

Université de Montréal

PARAMETERS ESTIMATION OF THE  
DISCRETE STABLE DISTRIBUTION

par

Shu Mei Jiang

Département de mathématiques et de statistique

Faculté des arts et des sciences

Mémoire présenté à la Faculté des études supérieures

en vue de l'obtention du grade de

Maître ès sciences (M.Sc.)  
en mathématique

décembre 2006



QA

3

U54

2007

V.002



## **AVIS**

L'auteur a autorisé l'Université de Montréal à reproduire et diffuser, en totalité ou en partie, par quelque moyen que ce soit et sur quelque support que ce soit, et exclusivement à des fins non lucratives d'enseignement et de recherche, des copies de ce mémoire ou de cette thèse.

L'auteur et les coauteurs le cas échéant conservent la propriété du droit d'auteur et des droits moraux qui protègent ce document. Ni la thèse ou le mémoire, ni des extraits substantiels de ce document, ne doivent être imprimés ou autrement reproduits sans l'autorisation de l'auteur.

Afin de se conformer à la Loi canadienne sur la protection des renseignements personnels, quelques formulaires secondaires, coordonnées ou signatures intégrées au texte ont pu être enlevés de ce document. Bien que cela ait pu affecter la pagination, il n'y a aucun contenu manquant.

## **NOTICE**

The author of this thesis or dissertation has granted a nonexclusive license allowing Université de Montréal to reproduce and publish the document, in part or in whole, and in any format, solely for noncommercial educational and research purposes.

The author and co-authors if applicable retain copyright ownership and moral rights in this document. Neither the whole thesis or dissertation, nor substantial extracts from it, may be printed or otherwise reproduced without the author's permission.

In compliance with the Canadian Privacy Act some supporting forms, contact information or signatures may have been removed from the document. While this may affect the document page count, it does not represent any loss of content from the document.

Université de Montréal

Faculté des études supérieures

Ce mémoire intitulé

PARAMETERS ESTIMATION OF THE  
DISCRETE STABLE DISTRIBUTION

présenté par

Shu Mei Jiang

a été évalué par un jury composé des personnes suivantes :

*Manuel Morales*

---

(président-rapporteur)

*Louis G. Doray*

---

(directeur de recherche)

*Richard Duncan*

---

(membre du jury)

Mémoire accepté le:

*6 december 2006*

---



## ACKNOWLEDGMENT

---

I would like to express my heartfelt gratitude to my supervisor, Professor Louis G. Doray, for not only his guidance, patient help and constructive criticism throughout all the course of this thesis work, but also his trust, understanding and support. I am lucky to work with an expert like him in the field of Actuarial Mathematics.

Moreover, I would like to express my appreciation to Professor Roch Roy, Professor Martin Bilodeau, Post-doctoral Researcher Abdessamad Saidi for their excellent lectures and for their valuable help.

Thanks are also due to computer administrator Francis Forget, coadministrators Alexandre Girouard and Rony Touma. I have benefitted greatly from interactions with them.

## SUMMARY

---

In this thesis, we analyze the method of parameter estimation of the discrete two-parameter stable distribution. We present an estimation method based on minimizing the quadratic distance between the empirical and theoretical probability generating functions. This method makes it possible to use the discrete stable distribution model in a variety of practical problems.

Firstly, we introduce some of the properties of the discrete stable distribution and review some theorems. Secondly, we develop an expression for the variance-covariance matrix for the terms of errors between the empirical and theoretical probability generating functions, and we give the formulas of the estimators. Thirdly, numerical examples are provided and the asymptotic properties of the estimators are studied.

We simulate several samples of discrete stable distributed datasets with different parameters. The estimators obtained were quite good.

We also conduct inference about the parameters such as confidence intervals of the parameters and tests concerning the parameters.

## SOMMAIRE

---

Dans cette thèse, nous analysons une méthode d'estimation des paramètres de la distribution discrète stable avec deux paramètres. Nous présentons la méthode d'estimation basée sur la minimisation de la distance quadratique entre les fonctions génératrices des probabilités empiriques et théoriques. La méthode permet d'utiliser le modèle de la distribution discrète stable pour une diversité de problèmes pratiques.

Premièrement, nous introduisons quelques propriétés de la distribution discrète stable et revisons certains théorèmes. Deuxièmement, nous développons une expression pour la matrice de variance-covariance des erreurs entre les fonctions génératrices des probabilités empiriques et théoriques et nous donnons les formules des estimateurs. Troisièmement, des exemples numériques sont fournis et les comportements asymptotiques des estimateurs sont étudiés.

Nous simulons plusieurs échantillons de jeux de données suivant une loi discrète stable avec des paramètres différents. Les estimateurs obtenus sont bons.

Nous faisons aussi l'inférence sur les paramètres, construisons les intervalles de confiance et nous faisons des tests d'hypothèse sur les paramètres.

# CONTENTS

---

<b>Acknowledgment</b> .....	iii
<b>Summary</b> .....	iv
<b>Sommaire</b> .....	v
<b>List of figures</b> .....	ix
<b>List of tables</b> .....	x
<b>Chapter 1. Introduction</b> .....	1
1.1. Stability of a random variable .....	1
1.2. Contributions of this thesis .....	2
1.3. Organization of the thesis .....	3
<b>Chapter 2. Discrete Stable Distribution</b> .....	4
2.1. Poisson distribution as a special case .....	4
2.2. As a compound Poisson distribution .....	5
2.3. As a Poisson random variable .....	5
2.4. Infinite divisibility .....	6
2.5. Discrete self-decomposability .....	7
2.6. Some other representations with discrete stable distribution .....	8
2.7. Probabilities .....	11



2.8. Moment characteristics .....	16
2.8.1. Case $\alpha = 1$ .....	16
2.8.2. Case $\alpha \in (0, 1)$ .....	16
<b>Chapter 3. Statistical Review</b> .....	<b>18</b>
3.1. Linear regression .....	18
3.2. Empirical probability generating function .....	21
3.3. Moments of multinomial distribution .....	22
3.4. Delta theorem .....	24
3.5. The singular value decomposition (SVD) of a matrix and the Pseudo-Inverse matrix .....	26
3.5.1. The singular value decomposition .....	26
3.5.2. Computing the SVD .....	28
3.5.3. Rank deficiency and numerical rank determination .....	30
3.5.4. The pseudo-inverse matrix .....	31
<b>Chapter 4. Estimation and Hypothesis Testing of the Parameters</b> .....	<b>33</b>
4.1. The model .....	33
4.2. The variance-covariance matrix .....	36
4.3. The initial values of the parameters .....	38
4.4. The algorithm .....	40
4.5. Inferences concerning the vector $\theta$ .....	40
4.5.1. Sampling distribution of the standardized statistic .....	41
4.5.2. Confidence intervals for $\beta$ and $\alpha$ .....	41
4.5.3. Tests concerning $\alpha$ .....	42

4.5.4. Tests concerning $\lambda$ .....	43
<b>Chapter 5. Numerical Examples .....</b>	<b>45</b>
5.1. Effect of the number of points taken .....	45
5.2. Confidence intervals for the parameters .....	49
5.3. Tests concerning $\lambda$ and $\alpha$ .....	50
5.4. Effect of truncation .....	52
<b>Chapter 6. Conclusion .....</b>	<b>59</b>
<b>Bibliography .....</b>	<b>61</b>
<b>Appendix A. Several terms of the probability function .....</b>	<b>A-i</b>

## LIST OF FIGURES

---

2.1	Probabilities with $\lambda = 4.5$ and different $\alpha$ 's.....	13
2.2	Probabilities with $\lambda = 1$ and different $\alpha$ 's.....	14
2.3	Probabilities with $\alpha = 0.4$ and different $\lambda$ 's.....	15
2.4	Probabilities with $\alpha = 0.8$ and different $\lambda$ 's.....	17

## LIST OF TABLES

---

5.1	Effect of the numbers of the points of $z$ .....	48
5.2	Confidence interval for $\beta$ and $\lambda$ with 10 points .....	50
5.3	Confidence interval for $\alpha$ with 10 points .....	51
5.4	Test concerning $\alpha$ with 10 points .....	51
5.5	Test concerning $\lambda$ with 10 points .....	52
5.6	The effect of truncation on ( $n=2000$ ) .....	55
5.7	The effect of truncation ( $n=1000$ ) .....	56
5.8	The effect of truncation ( $n=500$ ) .....	57
5.9	The effect of truncation ( $n=100$ ) .....	58

# Chapter 1

---

## INTRODUCTION

The discrete two-parameter stable distribution is a special case of certain mixtures of Poisson distributions. It was first introduced by Steutel and van Harn in 1979. It is a distribution that allows skewness and heavy tails and has many intriguing mathematical properties. The lack of closed formulas for the probability and distribution functions has been a major drawback to the use of discrete stable distribution by practitioners. For example, it is difficult to estimate the two parameters, to compute probabilities or quantiles.

In this thesis, we will develop a method to estimate the parameters by minimizing the quadratic distance between the empirical and the theoretical probability generating functions. This method makes it possible to use the discrete stable distribution model in a variety of practical problems.

### 1.1. STABILITY OF A RANDOM VARIABLE

Nolan (2004) defined a stable random variable as follows.

**Definition 1.1.1.** *A random variable  $X$  is stable or stable in the broad sense if for  $X_1$  and  $X_2$  independent copies of  $X$  and any positive constants  $a$  and  $b$ ,*

$$aX_1 + bX_2 \stackrel{D}{=} cX + d \tag{1.1.1}$$

*holds for some positive  $c$  and  $d \in R$ , where the symbol  $\stackrel{D}{=}$  means equality in distribution, i.e. both expressions have the same probability law.*

*The random variable is strictly stable or stable in the narrow sense if the equation (1.1.1) holds with  $d = 0$  for all choices of  $a$  and  $b$ . A random variable is symmetric stable if it is stable and symmetrically distributed around 0, e.g.*

$$X \stackrel{D}{=} -X$$

The word "stable" is used because the shape is stable or unchanged under sums of the type (1.1.1).

Examples of stable distribution include normal distribution, Cauchy distribution and Lévy distribution.

## 1.2. CONTRIBUTIONS OF THIS THESIS

When no explicit expression for the probability function exists, it will not be possible to use a method like the maximum likelihood estimation method to estimate the parameters. We will present an alternative estimation method based on minimizing the quadratic distance between the empirical and theoretical probability generating functions. The quadratic distance method is a useful tool which uses theory developed for the classical linear regression model. In order to obtain the estimators of the parameters, we need to minimize numerically the quadratic distance between the empirical and theoretical probability generating functions.

Secondly, the asymptotic properties of the estimators are studied. The consistency, asymptotic normality and robustness of the estimators will be investigated.

Thirdly, numerical examples are provided. We will conduct numerous calculations using Kanter's (1975) simulation method to generate groups of discrete stable datasets, and will estimate the parameters based on these datasets. Luong and Doray (2002) found that the quadratic estimators are robust and our results showed the same thing for the stable distribution. So, we can use this method to deal with truncated datasets. We will also give the effect of the percentage of truncation on the bias of the estimators.

### 1.3. ORGANIZATION OF THE THESIS

The thesis is organized as follows. In chapter 2, we introduce some of the properties of the discrete stable distribution. In chapter 3, we review some results that will be used later, such as the  $\delta$ -theorem, the moments of multinomial random variables, the quadratic distance estimation method, the singular value decomposition method for real matrix and the pseudo-inverse of a matrix. In chapter 4, we develop an expression of the variance-covariance matrix for the error terms between the empirical and theoretical probability generating functions. We also give the formulas of the estimators. In chapter 5, we use examples to illustrate the estimation method we produced. In chapter 6, we will summarize the main conclusions.

## Chapter 2

---

### DISCRETE STABLE 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, such as infinitely divisibility and self-decomposability.

The discrete stable distribution was introduced via its probability generating function. If  $X$  is a discrete random variable taking values on some subset of the non-negative integers  $\{0, 1, \dots\}$ , then the probability-generating function of  $X$  is defined as

$$P_X(z) = E(z^X) = \sum_{i=0}^{\infty} f(i)z^i,$$

where  $f$  is the probability mass function of  $X$ .

For  $\alpha \in (0, 1]$  and  $\lambda > 0$ , let  $X(\lambda, \alpha)$  be a discrete stable random variable, with probability generating function given by

$$P_X(z) = \exp[-\lambda(1 - z)^\alpha], \quad |z| \leq 1 \quad (2.0.1)$$

#### 2.1. POISSON DISTRIBUTION AS A SPECIAL CASE

Obviously, with  $\alpha = 1$ , we obtain

$$\begin{aligned} P_X(z) &= \exp[-\lambda(1 - z)] \\ &= \exp[\lambda(z - 1)], \quad |z| \leq 1, \quad \lambda > 0. \end{aligned}$$

It is the probability generating function of Poisson distribution with parameter  $\lambda$ .



## 2.2. AS A COMPOUND POISSON DISTRIBUTION

The discrete stable random variables can be obtained as

$$X \stackrel{D}{=} M_1 + M_2 + \cdots + M_Y \quad (2.2.1)$$

where  $Y \sim \text{Poisson}(\lambda)$ , and  $Y$  is independent of  $M_i$ , where  $M_1, M_2, \dots$ , are i.i.d. random variables with probability generating function

$$P_{M_i}(z) = 1 - (1 - z)^\alpha. \quad (2.2.2)$$

The discrete random variables  $M_i$  follow the Sibuya distribution with parameter  $\alpha$  (introduced by Sibuya (1979)). Note that

$$\begin{aligned} P_X(z) &= E(z^X) \\ &= E_Y[E(z^{M_1+M_2+\cdots+M_Y} | Y)] \\ &= E_Y\{E[(z^{M_i})^Y | Y]\} \\ &= E_Y[E(P_{M_i}(z)^Y | Y)] \\ &= E[P_{M_i}(z)^Y] \\ &= \exp\{\lambda[P_{M_i}(z) - 1]\} \\ &= \exp\{\lambda[1 - (1 - z)^\alpha - 1]\} \\ &= \exp[-\lambda(1 - z)^\alpha] \end{aligned}$$

This is the probability generating function of discrete stable distribution, hence the discrete stable distribution is a compound Poisson distribution.

## 2.3. AS A POISSON RANDOM VARIABLE

Devroye (1993) proved that a discrete stable random variable with parameters  $\lambda$  and  $\alpha$  is a conditional Poisson random variable with parameter  $\lambda^{1/\alpha} S_{\alpha,1}$ , where  $S_{\alpha,1}$  is a positive stable random variable with parameter  $\alpha$  and Laplace transform

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

See Devroye (1993).

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

$$S_{\alpha,1} \stackrel{L}{=} \left( \frac{\sin((1-\alpha)\pi U)}{E \sin(\alpha\pi U)} \right)^{\frac{1-\alpha}{\alpha}} \left( \frac{\sin(\alpha\pi U)}{\sin(\pi U)} \right)^{\frac{1}{\alpha}}. \quad (2.3.1)$$

where

$S_{\alpha,1}$  is the positive stable random variable with parameter  $\alpha$ .

$U \sim \text{Uniform}(0,1)$

$E \sim \text{Exponential}(1)$

$U$  and  $E$  are independent.

**Theorem 2.3.1.** *A discrete stable random variable  $X(\lambda, \alpha)$  is distributed as a conditional Poisson random variable with parameter  $\lambda^{1/\alpha} S_{\alpha,1}$ , where  $S_{\alpha,1}$  is a positive stable random variable with parameter  $\alpha$ .*

$$X(\lambda, \alpha) \stackrel{L}{=} \text{Poisson}(\lambda^{1/\alpha} S_{\alpha,1})$$

See Zolotarev (1986).

PROOF. The characteristic function of  $X$  is obtained as

$$E(e^{itX}) = E[e^{\lambda^{1/\alpha} S_{\alpha,1} (e^{it}-1)}] = \exp[-\lambda(1 - e^{it})^\alpha].$$

We recognize that this is the characteristic function of the discrete stable distribution. □

**Remark 2.3.1.** *For  $\alpha = 1$ ,  $S_{\alpha,1}$  becomes the degenerate distribution with atom at  $x = 1$ .*

## 2.4. INFINITE DIVISIBILITY

Steutel and van Harn (1979) give the definition of infinite divisibility as follows.

**Definition 2.4.1.** *A discrete random variable with probability generating function  $P(z)$  is infinitely divisible if and only if  $P(z)$  has the following form*

$$P(z) = \exp[\lambda(G(z) - 1)], \quad (2.4.1)$$

where  $\lambda > 0$  and  $G(z)$  is a unique probability generating function with  $G(0) = 0$ .

Note that the probability generating function of a discrete stable distribution is given by

$$\begin{aligned} P_X(z) &= \exp[-\lambda(1-z)^\alpha] \\ &= \exp[\lambda(G(z) - 1)] \end{aligned}$$

We already know that  $G(z) = 1 - (1-z)^\alpha$  is the probability generating function of a Sibuya( $\alpha$ ) random variable and  $G(0) = 0$ . This is in accordance with definition 2.4.1. Therefore a discrete stable distribution is infinitely divisible.

## 2.5. DISCRETE SELF-DECOMPOSABILITY

Steutel and van Harn (1979) define discrete self-decomposability as follows.

**Definition 2.5.1.** *A discrete distribution is called discrete self-decomposable if its probability generating function satisfies*

$$P_X(z) = P_X(1 - \alpha + \alpha Z)P_\alpha(z), \quad |z| < 1, \quad \alpha \in (0, 1], \quad (2.5.1)$$

with  $P_\alpha(z)$  a probability generating function.

**Theorem 2.5.1.** *A probability generating function  $P(z)$  is discrete self-decomposable if and only if it has the following form*

$$P(z) = \exp \left\{ -\beta \int_z^1 \frac{1 - G(u)}{1 - u} du \right\} \quad (2.5.2)$$

where  $\lambda > 0$  and  $G$  is a unique probability generating function with  $G(0) = 0$ .

As a special case, consider the probability generating function of a Sibuya( $\alpha$ ) distribution

$$G(u) = 1 - (1 - u)^\alpha.$$

Since

$$\begin{aligned}
 \int_z^1 \frac{1 - G(u)}{1 - u} du &= \int_z^1 \frac{(1 - u)^\alpha}{1 - u} du \\
 &= \int_z^1 (1 - u)^{\alpha-1} du \\
 &= -\frac{1}{\alpha} (1 - u)|_z^1 \\
 &= \frac{1}{\alpha} (1 - z)^\alpha,
 \end{aligned}$$

hence

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

It is the probability generating function of a discrete stable distribution. We conclude that the Sibuya( $\alpha$ ) distribution is self-decomposable.

