A Thesis for the Degree of Ph.D. in Science

Extensions of Binomial and Negative Binomial Distributions

February 2014

Graduate School of Science and Technology
Keio University

Tomoaki Imoto

Contents

Chapter 1

Preliminaries

1.1 Introduction

In a variety of scientific fields such as atmospheric science, biology, ecology, linguistics, financial science and so on, counting observations often appear and discrete distributions play an important role in studying the fields through analyzing the count data. Various types of discrete distributions have been derived by taking account of the mechanism of phenomenon. Classical examples include the binomial, Poisson, negative binomial and hypergeometric distributions. For improving the fittings and analyzing the phenomenon more accurately, a lot of extensions and generalizations of classical distributions have been considered in the literature.

In statistical analysis of count data, the Poisson distribution is one of the most utilized distributions, since data in multiple research fields often fulfill the Poisson postulates. However, an important restriction of this model is that the mean and variance are equal. For many observed count data, it is common to have the sample mean to be greater or smaller than the sample variance which are referred to as under-dispersion and over-dispersion, respectively, relative to the Poisson distribution. They may arise due to one or more possible causes such as repulsion for under-dispersion and aggregation and hierarchy for over-dispersion. The negative binomial distribution was derived for allowing aggregation and hierarchy and is commonly used alternative to Poisson distribution when over-dispersion is present. Examples of the under-dispersed distribution includes the binomial distribution. The index of dispersion which is defined as the ratio of variance to the mean is used as the measure to detect such departures from the Poisson distribution in this article.

It is meaningful to consider the distribution which can be adapted for both under- and over-dispersion because it provides a unified approach to handle both under-dispersion and over-dispersion. The examples of such distributions include the generalized negative binomial distribution by Jain and Consul (1971), the generalized Poisson distribution by Consul and Jain (1973), the Conway–Maxwell–Poisson distribution, which was originally developed by Conway and Maxwell (1962) and revived by Shmueli et al. (2005), the Hurwitz–Lerch Zeta distribution, which was studied by various researchers including Panaretos (1989), Kulasekera and Tonkya (1992), Doray and Luong (1995, 1997) and Zörnig and Altmann (1995), the weighted Poisson distribution by Castillo and Pérez-Casany (1998), and its generalization as a family of distributions by Castillo and Pérez-Casany (2005) and $\mathrm{GIT}_{3,1}$ by Aoyama et al. (2008).

In this article, we introduce the flexible distributions to the dispersion by considering the distribution which includes the binomial and negative binomial distributions. In Chapter 2, we give theorems about the general Lagrangian distributions and their applications. In Chapter 3, we derive a distribution belonging to general Lagrangian distributions. This includes generalizations of binomial and negative binomial distributions. Moreover, this includes some well-studied distributions which belong to Lagrangian distributions and thus, plays the role of a full model which includes some different sub-models when we fit the distribution to real data. The theorem introduced in Chapter 2 leads a variety of properties for the distribution proposed in Chapter 3. Chapter 4 considers an extension of $\mathrm{GIT}_{3,1}$, which is generated from a convolution of binomial and negative binomial variables. The distribution plays the role of a continuous bridge between under- and over-dispersion. Moreover, various types of stochastic processes lead to the proposed distribution. In this thesis, we consider a three-dimensional random walk, a birth, death and immigration process and a thinned stochastic process. Chapter 5 provides a generalization of Conway–Maxwell–Poisson distribution, which is also a flexible distribution to the dispersion and has two modes with one mode at zero under some condition. This distribution is expressed by a simple form and is thus easy to use for various types of count data. In Chapter 6, we give the conclusion of this thesis.

The remainder of this chapter provides some basic concepts and convenient tools for studying discrete distributions and three important classical distributions; binomial, Poisson and negative binomial distributions.

1.2 Basic concepts

1.2.1 Probability

A σ-field is a collection $\mathcal{F}$ of subsets of a set Ω that contains the empty set as a member and is closed under countable unions and complements. A probability measure P on a σ-field of subsets of a set Ω is a function from $\mathcal{F}$ to the unit interval $[0, 1]$ such that $\mathrm{P}(\Omega) = 1$ and the probability measure of a countable union of disjoint sets $\{E_i\}$ is eqaul to $\sum_i \mathrm{P}(E_i)$.

For $\mathrm{P}(E)$ to be a probability measure, we require the following probability axioms to be satisfied:

1. $0 \leq \mathrm{P}(E) \leq 1$.

2. $\mathrm{P}(\Omega) = 1$.

3. If the events E_i are mutually exclusive, then $\mathrm{P}\left(\bigcup_i E_i\right) = \sum_i \mathrm{P}(E_i)$.

1.2.2 Discrete distributions

A random variable X is a mapping from a sample space into the real numbers, with the property that for every outcome there is an associated probability $\mathrm{P}(X \leq x)$ which exists for all real values of x. The cumulative distribution function of X is defined as $\mathrm{P}(X \leq x)$ and regarded as a function of x. Clearly $\mathrm{P}(X \leq x)$ is a non-decreasing function of x and $0 \leq \mathrm{P}(X \leq x) \leq 1$. If $\lim_{x \to -\infty} \mathrm{P}(X \leq x) = 0$ and $\lim_{x \to +\infty} \mathrm{P}(X \leq x) = 1$, the distribution is proper.

For discrete distributions, $\mathrm{P}(X \leq x)$ is a step function with only an enumerable number of steps. If the height of the step at x_i is $p(i)$, then $\mathrm{P}(X = x_i) = p(i)$. We call $p(i)$ a probability mass function (pmf). If the distribution is proper, $\sum_i p(i) = 1$. The discrete distributions of interest in this thesis is defined over non-negative integers, or $p(i) = \mathrm{P}(X = i)$ for $i = 0, 1, \ldots$.

From the axioms introduced in Section 1.2.1, for $p(i)$ to be a pmf, it must satisfy the conditions that $p(i) \geq 0$ for $i = 0, 1, \ldots$ and $\sum_{i=0}^{\infty} p(i) = 1$.

1.2.3 Expected values and moments

The expected value of a function $g(X)$ of X is defined as

$$\mathrm{E}[g(X)] = \sum_{x=0}^{\infty} g(x)\mathrm{P}(X = x)$$

for discrete distributions. If this value is finite, we say $\mathrm{E}[g(X)]$ exists.

The expected value of X^r for any real number r is referred to as the r-th moment about zero:

$$\mu'_r = \mathrm{E}[X^r].$$

The first moment about zero, μ'_1, is called the mean of X.

The expected value of $(X - \mu'_1)^r$ for any real number r is referred to as the r-th moment about the mean:

$$\mu_r = \mathrm{E}[(X - \mu'_1)^r].$$

The second moment about the mean, μ_2, is called the variance of X and it is written as $\mathrm{Var}[X]$.

The descending factorial moments is also useful to study discrete distributions. The r-th descending factorial moment of X is the expected value of $X!/(X-r)!$:

$$\mu'_{[r]} = \mathrm{E}\left[\frac{X!}{(X-r)!}\right].$$

1.2.4 Probability generating functions

When studying discrete distributions, it is often advantageous to use the probability generating function (pgf). The pgf of a random variable X, or equivalently, of the distribution with pmf $p(i)$ is defined as

$$G_X(t) = \mathrm{E}[t^X] = \sum_{i=0}^{\infty} p(i)t^i.$$

Since a proper distribution satisfies $\sum_{i=0}^{\infty} p(i) = 1$, the pgf always exists for $|t| \leq 1$.

For the function $G(t)$ to be the pgf of proper distribution, it must be satisfied that $G(t)$ is successively differentiable in $-1 < t < 1$, n-th derivative of $G(t)$ at $t = 0$ is non-negative and $G(1) = 1$.

Probability generating function has several properties.

1. From the definition, the pmf defines the pgf. Conversely, the pgf defines

the pmf uniquely as

$$p(i) = \frac{1}{i!} \frac{\partial^i G(t)}{\partial t^i} \bigg|_{t=0}.$$

2. The moment generating function (mgf), if it exists, is defined as $G(\mathrm{e}^t)$ and its r-th derivative at $t = 0$ gives the r-th moment about zero, or

$$\mu'_r = \frac{\partial^r G(\mathrm{e}^t)}{\partial t^r} \bigg|_{t=0}.$$

3. The factorial moment generating function, if it exists, is defined as $G(t+1)$ and its r-th derivative at $t = 0$ gives the r-th descending factorial moment, or

$$\mu'_{[r]} = \frac{\partial^r G(t+1)}{\partial t^r} \bigg|_{t=0}.$$

4. The cumulant generating function, if it exists, is defined as $\log G(\mathrm{e}^t)$. Put

$$\kappa_r = \frac{\partial^r \log G(\mathrm{e}^t)}{\partial t^r}.$$

   Then the cumulants κ_r are known to be functions of the moments. For example, first three cumulants have the relations $\kappa_1 = \mu'_1$, $\kappa_2 = \mu_2$ and $\kappa_2 = \mu_3$.

5. If X_1 and X_2 are two independent random variables with pgf's $G_1(t)$ and $G_2(t)$, respectively, then the distribution of their sum $X_1 + X_2$ has the pgf $G_1(t)G_2(t)$. This is called the convolution of two variables.

1.3 Some discrete distributions

1.3.1 Binomial distribution

The binomial distribution is the probability distribution of the number of the successes in a sequence of independent trials. When the number of the trials is n

and the probability of success is p in each trial, the probability of x successes is given by

$$P(X = x) = \binom{n}{x} p^x q^{n-x},\ x = 0, 1, \ldots, n, \tag{1.1}$$

where n is a non-negative integer and $0 < p = 1 - q < 1$. If the random variable X has the pmf (1.1), we say that X follows the binomial distribution with parameters n and p and write $X \sim \mathrm{Bin}(n, p)$. When $n = 1$, the distribution is known as the Bernoulli distribution.

The pgf of the binomial distribution with parameters n and p is

$$G(t) = (q + pt)^n$$

and the mean and variance are

$$\mu_1' = np \text{ and } \mu_2 = npq,$$

respectively. The index of dispersion is given by $q < 1$ and the binomial distribution is thus always under-dispersed. The binomial distribution has simple recursive formulas about the moments μ_r, and cumulants κ_r, for $r \geq 1$,

$$\mu_{r+1} = pq\left(nr\mu_{r-1} + \frac{\partial \mu_r}{\partial p}\right) \text{ and } \kappa_{r+1} = pq\frac{\partial \kappa_r}{\partial p}. \tag{1.2}$$

1.3.2 Poisson distribution

Consider the limit of the binomial distribution with pmf (1.1) as n tends to infinity and p tends to zero with $np = \lambda$, where $\lambda > 0$. Then the resultant distribution has the pmf

$$P(X = x) = \frac{e^{-\lambda}\lambda^x}{x!},\ x = 0, 1, \ldots. \tag{1.3}$$

This distribution is called the Poisson distribution with parameter λ. The Poisson distribution also arises from the counting process $\{N(\tau)\}$:

- $N(0) = 0$.
- The numbers of occurrences counted in disjoint intervals are independent from each other.
- The probability distribution of the number of occurrences counted in any time interval only depends on the length of the interval.
- No counted occurrences are simultaneous.

In this process, the random variable $N(\tau)$ follows the Poisson distribution with parameter $\lambda(\tau)$, where $\lambda(\tau)$ is the rate of the event occurring in the interval time $[t, t+\tau)$ for any time t.

The pgf of the Poisson distribution with parameter λ is

$$G(t) = \mathrm{e}^{\lambda(t-1)}$$

and the mean and variance are

$$\mu_1' = \mu_2 = \lambda.$$

The index of dispersion of the Poisson distribution is thus always one.

1.3.3 Negative binomial distribution

Consider the parameter λ of the Poisson distribution (1.3) follows the gamma distribution whose probability density function is

$$f(\lambda) = \frac{\beta^{\nu}}{\Gamma(\nu)} \lambda^{\nu-1} \mathrm{e}^{-\lambda\beta}, \ \lambda > 0,$$

where $\nu, \beta > 0$. Then, the resultant distribution has the pmf

$$\mathrm{P}(X = x) = \binom{\nu + x - 1}{x} p^x q^r, \; x = 0, 1, \ldots, \tag{1.4}$$

where $p = 1/(1 + \beta)$. If the random variable X has the pmf (1.4), we say that X follows the negative binomial distribution with parameters ν and p and write $X \sim \mathrm{NB}(\nu, p)$.

The pgf of the negative binomial distribution with parameters ν and p is

$$G(t) = \left(\frac{q}{1 - pt}\right)^{\nu}$$

and the mean and variance are

$$\mu_1' = \frac{\nu p}{q} \text{ and } \mu_2 = \frac{\nu p}{q^2},$$

respectively. The index of dispersion is given by $1/q > 1$ and the negative binomial distribution is thus always over-dispersed. The negative binomial distribution has simple recursive formulas about the moments μ_r, and cumulants κ_r, for $r \geq 1$,

$$\mu_{r+1} = q\frac{\partial \mu_r}{\partial q} + \frac{r\nu q}{p^2}\mu_{r-1} \text{ and } \kappa_{r+1} = q\frac{\partial \kappa_r}{\partial q}. \tag{1.5}$$

Chapter 2

General Lagrangian distributions and their properties

2.1 Introduction

Lagrange (1736–1813) gave two expansions for inverting an analytic function in terms of another analytic function. Let $f(t)$ and $g(t)$ be analytic functions around $t = 0$ such that $g(0) \neq 0$. Under the transformation $t = ug(t)$, Lagrange obtained the following expansions

$$f(t) = \sum_{x=0}^{\infty} \frac{u^x}{x!} [\mathrm{D}^{x-1}(g(t)^x \mathrm{D} f(t))]_{t=0} \tag{2.1}$$

and

$$\frac{f(t)}{1 - tg'(t)/g(t)} = \sum_{x=0}^{\infty} \frac{u^x}{x!} [\mathrm{D}^{x}(g(t)^x f(t))]_{t=0}, \tag{2.2}$$

where $\mathrm{D} = \partial/\partial t$ is a derivative operator and D^{-1} is an operator such that $\mathrm{D}^{-1}\mathrm{D} = \mathrm{I}$ with an identity operator I (see Riordan, 1968, p. 146). Using the expansion (2.1), Consul and Shenton (1972, 1975) defined the class of general Lagrangian distributions of the first kind (GLD_1). Similarly, Janardan and Rao (1983) and Janardan (1997) used the expansion (2.2) to define the class of general Lagrangian distributions of the second kind (GLD_2).

Lagrange expansions had been applied to branching and queueing processes before general Lagrangian distributions were defined. Good (1949) and Otter (1949) independently showed the importance of the transformation $t = ug(t)$ in the branching process and Otter (1949) applied the expansion (2.1) to derive the distribution of the number of vertices in the rooted tree. Benes (1957), Haight and Breuer (1960) and Takács (1967) also used the expansion (2.1) for development of queueing processes. The potential of these techniques for deriving distributions and their properties have been systematically explored (see Consul and Famoye, 2006). This chapter gives the definitions of general Lagrangian distributions, some of their properties and applications. Section 2.2 provides the propositions for the definitions. Sections 2.3 and 2.4 give new theorems about general Lagrangian distributions and their applications which are considered by Imoto (to appear, b)

2.2 General Lagrangian distributions

First, we prove the following propositions to provide the definitions of GLD_1 and GLD_2.

Proposition 2.1

1. Let $f(t)$ and $g(t)$ be analytic functions in the domain $|t| < 1$ such that $f(1) = g(1) = 1$, $g(0) > 0$, $g'(1) < 1$, $g(t)$ and $g'(t)$ are non-decreasing functions in $0 \le t \le 1$ and $[\mathrm{D}^{x-1}(g(t)^x \mathrm{D} f(t))]_{t=0} \ge 0$ for $x = 0, 1, \ldots$. Then the function

$$L(u) = f(t), \text{ where } u = \frac{t}{g(t)} \tag{2.3}$$

is a pgf.

2. Let $f(t)$ and $g(t)$ be analytic functions in the domain $|t| < 1$ such that $f(1) = g(1) = 1$, $g(0) > 0$, $g'(1) < 1$, $g(t)$ and $g'(t)$ are non-decreasing functions in $0 \leq t \leq 1$ and $[\mathrm{D}^x(g(t)^x f(t))]_{t=0} \geq 0$ for $x = 0, 1, 2, \ldots$. Then the function

$$L(u) = \frac{(1 - g'(1))f(t)}{1 - tg'(t)/g(t)}, \text{ where } u = \frac{t}{g(t)} \tag{2.4}$$

is a pgf.

