|z_s| \leq 1$, $s = 1, \ldots, k$ we should use to calculate the QDE $\hat{\theta}$.

We considered the following values:
A- $k = 19$ (no negative values)

$z = \{0.05, 0.10, 0.15, 0.20, 0.25, 0.30, 0.35, 0.40, 0.45, 0.50, 0.55, 0.60, 0.65, 0.70, 0.75, 0.80, 0.85, 0.90, 0.95\}$

B- $k = 18$
$z = \{-0.9, -0.8, -0.7, -0.6, -0.5, -0.4, -0.3, -0.2, -0.1, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9\}$

C- $k = 10$
$z = \{-0.9, -0.7, -0.5, -0.3, -0.1, 0.1, 0.3, 0.5, 0.7, 0.9\}$

D- $k = 9$ (no negative values)
$z = \{0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9\}$

E- $k = 4$
$z = \{-0.9, -0.3, 0.3, 0.9\}$

F- $k = 2$
$z = \{-0.5, 0.5\}$

G- $k = 2$
$z = \{-0.9, 0.9\}$

In case A, at the second iteration, because the selected values of z are too close, the variance-covariance matrix Σ is singular and its inverse does not exist (rank(Σ)= 12). The same thing happens in case B (rank(Σ)=15), and in case D (rank(Σ)=8). In these situations, we use the pseudo-inverse (see Golub and Van Loan (1989)) of Σ to obtain the results. Those results have been marked with a $*$ in Table 1.

The algorithm converged in only 2 iterations, except when we had to use the pseudo-inverse of Σ, which made the calculations much more time-consuming. Since the relative errors of the parameters do not decrease much with increasing value of k, it is suggested not to use too many values for z. On the other hand, using only 2 values for z may cause a large bias for the estimators of λ. See F and G in Table 1.

We therefore recommend the choice of C ($k = 10$) or E ($k = 4$), because the relative errors are smaller and the calculation speed is much faster.

Using more points for z_s does not change much the estimated value of Σ. In case B, we obtain

$$Var(\hat{\theta}) = \begin{pmatrix} 0.000251019 & 0.0000283142 \\ 0.0000283142 & 0.000034218 \end{pmatrix};$$

Table 1: Effect of the selection of the points z

Values of z	QDE	relative error
A^*	$\hat{\alpha} = 0.916731$	1.859%
	$\hat{\lambda} = 0.997756$	-0.224%
B^*	$\hat{\alpha} = 0.91825$	2.028%
	$\hat{\lambda} = 0.99742$	-0.258%
C	$\hat{\alpha}$=0.916447	1.827%
	$\hat{\lambda}$ =0.996336	-0.366%
D^*	$\hat{\alpha} = 0.916731$	1.859%
	$\hat{\lambda} = 0.997725$	-0.227%
E	$\hat{\alpha}$=0.914694	1.633%
	$\hat{\lambda}$=0.996986	-0.301%
F	$\hat{\alpha}$ =0.910087	1.121%
	$\hat{\lambda}$=0.99298	-0.707%
G	$\hat{\alpha}$=0.910871	1.208%
	$\hat{\lambda}$= 0.987717	-1.228%

in case C,

$$Var(\hat{\theta}) = \begin{pmatrix} 0.000251444 & 0.0000288046 \\ 0.0000288046 & 0.0000354222 \end{pmatrix},$$

and in case E,

$$Var(\hat{\theta}) = \begin{pmatrix} 0.00025702 & 0.000029364 \\ 0.000029364 & 0.0000388233 \end{pmatrix}.$$

5.2 Effect of truncation

When the dataset has some extreme values, truncating it speeds up the calculations of the QDE.

We used the parameters α=0.4 and λ=4.5 to generate samples of discrete stable random variables with different sample sizes. The samples contain some very large values (the largest value with $n = 2000$ is 446,630,588; it is 47,287,674 with $n = 1000$, 1.24×10^8 with $n = 500$ and 149,289 with $n = 100$).

Table 2 contains the QDE of α and λ. To calculate them, we used $k = 4$ and $z = \{-0.9, -0.3, 0.3, 0.9\}$.