## 2.6. SOME OTHER REPRESENTATIONS WITH DISCRETE STABLE DISTRIBUTION

Pakes (1998) gave out some other properties of discrete stable distributions. He found that some other discrete distributions such as the discrete Linnik distribution can be formed from discrete stable distribution.

Bouzar (2002) presented four other distributions derived from the discrete stable distribution.

1. The following representation is the discrete analogue of a result obtained in the continuous case. Let  $X(\alpha, \lambda)$  be a discrete stable random variable with parameters  $\alpha$ ,  $\lambda$ , and  $Y(\delta, \lambda)$  be a stable continuous counterpart of  $X(\alpha, \lambda)$  with Laplace transform:

$$\psi_{\delta, \lambda}(\tau) = \exp(-\lambda \tau^\delta), \quad \tau \geq 0 \quad (2.6.1)$$

then

$$X(\alpha, \lambda) \stackrel{D}{=} X(\beta, Y(\delta, \lambda)) \quad (2.6.2)$$

where  $0 < \alpha < \beta \leq 1$  and  $\lambda > 0$ , and  $\delta = \alpha/\beta$ .

We can prove it in the following way.

If  $Q(z)$  and  $F(x)$  denote the pgf of the right-hand side of (2.6.2) and the distribution function of  $Y(\delta, \lambda)$  respectively, then, since  $\beta\delta = \alpha$ ,

$$Q(z) = \int_0^\infty e^{-x(1-z)^\beta} dF(x) = \psi_{\delta, \lambda}[(1-z)^\beta] = e^{-\lambda(1-z)^{\beta\delta}} = P_X(z).$$

2. Let  $L_{\alpha, \lambda}(\nu)$  denote the discrete Linnik distribution with probability generating function

$$P_L(z) = [1 + \lambda(1-z)^\alpha]^{-\nu}, \quad |z| \leq 1 \quad (2.6.3)$$

where  $\alpha \in (0, 1]$ ,  $\lambda > 0$  and  $\nu > 0$ , and let  $M_{1, \lambda}(\nu)$  denote the Gamma distribution with density

$$f_{\lambda, \nu}(x) = \frac{1}{\lambda^\nu \Gamma(\nu)} x^{\nu-1} e^{-x/\lambda}, \quad x > 0 \quad (2.6.4)$$

and Laplace transform

$$\phi_{1, \lambda}(\tau) = (1 + \lambda\tau)^{-\nu}; \quad (2.6.5)$$

then, for  $\alpha \in (0, 1]$  and  $\lambda, \nu > 0$

$$L_{\alpha, \lambda}(\nu) \stackrel{D}{=} X(\alpha, M_{1, \lambda}(\nu)) \quad (2.6.6)$$

where  $X(\alpha, M_{1, \lambda}(\nu))$  is the discrete stable distribution with parameters  $\alpha$  and  $M_{1, \lambda}(\nu)$ .

We can prove it in following way.

Let  $Q(z)$  be the pgf of the right-hand side of equation (2.6.5), we have:

$$Q(z) = \int_0^\infty e^{-x(1-z)^\alpha} f_{\lambda, \nu}(x) dx = \phi_{1, \lambda}[(1-z)^\alpha] = [1 + \lambda(1-z)^\alpha]^{-\nu}$$

3. Let  $L_{\alpha, \lambda}(\nu)$  be a discrete Linnik distribution with parameters  $\alpha$  and  $\lambda$  and probability generating function (2.6.3) and let  $M_{\delta, \lambda}(\nu)$  be the positive continuous counterpart of the Linnik distribution with parameters  $\delta$  and  $\lambda$ , and Laplace transform

$$\phi_{\delta, \lambda}(\tau) = (1 + \lambda\tau^\delta)^{-\nu}, \quad \tau \geq 0 \quad (2.6.7)$$

then, we have

$$L_{\alpha,\lambda}(\nu) \stackrel{D}{=} X(\beta, M_{\delta,\lambda}(\nu)) \quad (2.6.8)$$

where  $0 < \alpha < \beta \leq 1$  and  $\lambda, \nu > 0$ , and  $\delta = \alpha/\beta$ , and  $X(\beta, M_{\delta,\lambda}(\nu))$  is a discrete stable distribution with parameters  $\beta$  and  $M_{\delta,\lambda}(\nu)$ .

We can prove it in following way. Note that  $M_{1,\lambda}(\nu)$  denotes the Gamma distribution with density (2.6.4) and Laplace transform (2.6.5), and  $M_{\delta,\lambda}(\nu)$  is the positive continuous counterpart of the Linnik distribution with parameters  $\delta$  and  $\lambda$ , and Laplace transform (2.6.7). Let  $Y(\delta, M_{1,\lambda}(\nu))$  be the positive continuous counterpart of the discrete stable random variable  $X$ . By (2.6.2) and (2.6.6) we have

$$L_{\delta,\lambda}(\nu) \stackrel{D}{=} X(\alpha, M_{1,\lambda}(\nu)) \stackrel{D}{=} X(\beta, Y(\delta, M_{1,\lambda}(\nu))) \quad (2.6.9)$$

If  $k(\tau)$  denotes the Laplace transform of  $Y(\delta, M_{1,\lambda}(\nu))$ , then

$$k(\tau) = \int_0^\infty e^{-x\tau^\delta} f_{\lambda,\nu}(x) dx = (1 + \lambda\tau^\delta)^{-\nu}$$

Hence,

$$Y(\delta, M_{1,\lambda}(\nu)) \stackrel{D}{=} M_{\delta,\lambda}(\nu)$$

which, combined with (2.6.9), implies (2.6.8).

4. Let  $M_{1,\lambda}(1)$  be an exponentially distributed random variable, and  $V_{\alpha,1}(1)$  be a special case of the classical Linnik distribution with  $\nu = 1$ . It was first established by Kotz and Ostrovskii (1996) that  $V_{\alpha,1}(1)$  has density function

$$g(x; \alpha, 1) = \left( \frac{1}{\pi} \sin(\pi\alpha) \right) \frac{x^{\alpha-1}}{1 + x^{2\alpha} + 2x^\alpha \cos(\pi\alpha)}, \quad x > 0 \quad (2.6.10)$$

and note that Pillai and Jayakumar (1995) give a mixture representation for the discrete Mittag-Leffler distribution. The mixing random variable  $L_{\alpha,\lambda}(1)$  follows the Mittag-Leffler distribution and density function given by

$$f_\lambda(x; \alpha) = (\lambda^{1/\alpha})^{-1} g(x; \alpha, 1), \quad x > 0. \quad (2.6.11)$$

Then

$$L_{\alpha,\lambda}(1) \stackrel{D}{=} X(1, M_{1,\lambda}(1)V_{\alpha,1}(1)) \quad (2.6.12)$$

where  $\alpha \in (0, 1]$  and  $\lambda > 0$ ,  $\nu = 1$ .

$X(1, M_{1,\lambda}(1)V_{\alpha,1}(1))$  is a discrete stable distribution with parameters 1 and  $M_{1,\lambda}(1)V_{\alpha,1}(1)$ .

## 2.7. PROBABILITIES

Expanding the probability generating function  $P_X(z)$  in a power series (first the exponential function and then  $(1 - z)^{\alpha j}$ , we obtain the expression of the probability distribution of the discrete stable random variable.

$$P(X = k) = (-1)^k \sum_{j=0}^{\infty} \binom{\alpha j}{k} \frac{(-\lambda)^j}{j!}, \quad (2.7.1)$$

where  $k = 0, 1, 2, \dots$ , and  $\alpha \in (0, 1]$ .

Christoph and Schreiber (1998) represented these probabilities with finite sums as follows.

$$P(X = k) = (-1)^k e^{-\lambda} \sum_{m=0}^k \sum_{j=0}^m \binom{m}{j} \binom{\alpha j}{k} (-1)^j \frac{\lambda^m}{m!}, \quad (2.7.2)$$

where  $k = 0, 1, 2, \dots$  and  $\alpha \in (0, 1]$

Another representation of these probabilities is given by

$$P(X = 0) = e^{-\lambda} \quad (2.7.3)$$

$$P(X = k) = (-1)^k e^{-\lambda} \sum \prod_{m=1}^k \frac{(-\lambda)^{v_m}}{v_m!} \binom{\alpha}{m}^{v_m}, \quad (2.7.4)$$

where  $k = 1, 2, \dots$ , and  $\alpha \in (0, 1]$

The summation is carried over all non-negative solutions  $(v_1, v_2, \dots, v_k)$  of the equation  $v_1 + v_2 + \dots + v_k = k$ .

Christoph and Schreiber (1998) also present the following recursion formula.

Let  $X$  be a discrete stable random variable with exponent  $\alpha$  and parameter  $\lambda$ . Then

$$(k+1)P(X = k+1) = \lambda \sum_{m=1}^k P(X = k-m)(m+1)(-1)^m \binom{\alpha}{m+1} \quad (2.7.5)$$

for  $k = 0, 1, 2, \dots$  and  $P(X = 0) = e^{-\lambda}$ .

These forms of the probability distribution of the discrete stable distribution are mathematical expressions, difficult to use to estimate the parameters.

In Appendix A, we give several terms of the probability function by expanding the probability generating function; it shows that the expressions of the probability distribution of the discrete stable distribution are difficult to deal with in practice.

In figures 2.1 to 2.4, we use the expression(2.7.1) of the probability function of the discrete stable random variable and give different values of  $\alpha$  and  $\lambda$  to see the changes of the probability distribution of the discrete stable distribution with parameters  $\alpha$  and  $\lambda$ .



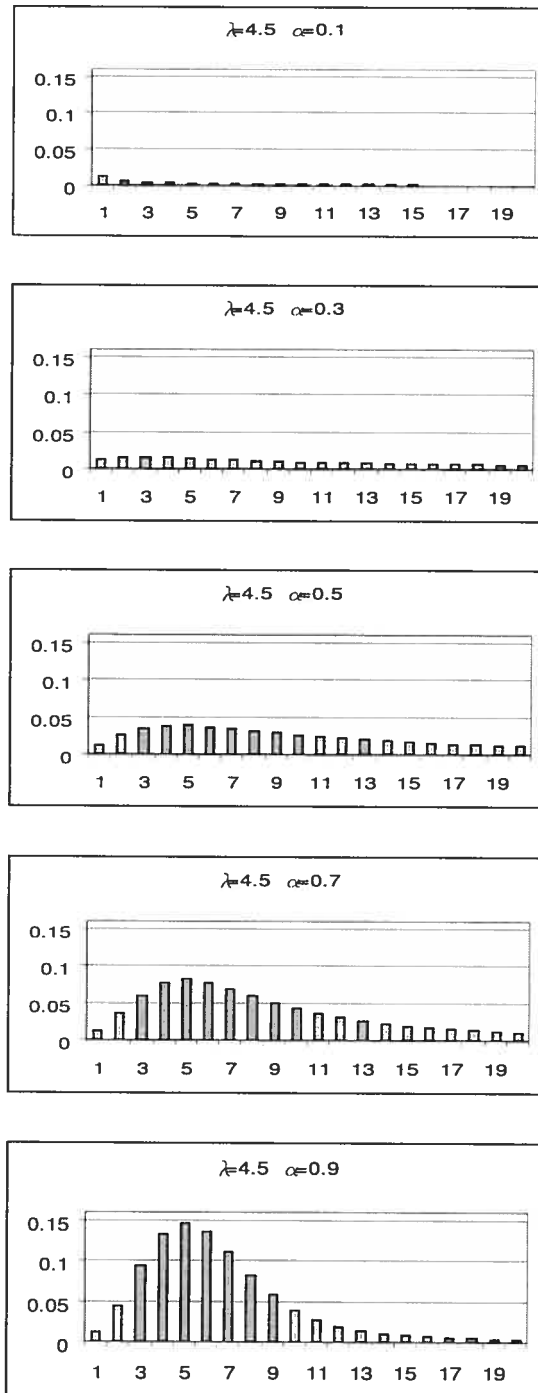


FIG. 2.1. Probabilities with  $\lambda = 4.5$  and different  $\alpha$ 's

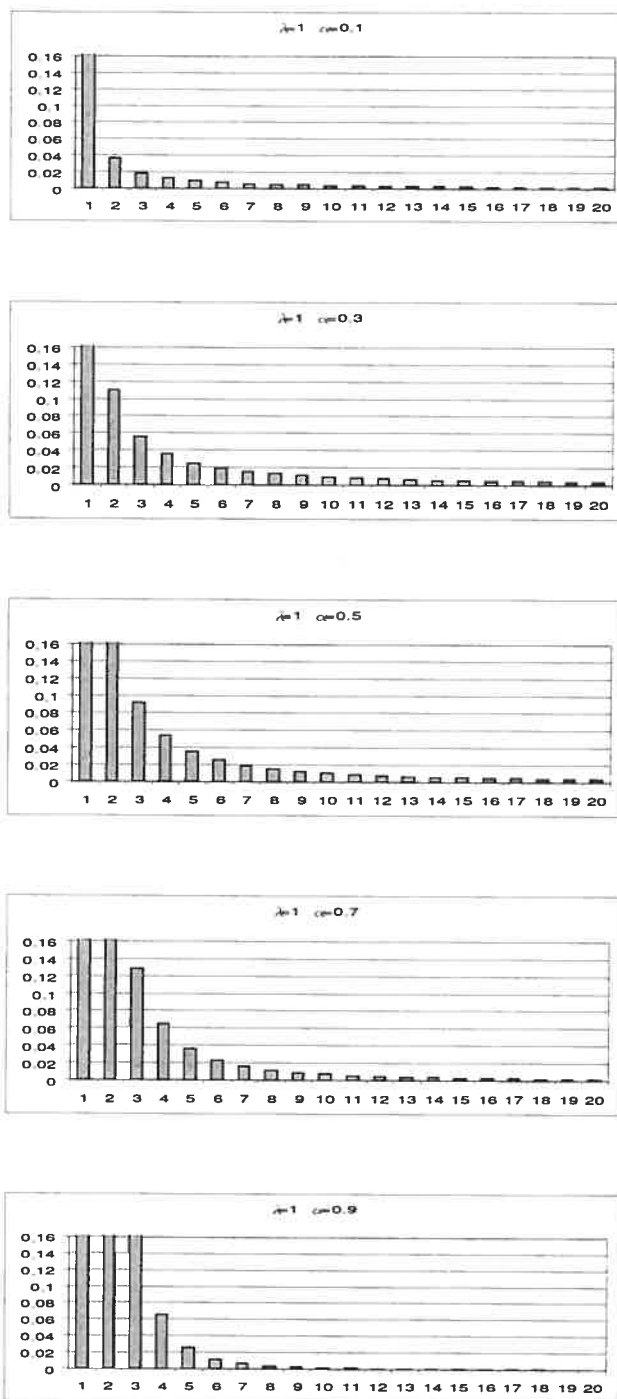


FIG. 2.2. Probabilities with  $\lambda = 1$  and different  $\alpha$ 's

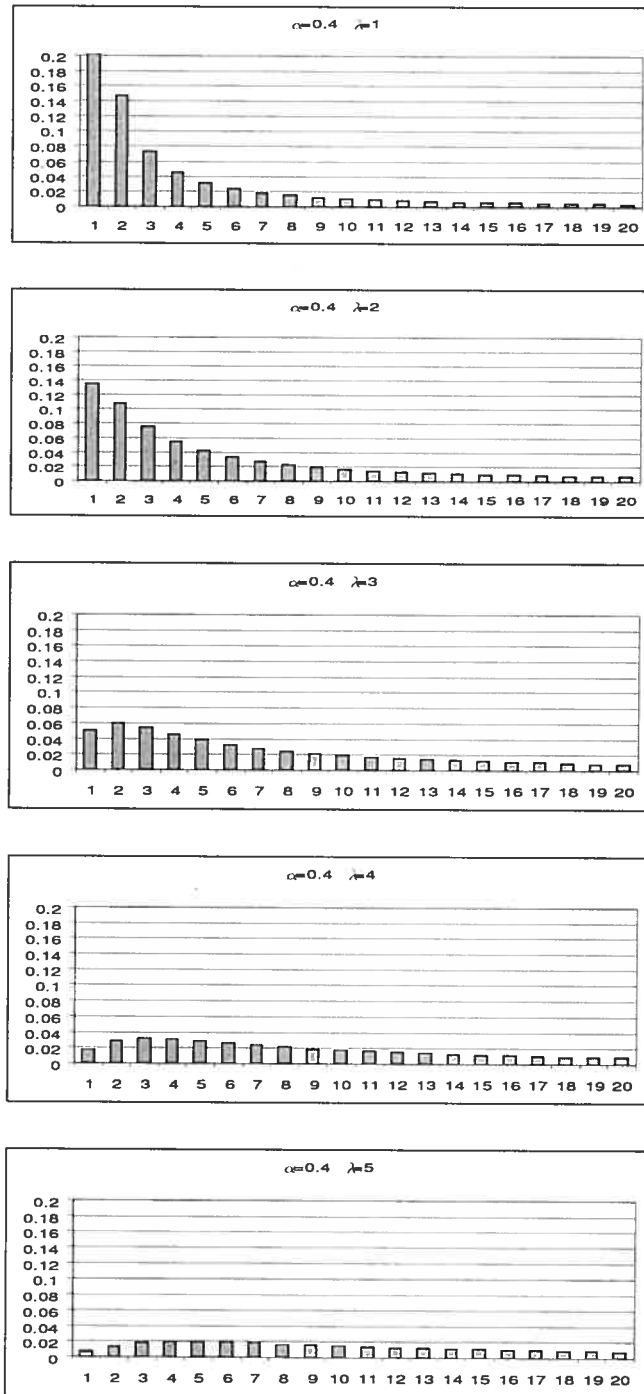


FIG. 2.3. Probabilities with  $\alpha = 0.4$  and different  $\lambda$ 's

## 2.8. MOMENT CHARACTERISTICS

For the moments  $E(X^r)$  with  $r$  an integer, we consider this problem in two cases.

### 2.8.1. Case $\alpha = 1$

With  $\alpha = 1$ , the discrete stable random variable  $X(\lambda, \alpha)$  follows a Poisson distribution, and all moments exist.