To prove these propositions, we need two lemmas, which are the extensions of Takács' theorems (cf. Theorems 3 and 4 in Section 5 of Takács, 1967). Assume that $\pi(t)$ is analytic function in the domain $|t| < 1$ such that $\pi(0) > 0$, $\pi(1) = 1$, and $\pi(t)$ and $\pi'(t)$ are non-decreasing functions in $0 \leq t \leq 1$.

Lemma 2.1

Let $t = \delta$ be the smallest non-negative real root of the equation $\pi(t) = t$. If $\pi'(1) \leq 1$, then $\delta = 1$. If $\pi'(1) > 1$, then $0 \leq \delta < 1$. The equation has no other root in the domain $|t| \leq \delta$.

Proof

Put $H(t) = \pi(t) - t$. Then $H(0) > 0$, $H(1) = 0$ and $H'(t) = \pi'(t) - 1$. If $\pi'(1) \leq 1$, then the equation $H(t) = 0$ has one and only one root $t = 1$ in $0 \leq t \leq 1$ because $\pi(t)$ and $\pi'(t)$ are non-decreasing functions in $0 \leq t \leq 1$. For the same reason, if $\pi'(1) > 1$, then the equation $H(t) = 0$ has exactly two roots in $0 \leq t \leq 1$, $t = 1$ and $t = \delta$, where $0 \leq \delta < 1$. This proves the first part of the lemma. To prove the second part, we note that always $\pi'(\delta) \leq 1$ and hence $|\pi'(t)| < \pi'(\delta) \leq 1$ if $|t| < \delta$. Accordingly, if $|t| \leq \delta$ and $t \neq \delta$, we have

$$|\pi(t) - \pi(\delta)| = \left| \int_\delta^t \pi'(z) \mathrm{d}z \right| < |t - \delta|$$

which shows that $\pi(t) = t$ is impossible if $|t| \leq \delta$ and $t \neq \delta$.

$\square$

Lemma 2.2

If $0 \leq u < 1$, then $t = u\pi(t)$ has exactly one real root $t = \delta(u)$ in $0 \leq t < \delta$ and $\lim_{u \to 1-0} \delta(u) = \delta$ is the smallest non-negative real root of $\pi(t) = t$.

Proof

Since $\pi'(t) \leq 1$ in $0 \leq t \leq \delta$, about the equation $u = t/\pi(t)$,

$$\frac{\partial u}{\partial t} = \frac{\pi(t) - t\pi'(t)}{\pi(t)^2} \geq \frac{\pi(t) - t}{\pi(t)^2} > 0$$

in $0 \leq t \leq \delta$. Hence, from the inverse function theorem, the equation $u = t/\pi(t)$ has exactly only one root $t = \delta(u)$ in $0 \leq t \leq \delta$ and $\delta(u)$ is a non-decreasing function. Since $\delta(u) < \delta$ for $0 \leq u < 1$ and $\lim_{u \to 1-0} \delta(u) = \delta^*$ is a root of $\pi(t) = t$, it follows that $\delta^* = \delta$ from Lemma 2.1.

$\square$

Proof of Proposition 2.1

From Lemmas 2.1 and 2.2, under the assumption of $g(t)$, we see that $u = t/g(t)$ has only one solution $t = l(u)$ in $0 \leq u \leq 1$ and $l(1) = 1$. Hence, the functions (2.3) and (2.4) satisfy the conditions to be pgf under the assumption of $f(t)$ and $g(t)$.

$\square$

The distribution with pgf (2.3) is the general Lagrangian distribution of the first kind generated through $f(t)$ and $g(t)$ and the pmf is obtained from the coefficient of u^x in (2.1) as

$$\mathrm{P}(X = x) = \frac{1}{x!}[\mathrm{D}^{x-1}(g(t)^x \mathrm{D} f(t))]_{t=0}, \ x = 0, 1, \ldots . \tag{2.5}$$

This class of general Lagrangian distributions of the first kind will be denoted by $\mathrm{GLD}_1(f, g)$. The mean and variance of $\mathrm{GLD}_1(f, g)$ can be expressed as

$$\mathrm{E}[X] = \frac{f_1}{1 - g_1} \text{ and } \mathrm{Var}[X] = \frac{f_2}{(1 - g_1)^2} + \frac{f_1 g_2}{(1 - g_1)^3}, \tag{2.6}$$

respectively, where $f_i = \mathrm{D}^i \log f(\mathrm{e}^t)$ and $g_i = \mathrm{D}^i \log g(\mathrm{e}^t)$. Suppose that there are some customers (initial customers) waiting for service in a queue at a counter when the service is initially started. If the number of the initial customers follows the distribution with pgf $f(t)$ and the number of arrivals during a service follows the distribution with pgf $g(t)$, then the number of customers served before the queue become empty for the first time, i.e., during a busy period, follows the distribution with pmf (2.5) (Consul and Shenton, 1973).

The distribution with pgf (2.4) is the general Lagrangian distribution of the second kind generated through $f(t)$ and $g(t)$ and the pmf is obtained from the expansion (2.2) as

$$\mathrm{P}(X = x) = \frac{1 - g'(1)}{x!}[\mathrm{D}^x(g(t)^x f(t))]_{t=0}, \ x = 0, 1, \ldots. \tag{2.7}$$

This class of general Lagrangian distributions of the second kind is denoted by $\mathrm{GLD}_2(f, g)$. The mean and variance of $\mathrm{GLD}_2(f, g)$ can be expressed as

$$\mathrm{E}[X] = \frac{f_1 - f_1 g_1 + g_2}{(1 - g_1)^2} \text{ and } \mathrm{Var}[X] = \frac{f_2}{(1 - g_1)^2} + \frac{f_1 g_2 + g_3}{(1 - g_1)^3} + \frac{2 g_2^2}{(1 - g_1)^4},$$

respectively.

A special case of $\mathrm{GLD}_1(f, g)$ includes the generalized negative binomial distribution (GNBD) defined by Jain and Consul (1971) which has the pmf

$$\mathrm{P}(X = x) = \frac{\nu}{\nu + \beta x}\binom{\nu + \beta x}{x} p^x q^{\nu + \beta x - x}, \ x = 0, 1, \ldots, \tag{2.8}$$

where $\nu > 0$, $0 < p = 1 - q < 1$ and $p \leq \beta p < 1$ or $\beta = 0$ and ν is a non-negative integer. This distribution is generated through different sets of functions. Examples include the case $f(t) = (q + pt)^\nu$ and $g(t) = (q + pt)^\beta$ and the case $f(t) = (q/(1 - pt))^\nu$ and $g(t) = (q/(1 - pt))^{\beta-1}$. Likewise the linear function negative binomial distribution (LFNBD) defined by Charalambides (1986) which has the pmf

$$\mathrm{P}(X = x) = (1 - \beta p)\binom{\nu + \beta x}{x} p^x q^{\nu+\beta x - x}, \ x = 0, 1, \ldots, \tag{2.9}$$

where $\nu > 0$, $0 < p = 1 - q < 1$ and $p \leq \beta p < 1$ or $\beta = 0$ and ν is a non-negative integer, is a special case of $\mathrm{GLD}_2(f, g)$ and generated through different sets of functions. Examples include the case $f(t) = (q + pt)^\nu$ and $g(t) = (q + pt)^\beta$ and the case $f(t) = (q/(1 - pt))^{\nu+1}$ and $g(t) = (q/(1 - pt))^{\beta-1}$. In the next section, we study the extensions of these facts. The theorems show that any Lagrangian distribution can be generated through different sets of functions. These theorems have a merit that $\mathrm{GLD}_1(f, g)$ and $\mathrm{GLD}_2(f, g)$ with complicated functions $f(t)$ and $g(t)$ may be expressed by $\mathrm{GLD}_1(f, g)$ and $\mathrm{GLD}_2(f, g)$ with simple functions $f(t)$ and $g(t)$, respectively. This merit enables us to give practical situations for the processes associated with general Lagrangian distributions.

2.3 Properties of GLD_1 and GLD_2

2.3.1 Formulations through different sets of functions

As stated in Section 2.2, the GNBD is generated by $\mathrm{GLD}_1(f, g)$ with $f(t) = (q+pt)^\nu$ and $g(t) = (1+pt)^\beta$ or with $f(t) = (q/(1-pt))^\nu$ and $g(t) = (q/(1-pt))^{\beta-1}$. This is easily seen by calculating the pmf from the definition (2.5). The GNBD is also generated by $\mathrm{GLD}_1(f, g)$ with $f(t) = ((1 - \sqrt{1 - 4pqt})/(2pt))^\nu$ and $g(t) = ((1 - \sqrt{1 - 4pqt})/(2pt))^{\beta-2}$, where $((1 - \sqrt{1 - 4pqt})/(2pt))^n$ is the pgf of a special case of GNBD with $\nu = n$ and $\beta = 2$ in (2.8), known as the lost-games distribution (Kemp and Kemp, 1968) or inverse binomial distribution (Yanagimoto, 1989). However, it does not look that this fact is easily obtained by calculating the pmf from (2.5) because of the complexity of the pgf of inverse binomial distribution. We can get the result as an example below using the following theorem.

Theorem 2.1

Suppose $G(t)$, $h_1(t)$ and $h_2(t)$ are analytic functions in the domain $|t| < 1$ such that $G(1) = h_1(1) = h_2(1) = 1$, $h_1(0) > 0$, $h_2(0) > 0$, $h_2'(1) < 1$, $h_1'(1)+h_2'(1) < 1$, $h_1(t)$, $h_2(t)$, $h_1'(t)$ and $h_2'(t)$ are non-decreasing functions in $0 \leq t \leq 1$ and $[\mathrm{D}^{x-1}((h_1(t)h_2(t))^x \mathrm{D}G(t))]_{t=0} \geq 0$ for $x = 0, 1, \ldots$. Let $f(t)$ and $g(t)$ be the pgf's of $\mathrm{GLD}_1(G, h_2)$ and $\mathrm{GLD}_1(h_1, h_2)$, respectively. Then $\mathrm{GLD}_1(f, g)$ is equal to $\mathrm{GLD}_1(G, h_1 h_2)$.

Proof

Since $f(t)$ and $g(t)$ are the pgf's of $\mathrm{GLD}_1(G, h_2)$ and $\mathrm{GLD}_1(h_1, h_2)$, respectively, $f(t) = G(z)$ and $g(t) = h_1(z)$, where $t = z/h_2(z)$, and, under the assumption, it is satisfied that $f(t)$ and $g(t)$ are analytic functions in the domain $|t| < 1$ such that $f(1) = g(1) = 1$, $g(0) > 0$, $g'(1) < 1$, and $g(t)$ and $g'(t)$ are non-decreasing functions in $0 \leq t \leq 1$. The pgf $L(u)$ of $\mathrm{GLD}_1(f, g)$ is expressed as $L(u) = f(t)$, where $u = t/g(t)$. Therefore, it holds that

$$L(u) = G(z), \text{ where } u = \frac{t}{h_1(z)} \text{ and } t = \frac{z}{h_2(z)},$$

or

$$L(u) = G(z), \text{ where } u = \frac{z}{h_1(z)h_2(z)}.$$

The function $L(u)$ is seen to be the pgf of $\mathrm{GLD}_1(G, h_1 h_2)$.

□

We give an example of Theorem 2.1. Put $G(t) = (q + pt)^\nu$, $h_1(t) = (q + pt)^{\beta-2}$ and $h_2(t) = (q + pt)^2$. Then $f(t) = ((1 - \sqrt{1 - 4pqt})/(2pt))^\nu$ and $g(t) = ((1 - \sqrt{1 - 4pqt})/(2pt))^{\beta-2}$. From Theorem 2.1, we can easily see that $\mathrm{GLD}_1(f, g)$ with $f(t) = ((1 - \sqrt{1 - 4pqt})/(2pt))^\nu$ and $g(t) = ((1 - \sqrt{1 - 4pqt})/(2pt))^{\beta-2}$ is equal to $\mathrm{GLD}_1(f, g)$ with $f(t) = (q + pt)^\nu$ and $g(t) = (q + pt)^\beta$.

The following theorem for GLD_2 provides a similar result to Theorem 2.1.

Theorem 2.2

Suppose $G(t)$, $h_1(t)$ and $h_2(t)$ are analytic functions in the domain $|t| < 1$ such that $G(1) = h_1(1) = h_2(1) = 1$, $h_1(0) > 0$, $h_2(0) > 0$, $h_2'(1) < 1$, $h_1'(1)+h_2'(1) < 1$, $h_1(t)$, $h_2(t)$, $h_1'(t)$ and $h_2'(t)$ are non-decreasing functions in $0 \leq t \leq 1$ and $[\mathrm{D}^x((h_1(t)h_2(t))^x f(t))]_{t=0} \geq 0$ for $x = 0, 1, 2, \ldots$. Let $f(t)$ and $g(t)$ be the pgf's of $\mathrm{GLD}_2(G, h_2)$ and $\mathrm{GLD}_1(h_1, h_2)$, respectively. Then $\mathrm{GLD}_2(f, g)$ is equal to $\mathrm{GLD}_2(G, h_1h_2)$.

Proof.

Since $f(t)$ and $g(t)$ are the pgf's of $\mathrm{GLD}_2(G, h_2)$ and $\mathrm{GLD}_1(h_1, h_2)$, respectively, $f(u) = (1 - h_2'(1))G(t)/(1 - th_2'(t)/h_2(t))$ and $g(u) = h_1(t)$, where $u = t/h_2(t)$, and, under the assumption, it is satisfied that $f(t)$ and $g(t)$ are analytic functions in the domain $|t| < 1$ such that $f(1) = g(1) = 1$, $g(0) > 0$, $g'(1) < 1$, and $g(t)$ and $g'(t)$ are non-decreasing functions in $0 \leq t \leq 1$. From the expressions of pgf's (2.3) and (2.4), we see that $\mathrm{GLD}_2(f, g)$ is equal to $\mathrm{GLD}_1(L, g)$, where $L(t) = (1 - g'(1))f(t)/(1 - tg'(t)/g(t))$. It follows from the expression of $g(u)$ that

$$\frac{\partial}{\partial u} g(u) = h_1'(t) \frac{\partial t}{\partial u} = \frac{h_1'(t)(h_2(t))^2}{h_2(t) - th_2'(t)}, \text{ where } u = \frac{t}{h_2(t)}.$$

From Lemma 2.1, under the assumption, $u = 1$ if and only if $t = 1$ for the equation $u = t/h_2(t)$. Therefore, it holds that

$$\begin{aligned}
L(u) &= \frac{1-g'(1)}{1-ug'(u)/g(u)} f(u) \\
&= \frac{1-h_1'(1)/(1-h_2'(1))}{1-th_1'(t)h_2(t)^2/(h_1(t)h_2(t)(h_2(t)-th_2'(t)))} \frac{1-h_2'(1)}{1-th_2'(t)/h_2(t)} G(t) \\
&= \frac{(1-h_1'(1)-h_2'(1))G(t)}{1-t\mathrm{D}(h_1(t)h_2(t))/(h_1(t)h_2(t))}, \text{ where } u = \frac{t}{h_2(t)}.
\end{aligned}$$

The function $L(u)$, where $u = t/h_2(t)$, is seen to be the pgf of $\mathrm{GLD}_1(H, h_2)$, where

$$H(t) = \frac{(1-h_1'(1)-h_2'(1))G(t)}{1-t\mathrm{D}(h_1(t)h_2(t))/(h_1(t)h_2(t))}.$$

From Theorem 2.1, $\mathrm{GLD}_1(L, g)$ is equal to $\mathrm{GLD}_1(H, h_1h_2)$ and this is equal to $\mathrm{GLD}_2(G, h_1h_2)$. Therefore, we get the result that $\mathrm{GLD}_2(f, g)$ is equal to $\mathrm{GLD}_2(G, h_1h_2)$.

□

2.3.2 Some other results

Suppose that ν and β are non-negative integers and $h(t)$ is analytic function in the domain $|t| < 1$ such that $h(1) = 1$, $h(0) > 0$, and $h(t)$ and $h'(t)$ are non-decreasing functions. Put $h(t)^n = \sum_{x=0}^{\infty} p(x; n)t^x$ in this section.

Theorem 2.3
Put $A(t) = t^\nu$ and $B(t) = h(t^\beta)$.

- Let $X_1 \sim \mathrm{GLD}_1(h^\nu, h^\beta)$ and $Y_1 \sim \mathrm{GLD}_1(A, B)$. Then the distribution of Y_1 is identical with that of $\nu + \beta X_1$.