In Tables 3 to 6, we compare the effect of truncating different percentages of the largest observations on the relative errors of the parameters. Note that

Table 2: Estimates and standard errors for α and λ ($z \in$ C)

n	$\hat{\alpha}$ (s.e.)	$\hat{\lambda}$ (s.e.)
2000	0.41506 (0.00896627)	4.63049 (0.133835)
1000	0.383514 (0.0190303)	4.32642 (0.205090)
500	0.380229 (0.014749)	4.24766 (0.179035)
100	0.39524 (0.0299940)	4.05186 (0.340105)

Table 3: Effect of truncation (n=2000)

Truncation	QDE	relative error
none	$\hat{\alpha}$ =0.41506	3.765%
	$\hat{\lambda}$ =4.63049	2.90%
8%	$\hat{\alpha}$= 0.424454	6.11%
	$\hat{\lambda}$=4.55938	1.320%
15%	$\hat{\alpha}$=0.438582	9.646%
	$\hat{\lambda}$=4.49284	-0.159%
20%	$\hat{\alpha}$=0.450093	12.523%
	$\hat{\lambda}$=4.44256	-1.276%

at the same percentage of truncation, the absolute value of the relative error of estimator $\hat{\lambda}$ increases when the sample size decreases. Also notice that the sum of the absolute relative error of $\hat{\alpha}$ and $\hat{\lambda}$ increases when the percentage of truncation increases.

After using many different truncation percentages to estimate the parameters, we conclude that when it is less than 15% and the sample size $n \geq 500$, the relative errors of the parameters will be less then 10%.

6 Conclusion

Minimizing the quadratic distance between the empirical and theoretical probability generating function is a powerful technique to estimate the parameters of a discrete distribution when no explicit expression for its probability function exists.

For the discrete stable distribution, the estimators we obtained are con-

Table 4: Effect of truncation (n=1000)

Truncation	QDE	relative error
none	$\hat{\alpha}$ =0.383514	-4.122%
	$\hat{\lambda}$ =4.32642	-3.857%
9%	$\hat{\alpha}$= 0.398745	-0.314%
	$\hat{\lambda}$=4.34436	-3.459%
15%	$\hat{\alpha}$=0.410567	2.642%
	$\hat{\lambda}$=4.18573	-6.984%
20%	$\hat{\alpha}$=0.421705	5.426%
	$\hat{\lambda}$=4.1342	-8.129%

Table 5: Effect of truncation (n=500)

Truncation	QDE	relative errors
none	$\hat{\alpha}$ =0.380229	-4.943%
	$\hat{\lambda}$ =4.24766	-5.608%
10%	$\hat{\alpha}$= 0.397106	-0.724%
	$\hat{\lambda}$=4.15322	-7.706%
15%	$\hat{\alpha}$=0.406938	1.735%
	$\hat{\lambda}$=4.10246	-8.834%
20%	$\hat{\alpha}$=0.417944	4.486%
	$\hat{\lambda}$=4.04903	-10.022%

Table 6: Effect of truncation (n=100)

Truncation	QDE	relative errors
none	$\hat{\alpha}$ =0.39524	-1.19%
	$\hat{\lambda}$ =4.05186	-9.959%
10%	$\hat{\alpha}$= 0.414054	3.515%
	$\hat{\lambda}$=3.95739	-12.058%
15%	$\hat{\alpha}$=0.42508	6.27%
	$\hat{\lambda}$=3.90667	-13.185%
20%	$\hat{\alpha}$=0.437482	9.371%
	$\hat{\lambda}$=3.85335	-14.37%

sistent, asymptotically unbiased and normally distributed, with a variance-convariance matrix easy to calculate .

We simulated several samples of discrete stable random variables with different parameter values. We analyzed the effect of the choice of values of $z_s, s = 1, \ldots, k$ on the estimated parameters, and concluded that $k = 4$ or 10 is a good choice since the bias was smaller and the calculations were faster. When the percentage of data truncation was less than 15% in the right tail and the sample size n greater than 500, the bias of the QDE was small.

A method analogous to the one developed in this paper for the discrete stable distribution could be applied to the discrete Linnik (DL) distribution with pgf (see Bouzar (2002))

$$P(z) = \begin{cases} [1 + \lambda(1-z)^\gamma/\beta]^{-\beta} & \text{for } \beta > 0 \\ \exp[-\lambda(1-z)^\gamma] & \text{for } \beta = \infty, \end{cases}$$

since the DL distribution also does not possess a simple form for its probability function. The discrete stable distribution is a limiting case of the DL distribution.

Minimizing the distance between the empirical and theoretical pgf's with a non-linear least-squares method would yield consistent estimators, with an asymptotic normal distribution. However, no transformation of the pgf can give a linear function of the 3 parameters for the DL distribution, contrarily to the discrete stable distribution where we used the log-log transformation; the expression for the variance-covariance matrix of the QDE would then be more complicated. It would also be more difficult to find initial values for the parameters.

ACKNOWLEDGEMENTS

The authors gratefully acknowledge the financial support of the Natural Sciences and Engineering Research Council of Canada.

References

[1] Bouzar, N. (2002). Mixture representations for discrete Linnik laws, *Statistica Neerlandica*, **56**, 295-300.