They are equal to

$$E(X^r) = \left. \frac{d^r P_X(z)}{dz^r} \right|_{z=1} = \left. \frac{d^r e^{-\lambda(1-z)}}{dz^r} \right|_{z=1}$$

where  $r=1,2,\dots$

### 2.8.2. Case $\alpha \in (0, 1)$

Steutel and van Harn (1979) mentioned that if we define a probability generating function  $P$  to be in the domain of discrete attraction of a stable probability generating function  $P_\gamma$ , and if there exist a  $\alpha_n$  such that

$$\lim_{n \rightarrow \infty} \{P(1 - \alpha_n + \alpha_n z)\}^n = P_\gamma(z),$$

then it follows that all distributions with finite first moment are attracted by the Poisson distribution by taking  $\alpha_n = 1/n$ . As for the discrete stable random variable with  $0 < \alpha < 1$ , it belongs to the domain of normal attraction of a (strictly) stable random variable  $S_\alpha^\lambda$ . That is the discrete stable random variable  $X$  with  $0 < \alpha < 1$  is in the domain of discrete attraction of  $P_\gamma$  if and only if it is in the domain of attraction of  $S_\alpha^\lambda$ . Hence  $E(X^r) < \infty$  only for  $0 \leq r < \alpha < 1$ . For  $r > \alpha, \alpha < 1$ ,  $E(X^r) = \infty$ ; which is consistent with the results of Christoph and Schreiber (1998).

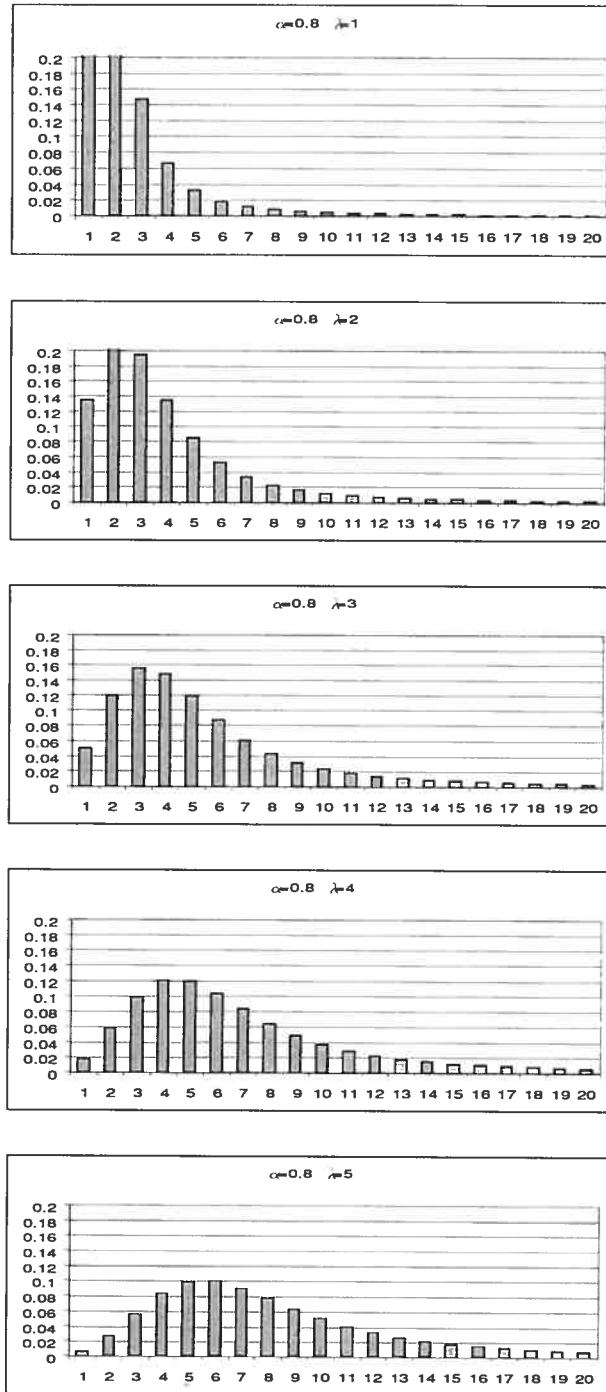


FIG. 2.4. Probabilities with  $\alpha = 0.8$  and different  $\lambda$ 's

## Chapter 3

---

### STATISTICAL REVIEW

#### 3.1. LINEAR REGRESSION

Standard parameter estimation methods such as maximum likelihood or the method of moments are not applicable to the discrete stable distribution since its density function cannot be written in a simple form, except for special cases, or its moments may not exist. We will use the probability generating function and some technique such as quadratic distance method to formulate our model and to estimate the parameters  $\alpha, \lambda$ . For this purpose, we review some theory first.

Recall the classical multiple linear normal regression model, (see Weisberg (1985) or Montgomery and Peck (1992)).

$$Y = X\theta + \epsilon \quad (3.1.1)$$

where the vectors  $Y, \epsilon, \theta$  and matrix  $X$  are defined

$$Y_{n \times 1} = (Y_1 \ Y_2 \ \dots \ Y_n)'$$

$$X_{n \times p} = \begin{pmatrix} 1 & X_{11} & X_{12} & \dots & X_{1p} \\ 1 & X_{21} & X_{22} & \dots & X_{2p} \\ \dots & \dots & \dots & \dots & \dots \\ 1 & X_{n1} & X_{n2} & \dots & X_{np} \end{pmatrix} \quad (3.1.3)$$

$$\epsilon_{n \times 1} = (\epsilon_1 \ \epsilon_2 \ \dots \ \epsilon_n)'$$

$$\theta_{p \times 1} = (\theta_1 \ \theta_2 \ \dots \ \theta_p)'$$
(3.1.5)

and where the  $\epsilon_i$ 's are independent errors distributed with a normal distribution  $N(0, \sigma^2)$ , so that  $E(\epsilon) = 0$ ,  $Var(\epsilon) = \sigma^2 I$ ,

$\sigma^2$  is an unknown parameter that needs to be estimated,

$Y_i$  is the response variable,

$X_{i,j}$  are explanatory variables (known and fixed),  $i = 1, 2, \dots, n$ ,  $j = 1, 2, \dots, p$ ,

$\theta$  is an unknown parameter vector of dimension  $p$  and needs to be estimated.

With the least squares method, we obtain an estimator which is also the maximum likelihood estimator of  $\theta$

$$\hat{\theta} = (X'X)^{-1}X'Y,$$
(3.1.6)

and we have

$$\begin{aligned} E(\hat{\theta}) &= E[(X'X)^{-1}X'Y] \\ &= E[(X'X)^{-1}X'(X\theta + \epsilon)] \\ &= E[(X'X)^{-1}X'X\theta + (X'X)^{-1}X'\epsilon] \\ &= \theta \end{aligned}$$

$$\begin{aligned} Var(\hat{\theta}) &= Var[(X'X)^{-1}X'Y] \\ &= [(X'X)^{-1}X'] [Var(Y)] [(X'X)^{-1}X']' \\ &= \sigma^2 [(X'X)^{-1}X'] [(X'X)^{-1}X']' \\ &= \sigma^2 (X'X)^{-1} (X'X) (X'X)^{-1} \\ &= \sigma^2 (X'X)^{-1} \end{aligned}$$

The estimator of  $\sigma^2$ ,

$$\hat{\sigma}^2 = \frac{SSE}{n-p} = \frac{[Y - X\hat{\theta}]'[Y - X\hat{\theta}]}{n-p}$$
(3.1.7)

is an unbiased estimator of  $\sigma^2$ , where  $SSE$  is the residual sum of squares

$$SSE = \sum_{i=1}^n (y_i - \hat{y}_i)^2.$$

Sometimes the assumption  $Var(\epsilon) = \sigma^2 I$  is unreasonable. We then need to modify the ordinary least squares procedure. Suppose that we know the value of a symmetric positive definite matrix  $\Sigma$ , such that the covariance matrix for the error vector  $\epsilon$  is given by  $Var(\epsilon) = \sigma^2 \Sigma$ , with  $\sigma^2 > 0$ , but not necessarily known. The model will be

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

where  $E(\epsilon) = 0$  and  $Var(\epsilon) = \sigma^2 \Sigma$ .  $\epsilon$  is an error vector distributed with a normal distribution  $N(0, \sigma^2 \Sigma)$ .

We can estimate  $\theta$  by minimizing the generalized quadratic distance

$$S(\theta) = (Y - X\theta)' \Sigma^{-1} (Y - X\theta). \quad (3.1.8)$$

Minimizing this expression, we get the estimator

$$\hat{\theta} = (X' \Sigma^{-1} X)^{-1} X' \Sigma^{-1} Y. \quad (3.1.9)$$

We can show that

$$E(\hat{\theta}) = \theta \quad (3.1.10)$$

and

$$Var(\hat{\theta}) = (X' \Sigma^{-1} X)^{-1}. \quad (3.1.11)$$

But in some circumstances, the covariance matrix  $\Sigma$  will be a function of the parameter  $\theta$  and needs to be estimated. Luong and Doray (2002) present examples where this happens and use the following procedure to estimate the parameter vector  $\theta$  and the variance-covariance matrix  $\Sigma(\theta)$ , where  $\Sigma(\theta)$  is a function of parameter  $\theta$ .

The algorithm is the following. By choosing  $\Sigma^{-1}(\theta) = I$ , the identity matrix, and by minimizing the generalized quadratic distance  $S(\theta) = (Y - X\theta)' \Sigma^{-1}(\theta) (Y -$



$X\theta$ ) we obtain  $\tilde{\theta}$ . Despite the fact that  $\tilde{\theta}$  is less efficient, it can be used to estimate  $\Sigma^{-1}(\theta)$  by letting  $\Sigma^{-1}(\theta) = \Sigma^{-1}(\tilde{\theta})$ . We then can use  $\Sigma^{-1}(\theta)$  to obtain the second iteration for  $\hat{\theta}$  and this procedure can be repeated with  $\Sigma^{-1}(\theta)$  re-estimated at each step; and  $\hat{\theta}$  is defined as the convergent vector value of the procedure.

Luong and Doray (2002) also studied the asymptotic properties of the quadratic distance estimator  $\hat{\theta}$ .

1.  $\hat{\theta}$  is a consistent estimator.
2.  $\hat{\theta}$  is asymptotically distributed as normal distribution with mean  $\theta$  and variance  $(X'\Sigma^{-1}X)^{-1}$ .
3. For certain parametric families,  $\hat{\theta}$  has high efficiency.
4. For protection against misspecification of the parametric family as regards to truncation, the quadratic distance estimator  $\hat{\theta}$  has clear advantages over the maximum likelihood estimator.

### 3.2. EMPIRICAL PROBABILITY GENERATING FUNCTION

Since we will estimate the parameters using the empirical probability generating function, we need first to consider its asymptotic behaviour.

Nakamura and Pérez-Abreu (1993) give the definition of the empirical probability generating function as follows.

Let  $X_1, X_2, \dots, X_n$  be a random sample from a discrete distribution over  $0, 1, 2, \dots$ , with corresponding probabilities  $p_k$ ,  $k = 0, 1, 2, \dots$ . The empirical probability generating function is defined as

$$P_n(z) = \frac{1}{n} \sum_{i=1}^n z^{X_i}, \quad (3.2.1)$$

for  $z \in (0, 1]$ . It is an estimator of the theoretical probability generating function

$$P_X(z) = E(z^X) = \sum_{k=0}^{\infty} p_k z^k, \quad |z| \leq 1. \quad (3.2.2)$$

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

$$P \left( \lim_{n \rightarrow \infty} \sup_{z \in (0,1]} |P_n(z) - P_X(z)| = 0 \right) = 1. \quad (3.2.3)$$

For the discrete stable random variable  $X$  and for a fixed  $z$ , call it  $z_0$ , we have

$$E(z_0^X) = P_X(z_0) = e^{-\lambda(1-z_0)^\alpha},$$

which exists for  $|z_0| \leq 1$ . Since  $|z_0| \leq 1$ , we have  $|z_0^2| \leq 1$ , so  $E(z_0^{2X}) = e^{-\lambda(1-z_0^2)^\alpha}$  also exists, where  $E(z_0^{2X}) \neq 0$ . By the central limit theorem, the standardized empirical probability generating function will converge to a standard normal distribution  $N(0, 1)$ , and the mean of the empirical probability generating function will be

$$E(z_0^X) = P_X(z_0) = e^{-\lambda(1-z_0)^\alpha},$$

the theoretical probability generating function.

We can use the empirical probability generating function and the minimum quadratic distance method to estimate the two parameters  $\lambda$  and  $\alpha$  of the discrete stable distribution.

### 3.3. MOMENTS OF MULTINOMIAL DISTRIBUTION

Johnson, Kotz and Balakrishnan (1997) introduce the definition and the properties of the multinomial distribution.

Consider a series of  $n$  independent trials, in each of which just one of  $k$  mutually exclusive events  $E_1, E_2, \dots, E_k$  can be observed, and in which the probability of occurrence of event  $E_j$  in any trial is equal to  $p_j$  (with, of course,  $p_1 + p_2 + \dots + p_k = 1$ ). Let  $f_1, f_2, \dots, f_k$  be the random variables denoting the numbers of occurrences of the events  $E_1, E_2, \dots, E_k$ , respectively, in these  $n$  trials, with  $f_1 + f_2 + \dots + f_k = n$ . Then the joint distribution of  $f_1, f_2, \dots, f_k$  is given by

$$P \left[ \bigcap_{i=1}^k (f_i = n_i) \right] = P(n_1, n_2, \dots, n_k) = \binom{n}{n_1, n_2, \dots, n_k} \prod_{i=1}^k p_i^{n_i}. \quad (3.3.1)$$

This is the probability function of a multinomial distribution with parameters  $(n; p_1, p_2, \dots, p_k)$ .

Note that if  $k = 2$ , the distribution reduces to a binomial distribution (for either  $f_1$  or  $f_2$ ). The marginal distribution of  $f_i$  is binomial with parameter  $(n, p_i)$ .

If we define the  $b^{th}$  descending factorial of  $a$  as  $a^{(b)} = a(a-1)(a-2)\dots(a-b+1)$ , with  $a^{(0)} = 1$ , the mixed factorial  $(r_1, r_2, \dots, r_k)$  moments of a multinomial distribution are given by

$$\begin{aligned} E(f_1^{(r_1)} f_2^{(r_2)} \dots f_k^{(r_k)}) &= \sum n_1^{(r_1)} n_2^{(r_2)} \dots n_k^{(r_k)} \binom{n}{n_1, n_2, \dots, n_k} \prod_{i=1}^k p_i^{n_i} \\ &= n^{(\sum_{i=1}^k r_i)} \prod_{i=1}^k p_i^{r_i} \end{aligned}$$

From the above equation we obtain, in particular,

$$E(f_i) = np_i \quad (3.3.2)$$

$$Var(f_i) = np_i(1 - p_i). \quad (3.3.3)$$

In terms of the relative frequency,

$$E(f_i/n) = p_i \quad (3.3.4)$$

$$Var(f_i/n) = \frac{1}{n} p_i(1 - p_i) \quad (3.3.5)$$

because  $f_i$  has a binomial  $(n, p_i)$  distribution. More generally, from the equation of mixed factorial moments, we also obtain

$$E(f_i f_j) = n(n-1)p_i p_j. \quad (3.3.6)$$

Thus, we have

$$\begin{aligned} Cov(f_i, f_j) &= E(f_i f_j) - E(f_i)E(f_j) \\ &= n(n-1)p_i p_j - n^2 p_i p_j \\ &= -n p_i p_j \end{aligned}$$

and

$$\begin{aligned} \text{Cov}(f_i/n, f_j/n) &= \frac{1}{n^2} \text{Cov}(f_i, f_j) \\ &= -\frac{1}{n} p_i p_j \end{aligned}$$

### 3.4. DELTA THEOREM

In our later study we need to estimate the variance of a function of an estimator by using the delta theorem. Rao (1973) presents the multivariate Delta theorem and Rice (1995) gives the univariate version of it.

It is of interest to estimate a nonlinear function  $g(\theta)$  of  $\theta$ . The variance of  $g(\hat{\theta})$  can be approximated from the variance of  $\hat{\theta}$  by expanding the function  $g(\theta)$  about its mean, usually with a one-step Taylor approximation, and then by taking the limiting distribution.

**Theorem 3.4.1.** (Multivariate delta theorem) *Let  $X_n$  be a  $k$ -dimensional random variables  $(X_{1n}, X_{2n}, \dots, X_{kn})$  and  $\mu$  be a vector  $(\mu_1, \mu_2, \dots, \mu_k)$ , such that the joint asymptotical distribution of  $\sqrt{n}(X_{1n} - \mu_1), \sqrt{n}(X_{2n} - \mu_2), \dots, \sqrt{n}(X_{kn} - \mu_k)$  is a  $k$ -variate normal with mean zero and variance-covariance matrix  $\Sigma = (\sigma_{ij})$ . Further let  $g$  be a function of  $k$ -variables ( $g : n \rightarrow k$ ) which is totally differentiable, that is, all  $\frac{\partial g}{\partial \mu_1}, \frac{\partial g}{\partial \mu_2}, \dots, \frac{\partial g}{\partial \mu_k}$  exist and not equal to zero. Then the asymptotical distribution of  $\sqrt{n}[g(X_{1n}, \dots, X_{kn}) - g(\mu_1, \dots, \mu_k)]$  is normal with mean zero and variance*

$$\text{Var} = \sum \sum \sigma_{ij} \frac{\partial g}{\partial \mu_i} \frac{\partial g}{\partial \mu_j} \quad (3.4.1)$$

provided  $\text{Var} \neq 0$ .