- Let $X_2 \sim \mathrm{GLD}_2(h^\nu, h^\beta)$ and $Y_2 \sim \mathrm{GLD}_2(A, B)$. Then the distribution of Y_2 is identical with that of $\nu + \beta X_2$.

The first statement can be seen in Sibuya et al. (1994), whose result was obtained through the viewpoint of a queueing process. A proof using general Lagrangian distributions is given here.

Proof.
From (2.5),

$$\mathrm{P}(Y_1 = x) = \frac{1}{x!}[\mathrm{D}^{x-1}(\nu t^{\nu-1} h(t^\beta)^x)]_{t=0} = \frac{\nu}{x} p\left(\frac{x-\nu}{\beta}; x\right).$$

Hence, it can be seen that

$$\mathrm{P}(Y_1 = \nu + \beta x) = \frac{\nu}{\nu + \beta x} p(x; \nu + \beta x) = \mathrm{P}(X_1 = x).$$

Similarly, from (2.7),

$$\mathrm{P}(Y_2 = x) = \frac{1 - \beta h'(1)}{x!}[\mathrm{D}^{x}(t^{\nu}(h(t^\beta)^x)]_{t=0} = (1 - \beta h'(1)) p\left(\frac{x-\nu}{\beta}; x\right).$$

Hence, it can be seen that

$$\mathrm{P}(Y_2 = \nu + \beta x) = (1 - \beta h'(1)) p(x; \nu + \beta x) = \mathrm{P}(X_2 = x).$$

From these results, the statements are proved.

□

Theorem 2.3 also states that a Lagrangian distribution, except for scale and location, can be generated through different sets of functions. When $\nu = \beta = 1$, this theorem reduces to Theorem 2.6 in Consul and Famoye (2006). The following is an application of Theorem 2.3.

Theorem 2.4
Assume that $X_1/\beta, X_2/\beta, \ldots$ are independent and identically distributed (i.i.d.) with pgf $h(t)$. Put $N_n := \nu + X_1 + \cdots + X_n$, $I := (\inf\{n : N_n = n\} - \nu)/\beta$ and

$S := (\sup\{n : N_n = n\} - \nu)/\beta$. If $\beta h'(1) \leq 1$, then

$$\mathrm{P}(I = x) = \frac{\nu}{\nu + \beta x} p(x; \nu + \beta x),\ x = 0, 1, \ldots$$

and

$$\mathrm{P}(S = x) = (1 - \beta h'(1)) p(x; \nu + \beta x),\ x = 0, 1, \ldots.$$

Proof.

Put $A(t) = t^{\nu}$ and $B(t) = h(t^{\beta})$. The random variable $\inf\{n : N_n = n\}$ is distributed as $\mathrm{GLD}_1(A, B)$ from Theorem 4 in Section 8 in Takács (1967). From Theorem 1 in Section 5 in Takács (1967), the random variable $\sup\{n : N_n = n\}$ is distributed as $\mathrm{GLD}_2(A, B)$. Hence, from Theorem 2.3, it can be seen that the random variables I and S are distributed as $\mathrm{GLD}_1(h^{\nu}, h^{\beta})$ and $\mathrm{GLD}_2(h^{\nu}, h^{\beta})$, respectively.

$\square$

These distributions, $\mathrm{GLD}_1(h^{\nu}, h^{\beta})$ and $\mathrm{GLD}_2(h^{\nu}, h^{\beta})$, have several applications as models in the field of dam, storage and insurance risk processes. Special cases $h(t) = q + pt$, $h(t) = q/(1 - pt)$ and $h(t) = \mathrm{e}^{\lambda(t-1)}$ are mentioned in Charalambides (1986).

2.4 Tandem queueing system

The two counter tandem queues attended by a single moving server were introduced by Nelson (1968) and Nair (1971). There are several practical applications of the tandem queues like robotic systems, network systems and telecommunication systems (cf. Nelson, 1968; Katayama, 1992; Van Oyen and Teneketzis, 1994). In such queues, the server attends to the counters according to some service policies. In this section, we consider the system where customers at counter 1 have priority over customers at counter 2, defined as follows.

(i) the number of initial customers at counter 1 follows the distribution with pgf $G(t)$ and that at counter 2 is zero.

(ii) the number of arrivals at counter 1 during a service at counter 1 follows the distribution with pgf $h_1(t)$ and customers served at counter 1 go to counter 2.

(iii) if there is no customer at counter 1, the server goes to counter 2.

(iv) the number of arrivals at counter 1 during a service at counter 2 follows the distribution with pgf $h_2(t)$.

(v) the server comes back to counter 1 after a service at counter 2.

(vi) this process continues until there are no customers at both counters.

When we look at this process from the side of counter 2, the number of initial customers is distributed as $\mathrm{GLD}_1(G, h_1)$ and that of arrivals during a service as $\mathrm{GLD}_1(h_2, h_1)$. Therefore, the number of customers served during a busy period is distributed as $\mathrm{GLD}_1(f, g)$, where $f(t)$ and $g(t)$ are the pgf's of $\mathrm{GLD}_1(G, h_1)$ and $\mathrm{GLD}_1(h_2, h_1)$, respectively, and from Theorem 2.1, this distribution is equal to $\mathrm{GLD}_1(G, h_1h_2)$.

The GLD_1 also has the relation with the M/G/1 queue which is a single server queue with Poisson arrivals and arbitrary service time distribution. For such a queue, when the number of initial customers follows the distribution with pgf $L(t)$, the number of customers served during a busy period is distributed as $\mathrm{GLD}_1(L, M)$, where $M(t) = m(\lambda(t-1))$, λ is the mean arrival rate and $m(t)$ is the mgf of service time distribution. Using this fact, we can consider the following steps, alternative to (ii) or (iv).

(ii)' the mgf of the service time distribution at counter 1 is $m_1(t)$ and then the mean arrival rate at counter 1 is λ_1 and customers served at counter 1 go to counter 2.

(iv)' the mgf of the service time distribution at counter 2 is $m_2(t)$ and then the mean arrival rate at counter 1 is λ_2.

Put $M_1(t) = m_1(\lambda_1(t-1))$ and $M_2(t) = m_2(\lambda_2(t-1))$. In the case with changing both (ii) to (ii)' and (iv) to (iv)', the number of customers served during a busy period is distributed as $\mathrm{GLD}_1(f,g)$, where $f(t)$ and $g(t)$ are the pgf's of $\mathrm{GLD}_1(G, M_1)$ and $\mathrm{GLD}_1(M_2, M_1)$, respectively. From Theorem 2.1, this distribution is equal to $\mathrm{GLD}_1(G, M_1M_2)$.

For example, assume that Poisson arrivals with mean arrival rate 1 during the services and service time at counter 1 is distributed as an exponential distribution with mean ρ_1 and that at counter 2 is distributed as an exponential distribution with mean ρ_2, or $G(t) = t$, $\lambda_1 = \lambda_2 = 1$ and $m_i(t) = 1/(1-\rho_i t)$ for $i = 1, 2$. This assumption is analogous to that given by Nelson (1968). In this case, the number of customers served during a busy period is distributed as $\mathrm{GLD}_1(f,g)$ with

$$f(t) = \frac{1+\rho_1 - \sqrt{(1+\rho_1)^2 - 4\rho_1 t}}{2\rho_1}$$

and

$$g(t) = \frac{2\rho_1}{2\rho_1 + \rho_1\rho_2 - \rho_2 + \rho_2\sqrt{(1+\rho_1)^2 - 4\rho_1 t}}.$$

It is hard to calculate the pmf from the above expression. However, by using Theorem 2.1, we can see that this distribution is equal to $\mathrm{GLD}_1(f,g)$ with $f(t) = t$ and $g(t) = 1/((1-\rho_1 t)(1-\rho_2 t))$. From the formula (4.6) in Kemp (1979):

$$\frac{(1-Q_1 z)^{U_1}(1-Q_2 z)^{U_2}}{(1-Q_1)^{U_1}(1-Q_2)^{U_2}} = (1-Q_1)^{-U_1}(1-Q_2)^{-U_2}\sum_{r\geq 0}\binom{U_1+U_2}{r}$$
$$\times {}_2F_1[-U_1, -U_1-U_2+r; -U_1-U_2; (Q_1-Q_2)/Q_1](Q_2/Q_1)^{r-U_1}(-Q_1)^r,$$

the number of customers, X, served during a busy period is seen to have the pmf

$$\mathrm{P}(X = x) = \frac{1}{3x-1}\binom{3x-1}{x}\frac{P_2^{2x-1}}{P_1^x}(1-P_1)^x(1-P_2)^x \\ \times {}_2F_1[x, 3x-1; 2x; (P_1 - P_2)/P_1], \;\; x = 0, 1, \ldots,$$

where $P_i = \rho_i/(1-\rho_i)$ for $i = 1, 2$.

Chapter 3

Lagrangian non-central negative binomial distribution

3.1 Introduction

In this chapter, the non-central negative binomial distribution (Ong and Lee, 1979) and Charlier series distribution (Ong, 1988) are formulated as general Lagrangian distributions and their generalization is considered through the concept of general Lagrangian distributions.

The non-central negative binomial distribution (NNBD) arises as a model in photon and neural counting, birth and death process and a mixture of Poisson distribution with the parameter distributed as non-central gamma distribution. It has pgf

$$G(t) = \left(\frac{q}{1-pt}\right)^{\nu} \exp\left(\lambda\left(\frac{q}{1-pt} - 1\right)\right),$$

where ν, $\lambda > 0$ and $0 < p = 1 - q < 1$. The pmf is given by

$$\mathrm{P}(X = x) = \mathrm{e}^{-\lambda p} p^x q^{\nu} L_x^{(\nu-1)}(-\lambda q),\ \ x = 0, 1, \ldots, \tag{3.1}$$

where $L_n^{(\alpha)}(z) = \binom{n+\alpha}{n} {}_1F_1[-n; \alpha+1; z]$ is the generalized Laguerre polynomial. The pmf (3.1) satisfies the three-terms recursive formula

$$(x+1)P_{x+1} = (2x + \nu + \lambda q)pP_x - p^2(x + \nu - 1)P_{x-1}, \ x \geq 0, \tag{3.2}$$

with $P_{-1} = 0$ and $P_0 = \mathrm{e}^{-\lambda q} q^\nu$, where $P_x = \mathrm{P}(X = x)$ in (3.1). When the random variable X has the pmf (3.1), we write $X \sim \mathrm{NNBD}(\nu, \lambda, p)$.

The Charlier series distribution (CSD) has pgf

$$G(t) = (q + pt)^N \exp(\lambda(q + pt - 1)),$$

where N is a non-negative integer, $\lambda > 0$ and $0 < p = 1 - q < 1$. The pmf is given by

$$\begin{aligned} \mathrm{P}(X = x) &= \mathrm{e}^{-\lambda p}(\lambda p)^x q^N C_x(N; -\lambda q)/x! \\ &= \mathrm{e}^{-\lambda p} p^x q^{N-x} L_x^{(N-x)}(-\lambda q), \ x = 0, 1, \ldots, N, \end{aligned} \tag{3.3}$$

where $C_r(N; -\lambda) = \dfrac{N!\lambda^r}{(N-r)!} {}_1F_1[-r; N - r + 1; -\lambda]$ is the Charier polynomial. The pmf (3.3) satisfies the three-terms recursive formula

$$(x+1)qP_{x+1} = p(N + \lambda q - x)P_x + \lambda p^2 P_{x-1}, \quad x \geq 0, \tag{3.4}$$

with $P_{-1} = 0$ and $P_0 = \mathrm{e}^{-\lambda q} q^N$, where $P_x = \mathrm{P}(X = x)$ in (3.3). When the random variable X has the pmf (3.3), we write $X \sim \mathrm{CSD}(N, \lambda, p)$.

Put $f_n(t) = (q + pt)^n \exp(\lambda(q + pt - 1))$, $g_n(t) = (q + pt)^n$, $v_n(t) = (q/(1 - pt))^{n+1} \exp(\lambda(q/(1 - pt) - 1))$ and $w_n(t) = (q/(1 - pt))^n)$ in this chapter. From (2.3) and (2.4), the pgf of $\mathrm{NNBD}(\nu, \lambda, p)$ is correspond with pgf of $\mathrm{GLD}_1(f_\nu, g_1)$ and that of $\mathrm{GLD}_2(f_{\nu-1}, g_1)$. The pgf of $\mathrm{CSD}(N, \lambda, p)$ obviously corresponds with pgf of $\mathrm{GLD}_1(f_N, g_0)$ and that of $\mathrm{GLD}_2(f_N, g_0)$. It is of interest to consider the Lagrangian distributions which include the NNBD and CSD. Ong et al. (2012) considered the Lagrangian distribution of the second kind that includes the NNBD and CSD, defined as $\mathrm{GLD}_2(f_N, g_M)$ for N, $\lambda > 0$, $0 < p = 1 - q < 1$ and $p \leq Mp < 1$ or $M = 0$ and a non-negative integer N. Obviously, this distribution reduces to $\mathrm{NNBD}(\nu, \lambda, p)$ when $N = \nu - 1$ and $M = 1$, $\mathrm{CSD}(N, \lambda, p)$ when $M = 0$ and LFNBD with pmf (2.9) when $N = \nu$, $M = \beta$ and $\lambda \to 0$. This distribution is also generated through the pgf's of $\mathrm{GLD}_2(f_N, g_\alpha)$ and $\mathrm{GLD}_1(g_{M-\alpha}, g_\alpha)$. This is seen by putting $G(t) = (q + pt)^N \exp(\lambda(q + pt - 1))$, $h_1(t) = (q + pt)^{M-\alpha}$ and $h_2(t) = (q + pt)^\alpha$ in Theorem 2.2. Ong et al. (2012) showed the fact for a special case when $\alpha = 1$, i.e., $\mathrm{GLD}_2(f_N, g_M)$ is equal to $\mathrm{GLD}_2(v_{N+1}, w_{M-1})$. Other properties like the generations as mixture distributions, some expressions of pmf, moments and its generalizations are studied in Ong et al. (2012).

In subsequent sections, we consider the Lagrangian distribution of the first kind that includes the NNBD and CSD, defined as $\mathrm{GLD}_1(f_\nu, g_\beta)$ for ν, $\lambda > 0$, $0 < p = 1 - q < 1$ and $p \leq \beta p < 1$ or $\beta = 0$ and a non-negative integer ν (Imoto and Ong, 2013). It is clear that this distribution reduces to $\mathrm{NNBD}(\nu, \lambda, p)$ when $\beta = 1$, $\mathrm{CSD}(N, \lambda, p)$ when $\beta = 0$ and GNBD with pmf (2.8) when $\lambda \to 0$. Properties of the proposed distribution are also shown. Section 3.2 provides four expressions of the pmf from the direct use of (2.5), three of which is expressed as the sum of two generalized Laguerre polynomial and the other is expressed by a generalized hypergeometric function. Section 3.3 shows that insight about the stopped-sum distributions (Johnson et al., 2005, p. 382) leads to the expressions of the pmf in terms of non-central negative binomial distributions and those of Charlier series distributions. The relationships between the NNBD and CSD are also studied in this section. In Section 3.4, we study about mixture, the index of dispersion, recursive formulas of the pmf and the relation with queueing systems. The fitting example to real count data set is given in Section 3.5.

3.2 Lagrangian non-central negative binomial distribution

From (2.5), we obtain the pmf of $\mathrm{GLD}_1(f_\nu, g_\beta)$ as

$$\mathrm{P}(X=x)=\begin{cases} \mathrm{e}^{-\lambda p}q^{\nu}, \ x=0 \\ \dfrac{\mathrm{e}^{-\lambda p}p^x q^{\nu+\beta x-x}}{x}\left(\nu L_{x-1}^{(\nu+\beta x-x)}(-\lambda q)+\lambda q L_{x-1}^{(\nu+\beta x-x+1)}(-\lambda q)\right), \\ \qquad\qquad\qquad\qquad\qquad\qquad x=1,2,\ldots. \end{cases} \tag{3.5}$$

The recursive formulas of the generalized Laguerre polynomials,

$$L_n^{(\alpha)}(x)=L_n^{(\alpha+1)}(x)-L_{n-1}^{(\alpha+1)}(x) \text{ and } nL_n^{(\alpha)}(x)=(n+\alpha)L_{n-1}^{(\alpha)}(x)-xL_{n-1}^{(\alpha+1)}(x),$$

lead to two expressions

$$\mathrm{P}(X=x)=\mathrm{e}^{-\lambda p}p^x q^{\nu+\beta x-x}\left(L_x^{(\nu+\beta x-x-1)}(-\lambda q)-(\beta-1)L_{x-1}^{(\nu+\beta x-x)}(-\lambda q)\right) \\ x=0,1,\ldots \tag{3.6}$$

and

$$\mathrm{P}(X=x)=\mathrm{e}^{-\lambda p}p^x q^{\nu+\beta x-x}\left(L_x^{(\nu+\beta x-x)}(-\lambda q)-\beta L_{x-1}^{(\nu+\beta x-x)}(-\lambda q)\right) \\ x=0,1,\ldots. \tag{3.7}$$

It is easily seen that (3.6) reduces to the NNBD's pmf (3.1) when $\beta=1$ and (3.7) reduces to the CSD's pmf (3.3) when $\beta=0$ and ν is a non-negative integer.