[2] Christoph, G. and Schreiber, K. (1998). Discrete stable random variables, *Statistics & Probability Letters*, **37**, 243-247.

[3] Christoph, G. and Schreiber, K. (2000). Scaled Sibuya distribution and discrete self-decomposability, *Statistics & Probability Letters*, **48**, 181-187

[4] Devroye, L. (1993). A triptych of discrete distributions related to the stable law, *Statistics & Probability Letters*, **18**, 349-351.

[5] Golub, G.H. and Van Loan, C.F. (1989). *Matrix computations*, Second Edition, John Wiley and Sons, Inc., New York.

[6] Johnson, N.L. and Kotz,S. and Balakrishnan, N. (1997). *Discrete multivariate distributions*, John Wiley and Sons, Inc., New York.

[7] Kanter, M. (1975). Stable densities under change of scale and total variation inequalities, *The Annals of Probability*, **3**, 697-707.

[8] Luong, A. and Doray, L.G. (2009). Inference for the Positive Stable Laws Based on a Special Quadratic Distance, *Statistical Methodology*, **6**, 147-156.

[9] Marcheselli, M., Baccini, A. and Barabesi, L. (2008). Parameter Estimation for the Inference for Discrete Stable Family, *Comm. in Stat.: Theory and methods*, **37**, 815-830.

[10] Nakamura, M. and Pérez-Abreu, V. (1993). Empirical probability generating function: An overview, *Insurance: Mathematics and Economics*, **12**, 287-295.

[11] Pakes, A.G. (1998), Mixture representations for symmetric generalized Linnik laws, *Statistics & Probability Letters*, **37**, 213-221.

[12] Press, S.J. (1972), Estimation in univariate and multivariate stable distribution, *Journal of the American Statistical Association*, **67**, 842-846.

[13] Rémillard, B. and Theodorescu, R. (2001). Inference based on the empirical probability generating function for mixtures of Poisson distributions, *Statistics & Decisions*, **18**, 349-366.

[14] Roussas, G.G. (1997), *A course in mathematical statistics*, Second Edition, Academic Press, San Diego.

[15] Sibuya, M. (1979). Generalized hypergeometric, digamma and trigamma distributions, *Annals of the Institute of Statistical Mathematics*, **31**, 373-390.

[16] Steutel, F.W. and van Harn, K. (1979). Discrete analogues of self-decomposability and stability, *The Annals of Probability*, **7**, 893-899.

[17] Steutel, F.W. and van Harn, K. (2004). *Infinite Divisibility of Probability Distributions on the Real Line*, Marcel Dekker, Inc., New York-Basel.

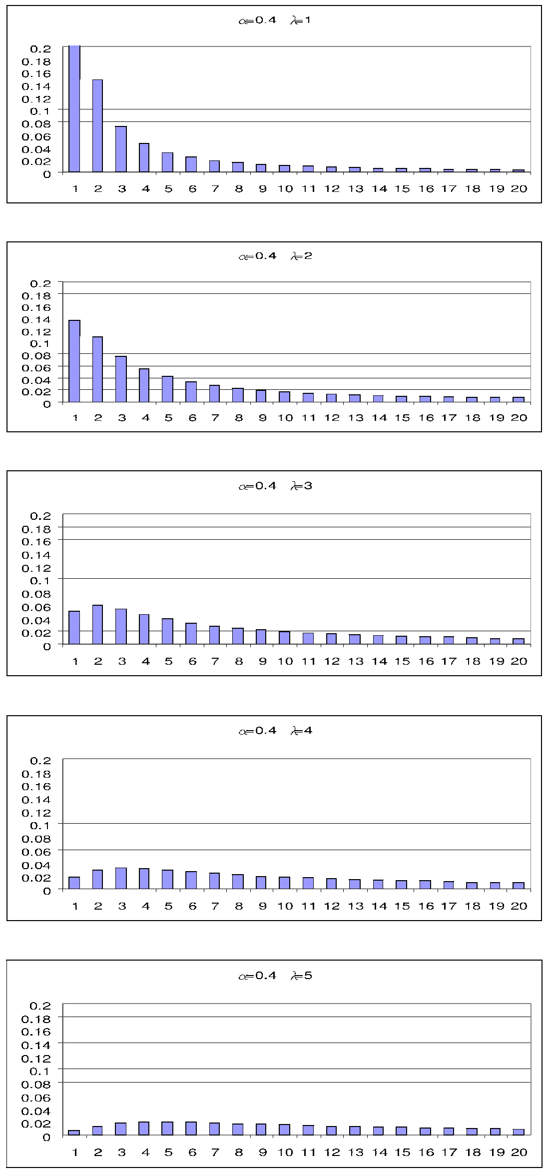

Figure 1: Probabilities with α=0.4 and different λ's