PROOF. Since  $g$  is a totally differentiable function, then

$$g(X_{1n}, \dots, X_{kn}) - g(\mu_1, \dots, \mu_k) = \sum_{i=1}^k (X_{in} - \mu_i) \left( \frac{\partial g}{\partial \mu_i} \right) + \epsilon_n \|X_n - \mu\|,$$

where  $\epsilon_n \rightarrow 0$  as  $X_{in} \rightarrow \mu_i$ . This implies that for any small  $\delta > 0$ ,  $|\epsilon_n| < \delta$  whenever  $|X_n - \mu| < \delta$ . Hence  $P(|\epsilon_n| < \delta) \rightarrow 1$  as  $n \rightarrow \infty$ . Since  $\delta$  is arbitrary,

$\epsilon_n \xrightarrow{p} 0$ . And since  $\sqrt{n} \|X_n - \mu\| = [\sum_{i=1}^k n(X_{in} - \mu_i)^2]^{1/2}$  has an asymptotic distribution

$$\left| \sqrt{n}[g(X_{1n}, \dots, X_{kn}) - g(\mu_1, \dots, \mu_k)] - \sqrt{n} \sum (X_{in} - \mu_i) \frac{\partial g}{\partial \mu_i} \right| \xrightarrow{p} 0.$$

But the asymptotic distribution of  $\sqrt{n} \sum (X_{in} - \mu_i) \frac{\partial g}{\partial \mu_i}$ , being a linear function of limiting normal variables is normal with zero mean and variance as given in (3.4.1). By the limiting distribution theorem (if  $Y_n \xrightarrow{L} Y$  and  $|X_n - Y_n| \xrightarrow{p} 0$ , then  $X_n \xrightarrow{L} Y$ ), the asymptotic distribution of  $\sqrt{n}[g(X_{1n}, \dots, X_{kn}) - g(\mu_1, \dots, \mu_k)]$  is the same as the asymptotic distribution of  $\sqrt{n} \sum (X_{in} - \mu_i) \frac{\partial g}{\partial \mu_i}$ .  $\square$

**Theorem 3.4.2.** (Univariate delta theorem) *Let  $X_n$  be a sequence of real-valued random variables such that for some  $\mu$  and  $\sigma$ ,  $\sqrt{n}(X_n - \mu)$  converges in distribution as  $n \rightarrow \infty$  to  $N(0, \sigma^2)$ . Let  $g(\cdot)$  be a real continuous differentiable function from  $\mathbf{R}$  to  $\mathbf{R}$  having a derivative  $g'(\mu)$  at  $\mu$ , and  $g'(\mu) \neq 0$ . Then  $\sqrt{n}[g(X_n) - g(\mu)]$  converges in distribution as  $n \rightarrow \infty$  to  $N(0, g'(\mu)^2 \sigma^2)$ .*

PROOF. We have  $X_n - \mu = o_p(1/\sqrt{n})$  as  $n \rightarrow \infty$ . Also by Taylor-series expansion of the function  $g(x)$  in a neighborhood of  $\mu$ ,  $|x - \mu| < \delta$ , we have

$$g(x) = g(\mu) + (x - \mu)g'(\mu) + o_p(|x - \mu|)$$

as  $x \rightarrow \mu$  by definition of derivative. Thus

$$g(X_n) = g(\mu) + g'(\mu)(X_n - \mu) + o_p(|X_n - \mu|),$$

so

$$\sqrt{n}[g(X_n) - g(\mu)] = g'(\mu)\sqrt{n}(X_n - \mu) + \sqrt{n}o_p(1/\sqrt{n}).$$

The last term is  $o_p(1)$ , so the conclusion follows.  $\square$

### 3.5. THE SINGULAR VALUE DECOMPOSITION (SVD) OF A MATRIX AND THE PSEUDO-INVERSE MATRIX

#### 3.5.1. The singular value decomposition

In our calculation example in chapter 5, we encounter the case where the variance-covariance matrix is nearly singular. We need to use the pseudo-inverse matrix to replace the variance-covariance matrix when the number of points of  $z$  of the probability function we take is large. So we first need to review some theory about pseudo-inverse. Golub and Van Loan (1989) and Watkins (2002) introduce the method of singular value decomposition as follows. Let  $A \in \mathbf{R}^{n \times m}$ , where  $A$  is a matrix and  $n$  and  $m$  are positive integers. We make no assumption about which of  $n$  or  $m$  is larger. The rank of  $A$  is the dimension of  $\text{range}(A)$ , and the range of  $A$  is defined by  $\text{range}(A) = \{y \in \mathbf{R}^n : y = Ax \text{ for some } x \in \mathbf{R}^m\}$ .

**Theorem 3.5.1.** (*SVD Theorem*) *If  $A \in \mathbf{R}^{n \times m}$  is a real nonzero matrix with rank  $r$ , then  $A$  can be expressed as the product*

$$A = U\Sigma V' \quad (3.5.1)$$

where  $U \in \mathbf{R}^{n \times n}$  and  $V \in \mathbf{R}^{m \times m}$  are orthogonal matrices, and  $\Sigma \in \mathbf{R}^{n \times m}$  is a nonsquare "diagonal" matrix as

$$\begin{pmatrix} \sigma_1 & 0 & \cdot & \cdot & \cdot & 0 \\ 0 & \sigma_2 & & & & \\ \cdot & & \sigma_3 & & & \\ \cdot & & & \dots & & \\ \cdot & & & & \sigma_r & \\ & & & & & 0 \\ 0 & & & & & \dots \end{pmatrix}$$

where  $\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_r > 0$  such that

$$U'AV = \text{diag}(\sigma_1, \dots, \sigma_r, 0, \dots, 0) \in \mathbf{R}^{n \times m}.$$

The coefficients  $\sigma_1, \sigma_2, \dots, \sigma_r$  are the singular values of  $A$  and they are uniquely determined. The columns of  $U$ ,  $u_1, u_2, \dots, u_n$  are orthonormal vectors called right singular vectors of  $A$ , and the columns of  $V$ ,  $v_1, v_2, \dots, v_m$  are called left singular vectors of  $A$ . The transpose of  $A$  has the SVD

$$A' = V\Sigma'U'.$$

It is easy to verify by comparing columns in the equations  $AV = \Sigma U$  and  $A'U = \Sigma'V$  that

$$Av_i = \sigma_i u_i \quad (3.5.2)$$

$$A'u_i = \sigma_i v_i \quad (3.5.3)$$

where  $i = 1, 2, \dots, \min\{n, m\}$ .

It is convenient to have the following notation for designating singular values:

$\sigma_i(A)$  = the  $i$ th largest singular value of  $A$ ,

$\sigma_{\max}(A)$  = the largest singular value of  $A$ ,

$\sigma_{\min}(A)$  = the smallest singular value of  $A$ .

The SVD reveals a great deal about the structure of a matrix, it is a powerful tool. The SVD may be the most important matrix decomposition of all, for both theoretical and computational purposes.

Moreover, if the associated right and left singular vectors of  $A$  are  $v_1, \dots, v_r$  and  $u_1, \dots, u_r$ , respectively, then, from equation (3.4.2) we have

$$A = \sum_{j=1}^r \sigma_j u_j v_j'. \quad (3.5.4)$$

Finally, from the definition of the 2-norm,  $\|A\|_2 = \sup_{x \neq 0} \frac{\|Ax\|_2}{\|x\|_2}$ , where  $x \in \mathbf{R}^m$ , and the definition of the Frobenius norm,  $\|A\|_F = \sqrt{\sum_{i=1}^n \sum_{j=1}^m |a_{ij}|^2}$ , where  $a_{ij}$  are elements of the matrix  $A$ , both the 2-norm and the Frobenius norm are neatly characterized in terms of the SVD as follows

$$\|A\|_F^2 = \sigma_1^2 + \sigma_2^2 + \dots + \sigma_r^2 \quad (3.5.5)$$

and

$$\|A\|_2 = \|A'\|_2 = \sigma_1. \quad (3.5.6)$$

### 3.5.2. Computing the SVD

One way to compute the SVD of  $A$  is simply to calculate the eigenvalues and eigenvectors of  $A'A$  and  $AA'$ .

**Example 3.5.1.** *Find the singular values and right and left singular vectors of the matrix  $A$  defined as*

$$A = \begin{pmatrix} 1 & 2 & 0 \\ 2 & 0 & 2 \end{pmatrix}$$

*Since  $A'A$  is  $3 \times 3$  and  $AA'$  is  $2 \times 2$ , it seems reasonable to work with the latter.*

*We easily compute*

$$AA' = \begin{pmatrix} 5 & 2 \\ 2 & 8 \end{pmatrix},$$

*so the characteristic polynomial is*

$$(\lambda - 5)(\lambda - 8) - 4 = \lambda^2 - 13\lambda + 36 = (\lambda - 9)(\lambda - 4),$$

*and the eigenvalues of  $AA'$  are  $\lambda_1 = 9$  and  $\lambda_2 = 4$ . The singular values of  $A$  are therefore*

$$\sigma_1 = 3$$

$$\sigma_2 = 2.$$

*The left singular vectors of  $A$  are eigenvectors of  $AA'$ . Solving  $(\lambda_1 I - AA')u = 0$ , we find that multiples of  $[1, 2]'$  are eigenvectors of  $AA'$  associated with  $\lambda_1$ . Then solving  $(\lambda_2 I - AA')u = 0$ , we find that the eigenvectors of  $AA'$  corresponding to  $\lambda_2$  are multiples of  $[2, -1]'$ . Since we want representatives with unit Euclidean norm, we take*

$$u_1 = \frac{1}{\sqrt{5}} \begin{pmatrix} 1 \\ 2 \end{pmatrix}$$



$$u_2 = \frac{1}{\sqrt{5}} \begin{pmatrix} 2 \\ -1 \end{pmatrix}.$$

These are the left singular vectors of  $A$ . Notice that they are orthogonal. We can find the right singular vectors  $v_1$ ,  $v_2$  and  $v_3$  by calculating the eigenvectors of  $A'A$ . However,  $v_1$  and  $v_2$  are more easily found by the formula  $v_i = \sigma_i^{-1} A' u_i$ ,  $i = 1, 2$ , thus

$$v_1 = \frac{1}{3\sqrt{5}} \begin{pmatrix} 5 \\ 2 \\ 4 \end{pmatrix}$$

$$v_2 = \frac{1}{\sqrt{5}} \begin{pmatrix} 0 \\ 2 \\ -1 \end{pmatrix}.$$

Notice that these vectors are orthonormal. The third vector must satisfy  $Av_3 = 0$ . Solving the equation  $Av = 0$  and normalizing the solution, we get

$$v_3 = \frac{1}{3} \begin{pmatrix} -2 \\ 1 \\ 2 \end{pmatrix}.$$

Now that we have the singular values and singular vectors of  $A$ , we can construct the SVD of  $A$  as  $A = U\Sigma V'$  with  $U \in \mathbf{R}^{2 \times 2}$  and  $V \in \mathbf{R}^{3 \times 3}$  orthogonal and  $\Sigma \in \mathbf{R}^{2 \times 3}$  and get

$$U = \begin{pmatrix} u_1 & u_2 \end{pmatrix} = \frac{1}{\sqrt{5}} \begin{pmatrix} 1 & 2 \\ 2 & -1 \end{pmatrix}$$

$$\Sigma = \begin{pmatrix} \sigma_1 & 0 & 0 \\ 0 & \sigma_2 & 0 \end{pmatrix} = \begin{pmatrix} 3 & 0 & 0 \\ 0 & 2 & 0 \end{pmatrix}$$

and

$$V = \begin{pmatrix} v_1 & v_2 & v_3 \end{pmatrix} = \frac{1}{3\sqrt{5}} \begin{pmatrix} 5 & 0 & -2\sqrt{5} \\ 2 & 6 & \sqrt{5} \\ 4 & -3 & 2\sqrt{5} \end{pmatrix}.$$

We can check that  $A = U\Sigma V'$ .

In MATHEMATICA we can use the command "SingularValueDecomposition" to compute the singular values or the singular value decomposition of a matrix.

### 3.5.3. Rank deficiency and numerical rank determination

One of the most valuable aspects of the SVD is that it enables us to deal sensibly with the concept of matrix rank. Rounding errors and fuzzy data make rank determination a nontrivial job. For example

$$A = \begin{pmatrix} \frac{1}{3} & \frac{1}{3} & \frac{2}{3} \\ \frac{2}{3} & \frac{2}{3} & \frac{4}{3} \\ \frac{1}{3} & \frac{2}{3} & \frac{3}{3} \\ \frac{2}{5} & \frac{2}{5} & \frac{4}{5} \\ \frac{3}{5} & \frac{1}{5} & \frac{4}{5} \end{pmatrix},$$

we note that the third column is the sum of the first two.  $A$  has rank 2. However, if we compute its rank with MATLAB, using IEEE standard double precision floating point arithmetic, we obtain

$$\sigma_1 = 2.5987$$

$$\sigma_2 = 0.3682$$

and

$$\sigma_3 = 8.66 \times 10^{-17}.$$

Since there are 3 nonzero singular values, we must conclude that the matrix has rank 3. But it is wrong! For this reason we introduce the notion of numerical rank.

We may consider the matrix that has  $k$  "large" singular values, the other being "tiny", has numerical rank  $k$ . For the purpose of determining which singular values are "tiny", we need to introduce a tolerance  $\epsilon$  that is roughly on the level of uncertainty in the data in the matrix.

Indeed, for some small  $\epsilon$  we may be interested in the  $\epsilon$ -rank of a matrix which we define by

$$\text{rank}(A, \epsilon) = \min_{\|A-B\|_2 \leq \epsilon} \text{rank}(B)$$

where  $\epsilon$  can be  $\epsilon = 10u\|A\|$ ,  $u$  is the unit roundoff error. Then, we say that  $A$  has numerical rank  $k$  if  $A$  has  $k$  singular values that are substantially larger than  $\epsilon$ , and all other singular values are smaller than  $\epsilon$ , that is

$$\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_k \gg \epsilon \geq \sigma_{k+1} \geq \dots$$

Thus, if  $A \in \mathbf{R}^{n \times n}$  has rank  $r$ , then we can expect  $n - r$  of the numerical singular values to be small.

In MATHEMATICA, there is a command "MatrixRank[m,Tolerance->t]" that gives the minimum rank with each element in a numerical matrix assumed to be correct only within tolerance  $t$ .

#### 3.5.4. The pseudo-inverse matrix

Watkins (2002) present the method to construct the pseudo-inverse matrix, also known as the Moore-Penrose generalized inverse. It is a generalization of the ordinary inverse. Note that if we define the matrix  $A^+ \in \mathbf{R}^{m \times n}$  by

$$A^+ = V\Sigma^+U'$$

where

$$\Sigma^+ = \begin{pmatrix} \frac{1}{\sigma_1} & 0 & \cdot & \cdot & \cdot & 0 \\ 0 & \frac{1}{\sigma_2} & & & & \\ \cdot & & \frac{1}{\sigma_3} & & & \\ \cdot & & & \dots & & \\ \cdot & & & & \frac{1}{\sigma_r} & \\ & & & & & 0 \\ 0 & & & & & \dots \end{pmatrix} \in \mathbf{R}^{m \times n},$$

$A^+$  is referred to as the pseudo-inverse of  $A$ . It is the unique minimal  $F$ -norm solution to the problem

$$\min_{X \in \mathbf{R}^{m \times n}} \|AX - T_n\|_F.$$

We see immediately by SVD

$$\text{rank}(A^+) = \text{rank}(A),$$

and  $u_1, u_2, \dots, u_n, v_1, v_2, \dots, v_m$  are left and right singular vectors of  $A^+$ , respectively, and  $\sigma_1^{-1}, \sigma_2^{-1}, \dots, \sigma_r^{-1}$  are the nonzero singular values.

The pseudo-inverse  $A^+$  satisfies the following four Moore-Penrose conditions:

$$(i) \quad AA^+A = A$$

$$(ii) \quad A^+AA^+ = A^+$$

$$(iii) \quad (AA^+)' = AA^+$$

$$(iv) \quad (A^+A)' = A^+A$$

Especially, if

$$m = n = \text{rank}(A),$$

then

$$A^+ = A^{-1}.$$

In MATHEMATICA, for numerical matrices, the command "PseudoInverse[m]" is based on the method of singular value decomposition.

## Chapter 4

---

# ESTIMATION AND HYPOTHESIS TESTING OF THE PARAMETERS

In this chapter, we will develop the methods to estimate the parameters based on minimizing the quadratic distance (see Doray and Luong (1997)) between the empirical and the theoretical probability generating functions of the discrete stable distribution.

### 4.1. THE MODEL

Let the theoretical and empirical probability generating functions be denoted by  $P_X(z)$ ,  $P_n(z)$ , respectively,

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

and

$$P_n(z) = \frac{1}{n} \sum_{i=1}^n z^{X_i}, \quad |z| \leq 1.$$

In order to define the linear regression model, we take the logarithmic transformation of  $P_X(z)$ ,

$$\ln P_X(z) = -\lambda(1 - z)^\alpha.$$

Let us define the function  $g(\cdot)$  as

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

where  $\beta = \ln \lambda$ . It is a linear function of the parameters  $\beta$  and  $\alpha$ . Now we can define a linear model in terms of parameters  $\beta$ ,  $\alpha$ , and an error term  $\epsilon$ , with the empirical probability generating function.

The model is the following:

$$g(P_n(z_s)) = g(P_X(z_s)) + \epsilon_s, \quad s = 1, 2, \dots, k \quad (4.1.1)$$

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

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

Since  $\ln [-\ln P_X(z_s)]$  is not a random variable, from equation 3.2.2 and the delta-theorem we can prove that, asymptotically,

$$\begin{aligned}
 E(\epsilon_s) &= E[g(P_n(z_s)) - g(P_X(z_s))] \\
 &= E\{\ln [-\ln P_n(z_s)]\} - \ln [-\ln P_X(z_s)] \\
 &= \ln [-\ln P_X(z_s)] - \ln [-\ln P_X(z_s)] \\
 &= 0
 \end{aligned}$$

and

$$E(\epsilon\epsilon') = \Sigma = \text{Var}(\epsilon).$$

Here, the variance-covariance matrix  $\Sigma$  is a function of the parameters  $\beta$  and  $\alpha$  and needs to be estimated. The formula to estimate  $\Sigma$  is presented in section (4.2).

Let

$$Y_{k \times 1} = \left( \ln(-\ln P_X(z_1)) \quad \ln(-\ln P_X(z_2)) \quad \dots \quad \ln(-\ln P_X(z_k)) \right)' \quad (4.1.2)$$

$$X_{k \times 2} = \begin{pmatrix} 1 & \ln(1 - z_1) \\ 1 & \ln(1 - z_2) \\ \dots & \dots \\ 1 & \ln(1 - z_k) \end{pmatrix} \quad (4.1.3)$$

$$\theta_{2 \times 1} = \begin{pmatrix} \beta & \alpha \end{pmatrix}' \quad (4.1.4)$$

$$\epsilon_{k \times 1} = \begin{pmatrix} \epsilon_1 & \epsilon_2 & \dots & \epsilon_k \end{pmatrix}'. \quad (4.1.5)$$

The model written in matrix form becomes

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

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

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

Explicitly,  $\hat{\theta}$  can be expressed as:

$$\hat{\theta} = (X' \Sigma^{-1} X)^{-1} X' \Sigma^{-1} Y. \quad (4.1.6)$$

From section 3.1 we have

$$E(\hat{\theta}) = \theta$$

and

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

## 4.2. THE VARIANCE-COVARIANCE MATRIX

To find the variance-covariance matrix  $\Sigma$  of the error term  $\epsilon$ , we need to use the theory in section 3.3 and section 3.4, the moments of a multinomial distribution and the delta theorem.