Since we have

$$(q+pt)^{\nu}\exp(\lambda(q+pt-1))=\mathrm{e}^{-\lambda}(q+pt)^{\nu}{}_1F_1[\nu+1;\nu+1;\lambda(q+pt)],$$

using the formulas

$$\mathrm{D}^n(t^{c-1}{}_1F_1[a;c;t]) = (-1)^n(1-c)_n t^{c-1-n}{}_1F_1[a;c-n;t]$$

(Erdélyi et al., 1953, p. 255) and

$$\mathrm{D}^n(t^{\delta}{}_1F_1[a;c;t]) = (\delta-n+1)_n t^{\delta-n}{}_2F_2\left[a,\delta+1;c,\delta+1-n;t\right]$$

(Luke, 1969, p. 117) lead to another expression of the pmf

$$\begin{aligned}\mathrm{P}(X=x) = {} & \frac{\nu}{\nu+\beta x}\binom{\nu+\beta x}{x}\mathrm{e}^{-\lambda}p^x q^{\nu+\beta x-x} \\ & \times{}_2F_2[\nu+1,\nu+\beta x;\nu,\nu+\beta x-x+1;\lambda q],\ x=0,1,\ldots.\end{aligned} \tag{3.8}$$

It is easily seen that (3.8) reduces to GNBD's pmf (2.8) when $\lambda=0$.

Moreover, it is easily seen that this distribution reduces to the binomial-Poisson distribution (Consul and Shenton, 1972), generated by $\mathrm{GLD}_1(f,g)$ with $f(t)=\mathrm{e}^{\lambda(t-1)}$ and $g(t)=(q+pt)^{\beta}$, when $\nu=0$ and reduces to generalized Poisson distribution (Consul and Jain, 1973), generated by $\mathrm{GLD}_1(f,g)$ with $f(t)=\mathrm{e}^{\lambda(t-1)}$ and $g(t)=\mathrm{e}^{\mu(t-1)}$, when $\nu=0$ and $p\to 0$ with $\beta p=\mu$.

This distribution also arises from the following way. Letting $f(z)=(1+z)^{\nu}\mathrm{e}^{\lambda qz}$ and $g(z)=(1+z)^{\beta}$ in (2.1), dividing the equation by $(1+z)^{\beta}$ and putting $z=p/q$ lead to

$$\mathrm{e}^{-\lambda p}q^{\nu}+\sum_{x=1}^{\infty}\frac{\mathrm{e}^{-\lambda p}p^x q^{\nu+\beta x-x}}{x}(\nu L_{x-1}^{(\nu+\beta x-x)}(-\lambda q)+\lambda q L_{x-1}^{(\nu+\beta x-x+1)}(-\lambda q))=1.$$

Each term on the left hand side of the above equation coincides with pmf (3.5). We call this distribution the Lagrangian non-central negative binomial distribution of the first kind and denote it by $\mathrm{LNNBD}_1(\beta,\nu,\lambda,p)$ or LNNBD_1 for short.

$\mathrm{LNNBD}_1(\beta,\nu,\lambda,p)$ is also generated through the pgf's of $\mathrm{LNNBD}_1(\alpha,\nu,\lambda,p)$

and GNBD$(\alpha, \beta - \alpha, p)$, where GNBD$(\beta, \nu, p)$ means the GNBD with pmf (2.8). This is seen by putting $G(t) = (q + pt)^\nu \exp(\lambda(q + pt - 1))$, $h_1(t) = (q + pt)^{\beta-\alpha}$ and $h_2(t) = (q + pt)^\alpha$ in Theorem 2.1. From a special case when $\alpha = 1$, we see that LNNBD$_1(\beta, \nu, \lambda, p)$ is generated by GLD$_1(v_\nu, w_{\beta-1})$.

3.3 Stopped-sum distributions

Suppose that $Z_1, Z_2, \ldots \overset{\text{i.i.d.}}{\sim} \mathcal{F}_Z$, $X \sim \mathcal{F}_X$, and these random variables are independent, and let $S = Z_1 + \cdots + Z_X$ if $X = 1, 2, \ldots$ and $S = 0$ otherwise. Then $S \sim \mathcal{F}_S$ is called the stopped-sum distribution or $\mathcal{F}_X$ generalized by the generalizing $\mathcal{F}_Z$ (see Johnson et al., 2005, Chapter 9). Let $G_Z(t)$, $G_X(t)$ and $G_S(t)$ be the pgf's of $\mathcal{F}_Z$, $\mathcal{F}_X$ and $\mathcal{F}_S$, respectively. Then it is satisfied $G_S(t) = G_X(G_Z(t))$.

Since the LNNBD$_1$ is generated by GLD$_1(f_\nu, g_\beta)$, LNNBD$_1$ is a CSD generalized by the generalizing Consul distribution, where the Consul distribution (Consul and Shenton, 1975) is generated by GLD$_1(d_1, g_\beta)$, where $d_n(t) = t^n$. Therefore, letting $X \sim$ LNNBD$_1(\beta, N, \lambda, p)$, $Y_{CS} \sim$ CSD(N, λ, p) and C_k be a random variable distributed as GLD$_1(d_k, g_\beta)$, i.e.,

$$\mathrm{P}(C_k = x) = \frac{k}{x}\binom{\beta x}{x-k} p^{x-k} q^{\beta x - x + k}, \ x = k, k+1, \ldots \tag{3.9}$$

with $\mathrm{P}(C_0 = 0) = 1$, then we get the expression

$$\mathrm{P}(X = x) = \sum_{k=0}^{x} \mathrm{P}(C_k = x)\mathrm{P}(Y_{CS} = k), \ x = 0, 1, \ldots. \tag{3.10}$$

Here N is not necessarily a non-negative integer, and we may put $N = \nu$. This expression is useful for computing the pmf of LNNBD$_1$ by using the recursive formulas (3.4) and

$$\begin{cases} \mathrm{P}(C_1 = x) = \dfrac{1}{x}\dbinom{\beta x}{x-1} p^{x-1} q^{\beta x - x + 1}, \\ \mathrm{P}(C_k = x) = \dfrac{k}{k-1}\dfrac{x-k+1}{\beta x - x + k}\dfrac{q}{p}\mathrm{P}(C_{k-1} = x), \ k = 2, 3, \ldots. \end{cases}$$

Similarly, since the LNNBD_1 is generated by $\mathrm{GLD}_1(v_\nu, w_{\beta-1})$, LNNBD_1 is a NNBD generalized by the generalizing Geeta distribution, where the Geeta distribution (Consul, 1990) is generated by $\mathrm{GLD}_1(d_1, w_\beta)$. Therefore, letting $Y_{NNB} \sim \mathrm{NNBD}(\nu, \lambda, p)$ and G_k be a random variable distributed as $\mathrm{GLD}_1(d_k, w_\beta)$, i.e.,

$$\mathrm{P}(G_k = x) = \frac{k}{x}\binom{\beta x - k - 1}{x - k} p^{x-k} q^{\beta x - x}, \ x = k, k+1, \ldots \tag{3.11}$$

with $\mathrm{P}(G_0 = 0) = 1$, then we get the expression

$$\mathrm{P}(X = x) = \sum_{k=0}^{x} \mathrm{P}(G_k = x)\mathrm{P}(Y_{NNB} = k), \ x = 0, 1, \ldots. \tag{3.12}$$

This expression is also useful for computing the pmf of LNNBD_1 by using the recursive formulas (3.2) and

$$\begin{cases} \mathrm{P}(G_1 = x) = \dfrac{1}{\beta x - 1}\dbinom{\beta x - 1}{x} p^{x-1} q^{\beta x - x}, \\ \mathrm{P}(G_k = x) = \dfrac{k}{k-1}\dfrac{x-k+1}{\beta x - k}\dfrac{1}{p}\mathrm{P}(G_{k-1} = x), \quad k = 2, 3 \ldots. \end{cases}$$

In particular, it holds that

$$\mathrm{P}(Y_{NNB} = x) = \sum_{k=0}^{x} \binom{x-1}{k-1} p^{x-k} q^k \mathrm{P}(Y_{CS} = k) \tag{3.13}$$

when $\beta = 1$ in (3.10) and

$$\mathrm{P}(Y_{CS} = x) = \sum_{k=0}^{x} \binom{x-1}{k-1} (-p)^{x-k} q^{-x} \mathrm{P}(Y_{NNB} = k) \tag{3.14}$$

when $\beta = 0$ in (3.12). From (3.13), we see that NNBD is a CSD generalized by the generalizing shifted geometric distribution. About (3.14), the term multiplied by $\mathrm{P}(Y_{NNB} = k)$ is the pmf of an invalid distribution. This is a shifted pseudo-binomial distribution by Kemp (1979). Thus, CSD is a NNBD generalized by the generalizing shifted pseudo-binomial distribution.

3.4 Properties and characteristics

3.4.1 Generalized negative binomial mixture

Let $X|N \sim \mathrm{GNBD}(\beta, \nu + N, p)$ and N be a random variable distributed as a Poisson distribution with mean λ. Then the unconditional distribution of X is

$$\begin{aligned}
&\mathrm{P}(X = x) \\
= \;& \sum_{n=0}^{\infty} \frac{\nu + n}{\nu + n + \beta x} \binom{\nu + n + \beta x}{x} p^x q^{\nu + n + \beta x - x} \frac{\mathrm{e}^{-\lambda} \lambda^n}{n!} \\
= \;& \frac{\mathrm{e}^{-\lambda} p^x q^{\nu + \beta x - x}}{x!} \sum_{n=0}^{\infty} \frac{\nu(\nu + \beta x - 1)!}{(\nu + \beta x - x)!} \frac{(\nu + 1)_n (\nu + \beta x)_n}{(\nu)_n (\nu + \beta x - x + 1)_n} \frac{(\lambda q)^n}{n!} \\
= \;& \frac{\nu}{\nu + \beta x} \binom{\nu + \beta x}{x} \mathrm{e}^{-\lambda} p^x q^{\nu + \beta x - x} {}_2F_2 \left[\nu + 1, \nu + \beta x; \nu, \nu + \beta x - x + 1; \lambda q \right].
\end{aligned}$$

This coincides with pmf (3.8).

3.4.2 Index of dispersion

From (2.6), the mean and variance of $\mathrm{LNNBD}_1(\beta, \nu, \lambda, p)$ are obtained as

$$\mu_1' = \frac{(\nu+\lambda)p}{1-\beta p} \text{ and } \mu_2 = \frac{(\nu q + \lambda - \beta\lambda p^2)p}{(1-\beta p)^3},$$

respectively. For $\beta \geq 1$, $\mathrm{LNNBD}_1(\beta, \nu, \lambda, p)$ is over-dispersed, i.e., the variance is greater than the mean. This fact can be proved as follows.

Let $X \sim \mathrm{LNNBD}_1(\beta, \nu, \lambda, p)$, $K \sim \mathrm{NNBD}(\nu, \lambda, p)$ and $Z \sim \mathrm{Gee}(\beta, p)$, where $\mathrm{Gee}(\beta, p)$ means the Geeta distribution with pmf (3.11) for $k = 1$. Since LNNBD_1 is a NNBD generalized by the generalizing Geeta distribution, the mean and variance of X are obtained as

$$\mathrm{E}[X] = \mathrm{E}[Z]\mathrm{E}[K] \text{ and } \mathrm{Var}[X] = \mathrm{E}[K]\mathrm{Var}[Z] + \mathrm{Var}[K]\mathrm{E}[Z]^2.$$

Since the NNBD is known to be over-dispersed, $\mathrm{Var}[K] > \mathrm{E}[K]$, and the random variable of Geeta distribution takes values greater than one, it is seen that

$$\mathrm{Var}[X] > \mathrm{Var}[K]\mathrm{E}[Z]^2 > \mathrm{E}[K]\mathrm{E}[Z] = \mathrm{E}[X].$$

However, when $\beta = 0$ and ν is a non-negative integer, LNNBD_1 reduces to CSD and LNNBD_1 is thus under-dispersed.

3.4.3 Recursive formulas

The three-terms recursive formulas (3.2) and (3.4) belong to a special case of Sundt's (1992) recursion known in actuarial science,

$$\mathrm{P}(X = x) = \sum_{i=1}^{k} \left(a_i + \frac{b_i}{x}\right) \mathrm{P}(X = x - i). \tag{3.15}$$

Sundt (1992) showed that the pmf of $S = Z_1 + Z_2 + \cdots + Z_X$, where Z_1, Z_2, $\ldots$ are independent and identically distributed random variables and independent of X, can be recursively evaluated by

$$f_S(s) = \frac{1}{1 - \sum_{i=1}^{k} a_i f_Z(0)^i} \sum_{z=1}^{s} f_S(s-z) \sum_{i=1}^{k} \left(a_i + \frac{b_i}{i} \frac{z}{s} \right) f_Z^{i*}(z),$$

with $f_S(0) = \sum_{n=0}^{\infty} \mathrm{P}(X = n) f_Z(0)^n$, where $f_S(s)$, $f_Z^{i*}(z)$ are the pmf's of S and $Z_1 + Z_2 + \cdots + Z_i$, respectively.

As described in Section 3.3, if $X \sim$CSD(ν, λ, p) and $Z_1, Z_2, \ldots \overset{\text{i.i.d.}}{\sim}Con(\beta, p)$, where Con$(\beta, p)$ means the Consul distribution with pmf (3.9) for $k = 1$ or if $X \sim$NNBD(ν, λ, p) and $Z_1, Z_2, \ldots \overset{\text{i.i.d.}}{\sim}Gee(\beta, p)$, then $S = Z_1 + Z_2 + \cdots + Z_X \sim$LNNBD$_1(\beta, \nu, \lambda, p)$. Therefore, we can get the two recursive formulas

$$\begin{aligned} \mathrm{P}(S = s) = &\sum_{z=1}^{s} \mathrm{P}(S = s - z) \\ &\times \left(\left(-\frac{p}{q} + \frac{(\nu + \lambda q + 1) p z}{q s} \right) \mathrm{P}(C_1 = z) + \frac{\lambda p^2 z}{2 q s} \mathrm{P}(C_2 = z) \right) \end{aligned}$$

and

$$\begin{aligned} \mathrm{P}(S = s) = &\sum_{z=1}^{s} \mathrm{P}(S = s - z) \\ &\times \left(\left(2p + \frac{(\nu + \lambda q - 2) p z}{s} \right) \mathrm{P}(G_1 = z) - \left(p^2 + \frac{(\nu - 2) p^2 z}{s} \right) \mathrm{P}(G_2 = z) \right) \end{aligned}$$

with $\mathrm{P}(S = 0) = \mathrm{e}^{-\lambda p} q^{\nu}$, where C_k and G_k for $k = 1, 2$ are defined by (3.9) and (3.11), respectively.

3.4.4 Queueing systems

From the relation of GLD_1 with queueing system, we see that the LNNBD_1 arises as the number of customers served during a busy period when the initial customers follows the CSD and the number of arrivals during a service follows binomial distribution or when the initial customers follows the NNBD and the number of arrivals during a service follows negative binomial distribution.

From the relation of GLD_1 with tandem queueing system in which customers at counter 1 have priority over customers at counter 2, introduced in Section 2.4, we see that LNNBD_1 arises as the number of customers served during a busy period when initial customers follows the CSD and the number of customers during a service in each counter follows the binomial distribution or when initial customers follows the NNBD and the number of customers during a service in each counter follows the negative binomial distribution.

3.5 Numerical examples

In this section, LNNBD_1 is fitted to the data of the number of Corbet's Malayan Butterfly with zeros (Blumer, 1974) in Table 3.1 by employing maximum likelihood method. Here the log-likelihood function is $\sum_{r=0}^{n} f_r \log \mathrm{P}(X = r)$, where $X \sim \mathrm{LNNBD}_1(\beta, \nu, \lambda, p)$, f_r is the observed frequency of the value r and n is the highest value in observed data. The numerical optimization of the log-likelihood function is used since solving likelihood equations is difficult because of the complicated form of the pmf. We also demonstrate comparative fittings with GNBD and NNBD. The performances of the model fittings are compared by chi-squared goodness-of-fit statistic (χ_2) and maximized log-likelihood statistic (Log L). The cells whose expected number is less than 5 are grouped such as the expected number of grouped cell is not less than 5.