From the model (4.1.1), we have

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

Since  $\ln [-\ln P_X(z_s)]$  is not a random variable, we get

$$\Sigma = \text{Var}(\epsilon) = \text{Var}[\ln (-\ln P_n(z_s))],$$

where  $\Sigma$  is a function of the parameters  $\beta$  and  $\alpha$  and takes the following form

$$\Sigma = \begin{pmatrix} \text{Var}(\epsilon_1) & \text{Cov}(\epsilon_1, \epsilon_2) & \text{Cov}(\epsilon_1, \epsilon_3) & \dots & \text{Cov}(\epsilon_1, \epsilon_k) \\ \text{Cov}(\epsilon_2, \epsilon_1) & \text{Var}(\epsilon_2) & \text{Cov}(\epsilon_2, \epsilon_3) & \dots & \text{Cov}(\epsilon_2, \epsilon_k) \\ \text{Cov}(\epsilon_3, \epsilon_1) & \text{Cov}(\epsilon_3, \epsilon_2) & \text{Var}(\epsilon_3) & \dots & \text{Cov}(\epsilon_3, \epsilon_k) \\ \dots & \dots & \dots & \dots & \dots \\ \text{Cov}(\epsilon_k, \epsilon_1) & \text{Cov}(\epsilon_k, \epsilon_2) & \text{Cov}(\epsilon_k, \epsilon_3) & \dots & \text{Var}(\epsilon_k) \end{pmatrix}. \quad (4.2.1)$$

Now we need to define  $f_i$ , the frequency of the sample point. Let  $X_1, X_2, \dots, X_n$  be a random sample of  $X$ , we define

$$f_i = \sum_{j=1}^n 1_i(X_j), \quad j = 1, 2, \dots, n,$$

where  $1_i(X_j) = 1$ , if  $i = j$ ;  $1_i(X_j) = 0$ , if  $i \neq j$ .

Roussas (1997) presents a limit theorem which will be useful to us to find the estimator of the probability generating function.

**Theorem 4.2.1.** *Let  $X_n$ ,  $n \geq 1$ , and  $X$  be random variables, and let  $g : \mathbf{R} \rightarrow \mathbf{R}$  be bounded and continuous, so that  $g(X_n)$ ,  $n \geq 1$ , and  $g(X)$  are random variables. Suppose  $X_n \xrightarrow{P} X$ , as  $n \rightarrow \infty$  then  $g(X_n) \xrightarrow{P} g(X)$ , as  $n \rightarrow \infty$ .*

Since  $\hat{p}_i = f_i/n$ , we can prove that

$$\ln \hat{P}_X(z) \xrightarrow{P} \ln P_X(z),$$

and we can estimate  $P_X(z)$  by  $\hat{P}_X(z)$  and estimate  $\ln P_X(z)$  by  $\ln \hat{P}_X(z)$ .



In our calculations, we have

$$\hat{P}_X(z_s) = \sum_{i=1}^h \hat{p}_i z_s^i = \sum_{i=1}^n \frac{f_i}{n} z_s^i.$$

From section 3.2, we know that we can also use  $\hat{P}_X(z_s)$  to estimate  $P_n(z_s)$ .

Now suppose the largest value of the observations in the sample is  $h$ , replacing  $p_i$  by its estimator  $\hat{p}_i = f_i/n$  and by theorem (3.4.2)

$$\begin{aligned} \text{Var}(\epsilon_s) &= \text{Var}[\ln(-\ln P_n(z_s))] \\ &= \text{Var}[\ln(-\ln \hat{P}_X(z_s))] \\ &= \text{Var}[\ln(-\ln \sum_{i=1}^h \hat{p}_i z_s^i)] \\ &= \text{Var}[\ln(-\ln X)] \Big|_{X=\sum_{i=1}^h \hat{p}_i z_s^i} \\ &\simeq \left( \frac{-1/\mu}{-\ln \mu} \right)^2 \text{Var}\left[\sum_{i=1}^h \frac{f_i}{n} z_s^i\right] \Big|_{\mu=\sum_{i=1}^h p_i z_s^i} \\ &= \left( \frac{1}{\mu \ln \mu} \right)^2 \text{Var}\left[\sum_{i=1}^h \frac{f_i}{n} z_s^i\right] \\ &= \frac{1}{[(\sum_{i=1}^h \frac{f_i}{n} z_s^i) \ln(\sum_{i=1}^n \frac{f_i}{n} z_s^i)]^2} \text{Var}\left[\sum_{i=1}^n \frac{f_i}{n} z_s^i\right]. \end{aligned}$$

Now, we only consider the term  $\text{Var}[\sum_{i=1}^n \frac{f_i}{n} z_s^i]$  and get

$$\begin{aligned} \text{Var}\left[\sum_{i=1}^n \frac{f_i}{n} z_s^i\right] &= \sum_{i=1}^h (z_s^i)^2 \text{Var}\left(\frac{f_i}{n}\right) + 2 \sum_{i < j} z_s^i z_s^j \text{Cov}\left(\frac{f_i}{n}, \frac{f_j}{n}\right) \\ &= \sum_{i=1}^h (z_s^i)^2 \frac{1}{n} p_i (1 - p_i) + 2 \sum_{i < j} z_s^i z_s^j \left(-\frac{1}{n} p_i p_j\right). \end{aligned}$$

The variance of  $\epsilon_s$  is given by

$$\text{Var}(\epsilon_s) = \frac{\sum_{i=1}^h (z_s^i)^2 \frac{1}{n} p_i (1 - p_i) + 2 \sum_{i < j} z_s^i z_s^j \left(-\frac{1}{n} p_i p_j\right)}{[(\sum_{i=1}^h \frac{f_i}{n} z_s^i) \ln(\sum_{i=1}^n \frac{f_i}{n} z_s^i)]^2} \quad (4.2.2)$$

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

Similarly, we can also find the covariances of the error terms as follows:

$$\begin{aligned}
Cov(\epsilon_r, \epsilon_s) &= Cov\left[\ln\left(-\ln \sum_{i=1}^h \hat{p}_i z_r^i\right), \ln\left(-\ln \sum_{j=1}^h \hat{p}_j z_s^j\right)\right] \\
&= Cov\left[\ln(-\ln X), \ln(-\ln Y)\right] \Big|_{X=\sum_{i=1}^h \hat{p}_i z_r^i, \quad Y=\sum_{j=1}^h \hat{p}_j z_s^j} \\
&= \left[ \frac{d \ln(-\ln \mu_1)}{d \mu_1} \Big|_{\mu_1=\sum_{i=1}^h p_i z_r^i} \right] \left[ \frac{d \ln(-\ln \mu_2)}{d \mu_2} \Big|_{\mu_2=\sum_{j=1}^h p_j z_s^j} \right] \\
&\quad Cov\left(\sum_{i=1}^h \frac{f_i}{n} z_r^i, \sum_{j=1}^h \frac{f_j}{n} z_s^j\right) \\
&= \frac{Cov\left(\sum_{i=1}^h \frac{f_i}{n} z_r^i, \sum_{j=1}^h \frac{f_j}{n} z_s^j\right)}{\left[ \left(\sum_{i=1}^h \frac{f_i}{n} z_r^i\right) \ln\left(\sum_{i=1}^h \frac{f_i}{n} z_r^i\right) \right] \left[ \left(\sum_{j=1}^h \frac{f_j}{n} z_s^j\right) \ln\left(\sum_{j=1}^h \frac{f_j}{n} z_s^j\right) \right]} \\
&= \frac{\sum_{i=1}^h \left[(z_r z_s)^i \frac{1}{n} p_i (1 - p_i)\right] + \sum_{i < j} \left[(z_r^i z_s^j + z_r^j z_s^i) \left(-\frac{1}{n} p_i p_j\right)\right]}{\left[ \left(\sum_{i=1}^h \frac{f_i}{n} z_r^i\right) \ln\left(\sum_{i=1}^h \frac{f_i}{n} z_r^i\right) \right] \left[ \left(\sum_{j=1}^h \frac{f_j}{n} z_s^j\right) \ln\left(\sum_{j=1}^h \frac{f_j}{n} z_s^j\right) \right]} \\
&= Cov(\epsilon_s, \epsilon_r).
\end{aligned}$$

We have the terms to evaluate all the elements of the variance-covariance matrix  $\Sigma$ .

Since in the expression of the probability generating function

$$P_X(z) = \sum_{i=1}^{\infty} p_i z^i,$$

all  $p_i$ 's are correlated, the variance-covariance matrix must be a full matrix.

#### 4.3. THE INITIAL VALUES OF THE PARAMETERS

In order to estimate the parameter vector  $\theta$ , we need to determine the initial value of the parameter vector. We can use either of the following two methods to find the initial value of  $\theta$ , denoted  $\hat{\theta}_0 = (\hat{\beta}_0, \hat{\alpha}_0)'$ , where  $\hat{\beta}_0 = \ln \hat{\lambda}_0$ .

**Method 1.** By replacing  $\Sigma$  by the identity matrix, we obtain a consistent estimator of the parameter vector  $\theta$ ,

$$\hat{\theta}_0 = (X'X)^{-1}X'Y. \quad (4.3.1)$$

However, it is not a fully efficient estimator of  $\theta$ .

**Method 2.** Using  $f_i/n$  to estimate  $p_i$  in the probability generating function, we get

$$\hat{p}_i = f_i/n, \quad (4.3.2)$$

$$\begin{aligned} \hat{P}_X(z) &= E(z^X) \\ &= \sum_{i=0}^{\infty} \hat{p}_i z^i \\ &= \sum_{i=0}^{\infty} \frac{f_i}{n} z^i. \end{aligned}$$

For initial values, we take the logarithmic transformation of  $\hat{P}_X(z)$ , and use  $\ln \hat{P}_X(z)$  to estimate  $\ln P_X(z)$ , we get

$$\ln \hat{P}_X(z) = -\lambda(1-z)^\alpha,$$

or

$$\ln \left( \sum_{i=0}^{\infty} p_i z^i \right) = -\lambda(1-z)^\alpha.$$

By Rémillard and Theodorescu (1991), using only two points  $z_1$  and  $z_2$ , we have

$$\ln \left( \sum_{i=0}^{\infty} p_i z_1^i \right) = -\lambda(1-z_1)^\alpha \quad (4.3.3)$$

and

$$\ln \left( \sum_{i=0}^{\infty} p_i z_2^i \right) = -\lambda(1-z_2)^\alpha. \quad (4.3.4)$$

Dividing (4.3.3) by (4.3.4) and replacing  $p_i$  by its estimator  $\hat{p}_i = f_i/n$ , we obtain

$$\frac{\ln \left( \sum_{i=0}^{\infty} \frac{f_i}{n} z_1^i \right)}{\ln \left( \sum_{i=0}^{\infty} \frac{f_i}{n} z_2^i \right)} = \left( \frac{1-z_1}{1-z_2} \right)^\alpha.$$

Solving, we get

$$\hat{\alpha}_0 = \frac{\ln \left( \frac{\ln \left( \sum_{i=0}^{\infty} \frac{f_i}{n} z_1^i \right)}{\ln \left( \sum_{i=0}^{\infty} \frac{f_i}{n} z_2^i \right)} \right)}{\ln \left( \frac{1-z_1}{1-z_2} \right)}, \quad (4.3.5)$$

then from (4.3.3),

$$\hat{\lambda}_0 = -\frac{\ln(\sum_{i=0}^{\infty} p_i z_1^i)}{(1 - z_1)^{\alpha_0}}. \quad (4.3.6)$$

In order to get more precise initial value of the two parameters, we should take the two values of  $z$  far apart, for example,  $z_1 = 0.1$  and  $z_2 = 0.9$ .

#### 4.4. THE ALGORITHM

1. Calculate the initial value of  $\theta$ , denoted by  $\hat{\theta}_0 = (\hat{\beta}_0, \hat{\alpha}_0)$ , using either of the methods in section (4.3).
2. By the series expansion of the probability generating function in terms of  $z$

$$P_X(z) = \exp[-\lambda_0(1 - z)^{\alpha_0}] = \sum_{i=1}^{\infty} p_i z^i$$

to calculate  $p'_i$ s using  $\theta_0$  (see appendix A).

3. Estimate the variance-covariance matrix  $\hat{\Sigma}_1$  using the method provided in section (4.2). It is function of the  $p'_i$ s.
4. Use our model to obtain the new values of  $\hat{\lambda}_1$  and  $\hat{\alpha}_1$  by the equation (4.1.6).
5. For iteration, redo the steps 2, 3 and 4 to estimate new  $p'_i$ s,  $\hat{\Sigma}_j$  and  $\hat{\theta}_j = (\hat{\beta}_j, \hat{\alpha}_j)$ , where  $j = 2, 3, \dots$ , up to the desired accuracy.

#### 4.5. INFERENCES CONCERNING THE VECTOR $\theta$

Neter, Wasserman and Kutner (1989) describes the method for hypothesis testing on the estimators. When  $n \rightarrow \infty$ , the sampling distribution of the vector  $\hat{\theta} = (\hat{\beta}, \hat{\alpha})'$  will follow an asymptotically normal distribution

$$\sqrt{n}(\hat{\theta} - \theta_0) \xrightarrow{L} ASN(0, (X' \Sigma_{\theta_0}^{-1} X)^{-1}) \quad (4.5.1)$$

and, separately

$$\sqrt{n}(\hat{\beta} - \beta_0) \xrightarrow{L} ASN(0, (Var(\hat{\beta}))) \quad (4.5.2)$$

$$\sqrt{n}(\hat{\alpha} - \alpha_0) \xrightarrow{L} ASN(0, (Var(\hat{\alpha}))), \quad (4.5.3)$$

where  $\theta_0$  is the true value of the vector  $\theta$ , and  $\beta_0, \alpha_0$  are the true values of the parameters  $\beta$  and  $\alpha$ , respectively, and  $Var(\hat{\beta}), Var(\hat{\alpha})$  are the diagonal elements of the variance-covariance matrix  $(X'\Sigma_{\theta_0}^{-1}X)^{-1}$ .

#### 4.5.1. Sampling distribution of the standardized statistic

Since  $\hat{\beta}$  and  $\hat{\alpha}$  are asymptotically normally distributed, we know that the standardized statistic  $(\hat{\beta} - \beta)/\sqrt{Var(\hat{\beta})}$ , and  $(\hat{\alpha} - \alpha)/\sqrt{Var(\hat{\alpha})}$  are standard normal variables. Ordinarily, we need to estimate  $(\hat{\beta} - \beta)/\sqrt{Var(\hat{\beta})}$ , and  $(\hat{\alpha} - \alpha)/\sqrt{Var(\hat{\alpha})}$  by  $(\hat{\beta} - \beta)/\sqrt{\widehat{Var}(\hat{\beta})}$ , and  $(\hat{\alpha} - \alpha)/\sqrt{\widehat{Var}(\hat{\alpha})}$ , and hence are interested in the distribution of the statistics  $(\hat{\beta} - \beta)/\sqrt{\widehat{Var}(\hat{\beta})}$  and  $(\hat{\alpha} - \alpha)/\sqrt{\widehat{Var}(\hat{\alpha})}$ . When a statistic is standardized but the denominator is an estimated standard deviation rather than the true standard deviation, it is called a studentized statistic. An important theorem in statistics states the following about the studentized statistic (see Montgomery and Peck (1992)):

$$\begin{aligned} \left( (\hat{\beta} - \beta)/\sqrt{\widehat{Var}(\hat{\beta})} \right) &\sim t(n-2) \text{ and} \\ \left( (\hat{\alpha} - \alpha)/\sqrt{\widehat{Var}(\hat{\alpha})} \right) &\sim t(n-2), \end{aligned}$$

where  $n$  is the number of the selected points of  $z$ , i.e.  $n = s$ . The reason for the degrees of freedom is that two parameters ( $\beta$  and  $\alpha$ ) need to be estimated for the model, hence, two degrees of freedom are lost.

This result places us in a position to make inferences concerning  $\beta$  and  $\alpha$ .

#### 4.5.2. Confidence intervals for $\beta$ and $\alpha$

Since  $(\hat{\beta} - \beta)/\sqrt{\widehat{Var}(\hat{\beta})}$  and  $(\hat{\alpha} - \alpha)/\sqrt{\widehat{Var}(\hat{\alpha})}$  follow  $t$ -distributions, we can make the following probability statement with confidence  $1 - \alpha$ ,

$$P \left\{ t_{\alpha/2}(n-2) \leq \frac{\hat{\beta} - \beta}{\sqrt{\widehat{Var}(\hat{\beta})}} \leq t_{(1-\alpha/2)}(n-2) \right\} = 1 - \alpha \quad (4.5.4)$$

and

$$P \left\{ t_{\alpha/2}(n-2) \leq \frac{\hat{\alpha} - \alpha}{\sqrt{\widehat{Var}(\hat{\alpha})}} \leq t_{(1-\alpha/2)}(n-2) \right\} = 1 - \alpha. \quad (4.5.5)$$

Here,  $t_{\alpha/2}(n-2)$  denotes the  $(\alpha/2)100$  percentile of the  $t$ -distribution with  $n-2$  degrees of freedom.

Because of the symmetry of the  $t$ -distribution around its mean 0, it follows that

$$t_{\alpha/2}(n-2) = -t_{(1-\alpha/2)}(n-2). \quad (4.5.6)$$

Rearranging the probability inequalities, we obtain:

$$P \left\{ \hat{\beta} - t_{(1-\alpha/2)}(n-2) \sqrt{\widehat{Var}(\hat{\beta})} \leq \beta \leq \hat{\beta} + t_{(1-\alpha/2)}(n-2) \sqrt{\widehat{Var}(\hat{\beta})} \right\} \\ = 1 - \alpha$$

and

$$P \left\{ \hat{\alpha} - t_{(1-\alpha/2)}(n-2) \sqrt{\widehat{Var}(\hat{\alpha})} \leq \alpha \leq \hat{\alpha} + t_{(1-\alpha/2)}(n-2) \sqrt{\widehat{Var}(\hat{\alpha})} \right\} \\ = 1 - \alpha$$

Since the above equations hold for all possible values of  $\beta$  and  $\alpha$ , the  $1 - \alpha$  (this  $\alpha$  is the significance level) confidence intervals for  $\beta$  and  $\alpha$  are

$$\hat{\beta} \pm t_{(1-\alpha/2)}(n-2) \sqrt{\widehat{Var}(\hat{\beta})} \quad (4.5.7)$$

$$\hat{\alpha} \pm t_{(1-\alpha/2)}(n-2) \sqrt{\widehat{Var}(\hat{\alpha})}. \quad (4.5.8)$$

#### 4.5.3. Tests concerning $\alpha$

Neter, Wasserman and Kutner (1989) have shown that since  $\frac{\hat{\alpha} - \alpha}{\sqrt{\widehat{Var}(\hat{\alpha})}}$  is distributed as a  $t$ -distribution with  $n-2$  degrees of freedom, tests concerning  $\alpha$  can be set up in the ordinary fashion using the  $t$ -distribution.