In this data set, the LNNBD_1 provides a best fit in the sense of log-likelihood value and the GNBD provides a best fit in the sense of p-value. However, a comparison of the expected frequencies of LNNBD_1 and GNBD shows a closer fit by the LNNBD_1. The performance in the fit to the frequency count data exemplifies the viability of the LNNBD_1 as a model for count data.

Table 3.1: The number of Corbet's Malayan Butterfly with zeros (Blumer, 1974)

Count	Observed	GNBD	NNBD	LNNBD_1
0	304	307.17	315.36	306.91
1	118	106.99	94.23	107.96
2	74	66.70	59.75	66.66
3	44	48.26	44.57	47.99
4	24	37.50	35.73	37.22
5	29	30.42	29.85	30.17
6	22	25.39	25.60	25.20
7	20	21.65	22.36	21.51
8	19	18.75	19.81	18.66
9	20	16.45	17.72	16.39
10	15	14.58	15.99	14.55
11	12	13.04	14.53	13.03
12	14	11.74	13.27	11.75
13	6	10.64	12.18	10.67
14	12	9.70	11.23	9.73
15	6	8.88	10.38	8.92
16	9	8.16	9.63	8.21
17	9	7.54	8.96	7.59
18	6	6.98	8.35	7.04
19	10	6.48	7.80	6.54
20	10	6.04	7.30	6.10
21	11	5.64	6.84	5.71
22	5	5.28	6.42	5.35
23	3	4.96	6.04	5.02
24	3	4.66	5.69	4.72
25+	119	120.39	114.41	120.40
Total	924	924.00	924.00	924.00
Log L		-2253.12	-2256.12	-2252.80
χ^2		24.13	28.50	23.40
df		21	21	20
p-value		0.286	0.127	0.270

GNBD: $\hat{\beta} = 1.0441$, $\hat{\nu} = 0.4218$, $\hat{p} = 0.9265$.
NNBD: $\hat{\nu} = 0.3082$, $\hat{\lambda} = 0.0012$, $\hat{p} = 0.9693$.
LNNBD_1: $\hat{\beta} = 1.1630$, $\hat{\nu} = 0.5106$, $\hat{\lambda} = 0.3092$, $\hat{p} = 0.8112$.

Chapter 4

Extension of generalized inverse trinomial distribution

4.1 Introduction

Kemp and Kemp (1968) considered the distribution of the total number of games lost by the ruined gambler starting with some monetary units against an infinitely rich adversary, called the lost-games distribution and Yanagimoto (1989) independently considered the same distribution, called the inverse binomial distribution. The distribution is generated from a random walk on $y = \ldots, -1, 0, 1, \ldots$ with an absorbing barrier at $y = n$. On the random walk, a particle starting from $y = 0$ moves with steps $+1$ and -1 according to the transitional probabilities. When the random variable X represents the first passage time of reaching the absorbing barrier on the random walk, $(X-n)/2$ is the random variable of inverse binomial distribution. Shimizu and Yanagimoto (1991) generalized the inverse binomial distribution by adding a stay probability to the random walk, which is called the inverse trinomial distribution. The inverse trinomial and shifted inverse trinomial distributions are generated as general Lagrangian distributions, $\mathrm{GLD}_1(f, g)$ with $f(t) = t^n$ and $g(t) = p + qt + rt^2$ and $\mathrm{GLD}_1(f, g)$ with $f(t) = (p + qt + rt^2)^n$

and $g(t) = p + qt + rt^2$, respectively, under the assumption that a particle on the random walk moves with steps -1, 0 and $+1$ according to the probabilities p, q and $r = 1 - p - q$, respectively. Aoyama et al. (2008) proposed a generalization of the shifted inverse trinomial distribution, denoted by GIT, which is generated from a two-dimensional random walk. On the random walk, a particle starting from $(0, 0)$ moves from (x, y) to $(x, y + 1)$, $(x + 1.y + 1)$, $(x + 1, y)$, $(x + 1, y - 1)$ and $(x, y - 1)$ with probabilities p_1, p_2, p_3, p_4 and $p_5 = 1 - p_1 - p_2 - p_3 - p_4 - p_5$, respectively.

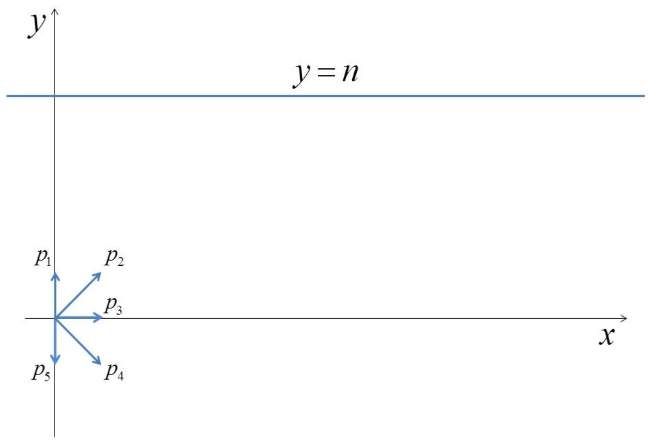

Figure 4.1: The random walk for GIT

The GIT is the first passage time distribution of reaching the absorbing barrier $y = n$ (see Fig 4.1). The subclass of the GIT when $p_4 = p_5 = 0$, denoted by $\mathrm{GIT}_{3,1}$, is interesting distribution. The distribution has the pmf

$$\mathrm{P}(X = x) = \binom{n + x - 1}{x} p_1^n p_3^x \; {}_2F_1\left[-n, -x; -n - x + 1; -\frac{p_2}{p_1 p_3}\right], \quad x = 0, 1, \ldots \tag{4.1}$$

and pgf

$$G(t) = \left(\frac{p_1 + p_2 t}{1 - p_3 t}\right)^n = \left(\frac{p_1 + p_2 t}{p_1 + p_2}\right)^n \left(\frac{1 - p_3}{1 - p_3 t}\right)^n. \tag{4.2}$$

The $\mathrm{GIT}_{3,1}$ includes the important distributions, shifted negative binomial distribution when $p_1 = 0$, negative binomial distribution when $p_2 = 0$ and binomial distribution when $p_3 = 0$. This means that the $\mathrm{GIT}_{3,1}$ has the flexibility to model under- and over-dispersion. Some properties of $\mathrm{GIT}_{3,1}$ like various expressions of pmf, recursive formulas of pmf and r-th cumulants and normal and Poisson approximations are studied in Aoyama et al. (2008).

This chapter considers an extension of $\mathrm{GIT}_{3,1}$, which has pgf

$$G(t) = (1 - P + Pt)^{\nu} \left(\frac{1 - p}{1 - pt}\right)^{\nu+\beta} \tag{4.3}$$

(Imoto, to appear, a). This is seen to be a convolution of binomial and negative binomial variables and reduces to $\mathrm{GIT}_{3,1}$ with pgf (4.2) when $\nu = n$, $\beta = 0$, $P = p_2/(p_1 + p_2)$ and $p = p_3$. The distribution with pgf (4.3) is denoted by $\mathrm{EGIT}_{3,1}(\nu, \beta, P, p)$ or $\mathrm{EGIT}_{3,1}$ for short. The parameter ν is a non-negative integer, $\nu + \beta > 0$, $0 < p < 1$ and $0 < P < 1$ or ν is a non-negative integer, $\beta > 0$, $0 < p < 1$ and $-p/(1 - p) < P < 0$. Kemp (1979) has considered a class of distributions formed by convolutions of binomial variables and binomial pseudo variables and quoted a number of physical models for members of the class. The $\mathrm{EGIT}_{3,1}$ belongs to Kemp's class and this chapter gives a more detailed study for this distribution. Section 4.2 gives the pmf, factorial moments and cumulants of $\mathrm{EGIT}_{3,1}$. These functions, including pgf, lead to the interesting properties and characteristics about $\mathrm{EGIT}_{3,1}$. In Section 4.3, we consider the stochastic processes leading to $\mathrm{EGIT}_{3,1}$, which have not been considered in Kemp (1979); (1) a three-dimensional random walk; (2) a birth, death and immigration process; (3) a thinned stochastic process. The profile maximum likelihood estimation and fitting examples to real count data sets are considered in Section 4.4. The improved fit and flexibility to the under- and over-dispersed data sets support the application of $\mathrm{EGIT}_{3,1}$ in empirical modeling.

4.2 Extension of $\mathrm{GIT}_{3,1}$

4.2.1 Probabilities and factorial moments

Consider the expansion of $(1-w/c)^x(1-w)^{-x-\beta}$ in powers of w. Following Chihara (1978, p. 176),

$$\left(1-\frac{w}{c}\right)^x(1-w)^{-x-\beta}=\sum_{n=0}^{\infty}M_n(x;\beta,c)\frac{w^n}{n!}, \tag{4.4}$$

where $M_n(x;\beta,c)=(-1)^x x!\sum_{k=0}^{x}\binom{n}{k}\binom{-n-\beta}{x-k}c^{-k}$ is a Meixner polynomial of the first kind. From the formula (4.4), the distribution with pgf (4.3) has the pmf

$$\mathrm{P}(X=x)=\frac{p^x(1-p)^{\nu+\beta}(1-P)^{\nu}}{x!}M_x\left(\nu;\beta,-\frac{(1-P)p}{P}\right),\ x=0,1,\ldots. \tag{4.5}$$

This distribution is obtained from a convolution of binomial and negative binomial variables when ν is a non-negative integer, $\nu+\beta>0$, $0<p<1$ and $0<P<1$, and a convolution of binomial variable and an improper variable when $-p/(1-p)<P<0$, which becomes a proper distribution from the idea of Kemp (1979).

The Meixner polynomials of the first kind belong to the class of orthogonal polynomials with respect to the negative binomial distribution and are characterized by a three-terms recursive formula

$$\begin{aligned}cM_{x+1}(n;\beta,c)&=((c-1)n+(1+c)x+c\beta)M_x(n;\beta,c)\\&\qquad-x(x+\beta-1)M_{x-1}(n,\beta,c),\ x\geq 0.\end{aligned}$$

Since the pmf (4.5) is expressed in term of Meixner polynomial of the first kind, we have the following three-terms recursive formula, which is a special case of Sundt's recursion (3.15),

$$\mathrm{P}(X = x) = \left(a + \frac{b}{x}\right) \mathrm{P}(X = x - 1) + \left(c + \frac{d}{x}\right) \mathrm{P}(X = x - 2),\ x \geq 1,$$

with $\mathrm{P}(X = -1) = 0$, $\mathrm{P}(X = 0) = (1 - P)^{\nu}(1 - p)^{\nu+\beta}$, where $a = (p(1 - P) - P)/(1 - P)$, $b = ((\nu + 1)P + (\nu + \beta - 1)(1 - P)p)/(1 - P)$, $c = Pp/(1 - P)$ and $d = Pp(\beta - 2)/(1 - P)$.

In term of the Gauss hypergeometric function, the pmf of $\mathrm{EGIT}_{3,1}$ is expressed as

$$\begin{aligned} P(X = x) = &\binom{\nu + \beta + x - 1}{x} p^x (1 - p)^{\nu+\beta} (1 - P)^{\nu} \\ &\times {}_2F_1 \left[-\nu, -x; -\nu - \beta - x + 1; -\frac{P}{(1 - P)p} \right],\ x = 0, 1, \ldots. \end{aligned} \tag{4.6}$$

When $\nu = n$, $\beta = 0$, $P = p_2/(p_1 + p_2)$ and $p = p_3$, the pmf (4.6) is seen to reduce to (4.1).

Returning to (4.3), we see that the factorial moment generating function is expressed as $G(t+1) = (1+Pt)^{\nu}(1-pt/(1-p))^{-\nu-\beta}$. Therefore, the r-th descending factorial moment $\mu_{[r]}$ is also expressed in term of the Meixner polynomials of the first kind as

$$\mu_{[r]} \left(\frac{p}{1 - p}\right)^r M_r \left(\nu; \beta, -\frac{p}{(1 - p)P}\right).$$

Also, the descending factorial moments have a three-term recursive formula

$$\mu_{[r+1]} = \left((\nu - r)P + (\nu + \beta + r)\frac{p}{1 - p}\right) \mu_{[r]} + r(r + \beta - 1)\frac{Pp}{1 - p}\mu_{[r-1]},\ r \geq 0,$$

with $\mu_{[-1]} = 0$ and $\mu_{[0]} = 1$, leading to easily enumeration of factorial moments.

4.2.2 Limiting distributions

The EGIT(ν, β, P, p) goes to the following distributions as limits:

1. Charlier series distribution when putting $\beta = m - \nu$ and letting $mp \to \lambda$ as $m \to \infty$.

2. Delaporte distribution (Delaporte, 1959), which is formulated as a convolution of negative binomial and Poisson variables, when letting $\nu P \to \lambda$ with $\alpha = \nu + \beta$ fixed as $\nu \to \infty$.

3. non-central negative binomial distribution when letting $(p+(1-p)P)\nu/p \to \lambda$ as $\nu \to \infty$.

This fact can be proved as follows.

1. $G(t) = (1-P+Pt)^{\nu}((1-p)/(1-pt))^{m} \to (1-P+Pt)^{\nu}\mathrm{e}^{\lambda(t-1)}$ as $m \to \infty$. This is the pgf of CSD$(\nu, \lambda/P, P)$.

2. $G(t) = (1-P+Pt)^{\nu}((1-p)/(1-pt))^{\alpha} \to ((1-p)/(1-pt))^{\alpha}\mathrm{e}^{\lambda(t-1)}$ as $\nu \to \infty$. This is the pgf of Delaporte distribution.

3.
$$\begin{aligned} G(t) &= \left(\frac{1-p}{1-pt}\right)^{\beta}\left(1+\frac{p+(1-p)P}{p}\frac{p(t-1)}{1-pt}\right)^{\nu} \\ &\to \left(\frac{1-p}{1-pt}\right)^{\beta}\exp\left(\frac{\lambda p(t-1)}{1-pt}\right) \text{ as } \nu \to \infty. \end{aligned}$$

   This is the pgf of NNBD(ν, λ, p).

Note. Both the non-central negative binomial and Delaporte distributions are generalizations of negative binomial distribution and suitable models for count data with high dispersion which means the index of dispersion is large (see Gupta and Ong, 2005). The EGIT$_{3,1}$ includes these generalized distributions as limiting distributions and therefore, can be adapted for count data with high dispersion.

4.2.3 Cumulants

The cumulant generating function of $\mathrm{EGIT}_{3,1}$ is expressed by the sum of those of binomial and negative binomial distributions. This means that $\kappa_r = \kappa_r^{(1)} + \kappa_r^{(2)}$, where κ_r, $\kappa_r^{(1)}$ and $\kappa_r^{(2)}$ are the r-cumulants of $\mathrm{EGIT}_{3,1}(\nu, \beta, P, p)$, $\mathrm{Bin}(\nu, P)$ and $\mathrm{NB}(\nu + \beta, p)$, respectively. Using the recursive formulas about cumulants of binomial and negative binomial distributions (1.2) and (1.5), we can easily get the cumulants of $\mathrm{EGIT}_{3,1}$. The first four cumulants of $\mathrm{EGIT}_{3,1}(\nu, \beta, P, p)$ are

$$\begin{cases} \kappa_1 = \nu P + (\nu+\beta)p/(1-p), \\ \kappa_2 = \nu P(1-P) + (\nu+\beta)p/(1-p)^2, \\ \kappa_3 = \nu P(1-P)(1-2P) + (\nu+\beta)p(1+p)/(1-p)^3, \\ \kappa_4 = \nu P(1-P)(1-6P+6P^2) + (\nu+\beta)p(1-4p+p^2)/(1-p)^4. \end{cases} \tag{4.7}$$

These cumulants are usable for the method of moments estimation for estimating four parameters of $\mathrm{EGIT}_{3,1}$.