##### 1. Two-Sided Test

To test

$$H_0 : \alpha = \alpha^*$$

$$\text{vs } H_a : \alpha \neq \alpha^*,$$

an explicit test of the alternatives  $H_a$  is based on the test statistic

$$t^* = \frac{\hat{\alpha} - \alpha^*}{\sqrt{\widehat{Var}(\hat{\alpha})}}. \quad (4.5.9)$$

The decision rule with this test statistic when controlling the significance level at  $\alpha$  is

If  $|t^*| \leq t_{(1-\alpha/2)}(n-2)$ , accept  $H_0$ , i.e.  $\alpha = \alpha^*$ ,

If  $|t^*| > t_{(1-\alpha/2)}(n-2)$ , reject  $H_0$ , i.e.  $\alpha \neq \alpha^*$ .

## 2. One-Sided Test

Suppose instead we had wished to test whether or not the parameter  $\alpha$  is greater than some specified value  $\alpha^*$ , controlling the significance level at  $\alpha$ . The alternative then would be:

$$H_0 : \alpha \leq \alpha^*$$

$$\text{vs } H_a : \alpha > \alpha^*.$$

The test statistic would still be

$$t^* = \frac{\hat{\alpha} - \alpha^*}{\sqrt{\widehat{Var}(\hat{\alpha})}},$$

and the decision rule based on the test statistic would be:

If  $|t^*| \leq t_{1-\alpha}(n-2)$ , accept  $H_0$ , i.e.  $\alpha = \alpha^*$ ,

If  $|t^*| > t_{1-\alpha}(n-2)$ , reject  $H_0$ , i.e.  $\alpha \neq \alpha^*$ .

### 4.5.4. Tests concerning $\lambda$

In section (4.1), we defined  $\beta$  as the logarithmic transformation of the parameter  $\lambda$ , so we have

$$\lambda = e^\beta.$$

To determine the sampling distribution of  $\frac{\hat{\lambda} - \lambda}{\sqrt{\widehat{Var}(\hat{\lambda})}}$ , we need first to calculate the estimated variance of  $\hat{\lambda}$  using the delta-theorem,

$$\widehat{Var}(\hat{\lambda}) = e^{2\beta} \widehat{Var}(\hat{\beta}). \quad (4.5.10)$$

By the delta-theorem, we know that  $\lambda$  is asymptotically normally distributed, then  $\frac{\hat{\lambda} - \lambda}{\sqrt{\widehat{Var}(\hat{\lambda})}}$  will be  $t$ -distributed,  $\frac{\hat{\lambda} - \lambda}{\sqrt{\widehat{Var}(\hat{\lambda})}} \sim t(n-2)$ .

Tests concerning  $\lambda$  can be set up in the following fashion using the  $t$ -distribution.

## Two-Sided Test

To test

$$H_0 : \lambda = \lambda^*$$

$$\text{vs } H_a : \lambda \neq \lambda^*,$$

an explicit test is based on the test statistic

$$t^* = \frac{\hat{\lambda} - \lambda^*}{\sqrt{\widehat{Var}(\hat{\lambda})}}. \quad (4.5.11)$$

The decision rule with this test statistic when controlling the significance level at  $\alpha$  is

If  $|t^*| \leq t_{(1-\alpha/2)}(n-2)$ , accept  $H_0$ , i.e.  $\lambda = \lambda^*$ ,

If  $|t^*| > t_{(1-\alpha/2)}(n-2)$ , reject  $H_0$ , i.e.  $\lambda \neq \lambda^*$ .

The one-sided test is easily defined.



# Chapter 5

---

## NUMERICAL EXAMPLES

In this chapter, we will use the method of Kanter (1975) (see section 2.3) to generate samples of discrete stable random variables and use the parameter estimation method provided in chapter 4 to estimate the two parameters of the distribution and test hypothesis on the parameters.

### 5.1. EFFECT OF THE NUMBER OF POINTS TAKEN

Considering the probability generating function of the discrete stable distribution

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

we select parameters  $\lambda = 1$  and  $\alpha = 0.9$  to generate 5000 discrete stable random variables, since when  $\alpha$  close to 1, the distribution is much like a Poisson distribution with parameter  $\lambda$ . With this set of data, we analyze the effect of the selected number of points of  $z$  that we should take in the process of the estimation. We also consider the situations in which  $z$  takes negative values with 18 points, 10 points, 4 points and 2 points.

We consider the following cases:

1.  $z$  takes 19 points without negative values

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

2.  $z$  takes 18 points with negative values of  $z$

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

3.  $z$  takes 10 points with negative values

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

4.  $z$  takes 9 points without negative values

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

5.  $z$  takes 4 points with negative values

$$z = \{-0.9, -0.3, 0.3, 0.9\}$$

6.  $z$  takes two points

$$z = \{-0.5, 0.5\}$$

7.  $z$  takes two points

$$z = \{-0.9, 0.9\}$$

When  $z$  takes 19 values, at the second iteration, the inverse of the variance-covariance matrix  $\Sigma$  does not exist since the inverse matrix is singular ( $\text{rank}(\Sigma)=12$ ). The reason is that the selected points of  $z$  are too close. The same thing happens when  $z$  takes 9 values, where  $\text{rank}(\Sigma)=8$ , and when  $z$  takes 18 values, where the  $\text{rank}(\Sigma)=15$ . In these situations we use the pseudo-inverse of  $\Sigma$  instead of the inverse of  $\Sigma$  and get the results. Those results have been marked with \* in table 5.1.

When the number of selected values of  $z$  is 18, the variance-covariance matrix  $\Sigma$  is a  $18 \times 18$  matrix and the variance-covariance matrix of the parameter vector  $\theta$ , denoted by  $\text{Var}(\hat{\theta})$  is a  $2 \times 2$  matrix. If we generate 5000 random variables,

then  $\hat{\Sigma}$  and  $Var(\hat{\theta})$  are given by

$$\hat{\Sigma} = \begin{pmatrix} 0.00170892 & 0.00149471 & 0.00131009 & 0.00115104 & \dots & 0.000198651 \\ 0.00149471 & 0.00132283 & 0.00117182 & 0.00103977 & \dots & 0.000201909 \\ 0.00131009 & 0.00117182 & 0.0010484 & 0.000939041 & \dots & 0.000203492 \\ 0.00115104 & 0.00103977 & 0.000939041 & 0.000848695 & \dots & 0.000204558 \\ 0.00101372 & 0.000924303 & 0.000842276 & 0.000767863 & \dots & 0.000205569 \\ \dots & \dots & \dots & \dots & \dots & \dots \\ 0.000225341 & 0.000226796 & 0.000226852 & 0.000226469 & \dots & 0.000286112 \\ 0.000198651 & 0.000201909 & 0.000203492 & 0.000204558 & \dots & 0.000335412 \end{pmatrix},$$

and

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

When the number of selected values of  $z$  is 10, the variance-covariance matrix  $\Sigma$  is a  $10 \times 10$  matrix and the variance-covariance matrix of the parameter vector  $\theta$ , denoted by  $Var(\hat{\theta})$  is a  $2 \times 2$  matrix. If we generate 5000 random variables, then  $\hat{\Sigma}$  and  $Var(\hat{\theta})$  are given by

$$\hat{\Sigma} = \begin{pmatrix} 0.00170892 & 0.00131009 & 0.00101372 & 0.000791541 & \dots & 0.000198651 \\ 0.00131009 & 0.0010484 & 0.000842276 & 0.000680596 & \dots & 0.000203492 \\ 0.00101372 & 0.000842276 & 0.000700581 & 0.000584763 & \dots & 0.000205569 \\ 0.000791541 & 0.000680596 & 0.000584763 & 0.00050339 & \dots & 0.000208052 \\ 0.000623099 & 0.000553014 & 0.000489813 & 0.000434216 & \dots & 0.000211567 \\ \dots & \dots & \dots & \dots & \dots & \dots \\ 0.000252964 & 0.000250874 & 0.000246846 & 0.000243059 & \dots & 0.000261848 \\ 0.000198651 & 0.000203492 & 0.000205569 & 0.000208052 & \dots & 0.000335412 \end{pmatrix},$$

and

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

TABLE 5.1. Effect of the numbers of the points of  $z$ 

points of $z$	initial value	first iteration	second iteration	relative error
19 points	$\alpha_0 = 0.913933$	$\alpha = 0.917313$	$\alpha = 0.916731^*$	1.859 %
	$\lambda_0 = 0.994707$	$\lambda = 0.998783$	$\lambda = 0.997756^*$	-0.224 %
18 points	$\alpha_0 = 0.913933$	$\alpha = 0.919324$	$\alpha = 0.91825^*$	2.028 %
	$\lambda_0 = 0.994707$	$\lambda = 0.997456$	$\lambda = 0.99742^*$	-0.258 %
10 points	$\alpha_0 = 0.913933$	$\alpha = 0.916447$	$\alpha = 0.916447$	1.827 %
	$\lambda_0 = 0.994707$	$\lambda = 0.996336$	$\lambda = 0.996336$	-0.366 %
9 points	$\alpha_0 = 0.913933$	$\alpha = 0.916731$	$\alpha = 0.916731^*$	1.859 %
	$\lambda_0 = 0.994707$	$\lambda = 0.997725$	$\lambda = 0.997725^*$	-0.227 %
4 points	$\alpha_0 = 0.913933$	$\alpha = 0.914694$	$\alpha = 0.914694$	1.633 %
	$\lambda_0 = 0.994707$	$\lambda = 0.996986$	$\lambda = 0.996986$	-0.301 %
{ -0.5, 0.5 }	$\alpha_0 = 0.913933$	$\alpha = 0.910087$	$\alpha = 0.910087$	1.121 %
	$\lambda_0 = 0.994707$	$\lambda = 0.99298$	$\lambda = 0.99298$	-0.707 %
{ -0.9, 0.9 }	$\alpha_0 = 0.913933$	$\alpha = 0.910871$	$\alpha = 0.910871$	1.208 %
	$\lambda_0 = 0.994707$	$\lambda = 0.987717$	$\lambda = 0.987717$	-1.228 %

When the number of selected values of  $z$  is 4, the variance-covariance matrix  $\Sigma$  is a  $4 \times 4$  matrix and the variance-covariance matrix of the parameter vector  $\theta$ , denoted by  $Var(\hat{\theta})$  is a  $2 \times 2$  matrix. If we generate 5000 random variables, then  $\hat{\Sigma}$  and  $Var(\hat{\theta})$  are given by

$$\hat{\Sigma} = \begin{pmatrix} 0.00170892 & 0.000791541 & 0.000394247 & 0.000198651 \\ 0.000791541 & 0.00050339 & 0.000324684 & 0.000208052 \\ 0.000394247 & 0.000324684 & 0.000267138 & 0.000224311 \\ 0.000198651 & 0.000208052 & 0.000224311 & 0.000335412 \end{pmatrix},$$

and

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

Also note that only with 2 iterations, the algorithm converged except when using the pseudo-inverse variance-covariance matrix  $\Sigma$ . Using values of  $z$  too close to calculate the estimators makes the variance-covariance matrix  $\Sigma$  singular and we have to use the pseudo-inverse matrix. It also makes the calculations much more time-consuming and since the relative errors of the parameters do not decrease with the number of selected values of  $z$ , it is not suggested to use values of  $z$  too close. 10 points of  $z$  with negative values

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

and 4 points with negative values

$$z = \{-0.9, -0.3, 0.3, 0.9\}$$

are recommended.

But too few points of  $z$  may cause a large bias of the estimators. To investigate the relative errors of the parameters and the variance-covariance matrix of the parameters we note that when the number of selected values of  $z$  equals 10 or 4, the results are quite good. The relative errors increase a lot (especially the relative error of  $\lambda$ ) as for the results obtained with only two points of  $z$ . (Refer to the last two lines of Table 5.1).

Note that there is no significant difference between the results if we use or not the negative value points of  $z$ .

We conclude that calculations with 10 or 4 values of  $z$  give the better estimation, the relative errors are smaller than that of the others, and the calculation speed is much faster.

## 5.2. CONFIDENCE INTERVALS FOR THE PARAMETERS

We have used many sets of data and have found that when the parameter  $\alpha$  becomes much smaller, the calculation speed is much slower. Thus to calculate the confidence intervals of the parameters  $\lambda$  and  $\alpha$ , we generated several datasets

TABLE 5.2. Confidence interval for  $\beta$  and  $\lambda$  with 10 points

Size	$\hat{\beta}$	$\hat{\lambda}$	$\sqrt{\widehat{Var}(\hat{\beta})}$	C.I. for $\beta$	C.I. for $\lambda$
$n = 2000$	1.532663	4.63049	0.028903	(1.46601, 1.599314)	(4.3319, 4.9496)
$n = 1000$	1.46474	4.32642	0.047404	(1.3554, 1.5741)	(3.8784, 4.8262)
$n = 500$	1.446368	4.24766	0.042149	(1.3492, 1.5436)	(3.8542, 4.6812)
$n = 100$	1.399176	4.05186	0.083938	(1.2056, 1.5927)	(3.3388, 4.9172)

with  $\lambda = 4.5$  and  $\alpha = 0.4$ . We use the results of Chapter 4 to calculate the confidence intervals of the two parameters  $\alpha$  and  $\lambda$ ,

$$\hat{\beta} \pm t_{(1-\alpha/2)}(n-2) \sqrt{\widehat{Var}(\hat{\beta})}$$

$$\hat{\alpha} \pm t_{(1-\alpha/2)}(n-2) \sqrt{\widehat{Var}(\hat{\alpha})}.$$

Assume the significance level  $\alpha$  is 5% and  $n = 10$ ,

$$t_{(1-\alpha/2)}(n-2) = t_{0.975}(8) = 2.306,$$

we get the C.I. for the parameters  $\alpha$  and  $\lambda$  in Tables 5.2 and 5.3. Notice that the confidence intervals of the parameters become wider when  $n$  decreases.

### 5.3. TESTS CONCERNING $\lambda$ AND $\alpha$

We use the estimation results of the previous section to conduct a two-sided test concerning parameters  $\alpha$  and  $\lambda$ . The results are found in Tables 5.4 and 5.5 respectively.

#### 1. To test

$$H_0 : \alpha = 0.4$$

TABLE 5.3. Confidence interval for  $\alpha$  with 10 points

Size	$\hat{\alpha}$	$\sqrt{\widehat{Var}(\hat{\alpha})}$	C.I. for $\alpha$
$n = 2000$	0.41506	0.00896627	(0.394384, 0.435736)
$n = 1000$	0.383514	0.0180303	(0.341936, 0.425092)
$n = 500$	0.380229	0.014749	(0.346217, 0.4414241)
$n = 100$	0.39524	0.0299940	(0.326120, 0.464281)

TABLE 5.4. Test concerning  $\alpha$  with 10 points

Size	$\hat{\alpha}$	$t^* = \frac{\hat{\alpha}-0.4}{\sqrt{\widehat{Var}(\hat{\alpha})}}$	$t_{0.975}(8) = 2.306$	conclusion
$n = 2000$	0.41506	1.6796	2.306	accept $H_0$
$n = 1000$	0.383514	-0.9144	2.306	accept $H_0$
$n = 500$	0.380229	-1.3405	2.306	accept $H_0$
$n = 100$	0.39524	-0.1587	2.306	accept $H_0$

vs  $H_a : \alpha \neq 0.4$ ,

the test statistic is

$$t^* = \frac{\hat{\alpha} - \alpha^*}{\sqrt{\widehat{Var}(\hat{\alpha})}} = \frac{\hat{\alpha} - 0.4}{\sqrt{\widehat{Var}(\hat{\alpha})}}.$$

The decision rule with this test statistic at the 5% significance level is:

If  $|t^*| \leq t_{0.975}(8) = 2.306$ , accept  $H_0$ , i.e.  $\alpha = 0.4$ ,

If  $|t^*| > t_{0.975}(8) = 2.306$ , reject  $H_0$ , i.e.  $\alpha \neq 0.4$ .

## 2. To test

$$H_0 : \lambda = 4.5$$

vs  $H_a : \lambda \neq 4.5$ ,

TABLE 5.5. Test concerning  $\lambda$  with 10 points

Size	$\hat{\lambda}$	$\sqrt{e^{2\beta}\widehat{Var}(\hat{\beta})}$	$\frac{\hat{\lambda}-4.5}{\sqrt{\widehat{Var}(\hat{\lambda})}}$	$t_{0.975}(8) = 2.306$	conclusion
$n = 2000$	4.63049	0.133835	0.975	2.306	accept $H_0$
$n = 1000$	4.32642	0.205090	-0.846	2.306	accept $H_0$
$n = 500$	4.24766	0.179035	-1.409	2.306	accept $H_0$
$n = 100$	4.05186	0.340105	-1.318	2.306	accept $H_0$

the test statistic is

$$t^* = \frac{\hat{\lambda} - \lambda^*}{\sqrt{\widehat{Var}(\hat{\lambda})}} = \frac{\hat{\lambda} - 4.5}{\sqrt{\widehat{Var}(\hat{\lambda})}};$$

note that we defined  $\lambda = e^\beta$  and by the delta-theorem  $\widehat{Var}(\hat{\lambda}) = e^{2\beta}\widehat{Var}(\hat{\beta})$ .

The decision rule with this test statistic at the 5% significance level is:

If  $|t^*| \leq t_{0.975}(8) = 2.306$ , accept  $H_0$ , i.e.  $\lambda = 4.5$ ,

If  $|t^*| > t_{0.975}(8) = 2.306$ , reject  $H_0$ , i.e.  $\lambda \neq 4.5$ .

#### 5.4. EFFECT OF TRUNCATION

In this section we will discuss the effect of data truncation. When the dataset is heavy tailed or with some extreme values, it must be truncated in order to obtain the estimators with the algorithm proposed.

We use the parameters  $\alpha=0.4$  and  $\lambda=4.5$  to generate samples of discrete stable random variables with different sample sizes ( $n=2000$ ,  $n=1000$ ,  $n=500$  and  $n=100$ ). These datasets are distributed with a heavy tail and the largest value of the observation in the sample are so large (when  $n = 2000$ , it is 446,630,588; when  $n = 1000$ , it is 47,287,674; when  $n = 500$ , it is  $1.24 \times 10^8$  and when  $n = 100$ , it is 149,289) that we must cut the datasets somewhere in order to estimate the parameters.



To conduct our calculation, we take:

$$z = \{-0.9, -0.3, 0.3, 0.9\}$$

We put all the calculation results in Tables 5.6 to 5.9 to compare the differences among the different situations.

Notice that at the same percentage of truncation, the absolute value of the relative errors of estimator  $\hat{\lambda}$  increases when the sample size decreases.

With 8% or 10% truncation, when  $n = 2000$ , the absolute value of the relative errors of estimator  $\hat{\lambda}$  is 1.3%; when  $n = 1000$ , it is 3.5%; when  $n = 500$ , it is 7.7% and when  $n = 100$ , it is 12.1%.

With 20% truncation, when  $n = 2000$ , the absolute value of the relative errors of estimator  $\hat{\lambda}$  is 1.3%; when  $n = 1000$ , it is 8.1%; when  $n = 500$ , it is 10.0% and when  $n = 100$ , it is 14.4%. We can see that the absolute value of the relative errors of estimator  $\hat{\lambda}$  increases a lot with the decrease of the sample size  $n$ .

In total, the sum of the absolute value of the relative errors of the two estimators increase with the decrease of the sample size  $n$ .

With 15% truncation, when  $n = 2000$ , the sum of the absolute value of the relative errors of the two estimators are 9.8%; when  $n = 1000$ , it is 9.3%; when  $n = 500$ , it is 10.6% and when  $n = 100$ , it is 19.5%.

With 30% truncation, when  $n = 2000$ , the sum of the absolute value of the relative errors of the two estimators are 23.1%; when  $n = 1000$ , it is 23.7%; when  $n = 500$ , it is 23.7% and when  $n = 100$ , it is 33.9%.

Also notice that the relative errors of estimators increase when the percentage of truncation increases.

With  $n = 2000$ , the relative errors of estimator  $\hat{\alpha}$  increase from 3.8% (without truncation) to 19.5% (with 30% truncation). At the same time, the absolute values of the relative errors of the parameter  $\lambda$ , fluctuate from 2.9% to  $-3.7\%$  with the percentage of truncation 8% to 30%.

With  $n = 100$ , the relative errors of estimator  $\hat{\alpha}$  increase from 3.5% to 16.9% when the percentage of truncations changes from 10% to 30%. And the absolute

values of the relative errors of the estimator  $\lambda$  increase from 12.1% to 16.9% when the percentage of truncation changed from 10% to 30%.

After using many different percentages of the truncation to estimate the parameters, we conclude that with the percentage of truncation less than 15% and the sample size  $n \geq 500$ , the estimation gives better results, the relative errors of the parameters will be less than 10%.

TABLE 5.6. The effect of truncation on ( $n=2000$ )

	estimators	relative errors
without truncation	$\alpha_0=0.41337$	
initial values	$\lambda_0=4.6587$	
first iteration	$\alpha=0.41506$	3.765 %
	$\lambda=4.63049$	2.90 %
with truncation	$\alpha_0=0.426388$	
off 8 %	$\lambda_0=4.57748$	
first iteration	$\alpha=0.424454$	
	$\lambda=4.55938$	
second iteration	$\alpha=0.424454$	6.11 %
	$\lambda=4.55938$	1.320 %
with truncation	$\alpha_0=0.439585$	
off 15 %	$\lambda_0=4.50096$	
first iteration	$\alpha=0.438582$	
	$\lambda=4.49284$	
second iteration	$\alpha=0.438582$	9.646 %
	$\lambda=4.49284$	-0.159 %
with truncation	$\alpha_0=0.450307$	
off 20 %	$\lambda_0=4.44248$	
first iteration	$\alpha=0.450093$	
	$\lambda=4.44256$	
second iteration	$\alpha=0.450093$	12.523 %
	$\lambda=4.44256$	-1.276 %
with truncation	$\alpha_0=0.476048$	
off 30 %	$\lambda_0=4.31414$	
first iteration	$\alpha=0.477842$	
	$\lambda=4.33432$	
second iteration	$\alpha=0.477842$	19.461 %
	$\lambda=4.33432$	-3.682 %

TABLE 5.7. The effect of truncation ( $n=1000$ )

	estimators	relative errors
without truncation initial values	$\alpha_0=0.384607$ $\lambda_0=4.33825$	
first iteration	$\alpha=0.383514$ $\lambda=4.32642$	-4.122 % -3.857 %
with truncation off 9 %	$\alpha_0=0.398828$ $\lambda_0=4.2464$	
first iteration	$\alpha=0.398745$ $\lambda=4.24436$	
second iteration	$\alpha=0.398745$ $\lambda=4.34436$	-0.314 % -3.459 %
with truncation off 15 %	$\alpha_0=0.409830$ $\lambda_0=4.18011$	
first iteration	$\alpha=0.410567$ $\lambda=4.18573$	
second iteration	$\alpha=0.410567$ $\lambda=4.18573$	2.642 % -6.984 %
with truncation off 20 %	$\alpha_0=0.42017$ $\lambda_0=4.12129$	
first iteration	$\alpha=0.421705$ $\lambda=4.1342$	
second iteration	$\alpha=0.421705$ $\lambda=4.1342$	5.426 % -8.129 %
with truncation off 30 %	$\alpha_0=0.444763$ $\lambda_0=3.99354$	
first iteration	$\alpha=0.44831$ $\lambda=4.02405$	
second iteration	$\alpha=0.44831$ $\lambda=4.02405$	12.078 % -10.577 %

TABLE 5.8. The effect of truncation ( $n=500$ )

	estimators	relative errors
without trunction initial values	$\alpha_0=0.430911$ $\lambda_0=4.78148$	
first iteration	$\alpha=0.380229$ $\lambda=4.24766$	-4.943 % -5.608 %
with trunction off 10 %	$\alpha_0=0.44818$ $\lambda_0=4.67973$	
first iteration	$\alpha=0.397106$ $\lambda=4.15322$	
second iteration	$\alpha=0.397106$ $\lambda=4.15322$	-0.724 % -7.706 %
with trunction off 15 %	$\alpha_0=4.5819$ $\lambda_0=4.62468$	
first iteration	$\alpha=0.406938$ $\lambda=4.10246$	
second iteration	$\alpha=0.406938$ $\lambda=4.10246$	1.735 % -8.834 %
with trunction off 20 %	$\alpha_0=0.469351$ $\lambda_0=4.56642$	
first iteration	$\alpha=0.417944$ $\lambda=4.04903$	
second iteration	$\alpha=0.417944$ $\lambda=4.04903$	4.486 % -10.022 %
with trunction off 30 %	$\alpha_0=0.49616$ $\lambda_0=4.43864$	
first iteration	$\alpha=0.444568$ $\lambda=3.93308$	
second iteration	$\alpha=0.444568$ $\lambda=3.93308$	11.142 % -12.598 %

TABLE 5.9. The effect of truncation ( $n=100$ )

	estimators	relative errors
without truncation initial values	$\alpha_0=0.381712$ $\lambda_0=3.95355$	
first iteration	$\alpha=0.39524$ $\lambda=4.05186$	-1.19 % -9.959 %
with truncation off 10 %	$\alpha_0=0.399096$ $\lambda_0=3.85091$	
first iteration	$\alpha=0.414054$ $\lambda=3.95739$	
second iteration	$\alpha=0.414054$ $\lambda=3.95739$	3.515 % -12.058 %
with truncation off 15 %	$\alpha_0=0.409252$ $\lambda_0=3.79536$	
first iteration	$\alpha=0.42508$ $\lambda=3.90667$	
second iteration	$\alpha=0.42508$ $\lambda=3.90667$	6.27 % -13.185 %
with truncation off 20 %	$\alpha_0=0.420645$ $\lambda_0=3.73654$	
first iteration	$\alpha=0.437482$ $\lambda=3.85335$	
second iteration	$\alpha=0.437482$ $\lambda=3.85335$	9.371 % -14.37 %
with truncation off 30 %	$\alpha_0=0.448314$ $\lambda_0=3.60746$	
first iteration	$\alpha=0.467762$ $\lambda=3.73793$	
second iteration	$\alpha=0.467762$ $\lambda=3.73793$	16.941 % -16.935 %

## Chapter 6

---

### CONCLUSION

In this chapter, we will draw some conclusions from our work.

To estimate the two parameters of the discrete stable distribution, we employed the method of minimizing the quadratic distance between the empirical and theoretical probability generating function. The results show that this technique is powerful when the distribution that we worked on has no explicit expression for the probability distribution function.

We calculated the variance-covariance matrix of the difference between empirical and theoretical probability generating functions and we gave out the formulas for the quadratic distance estimators of the discrete stable distribution.

The estimators we got are consistent estimators and asymptotically have a normal distribution with variance-covariance matrix

$$Var(\hat{\theta}) = (X'\Sigma^{-1}X)^{-1}$$

We simulated several samples of discrete stable random distributed datasets with different parameters. The estimators obtained were quite good.

We analyzed the effect of the selected number of values of  $z$  on the results of estimation, and we found that 10 or 4 points of  $z$  is a better choice since it gave us better estimators and it is more time-saving in calculation, because of the smaller size of the variance-covariance matrix.

We also conducted inference about the parameters such as confidence intervals constructing and hypothesis testing.

Luong and Doray (2002) indicate that the quadratic distance estimator protects against a certain form of misspecification of the distribution, which makes the maximum likelihood estimator biased, while keeping the quadratic distance estimator unbiased. Therefore, the quadratic distance estimator can be considered as a robust semi-parametric estimator, offering protection against misspecification of the parametric family, while the maximum likelihood estimator, strictly a parametric estimator, is less robust.

Overall, the estimation results we got are quite good, especially for parameter  $\alpha$  close to 1. As for data truncation, when the percentage of truncation is less than 15% and the sample size  $n$  greater than 500, the estimation results are better.

The method to estimate the parameters by minimizing the quadratic distance between the empirical and the theoretical probability generating functions is good to use to estimate the parameters for certain distributions, especially for distributions that lack a closed formula for the probability and distribution functions, such as the discrete stable distribution.



## BIBLIOGRAPHY

---

- [1] BOUZAR,N. (2002), *Mixture representations for discrete Linnik laws*, Statistica Neerlandica (2002) Vol.56, nr.3, 295-300
- [2] CHRISTOPH,G. AND SCHREIBER,K. (1998), *Discrete stable random variables*, Stat. and Probab. Lett. 37, 243-247
- [3] DEVROYE,L. (1993), *A triptych of discrete distributions related to the stable law*, Statist. Probab. Lett. 18, 349-351.
- [4] DORAY,L.G. AND LUONG,A. (1997), *Efficient estimators for the Good family*, Communication in Statistics: Simulation and Computation, 26, 1075-1088.
- [5] GOLUB,G.H. AND VAN LOAN,C.F. (1989), *Matrix computations*, Second Edition, John Wiley and Sons, INC.
- [6] JOHNSON,N.L. AND KOTZ,S. AND BALAKRISHNAN,N. (1997), *Discrete multivariate distributions*, The Johns Hopkins University Press
- [7] KANTER,M. (1975), *Stable densities under change of scale and total variation inequalities*, Annals of Probability, vol.3, 697-707
- [8] KOTZ,S. AND OSTROVSKII,I.V. (1996), *A mixture representation of the Linnik distribution*, Statistics and Probability Letters 26,61-64
- [9] LUONG,A. AND DORAY,L.G. (2002), *General quadratic distance methods for discrete distributions definable recursively*, Insurance: Mathematics and Economics 30, 255-267
- [10] MONTGOMERY,D.C. AND PECK,E.A. (1992), *Introduction to linear regression analysis*, Second Edition John Wiley and Sons, INC.
- [11] NAKAMURA,M. AND PÉREZ-ABREU,V. (1993), *Empirical probability generating function: An overview*, Insurance: Mathematics and Economics 12, 287-295 North-Holland

- [12] NETER,J, WASSERMAN,W. AND KUTNER,M.H. (1989), *Applied linear regression models*,Second Edition Irwin, Homewood, IL 60430, Boston, MA 02116.
- [13] NOLAN,J.P. (2004), *Stable distributions, models for heavy tailed data*, <http://academic2.american.edu/~jpnolan/stable/chap1.pdf>
- [14] PAKES,A.G. (1998), *Mixture representations for symmetric generalized Linnik laws*, Statistics and Probability Letters 37,213-221
- [15] PILLAI,R.N. AND JAYAKUMAR,K. (1995), *Discrete Mittag-Leffler distribution*, Statistics and Probability Letters 23, 271-274
- [16] RAO,R.C. (1973), *Linear statistical inference and its applications*, Second Edition John Wiley and Sons, INC.
- [17] RÉMILLARD,B. AND THEODORESCU,R. (1991), *Inference based on the empirical probability generating function for mixtures of Poisson distributions*, AMS 1991 subject classifications. Primary 62E10, 62G05: Secondary 65C10
- [18] RICE,J.A. (1995), *Mathematical statistics and data analysis*, Duxbury Press, Belmont, CA.
- [19] ROUSSAS,G.G. (1997), *A course in mathematical statistics*, Second Edition Academic Press.
- [20] SIBUYA,M. (1979), *Generalized hypergeometric distributions*, Proceedings of the Institute of Statistical Mathematics, vol.31, pp.373-390
- [21] STEUTEL,F.W. AND VAN HARN,K. (1979), *Discrete analogues of self-decomposability and stability*, Ann. Probab. 7, 893-899.
- [22] WATKINS,D.S. (2002), *Fundamentals of matrix computations*, Second Edition, John Wiley and Sons, INC.
- [23] WEISBERG,S. (1985), *Applied linear regression*, Second Edition, John Wiley and Sons, INC.
- [24] ZOLOTAREV,V.M. (1986), *One-dimensional stable distributions*, American Mathematical Society, Providence, R.I.

## Appendix A

---

### SEVERAL TERMS OF THE PROBABILITY FUNCTION

By expanding the probability generating function

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

we obtain the first several terms of the probability function,  $p_0, p_1, p_2, \dots, p_8$ . It shows that the expressions for the terms of the probability distribution function of the discrete stable distribution are difficult to deal with in practice.

$$p_0 = e^{-\lambda}$$

$$p_1 = e^{-\lambda} \alpha \lambda$$

$$p_2 = \frac{1}{2} e^{-\lambda} (-(-1+\alpha) \alpha \lambda + \alpha^2 \lambda^2)$$

$$p_3 = \frac{1}{3} e^{-\lambda}$$

$$\left( \frac{1}{2} (-2+\alpha) (-1+\alpha) \alpha \lambda - (-1+\alpha) \alpha^2 \lambda^2 + \frac{1}{2} \alpha \lambda (-(-1+\alpha) \alpha \lambda + \alpha^2 \lambda^2) \right)$$

$$p_4 = \frac{1}{4} e^{-\lambda} \left( -\frac{1}{6} (-3+\alpha) (-2+\alpha) (-1+\alpha) \alpha \lambda + \right.$$

$$\frac{1}{2} (-2+\alpha) (-1+\alpha) \alpha^2 \lambda^2 - \frac{1}{2} (-1+\alpha) \alpha \lambda (-(-1+\alpha) \alpha \lambda + \alpha^2 \lambda^2) + \frac{1}{3} \alpha \lambda$$

$$\left( \frac{1}{2} (-2+\alpha) (-1+\alpha) \alpha \lambda - (-1+\alpha) \alpha^2 \lambda^2 + \frac{1}{2} \alpha \lambda (-(-1+\alpha) \alpha \lambda + \alpha^2 \lambda^2) \right) \Bigg)$$

$$p_5 = \frac{1}{5} e^{-\lambda}$$

$$\left( \frac{1}{24} (-4+\alpha) (-3+\alpha) (-2+\alpha) (-1+\alpha) \alpha \lambda - \frac{1}{6} (-3+\alpha) (-2+\alpha) (-1+\alpha) \alpha^2 \right.$$

$$\lambda^2 + \frac{1}{4} (-2+\alpha) (-1+\alpha) \alpha \lambda (-(-1+\alpha) \alpha \lambda + \alpha^2 \lambda^2) - \frac{1}{3} (-1+\alpha) \alpha \lambda$$

$$\left( \frac{1}{2} (-2+\alpha) (-1+\alpha) \alpha \lambda - (-1+\alpha) \alpha^2 \lambda^2 + \frac{1}{2} \alpha \lambda (-(-1+\alpha) \alpha \lambda + \alpha^2 \lambda^2) \right) +$$

$$\frac{1}{4} \alpha \lambda \left( -\frac{1}{6} (-3+\alpha) (-2+\alpha) (-1+\alpha) \alpha \lambda + \frac{1}{2} (-2+\alpha) (-1+\alpha) \alpha^2 \lambda^2 - \right.$$

$$\frac{1}{2} (-1+\alpha) \alpha \lambda (-(-1+\alpha) \alpha \lambda + \alpha^2 \lambda^2) + \frac{1}{3} \alpha \lambda \left( \frac{1}{2} (-2+\alpha) (-1+\alpha) \alpha \lambda - \right.$$

$$\left. \left. (-1+\alpha) \alpha^2 \lambda^2 + \frac{1}{2} \alpha \lambda (-(-1+\alpha) \alpha \lambda + \alpha^2 \lambda^2) \right) \right) \Bigg)$$

$$p_6 = \frac{1}{6} e^{-\lambda} \left( -\frac{1}{120} (-5+\alpha) (-4+\alpha) (-3+\alpha) (-2+\alpha) (-1+\alpha) \alpha \lambda + \right.$$

$$\frac{1}{24} (-4+\alpha) (-3+\alpha) (-2+\alpha) (-1+\alpha) \alpha^2 \lambda^2 -$$

$$\frac{1}{12} (-3+\alpha) (-2+\alpha) (-1+\alpha) \alpha \lambda (-(-1+\alpha) \alpha \lambda + \alpha^2 \lambda^2) +$$

$$\frac{1}{6} (-2+\alpha) (-1+\alpha) \alpha \lambda \left( \frac{1}{2} (-2+\alpha) (-1+\alpha) \alpha \lambda - \right.$$

$$\left. (-1+\alpha) \alpha^2 \lambda^2 + \frac{1}{2} \alpha \lambda (-(-1+\alpha) \alpha \lambda + \alpha^2 \lambda^2) \right) - \frac{1}{4} (-1+\alpha) \alpha \lambda$$

$$\left( -\frac{1}{6} (-3+\alpha) (-2+\alpha) (-1+\alpha) \alpha \lambda + \frac{1}{2} (-2+\alpha) (-1+\alpha) \alpha^2 \lambda^2 - \right.$$

$$\frac{1}{2} (-1+\alpha) \alpha \lambda (-(-1+\alpha) \alpha \lambda + \alpha^2 \lambda^2) + \frac{1}{3} \alpha \lambda \left( \frac{1}{2} (-2+\alpha) (-1+\alpha) \alpha \lambda - \right.$$

$$\left. \left. (-1+\alpha) \alpha^2 \lambda^2 + \frac{1}{2} \alpha \lambda (-(-1+\alpha) \alpha \lambda + \alpha^2 \lambda^2) \right) \right) \Bigg) +$$