From (4.7), we can get three indices measuring dispersion, skewness and kurtosis

$$\begin{cases} \dfrac{\kappa_2}{\kappa_1} = \dfrac{\nu P(1-P)(1-p)^2 + (\nu+\beta)p}{\nu P(1-p)^2 + (\nu+\beta)p(1-p)}, \\ \dfrac{\kappa_3}{\kappa_2^{3/2}} = \dfrac{\nu P(1-P)(1-2P)(1-p)^3 + (\nu+\beta)p(1+p)}{(\nu P(1-P)(1-p)^2 + (\nu+\beta)p)^{3/2}}, \\ \dfrac{\kappa_4}{\kappa_2^2} = \dfrac{\nu P(1-P)(1-6p+6P^2)(1-p)^4 + (\nu+\beta)p(1-4p+p^2)}{(\nu P(1-P)(1-p)^2 + (\nu+\beta)p)^2}, \end{cases}$$

respectively. The index of dispersion is seen to be greater (smaller) than one for $p/((1-p)P) > (<) \sqrt{\nu/(\nu+\beta)}$ when ν is a non-negative integer, $\nu + \beta > 0$, $0 < p < 1$ and $0 < P < 1$, and always greater than one when ν is a non-negative integer, $\beta > 0$, $0 < p < 1$ and $-p/(1-p) < P < 0$.

4.2.4 Mixture distributions

The $\text{EGIT}_{3,1}$ is generated from various types of mixture distributions.

1. Assume that $X|Y$ $\sim$NB$(\nu + Y, p)$ and Y $\sim$Bin(β, p') with non-negative integers ν and β. Then X $\sim\text{EGIT}_{3,1}(\beta, \nu, -p(1-p')/(1-p), p)$.

2. Assume that $X|Y$ $\sim$NB$(\nu + Y, p)$ and Y $\sim$NB$(\nu + \beta, p')$ with non-negative integers ν and β. Then X $\sim\text{EGIT}_{3,1}(\beta, \nu, -p/(1-p), p/(1-(1-p)p'))$.

3. Assume that $X|Y$ $\sim$Bin$(\nu + Y, p)$ and Y $\sim$NB$(\nu + \beta, p')$ with non-negative integers ν and β. Then X $\sim\text{EGIT}_{3,1}(\nu, \beta, p, pp'/(1-(1-p)p'))$.

4. Assume that $X|Y$ $\sim\text{EGIT}_{3,1}(Y, \beta, P, p)$ and Y $\sim\text{EGIT}_{3,1}(\nu, \beta, P', p')$ with a non-negative integer ν. Then X $\sim\text{EGIT}_{3,1}(\nu, \beta, PP' - p(1-P')/(1-p), (p + p'(1-p)P)/(1-p'(1-p)(1-P)))$.

This fact can be proved as follows.

1. The pgf of unconditional distribution of X is given by

$$\begin{aligned}
\sum_{x=0}^{\infty} \mathrm{P}(X = x)t^x &= \sum_{x=0}^{\infty}\sum_{y=0}^{\infty} \mathrm{P}(X = x|Y = y)\mathrm{P}(Y = y)t^x \\
&= \sum_{y=0}^{\infty} \left(\frac{1-p}{1-pt}\right)^{\nu+y} \binom{\beta}{y} p'^y (1-p')^{\beta-y} \\
&= \left(1 - p' + p\frac{1-p}{1-pt}\right)^{\beta} \left(\frac{1-p}{1-pt}\right)^{\nu} \\
&= \left(p + \frac{1-p'}{1-p} - \frac{p(1-p')}{1-p}t\right)^{\beta} \left(\frac{1-p}{1-pt}\right)^{\nu+\beta}.
\end{aligned}$$

This is the pgf of $\text{EGIT}_{3,1}(\beta, \nu, -p(1-p')/(1-p), p)$.

The proofs of the other parts can be given by a similar way.

4.2.5 Conditional bivariate negative binomial distribution

Assume that the random vector (X, Y) has the pgf

$$G_{X,Y}(t_1, t_2) = \left(\frac{1 - p_1 - p_2 - p_3}{1 - p_1 t_1 - p_2 t_2 - p_3 t_1 t_2} \right)^{\beta}, \tag{4.8}$$

defined by Edwards and Gurland (see Kocherlakota and Kocherlakota, 1992, p. 128). Then the pgf of the conditional distribution given $Y = \nu$ is given by

$$\begin{aligned} G(t) &= \frac{\partial^{\nu}}{\partial t_2^{\nu}} G_{X,Y}(t_1, t_2)|_{t_1=t, t_2=0} \Big/ \frac{\partial^{\nu}}{\partial t_2^{\nu}} G_{X,Y}(t_1, t_2)|_{t_1=1, t_2=0} \\ &= \left(\frac{p_2 + p_3 t}{p_2 + p_3} \right)^{\nu} \left(\frac{1 - p_1}{1 - p_1 t} \right)^{\nu+\beta}. \end{aligned}$$

This is the pgf of $\mathrm{EGIT}_{3,1}(\nu, \beta, p_3/(p_2 + p_3), p_1)$.

4.2.6 Characterization by conditional distribution

Assume that the random variables X_1 and X_2 are independent. If these are distributed as binomial distributions, then the conditional distribution of X_1 given $X_1 + X_2 = n$ is distributed as a hypergeometric distribution and if distributed as negative binomial distributions, then the conditional distribution is a negative hypergeometric distribution. Here we assume that $X_i \sim \mathrm{EGIT}_{3,1}(\nu_i, \beta_i, P, p)$ for $i = 1, 2$. Then the conditional distribution of X_1 given $X_1 + X_2 = n$ has the pmf

$$\begin{aligned} &\mathrm{P}(X_1 = x | X_1 + X_2 = n) \\ &\qquad = \binom{n}{x} \frac{M_x(\nu_1, \beta_1, -(1-P)p/P) M_{n-x}(\nu_2, \beta_2, -(1-P)p/P)}{M_x(\nu_1 + \nu_2, \beta_1 + \beta_2, -(1-P)p/P)}. \end{aligned} \tag{4.9}$$

This is a generalized hypergeometric and negative hypergeometric distribution. Conversely, if the conditional distribution of X_1 given $X_1 + X_2 = n$ has the pmf (4.9), then $X_i \sim \mathrm{EGIT}_{3,1}(\nu_i, \beta_i, P', p')$, where $(1 - P')p'/P' = (1 - P)p/P$. This can be proved by using Menon's theorem (1966):

Let X and Y be two independent discrete random variables such that the conditional distribution of $X = x$ given $X+Y = z$ is of the form $\alpha(x)\beta(z-x)/\gamma(z)$ for some functions α, β and γ. If the probability mass functions of X and Y are $p(x)$ and $q(y)$ respectively, then

$$\begin{aligned} p(x) &= \lambda\alpha(x)\mathrm{e}^{\theta x} \\ q(y) &= \mu\beta(y)\mathrm{e}^{\theta y} \end{aligned}$$

for some arbitrary constants λ, μ and θ.

In this case, $\alpha(x) = M_x(\nu_1; \beta_1, -(1-P)p/P)/x!$. Set $\mathrm{e}^{\theta} = p'$. Then from (4.4), we have

$$\sum_{x=0}^{\infty} p(x) = \lambda \left(1 + \frac{Pp'}{(1-P)p}\right)^{\nu_1} (1-p')^{-\nu_1-\beta_1}.$$

Therefore, we obtain

$$p(x) = \frac{p'^x(1-p')^{\nu_1+\beta_1}(1-P')^{\nu_1}}{x!} M_x\left(\nu_1; \beta_1, -\frac{1-P'}{P'}' p\right),$$

where $P' = Pp'/((1-P)p + Pp')$, or equivalently, $(1-P')p'/P' = (1-P)p/P$. Similarly, we can get

$$q(y) = \frac{p'^y(1-p')^{\nu_2+\beta_2}(1-P')^{\nu_2}}{y!} M_y\left(\nu_2; \beta_2, -\frac{1-P'}{P'}p'\right).$$

4.3 Stochastic processes leading to EGIT$_{3,1}$

4.3.1 Three-dimensional random walk

Consider a particle moving from (x, y, z) to $(x+1, y, z)$, $(x, y+1, z)$, $(x+1, y+1, z)$ and $(x, y, z+1)$ with probabilities p_1, p_2, p_3 and $q = 1 - p_1 - p_2 - p_3$, respectively. When this particle starts from the origin and reaches the barrier $z =$

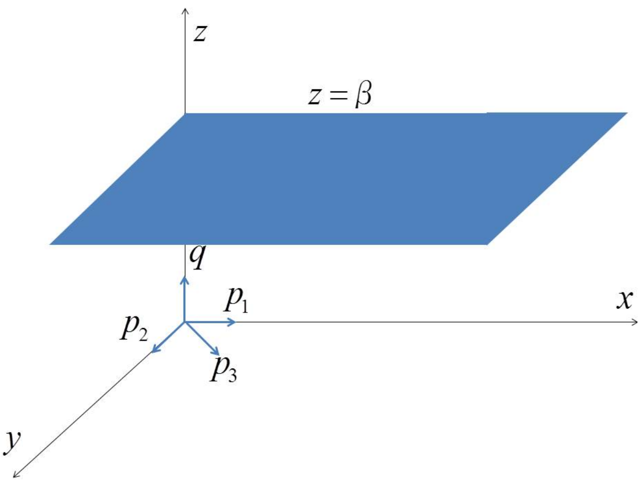

Figure 4.2: The random walk for bivariate negative binomial distribution

β, then the random vector (X, Y) representing the (x, y)-coordinate of reaching point is distributed as a bivariate negative binomial distribution with pgf (4.8). This fact can be proved as follows.

Assume that the pmf of random vector (X, Y) is $f_\beta(x, y)$. This pmf satisfies the difference equation

$$f_\beta(x, y) = p_1 f_\beta(x-1, y) + p_2 f_\beta(x, y-1) + p_3 f_\beta(x-1, y-1) + q f_{\beta-1}(x, y)$$

with boundary conditions $f_\beta(-1, y) = f_\beta(x, -1) = 0$ and $f_0(0, 0) = 1$. Putting $H_\beta(t_1, t_2) = \sum_{x=0}^{\infty}\sum_{y=0}^{\infty} f_\beta(x, y) t_1^x t_2^y$, we get the equation about $H_\beta(t_1, t_2)$

$$H_\beta(t_1, t_2) = pt_1 H_\beta(t_1, t_2) + p_2 t_2 H_\beta(t_1, t_2) + p_3 t_1 t_2 H_\beta(t_1, t_2) + q H_{\beta-1}(t_1, t_2)$$

with condition $H_0(t_1, t_2) = 1$. The solution is the pgf (4.8).

The conditional bivariate negative binomial distribution given $Y = \nu$ is shown to be $\mathrm{EGIT}_{3,1}$. Using this fact, when the given particle starts from the origin and reaches the line $y = \nu$ and $z = \beta$, the random variable X representing the x-coordinate of reaching point is seen to be distributed as $\mathrm{EGIT}_{3,1}(\nu, \beta, p_3/(p_2 +$

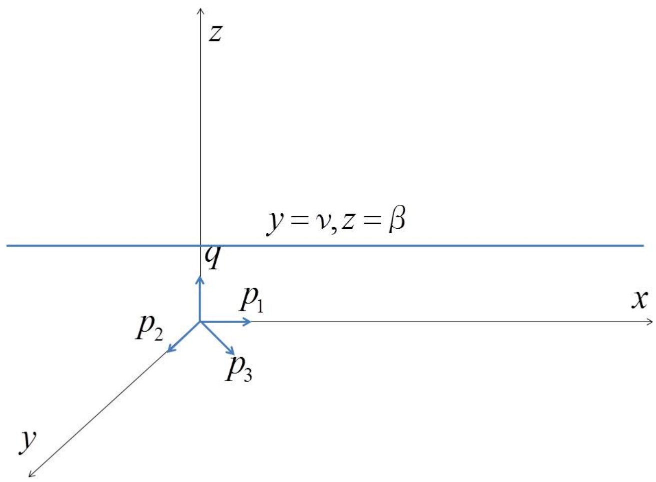

Figure 4.3: The random walk for EGIT

$p_3), p_1)$. When $\beta = 0$, this random walk reduces to a two-dimensional random walk leading to the $\text{GIT}_{3,1}$. Therefore, the random walk considered in this section generalizes that of $\text{GIT}_{3,1}$.

As mentioned in Section 4.1, the lost-games distribution is generated from a random walk and has applicability for the number of games lost by the ruined gambler. The three-dimensional random walk considered here also has applicability for gambler problem, in which a player might get two different types of win. For example, consider the game in which a player loses 1 dollar with probability p_1, wins A with probability p_2, wins B with probability q and loses 1 dollar and wins A with probability p_3. Then the money lost by getting ν units of A and β units of B is distributed as $\text{EGIT}(\nu, \beta, p_3/(p_2 + p_3), p_1)$.

4.3.2 Birth, death and immigration process

The number of individuals in a non-homogeneous birth and death process and that in a homogeneous birth, death and immigration process follow the distribution generated from a convolution of binomial and negative binomial variables. In this section, we consider a generalization of these two processes.

We consider a non-homogeneous birth, death and immigration process in which individuals in a colony are born at time s with birth rate $\lambda(s)$ and die at time s with death rate $\mu(s)$ and individuals from other colony immigrate at time s with immigration rate $\alpha(s)$. In this process, we assume that immigration rate is proportional to birth rate, i.e., $\alpha(s) = \beta\lambda(s)$ for a constant β. This assumption is natural because a high birth rate means that the colony is comfortable place to live in and therefore, leads to high immigration rate. Putting that $P(x, s)$ is a probability of the number of individuals in a colony being x at time s and $P(\nu, 0) = 1$, then we see that $P(x, s)$ satisfies the difference equation

$$
\begin{aligned}
P(x, s+h) &= (1 - \lambda(s)xh - \beta\lambda(s)h - \mu(s)xh)P(x, s) \\
&\quad + \lambda(s)(x-1)hP(x-1, s) + \beta\lambda(s)hP(x-1, s) + \mu(s)(x+1)hP(x+1, s).
\end{aligned}
$$

Putting $G(t, s) = \sum_{x=0}^{\infty} P(x, s)t^x$ and letting $h \to 0$, we can get the partial differential equation

$$
\begin{cases}
\dfrac{\partial}{\partial s}G(t, s) + (\mu(s) - \lambda(s)t)(t-1)\dfrac{\partial}{\partial t}G(t, s) = \beta\lambda(s)(t-1)G(t, s), \\
G(t, 0) = t^{\nu}.
\end{cases}
$$

From this partial differential equation, we can get the relations

$$
\mathrm{d}s = \frac{\mathrm{d}t}{(\mu(s) - t\lambda(s))(t-1)} = \frac{\mathrm{d}G}{\beta(t-1)\lambda(s)G}. \tag{4.10}
$$

Set $t = 1 + 1/y$ for the first equality in (4.10). Then the equation transforms to $\mathrm{d}y/\mathrm{d}s = (\lambda(s) - \mu(s))y + \lambda(s)$ and solving this differential equation and resetting $y = 1/(t-1)$, we get

$$
\frac{\mathrm{e}^{\rho(s)}}{t-1} - \int_0^s \lambda(u)\mathrm{e}^{\rho(u)}\mathrm{d}u = C_1, \tag{4.11}
$$

where $\rho(s) = \displaystyle\int_0^s (\mu(u) - \lambda(u))\mathrm{d}u$ and C_1 is a constant. From (4.11) and the second equality in (4.10), we obtain the differential equation

$$\frac{\mathrm{d}\log G}{\mathrm{d}s} = \frac{\beta\lambda(s)\mathrm{e}^{\rho(s)}}{C_1 + \displaystyle\int_0^s \lambda(u)\mathrm{e}^{\rho(u)}\mathrm{d}u} = \beta\frac{\mathrm{d}}{\mathrm{d}s}\log\left(C_1 + \int_0^s \lambda(u)\mathrm{e}^{\rho(u)}\mathrm{d}u\right)$$

Solving this differential equation, we get

$$G(t,s)\left(\frac{\mathrm{e}^{\rho(s)}}{t-1}\right)^{-\beta} = C_2, \tag{4.12}$$

where C_2 is a constant. Combining (4.11) and (4.12), we see that the function $\Psi(\cdot)$ satisfying

$$G(t,s)\left(\frac{\mathrm{e}^{\rho(s)}}{t-1}\right)^{-\beta} = \Psi\left(\frac{\mathrm{e}^{\rho(s)}}{t-1} - \int_0^s \lambda(u)\mathrm{e}^{\rho(u)}\mathrm{d}u\right)$$

exists. Substituting $t = 0$ and setting $s = 1 + 1/y$, we see $\Psi(y) = (1 + 1/y)^{\nu}y^{-\beta}$ and thus get the solution

$$\begin{aligned} G(t,s) &= \left(\frac{\mathrm{e}^{\rho(s)}}{t-1}\right)^{\beta}\left(\frac{\mathrm{e}^{\rho(s)}/(t-1) + 1 - \phi(s)}{\mathrm{e}^{\rho(s)}/(t-1) - \phi(s)}\right)^{\nu}\left(\frac{1}{\mathrm{e}^{\rho(s)}/(t-1) - \phi(s)}\right)^{\beta} \\ &= (1 - P(s) + P(s)t)^{\nu}\left(\frac{1-p(s)}{1-p(s)t}\right)^{\nu+\beta}, \end{aligned}$$

where $P(s) = (1-\phi(s))\mathrm{e}^{-\rho(s)}$, $p(s) = \phi(s)\mathrm{e}^{-\rho(s)}/(1+\phi(s)\mathrm{e}^{-\rho(s)})$, $\rho(s) = \displaystyle\int_0^s (\mu(u) - \lambda(u))\mathrm{d}u$ and $\phi(s) = \displaystyle\int_0^s \lambda(u)\mathrm{e}^{\rho(u)}\mathrm{d}u$. This is the pgf of $\mathrm{EGIT}_{3,1}(\nu, \beta, P(t), p(t))$.