$$\begin{aligned}
& \frac{1}{6} \alpha \lambda \left( -\frac{1}{120} (-5+\alpha) (-4+\alpha) (-3+\alpha) (-2+\alpha) (-1+\alpha) \alpha \lambda + \right. \\
& \quad \frac{1}{24} (-4+\alpha) (-3+\alpha) (-2+\alpha) (-1+\alpha) \alpha^2 \lambda^2 - \\
& \quad \frac{1}{12} (-3+\alpha) (-2+\alpha) (-1+\alpha) \alpha \lambda (-(-1+\alpha) \alpha \lambda + \alpha^2 \lambda^2) + \\
& \quad \frac{1}{6} (-2+\alpha) (-1+\alpha) \alpha \lambda \left( \frac{1}{2} (-2+\alpha) (-1+\alpha) \alpha \lambda - \right. \\
& \quad \quad \left. (-1+\alpha) \alpha^2 \lambda^2 + \frac{1}{2} \alpha \lambda (-(-1+\alpha) \alpha \lambda + \alpha^2 \lambda^2) \right) - \frac{1}{4} (-1+\alpha) \alpha \lambda \\
& \quad \left( -\frac{1}{6} (-3+\alpha) (-2+\alpha) (-1+\alpha) \alpha \lambda + \frac{1}{2} (-2+\alpha) (-1+\alpha) \alpha^2 \lambda^2 - \right. \\
& \quad \quad \frac{1}{2} (-1+\alpha) \alpha \lambda (-(-1+\alpha) \alpha \lambda + \alpha^2 \lambda^2) + \frac{1}{3} \alpha \lambda \left( \frac{1}{2} (-2+\alpha) (-1+\alpha) \alpha \lambda - \right. \\
& \quad \quad \quad \left. (-1+\alpha) \alpha^2 \lambda^2 + \frac{1}{2} \alpha \lambda (-(-1+\alpha) \alpha \lambda + \alpha^2 \lambda^2) \right) \left. \right) + \\
& \quad \frac{1}{5} \alpha \lambda \left( \frac{1}{24} (-4+\alpha) (-3+\alpha) (-2+\alpha) (-1+\alpha) \alpha \lambda - \right. \\
& \quad \quad \frac{1}{6} (-3+\alpha) (-2+\alpha) (-1+\alpha) \alpha^2 \lambda^2 + \frac{1}{4} (-2+\alpha) (-1+\alpha) \alpha \lambda \\
& \quad \quad \left. (-(-1+\alpha) \alpha \lambda + \alpha^2 \lambda^2) - \frac{1}{3} (-1+\alpha) \alpha \lambda \left( \frac{1}{2} (-2+\alpha) (-1+\alpha) \alpha \lambda - \right. \right. \\
& \quad \quad \quad \left. (-1+\alpha) \alpha^2 \lambda^2 + \frac{1}{2} \alpha \lambda (-(-1+\alpha) \alpha \lambda + \alpha^2 \lambda^2) \right) + \\
& \quad \quad \frac{1}{4} \alpha \lambda \left( -\frac{1}{6} (-3+\alpha) (-2+\alpha) (-1+\alpha) \alpha \lambda + \frac{1}{2} (-2+\alpha) (-1+\alpha) \alpha^2 \lambda^2 - \right. \\
& \quad \quad \quad \frac{1}{2} (-1+\alpha) \alpha \lambda (-(-1+\alpha) \alpha \lambda + \alpha^2 \lambda^2) + \frac{1}{3} \alpha \lambda \left( \frac{1}{2} (-2+\alpha) (-1+\alpha) \right. \\
& \quad \quad \quad \left. \left. \alpha \lambda - (-1+\alpha) \alpha^2 \lambda^2 + \frac{1}{2} \alpha \lambda (-(-1+\alpha) \alpha \lambda + \alpha^2 \lambda^2) \right) \right) \left. \right) \left. \right) \left. \right)
\end{aligned}$$

$$p_8 = \frac{1}{8} e^{-\lambda}$$

$$\begin{aligned}
& \left( -\frac{1}{5040} ((-7+\alpha) (-6+\alpha) (-5+\alpha) (-4+\alpha) (-3+\alpha) (-2+\alpha) (-1+\alpha) \alpha \lambda) + \right. \\
& \quad \frac{1}{720} (-6+\alpha) (-5+\alpha) (-4+\alpha) (-3+\alpha) (-2+\alpha) (-1+\alpha) \alpha^2 \lambda^2 - \\
& \quad \frac{1}{240} (-5+\alpha) (-4+\alpha) (-3+\alpha) (-2+\alpha) (-1+\alpha) \alpha \lambda (-(-1+\alpha) \alpha \lambda + \alpha^2 \lambda^2) + \\
& \quad \frac{1}{72} (-4+\alpha) (-3+\alpha) (-2+\alpha) (-1+\alpha) \alpha \lambda \\
& \quad \left( \frac{1}{2} (-2+\alpha) (-1+\alpha) \alpha \lambda - (-1+\alpha) \alpha^2 \lambda^2 + \frac{1}{2} \alpha \lambda (-(-1+\alpha) \alpha \lambda + \alpha^2 \lambda^2) \right) - \\
& \quad \frac{1}{24} (-3+\alpha) (-2+\alpha) (-1+\alpha) \alpha \lambda \\
& \quad \left( -\frac{1}{6} (-3+\alpha) (-2+\alpha) (-1+\alpha) \alpha \lambda + \frac{1}{2} (-2+\alpha) (-1+\alpha) \alpha^2 \lambda^2 - \right.
\end{aligned}$$

$$\begin{aligned}
& \frac{1}{2} (-1+\alpha) \alpha \lambda (-(-1+\alpha) \alpha \lambda + \alpha^2 \lambda^2) + \frac{1}{3} \alpha \lambda \\
& \left( \frac{1}{2} (-2+\alpha) (-1+\alpha) \alpha \lambda - (-1+\alpha) \alpha^2 \lambda^2 + \frac{1}{2} \alpha \lambda (-(-1+\alpha) \alpha \lambda + \alpha^2 \lambda^2) \right) + \\
& \frac{1}{10} (-2+\alpha) (-1+\alpha) \alpha \lambda \left( \frac{1}{24} (-4+\alpha) (-3+\alpha) (-2+\alpha) (-1+\alpha) \alpha \lambda - \right. \\
& \frac{1}{6} (-3+\alpha) (-2+\alpha) (-1+\alpha) \alpha^2 \lambda^2 + \\
& \frac{1}{4} (-2+\alpha) (-1+\alpha) \alpha \lambda (-(-1+\alpha) \alpha \lambda + \alpha^2 \lambda^2) - \frac{1}{3} (-1+\alpha) \alpha \lambda \\
& \left. \left( \frac{1}{2} (-2+\alpha) (-1+\alpha) \alpha \lambda - (-1+\alpha) \alpha^2 \lambda^2 + \frac{1}{2} \alpha \lambda (-(-1+\alpha) \alpha \lambda + \alpha^2 \lambda^2) \right) + \right. \\
& \frac{1}{4} \alpha \lambda \left( -\frac{1}{6} (-3+\alpha) (-2+\alpha) (-1+\alpha) \alpha \lambda + \frac{1}{2} (-2+\alpha) (-1+\alpha) \alpha^2 \lambda^2 - \right. \\
& \frac{1}{2} (-1+\alpha) \alpha \lambda (-(-1+\alpha) \alpha \lambda + \alpha^2 \lambda^2) + \frac{1}{3} \alpha \lambda \left( \frac{1}{2} (-2+\alpha) (-1+\alpha) \alpha \lambda - \right. \\
& \left. \left. (-1+\alpha) \alpha^2 \lambda^2 + \frac{1}{2} \alpha \lambda (-(-1+\alpha) \alpha \lambda + \alpha^2 \lambda^2) \right) \right) \Big) - \\
& \frac{1}{6} (-1+\alpha) \alpha \lambda \left( -\frac{1}{120} (-5+\alpha) (-4+\alpha) (-3+\alpha) (-2+\alpha) (-1+\alpha) \alpha \lambda + \right. \\
& \frac{1}{24} (-4+\alpha) (-3+\alpha) (-2+\alpha) (-1+\alpha) \alpha^2 \lambda^2 - \\
& \frac{1}{12} (-3+\alpha) (-2+\alpha) (-1+\alpha) \alpha \lambda (-(-1+\alpha) \alpha \lambda + \alpha^2 \lambda^2) + \\
& \frac{1}{6} (-2+\alpha) (-1+\alpha) \alpha \lambda \\
& \left. \left( \frac{1}{2} (-2+\alpha) (-1+\alpha) \alpha \lambda - (-1+\alpha) \alpha^2 \lambda^2 + \frac{1}{2} \alpha \lambda (-(-1+\alpha) \alpha \lambda + \alpha^2 \lambda^2) \right) - \right. \\
& \frac{1}{4} (-1+\alpha) \alpha \lambda \left( -\frac{1}{6} (-3+\alpha) (-2+\alpha) (-1+\alpha) \alpha \lambda + \frac{1}{2} (-2+\alpha) (-1+\alpha) \right. \\
& \alpha^2 \lambda^2 - \frac{1}{2} (-1+\alpha) \alpha \lambda (-(-1+\alpha) \alpha \lambda + \alpha^2 \lambda^2) + \frac{1}{3} \alpha \lambda \left( \frac{1}{2} (-2+\alpha) \right. \\
& \left. (-1+\alpha) \alpha \lambda - (-1+\alpha) \alpha^2 \lambda^2 + \frac{1}{2} \alpha \lambda (-(-1+\alpha) \alpha \lambda + \alpha^2 \lambda^2) \right) \Big) + \\
& \frac{1}{5} \alpha \lambda \left( \frac{1}{24} (-4+\alpha) (-3+\alpha) (-2+\alpha) (-1+\alpha) \alpha \lambda - \frac{1}{6} (-3+\alpha) (-2+\alpha) \right. \\
& (-1+\alpha) \alpha^2 \lambda^2 + \frac{1}{4} (-2+\alpha) (-1+\alpha) \alpha \lambda (-(-1+\alpha) \alpha \lambda + \alpha^2 \lambda^2) - \\
& \frac{1}{3} (-1+\alpha) \alpha \lambda \left( \frac{1}{2} (-2+\alpha) (-1+\alpha) \alpha \lambda - (-1+\alpha) \alpha^2 \lambda^2 + \right. \\
& \frac{1}{2} \alpha \lambda (-(-1+\alpha) \alpha \lambda + \alpha^2 \lambda^2) \Big) + \frac{1}{4} \alpha \lambda \left( -\frac{1}{6} (-3+\alpha) (-2+\alpha) \right. \\
& (-1+\alpha) \alpha \lambda + \frac{1}{2} (-2+\alpha) (-1+\alpha) \alpha^2 \lambda^2 - \frac{1}{2} (-1+\alpha) \alpha \lambda \\
& \left. (-(-1+\alpha) \alpha \lambda + \alpha^2 \lambda^2) + \frac{1}{3} \alpha \lambda \left( \frac{1}{2} (-2+\alpha) (-1+\alpha) \alpha \lambda - \right. \right.
\end{aligned}$$

$$\begin{aligned}
& (-1+\alpha) \alpha^2 \lambda^2 + \frac{1}{2} \alpha \lambda (-(-1+\alpha) \alpha \lambda + \alpha^2 \lambda^2) \Big) \Big) \Big) + \\
& \frac{1}{7} \alpha \lambda \Big( \frac{1}{720} (-6+\alpha) (-5+\alpha) (-4+\alpha) (-3+\alpha) (-2+\alpha) (-1+\alpha) \alpha \lambda - \\
& \frac{1}{120} (-5+\alpha) (-4+\alpha) (-3+\alpha) (-2+\alpha) (-1+\alpha) \alpha^2 \lambda^2 + \\
& \frac{1}{48} (-4+\alpha) (-3+\alpha) (-2+\alpha) (-1+\alpha) \alpha \lambda (-(-1+\alpha) \alpha \lambda + \alpha^2 \lambda^2) - \\
& \frac{1}{18} (-3+\alpha) (-2+\alpha) (-1+\alpha) \alpha \lambda \Big( \frac{1}{2} (-2+\alpha) (-1+\alpha) \alpha \lambda - \\
& (-1+\alpha) \alpha^2 \lambda^2 + \frac{1}{2} \alpha \lambda (-(-1+\alpha) \alpha \lambda + \alpha^2 \lambda^2) \Big) + \frac{1}{8} (-2+\alpha) (-1+\alpha) \\
& \alpha \lambda \Big( -\frac{1}{6} (-3+\alpha) (-2+\alpha) (-1+\alpha) \alpha \lambda + \frac{1}{2} (-2+\alpha) (-1+\alpha) \alpha^2 \lambda^2 - \\
& \frac{1}{2} (-1+\alpha) \alpha \lambda (-(-1+\alpha) \alpha \lambda + \alpha^2 \lambda^2) + \frac{1}{3} \alpha \lambda \Big( \frac{1}{2} (-2+\alpha) (-1+\alpha) \alpha \lambda - \\
& (-1+\alpha) \alpha^2 \lambda^2 + \frac{1}{2} \alpha \lambda (-(-1+\alpha) \alpha \lambda + \alpha^2 \lambda^2) \Big) \Big) - \\
& \frac{1}{5} (-1+\alpha) \alpha \lambda \Big( \frac{1}{24} (-4+\alpha) (-3+\alpha) (-2+\alpha) (-1+\alpha) \alpha \lambda - \\
& \frac{1}{6} (-3+\alpha) (-2+\alpha) (-1+\alpha) \alpha^2 \lambda^2 + \frac{1}{4} (-2+\alpha) (-1+\alpha) \alpha \lambda \\
& (-(-1+\alpha) \alpha \lambda + \alpha^2 \lambda^2) - \frac{1}{3} (-1+\alpha) \alpha \lambda \Big( \frac{1}{2} (-2+\alpha) (-1+\alpha) \alpha \lambda - \\
& (-1+\alpha) \alpha^2 \lambda^2 + \frac{1}{2} \alpha \lambda (-(-1+\alpha) \alpha \lambda + \alpha^2 \lambda^2) \Big) + \\
& \frac{1}{4} \alpha \lambda \Big( -\frac{1}{6} (-3+\alpha) (-2+\alpha) (-1+\alpha) \alpha \lambda + \frac{1}{2} (-2+\alpha) (-1+\alpha) \alpha^2 \lambda^2 - \\
& \frac{1}{2} (-1+\alpha) \alpha \lambda (-(-1+\alpha) \alpha \lambda + \alpha^2 \lambda^2) + \frac{1}{3} \alpha \lambda \Big( \frac{1}{2} (-2+\alpha) (-1+\alpha) \\
& \alpha \lambda - (-1+\alpha) \alpha^2 \lambda^2 + \frac{1}{2} \alpha \lambda (-(-1+\alpha) \alpha \lambda + \alpha^2 \lambda^2) \Big) \Big) \Big) + \\
& \frac{1}{6} \alpha \lambda \Big( -\frac{1}{120} (-5+\alpha) (-4+\alpha) (-3+\alpha) (-2+\alpha) (-1+\alpha) \alpha \lambda + \\
& \frac{1}{24} (-4+\alpha) (-3+\alpha) (-2+\alpha) (-1+\alpha) \alpha^2 \lambda^2 - \\
& \frac{1}{12} (-3+\alpha) (-2+\alpha) (-1+\alpha) \alpha \lambda (-(-1+\alpha) \alpha \lambda + \alpha^2 \lambda^2) + \\
& \frac{1}{6} (-2+\alpha) (-1+\alpha) \alpha \lambda \Big( \frac{1}{2} (-2+\alpha) (-1+\alpha) \alpha \lambda - (-1+\alpha) \alpha^2 \lambda^2 + \\
& \frac{1}{2} \alpha \lambda (-(-1+\alpha) \alpha \lambda + \alpha^2 \lambda^2) \Big) - \frac{1}{4} (-1+\alpha) \alpha \lambda \\
& \Big( -\frac{1}{6} (-3+\alpha) (-2+\alpha) (-1+\alpha) \alpha \lambda + \frac{1}{2} (-2+\alpha) (-1+\alpha) \alpha^2 \lambda^2 -
\end{aligned}$$



$$\begin{aligned}
& \frac{1}{2} (-1+\alpha) \alpha \lambda (-(-1+\alpha) \alpha \lambda + \alpha^2 \lambda^2) + \frac{1}{3} \alpha \lambda \left( \frac{1}{2} (-2+\alpha) (-1+\alpha) \alpha \lambda - \right. \\
& \quad \left. (-1+\alpha) \alpha^2 \lambda^2 + \frac{1}{2} \alpha \lambda (-(-1+\alpha) \alpha \lambda + \alpha^2 \lambda^2) \right) + \\
& \frac{1}{5} \alpha \lambda \left( \frac{1}{24} (-4+\alpha) (-3+\alpha) (-2+\alpha) (-1+\alpha) \alpha \lambda - \right. \\
& \quad \frac{1}{6} (-3+\alpha) (-2+\alpha) (-1+\alpha) \alpha^2 \lambda^2 + \frac{1}{4} (-2+\alpha) (-1+\alpha) \alpha \lambda \\
& \quad \left. (-(-1+\alpha) \alpha \lambda + \alpha^2 \lambda^2) - \frac{1}{3} (-1+\alpha) \alpha \lambda \left( \frac{1}{2} (-2+\alpha) (-1+\alpha) \alpha \lambda - \right. \right. \\
& \quad \left. \left. (-1+\alpha) \alpha^2 \lambda^2 + \frac{1}{2} \alpha \lambda (-(-1+\alpha) \alpha \lambda + \alpha^2 \lambda^2) \right) \right) + \\
& \frac{1}{4} \alpha \lambda \left( -\frac{1}{6} (-3+\alpha) (-2+\alpha) (-1+\alpha) \alpha \lambda + \frac{1}{2} (-2+\alpha) (-1+\alpha) \alpha^2 \lambda^2 - \right. \\
& \quad \frac{1}{2} (-1+\alpha) \alpha \lambda (-(-1+\alpha) \alpha \lambda + \alpha^2 \lambda^2) + \frac{1}{3} \alpha \lambda \left( \frac{1}{2} (-2+\alpha) (-1+\alpha) \right. \\
& \quad \left. \left. \alpha \lambda - (-1+\alpha) \alpha^2 \lambda^2 + \frac{1}{2} \alpha \lambda (-(-1+\alpha) \alpha \lambda + \alpha^2 \lambda^2) \right) \right) \right) \right) \right) \right)
\end{aligned}$$