The stochastic process considered in this section reduces to a non-homogeneous birth and death process giving rise to $\mathrm{GIT}_{3,1}$ when $\beta = 0$ and reduces to a homogeneous birth, death and immigration process giving rise to $\mathrm{EGIT}_{3,1}$ when $\lambda(t) = \lambda$ and $\mu(t) = \mu$, i.e., birth rate and death rate are independent from the time factor.

4.3.3 Thinned stochastic process

Ong (1995) has given the thinned stochastic process which leads to the distribution generated from a convolution of two binomial variables. This section introduces the thinned stochastic process which leads to $\mathrm{EGIT}_{3,1}$.

Theorem 4.1

Assume that $Y_1, Y_2, \ldots \overset{\text{i.i.d.}}{\sim} \mathrm{Bin}(1, 1-(1-\rho)/(1-\phi))$, $X_n \sim \mathrm{NB}(\beta, \phi)$ and $T(X_n) \sim \mathrm{NB}(X_n+\beta, \phi(1-\rho)/((1-\phi)\rho))$, where $0<\rho<1$ and $0<(1-\rho)/(1-\phi)<1$. Then, the linear stochastic model

$$X_{n+1} = \sum_{i=1}^{X_n+T(X_n)} Y_i \tag{4.13}$$

has a conditional pgf

$$G_{X_{n+k}|X_n=x}(t) = (1-P(k)+P(k)t)^x\,(1+p(k)-p(k)t)^{-x-\beta}, \tag{4.14}$$

which corresponds to the pgf of $\mathrm{EGIT}_{3,1}$ and the regression function is given by $\mathrm{E}[X_{n+k}|X_n=x] = (P(k)+p(k))x + \beta p(k)$, where $P(k) = (\rho^k-\phi)/(1-\phi)$ and $p(k) = \phi(1-\rho^k)/(1-\phi)$. The autocorrelation function is given by $\mathrm{Corr}[X_{n+k}, X_n] = \rho^k$.

Proof

The random variable X_{n+1} can be written by the sum of independent random variables as $X_{n+1} = U + V$, where $U = \sum_{i=1}^{X_n} Y_i$ and $V = \sum_{i=1}^{T(X_n)} Y_i$. Since the conditional pgf of V given $X_n = x$ is expressed as

$$\begin{aligned}
\mathrm{E}[t^V|X_n = x] &= \sum_{k=0}^{\infty} \mathrm{E}[t^V|T(X_n) = k, X_n = x]P(T(X_n) = k|X_n = x) \\
&= \sum_{k=0}^{\infty} \mathrm{E}[t^{Y_i}]^k P(T(X_n) = k|X_n = x) \\
&= \left(1 + \frac{\phi(1-\rho)}{1-\phi} - \frac{\phi(1-\rho)}{1-\phi}t\right)^{-x-\beta},
\end{aligned}$$

the conditional pgf of X_{n+1} given $X_n = x$ is calculated as

$$\begin{aligned}
&G_{X_{n+1}|X_n=x}(t) = \mathrm{E}[t^{X_{n+1}}|X_n = x] = \mathrm{E}[t^U|X_n = x]\mathrm{E}[t^V|X_n = x] \\
&= \left(\frac{1-\rho}{1-\phi} + \left(1 - \frac{1-\rho}{1-\phi}\right)t\right)^x \left(1 + \frac{\phi(1-\rho)}{1-\phi} - \frac{\phi(1-\rho)}{1-\phi}t\right)^{-x-\beta}.
\end{aligned}$$

Therefore, (4.14) is satisfied when $k = 1$. It is easily seen that (X_n, X_{n+1}) is distributed as a bivariate negative binomial distribution and X_{n+1} is identical with X_n. Suppose that (4.14) is satisfied for some $k \geq 1$. Then we can see

$$\begin{aligned}
&G_{X_{n+k+1}|X_n=x}(t) \\
&= \sum_{l_0,\ldots,l_{k+1}} t^{l_0} P(X_{n+k+1} = l_0|X_{n+k} = l_1) \cdots P(X_{n+1} = l_{k+1}|X_n = x) \\
&= \sum_{l_{k+1}} G_{X_{n+k+1}|X_{n+1}=l_{k+1}}(t) P(X_{n+1} = l_{k+1}|X_n = x) \\
&= (1 + p(k) - p(k)t)^{\beta} G_{X_{n+1}|X_n=x}\left((1 - P(k) + P(k)t)(1 + p(k) - p(k)t)\right) \\
&= (1 - P(k+1) + P(k+1)t)^x (1 + p(k+1) - p(k+1)t)^{-x-\beta}.
\end{aligned}$$

The regression function and autocorrelation function about (X_{n+k}, X_n) are easily obtained.

□

McKenzie (1986) considered the discrete AR(1) model which satisfy the stochastic equation

$$N_{n+1} = \alpha * N_{n-1} + R_n,$$

where $\alpha * N = \sum_{i=0}^{N} Y_i$ with $Y_1, Y_2, \ldots, Y_N \overset{\text{i.i.d.}}{\sim} \mathrm{Bin}(1, \alpha)$. Although the process (4.13) is not analogous to the AR(1) models by McKenzie (1986) because $R_n = \alpha * T(X_n)$ with $\alpha = 1 - (1-\rho)/(1-\phi)$ depends on X_n, this model has the first-order autoregressive properties, that is, the autocorrelation function has the form $\mathrm{Corr}[X_{n+k}, X_n] = \rho^k$ and the regression is linear. It seems that there are not many models for discrete variate processes although situations abound in which discrete data occur. Thus, the thinned stochastic process (4.13) should be useful as a model for discrete situations.

4.4 Numerical examples

To show the usefulness of $\mathrm{EGIT}_{3,1}$, we have fitted the $\mathrm{EGIT}_{3,1}$ to the real count data sets. Since the parameter ν in $\mathrm{EGIT}_{3,1}(\nu, \beta, P, p)$ can take only a non-negative integer and is complicated to be estimated by the maximum likelihood method, the profile maximum likelihood estimation is used. The log-likelihood function of $\mathrm{EGIT}_{3,1}$ is given by

$$\begin{aligned} L(\nu, \beta, P, p) = N\overline{x}\log p + N(\nu+\beta)\log(1-p) + N\nu\log(1-P) \\ + \sum_{r=0}^{n} f_r \log M_r\left(\nu; \beta, -\frac{1-P}{P}p\right) + \sum_{r=0}^{n} f_r \log(r!), \end{aligned}$$

where f_r is the observed frequency of value r, n is the highest value in observed data, $N = \sum_{r=0}^{n} f_r$ is the size of the sample and $\overline{x} = \sum_{r=1}^{n} r f_r / N$ is the sample mean of data set. From the relation

$$\frac{\partial}{\partial x}\frac{\Gamma(x+k)}{\Gamma(x)} = \frac{\Gamma(x+k)}{\Gamma(x)}\sum_{j=1}^{k}\frac{1}{x+j-1},$$

we can get the likelihood equations about β, P and p as

$$\begin{cases} \displaystyle\sum_{r=0}^{n}\frac{r!f_r}{M_r(\nu;\beta,-(1-P)p/P)}\sum_{k=0}^{\min(r,\nu)}\binom{\nu}{k}\binom{\nu+\beta+r-k-1}{r-k}\left(\frac{P}{(1-P)p}\right)^k \\ \qquad \displaystyle\times\sum_{j=1}^{r-k}\frac{1}{\nu+\beta+j-1} = -N\log(1-P), \\ \displaystyle(\nu+\beta)\sum_{r=0}^{n}\frac{rf_rM_{r-1}(\nu-1;\beta+1,-(1-P)p/P)}{M_r(\nu;\beta,-(1-P)p/P)} = Np(1-p), \\ \nu P+(\nu+\beta)p/(1-p) = \overline{x}. \end{cases}$$

The last equation has the same meaning as putting the mean of $\mathrm{EGIT}_{3,1}$ into the sample mean. For getting the profile maximum likelihood estimators, we solve the likelihood equations about β, P and p for a given value of ν and find the $\hat{\nu}$ such as $L(\hat{\nu},\hat{\beta}_{\hat{\nu}},\hat{P}_{\hat{\nu}},\hat{p}_{\hat{\nu}}) \geq L(\hat{\nu}+1,\hat{\beta}_{\hat{\nu}+1},\hat{P}_{\hat{\nu}+1},\hat{p}_{\hat{\nu}+1})$, where $\hat{\beta}_\nu$, $\hat{P}_\nu$ and $\hat{p}_\nu$ are maximum likelihood estimates for a given value of ν about β, P and p, respectively.

Using this method, the $\mathrm{EGIT}_{3,1}$ has been fitted to the under- and over-dispersed data sets: the number of children born by each 1,170 women who were 45–76 years old and living in Sweden in 1991 (Erikson and Åberg, 1987) in Table 4.1, which is a under-dispersed data and the quarterly sales of a well-known brand of a particular article of clothing at stores of a large national retailer (Shmueli et al., 2005) in Table 4.2, which is a over-dispersed data. Here, the $\mathrm{GIT}_{3,1}$ has also been fitted to these data sets for comparing the fits of $\mathrm{EGIT}_{3,1}$ with those of $\mathrm{GIT}_{3,1}$. Then, we treated the $\mathrm{EGIT}_{3,1}(\nu,0,P,p)$ as the $\mathrm{GIT}_{3,1}(\nu,P,p)$.

For both of data sets, the $\mathrm{EGIT}_{3,1}$ provides the best fitting in the sense of the chi-square values and p-values as well as that of the log-likelihood values. The $\mathrm{EGIT}_{3,1}$ includes generalizations of binomial and negative binomial distributions as limiting distributions while the $\mathrm{GIT}_{3,1}$ includes only binomial and negative binomial distributions. Therefore, the $\mathrm{EGIT}_{3,1}$ gives improved fitting.

Table 4.1: The number of children (Erikson and Åberg, 1987)

Count	Observed	$\text{GIT}_{3,1}$	$\text{EGIT}_{3,1}$
0	114	86.71	76.45
1	205	288.99	291.90
2	466	377.50	405.13
3	242	255.37	252.71
4	85	108.61	86.78
5	35	37.23	33.71
6	16	11.29	13.63
7	4	3.16	5.62
8	1	0.84	2.35
9	0	0.21	0.99
10	1	0.05	0.42
11	0	0.01	0.18
12+	1	0.00	0.13
Total	1170	1170.00	1170.00
Log L		−1928.98	−1921.77
χ^2		63.39	55.16
df		4	3
p-value		0.000	0.000

$\text{GIT}_{3,1}$: $\hat{\nu} = 3$, $\hat{P} = 0.4779$, $\hat{p} = 0.1954$.
$\text{EGIT}_{3,1}$: $\hat{\nu} = 3$, $\hat{\beta} = -2.3105$, $\hat{P} = 0.5394$, $\hat{p} = 0.4421$.

Table 4.2: The quarterly sales of a well-known brand of a particular article of clothing (Shmueli et al., 2005)

Count	Observed	$GIT_{3,1}$	$EGIT_{3,1}$
0	514	410.25	512.30
1	503	525.24	510.22
2	457	504.34	467.16
3	423	430.46	395.79
4	326	344.45	319.67
5	233	264.59	249.97
6	195	197.61	190.95
7	139	144.57	143.31
8	101	104.11	106.07
9	77	74.05	77.63
10	56	52.14	56.30
11	40	36.41	40.52
12	37	25.25	28.97
13	22	17.41	20.60
14	9	11.94	14.58
15	7	8.15	10.27
16	10	5.55	7.21
17	9	3.76	5.04
18	3	2.54	3.52
19	2	1.71	2.45
20	2	1.15	1.70
21	2	0.77	1.17
22+	1	1.55	2.60
Total	3168	3168.00	3168.00
Log L		-7539.30	-7520.32
χ^2		56.87	13.82
df		16	15
p-value		0.000	0.539

$GIT_{3,1}$: $\hat{\nu} = 2$, $\hat{P} = 0.000$, $\hat{p} = 0.6401$.
$EGIT_{3,1}$: $\hat{\nu} = 2$, $\hat{\beta} = 0.063$, $\hat{P} = -0.2238$, $\hat{p} = 0.6600$.

Chapter 5

A generalization of Conway–Maxwell–Poisson distribution

5.1 Introduction

The $\mathrm{EGIT}_{3,1}$ defined in the previous chapter is a flexible distribution to model under- and over-dispersion and is generated from various types of stochastic processes. In this chapter, we define a flexible distribution to model under- and over-dispersion with simple pmf as a generalization of the Conway–Maxwell–Poisson (COMP) distribution, which is considered by Imoto (submitted). The COMP distribution was originally developed by Conway and Maxwell (1962) as a solution to handling queueing systems with state-dependent arrival or service rates and revived as a flexible distribution to model to under- and over-dispersion by Shmueli et al. (2005). The pmf of the COMP distribution is given by

$$\mathrm{P}(X = x) = \frac{\theta^x}{(x!)^r} \frac{1}{Z(\theta, r)}, \ x = 0, 1, \ldots, \ \text{where} \quad Z(\theta, r) = \sum_{k=0}^{\infty} \frac{\theta^k}{(k!)^r} \tag{5.1}$$

for $r > 0$ and $\theta > 0$. The COMP distribution reduces to the geometric distribution when $r \to 0$ and $0 < \lambda < 1$ and the Bernoulli distribution when $r \to \infty$.

The negative binomial distribution is a generalized form of geometric distribution and more useful than the geometric distribution. For example, the mode of the negative binomial distribution takes on the any non-negative integer value while the mode of the geometric distribution always take on zero. The generalized COMP (GCOMP) distribution proposed in this chapter includes the negative binomial distribution as a special case and has the flexibility for zero-valued observations. Moreover, the GCOMP distribution can become a bimodal distribution with one mode at zero and thus, can be adapted to count data with many zeros. The flexibility for zero-valued observations will make the proposed distribution more versatile than the COMP distribution.

The definition of the GCOMP distribution and some properties are given in Section 5.2. We consider the methods of estimation for GCOMP distribution in Section 5.3 and fitting examples to real count data sets using the methods are given in Section 5.4.

5.2 Generalized Conway–Maxwell–Poisson distribution

5.2.1 Definition

A random variable X is said to have the GCOMP distribution with three parameters r, ν and θ if

$$\mathrm{P}(X = x) = \frac{\Gamma(\nu + x)^r \theta^x}{x! C(r, \nu, \theta)}, \; x = 0, 1, \ldots, \tag{5.2}$$

where the normalizing constant $C(r.\nu, \theta)$ is given by

$$C(r, \nu, \theta) = \sum_{k=0}^{\infty} \frac{\Gamma(\nu + k)^r \theta^k}{k!}. \tag{5.3}$$

The ratios of consecutive probabilities are formed as

$$\frac{\mathrm{P}(X = x)}{\mathrm{P}(X = x - 1)} = \frac{\theta(\nu - 1 + x)^r}{x} \tag{5.4}$$

and it can be seen from the ratios that $C(r, \nu, \theta)$ converges for $r < 1$ or $r = 1$ and $|\theta| < 1$. Hence, the parameter space of the GCOMP distribution is $r < 1$, $\nu > 0$ and $\theta > 0$ or $r = 1$, $\nu > 0$ and $0 < \theta < 1$. This distribution reduces to the COMP distribution with parameters $1 - r$ and θ when $\nu = 1$ and therefore, includes geometric and Bernoulli distributions. Moreover, this includes the negative binomial distribution for $r = 1$. Since $\Gamma(\nu + x)$ is a log-convex function, Corollary 4 in Castillo and Pérez-Casany (2005) confirms that the GCOMP distribution is over-dispersed for $0 < r < 1$ and under-dispersed for $r < 0$. The ratio of successive probabilities (5.4) with $x = 1$ is $\theta\nu^r$ and this depends on three parameters in which the parameter θ controls the mean and the parameter r controls the dispersion. This means that the parameters ν plays a role in controlling the value on $x = 0$.

5.2.2 Queueing process

Conway and Maxwell (1962) considered queueing systems with state-dependent arrival or service rates, which lead to the COMP distribution. The GCOMP distribution is also generated as a queueing model with the arrival rate $\lambda(\nu + x)^r$ and the service rate μx, where x is the size of the queue. This means that the service rate is directly proportional to the state and the arrival rate increases as queue becoming large for $r > 0$ and decreases as queue becoming large for $r < 0$.

Then $P(x,t)$, the probability of the size of the queue being x at time t, satisfies the difference equation for small h

$$\begin{aligned} P(x,t+h) &= (1-\lambda(\nu+x)^r h - \mu x h)P(x,t) \\ &\qquad + \lambda(\nu+x-1)^r h P(x-1,t) + \mu(x+1)hP(x+1,t). \end{aligned}$$

Putting $\theta = \lambda/\mu$ and letting $h \to 0$, we get the difference-differential equation

$$\frac{\partial P(x,t)}{\partial t} = (\nu+x-1)^r \theta P(x-1,t) + ((\nu+x)^r\theta + x)P(x,t) + (x+1)P(x+1,t).$$

Assuming a steady state and then putting $P(x,t) = P(x)$, we have the difference equation

$$P(x+1) = \frac{x}{x+1}P(x) + \frac{(\nu+x)^r\theta}{x+1}P(x) - \frac{(\nu+x-1)^r\theta}{x+1}P(x-1).$$

The solution of this equation is proved by induction to be

$$P(x) = \frac{\Gamma(\nu+x)^r\theta^x}{x!}\frac{P(0)}{\Gamma(\nu)^r}.$$

Note that $\Gamma(\nu)^r/P(0)$ is independent of the size x and is equal to the normalizing constant $C(r,\nu,\theta)$ in (5.3). This means $P(x)$ is the pmf of the GCOMP distribution. The system with $\nu = 1$, i.e., the arrival rate being $\lambda(x+1)^r$ which avoids to become zero when the size of the queue is zero, was considered by Conway and Maxwell (1962). Therefore, the system putting the arrival rate $\lambda(\nu+x)^r$ instead of $\lambda(x+1)^r$ is a natural generalization.

5.2.3 Moments

The k-th descending factorial moment of the GCOMP distribution is given by

$$\begin{aligned} \mathrm{E}[X(X-1)\cdots(X-k+1)] &= \frac{1}{C(r,\nu,\theta)}\sum_{x=0}^{\infty}\frac{\Gamma(\nu+x)\theta^x}{(x-k)!} \\ &= \frac{C(r,\nu+k,\theta)\theta^k}{C(r,\nu,\theta)}. \end{aligned} \tag{5.5}$$

From the first and second factorial moments, $k = 1, 2$ in (5.5), the mean and variance are obtained as

$$\mathrm{E[X]} = \frac{C(r,\nu+1,\theta)\theta}{C(r,\nu,\theta)}$$

and

$$\mathrm{Var}[X] = \frac{C(r,\nu+2,\theta)\theta^2}{C(r,\nu,\theta)} + \frac{C(r,\nu+1,\theta)\theta}{C(r,\nu,\theta)} - \frac{C(r,\nu+1,\theta)^2\theta^2}{C(r,\nu,\theta)^2},$$

respectively.

From the fact that the GCOMP distribution belongs to the class of power series distributions, we can get the recursive formulas about the moments as

$$\begin{cases} \mathrm{E}[X] = \theta\mathrm{E}[(\nu+X)^r], \\ \mathrm{E}[X^{k+1}] = \theta\dfrac{\partial}{\partial\theta}\mathrm{E}[X^k] + \mathrm{E}[X]\mathrm{E}[X^k], \\ \mathrm{E}[(X-\mathrm{E[X]})^{k+1}] = \theta\dfrac{\partial}{\partial\theta}\mathrm{E}[(X-\mathrm{E}[X])^k] + k\mathrm{E}[(X-\mathrm{E}[X])^{k-1}]\mathrm{Var}[X]. \end{cases} \tag{5.6}$$

The mean, variance and the third moment about the mean can be approximated by

$$\begin{aligned} \mathrm{E}[X] &\approx \theta^{1/(1-r)} + \frac{(2\nu-1)r}{2(1-r)}, \\ \mathrm{Var}[X] &\approx \frac{\theta^{1/(1-r)}}{1-r}, \\ \mathrm{E}[(X-\mathrm{E}[X])^3] &\approx \frac{\theta^{1/(1-r)}}{(1-r)^2}, \end{aligned} \tag{5.7}$$

respectively, by using the recursive formulas (5.6) and an approximating formula

$$C(r,\nu,\theta) = \frac{\theta^{(2\nu-1)r/\{2(1-r)\}}(2\pi)^{r/2}\exp((1-r)\mathrm{e}^{1/(1-r)})}{\sqrt{1-r}}(1+O(\theta^{-1/(1-r)})).$$

This approximating formula can be led by the relation

$$\frac{1}{2\pi}\int_{-\pi}^{\pi}\exp(\mathrm{e}^{\mathrm{i}z}-\mathrm{i}z(\nu-1))\mathrm{e}^{-\mathrm{i}zk}\mathrm{d}z = \frac{1}{\Gamma(\nu+k)},$$

where $\mathrm{i}=\sqrt{-1}$, and a similar argument to the Shmueli et al.'s (2005) for the normalizing constant $Z(\lambda,r)$ of the COMP distribution in (5.1).

5.2.4 Mode

In this section, we study the modality of the GCOMP distribution. From the ratios of consecutive probabilities (5.4), it is easily seen that the GCOMP distribution (5.2) is a unimodal distribution when $r<0$ or $r<1$ and $\nu>1$. Moreover, in these cases, the GCOMP distribution has strong unimodality, the property that its convolution with any unimodal distribution is unimodal. This fact is proved by showing the log-concavity of the function (5.2), or equivalently, $M := \mathrm{P}(X=x+1)\mathrm{P}(X=x-1)/\mathrm{P}(X=x)^2$ is smaller than one. Since

$$M = \frac{x}{x+1}\left(\frac{\nu+x}{\nu+x-1}\right)^r,$$

it is obvious that $M<1$ for $r\le 0$. When $0<r<1$ and $\nu\ge 1$, we see that

$$\left(\frac{x}{x+1}\right)^{1/r} < \frac{x}{x+1} \le \frac{x+\nu-1}{x+\nu}$$

and thus, $M<1$.

When $0<r<1$, $0<\nu<1$ and $\theta\nu^r<1$, however, the GCOMP distribution may become a bimodal distribution with one mode at zero as can be seen in

Figure 5.1 for $r = 0.5$, $\nu = 0.01$ and $\theta = 2.5$. From this fact, the GCOMP distribution is applicable to the count data which has two mode with one mode at zero.

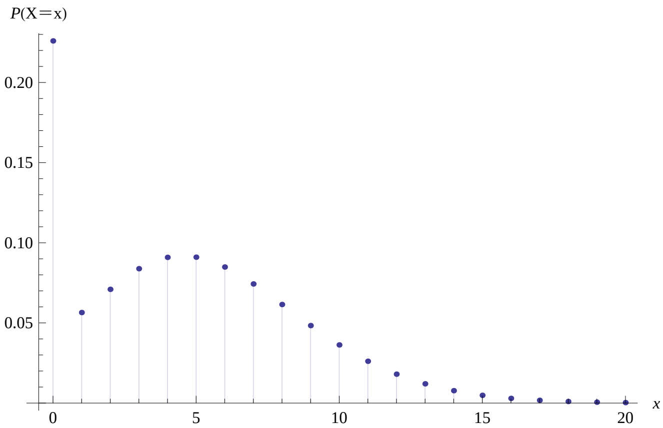

Figure 5.1: The pmf of the GCOMP distribution for $r = 0.5$, $\nu = 0.01$, $\theta = 2.5$

5.3 Estimation

In this section, we deal with the methods using first three moments or four consecutive probabilities for estimating the parameters of the GCOMP distribution. The estimated parameters obtained from these methods are crude and therefore, refined by feeding them as initial values into the maximum likelihood estimation, which is more accurate and the best way to do inference.

5.3.1 Estimation by moments

Let m_1, m_2 and m_3 be the sample mean, variance and the third sample moment about the mean, respectively. For estimates $(\tilde{r}, \tilde{\nu}, \tilde{\theta})$ of (r, ν, θ) by the method of moments, put equal to the mean, variance and the third moment about the mean into m_1, m_2 and m_3, respectively, or

$$\begin{cases} m_1 = \mathrm{E}[X] = \tilde{\theta}\dfrac{\partial}{\partial\theta}C(r,\nu,\theta)|_{r=\tilde{r},\nu=\tilde{\nu},\theta=\tilde{\theta}}, \\ m_2 = \mathrm{Var}[X] = \tilde{\theta}\dfrac{\partial}{\partial\theta}\mathrm{E}[X]|_{r=\tilde{r},\nu=\tilde{\nu},\theta=\tilde{\theta}}, \\ m_3 = \mathrm{E}[(X-\mathrm{E}[X])^3] = \tilde{\theta}\dfrac{\partial}{\partial\theta}\mathrm{Var}[X]|_{r=\tilde{r},\nu=\tilde{\nu},\theta=\tilde{\theta}}. \end{cases}$$

In these equations, using the approximating moments (5.7), we get the approximating equations

$$\begin{cases} m_1 \approx \tilde{\theta}^{1/(1-\tilde{r})} + \dfrac{(2\tilde{\nu}-1)\tilde{r}}{2(1-\tilde{r})}, \\ m_2 \approx \dfrac{\tilde{\theta}^{1/(1-\tilde{r})}}{1-\tilde{r}}, \\ m_3 \approx \dfrac{\theta^{1/(1-\tilde{r})}}{(1-\tilde{r})^2} \end{cases}$$

and the solutions of these equations are obtained as

$$\tilde{r} \approx 1 - \frac{m_2}{m_3},\ \tilde{\nu} \approx \frac{m_2(m_1m_3 - m_2^2)}{m_3(m_3-m_2)} + \frac{1}{2},\ \tilde{\theta} \approx \left(\frac{m_2^2}{m_3}\right)^{m_2/m_3}.$$

The main advantage of this method is so simple that we can carry it out by hand. As a disadvantage, however, this is not always applicable. For example, when $m_3 \leq 0$, the estimate $\tilde{r}$ becomes greater than one and $\tilde{\theta}$ might become a complex number. Even when $m_3 > 0$, the estimate $\tilde{\nu}$ is not always positive. For such a case, the method in the next subsection is recommended.

5.3.2 Estimation by four consecutive probabilities

Denote the pmf of the GCOMP distribution (5.2) by P_x for simplicity. We see the following equations

$$\log\left(\frac{x+2}{x+1}\frac{P_{x+2}P_x}{P_{x+1}^2}\right) = r\log\left(\frac{\nu+x+1}{\nu+x}\right) \tag{5.8}$$

and

$$\frac{\log\left(\dfrac{x+3}{x+2}\dfrac{P_{x+3}P_{x+1}}{P_{x+2}^2}\right)}{\log\left(\dfrac{x+2}{x+1}\dfrac{P_{x+2}P_{x}}{P_{x+1}^2}\right)} = \frac{\log\left(\dfrac{\nu+x+2}{\nu+x+1}\right)}{\log\left(\dfrac{\nu+x+1}{\nu+x}\right)}. \tag{5.9}$$

Putting the observed frequency of x events instead of P_x in (5.4), (5.8) and (5.9) and solving the equations, we can get the estimates $(\overline{r}, \overline{\nu}, \overline{\theta})$ of the GCOMP distribution. Although this estimation is rather rough due to the use of only four frequencies of the data, an advantage of this method is that we can choose the x so that the estimates are in the parameter space of the GCOMP distribution.

5.3.3 Maximum likelihood estimation

Let a random variable X be distributed as the GCOMP distribution (5.2) with parameters r, ν and θ. When ν is known, the GCOMP distribution belongs to the exponential family with natural parameters $(\nu, \log\theta)$ and therefore, the ML estimates about ν and $\log\theta$ are uniquely determined by using the minimal sufficient statistics $(\log\Gamma(\nu+X), X)$.

We consider the MLE in the case when ν is unknown. The log-likelihood function is

$$\begin{aligned} L(r,\nu,\theta) &= \log\left(\prod_{i=0}^{n} P(X=i)^{f_i}\right) \\ &= r\sum_{i=0}^{n} f_i \log\Gamma(\nu+i) + N\log\theta\sum_{i=0}^{n} i f_i - N\log C(r,\nu,\theta) - \sum_{i=0}^{n} f_i \log(i!), \end{aligned}$$

where f_i is the observed frequency of i events, n is the highest observed value and $N = \displaystyle\sum_{i=0}^{n} f_i$ is the size of the sample. Using the digamma function $\psi(y) =$

$\dfrac{\partial}{\partial y}\log\Gamma(y)$, we get the likelihood equations as

$$\begin{cases} \mathrm{E}[\log\Gamma(\nu+X)] = \displaystyle\sum_{i=0}^{n} \log\Gamma(\nu+i)\frac{f_i}{N}, \\ \mathrm{E}[\psi(\nu+X)] = \displaystyle\sum_{i=0}^{n} \psi(\nu+i)\frac{f_i}{N}, \\ \mathrm{E}[X] = \displaystyle\sum_{i=1}^{n} i\frac{f_i}{N}. \end{cases} \tag{5.10}$$

The solutions of these equations are not always unique and might not give a local maximum point for $L(r,\nu,\theta)$. However, it is easy to give a sufficient condition for the solution giving a local maximum point. We see that

$$\begin{aligned}
&\frac{\partial^2}{\partial r^2}L(r,\nu,\theta) = -N\mathrm{Var}[\log\Gamma(\nu+X)],\\
&\frac{\partial^2}{\partial r\partial\nu}L(r,\nu,\theta) = \sum_{i=0}^{n}\psi(\nu+X)f_i - N\mathrm{E}[\psi(\nu+X)]\\
&\qquad\qquad\qquad\qquad\qquad\qquad -N\mathrm{Cov}[r\psi(\nu+X),\log\Gamma(\nu+X)],\\
&\frac{\partial^2}{\partial r\partial\log\theta}L(r,\nu,\theta) = -N\mathrm{Cov}[\log\Gamma(\nu+X),X],\\
&\frac{\partial^2}{\partial\nu^2}L(r,\nu,\theta) = \sum_{i=0}^{n} r\psi'(\nu+i)f_i - N\mathrm{E}[r\psi'(\nu+X)] - N\mathrm{Var}[r\psi(\nu+X)],\\
&\frac{\partial^2}{\partial\nu\partial\log\theta}L(r,\nu,\theta) = -N\mathrm{Cov}[r\psi(\nu+X),X],\\
&\frac{\partial}{\partial(\log\theta)^2}L(r,\nu,\theta) = -N\mathrm{Var}[X],
\end{aligned}$$

and the Hesse matrix of the function $L(r,\nu,\theta)$ evaluated at the solution of the equations (5.10) is given by

$$\begin{aligned}
&-N\mathbf{Var}[\log\Gamma(\nu+X), r\psi(\nu+X), X]\\
&\qquad\qquad -N\begin{pmatrix} 0 & 0 & 0\\ 0 & \mathrm{E}[r\psi'(\nu+X)] - \sum_{i=0}^{n} r\psi'(\nu+i)f_i/N & 0\\ 0 & 0 & 0 \end{pmatrix},
\end{aligned}$$

where $\mathbf{Var}[\log\Gamma(\nu+X), r\psi(\nu+X), X]$ is the variance-covariance matrix of the random vector $(\log\Gamma(\nu+X), r\psi(\nu+X), X)$. Since the variance-covariance matrix is non-negative definite, the Hesse matrix is always negative definite if $\mathrm{E}[r\psi'(\nu+X)] > \sum_{i=0}^{n} r\psi'(\nu+i)f_i/N$. This is the assuring method for $L(r,\nu,\theta)$ becomes, at least, a local maximum at the solutions of the equations (5.10).

An alternative way to find a maximum point of $L(r,\nu,\theta)$ is to use the profile maximum likelihood estimation. For a given ν, consider the MLE for $(r,\log\theta)$, or solve the likelihood equations $\mathrm{E}[\log\Gamma(\nu+X)] = \sum_{i=0}^{n}\log\Gamma(\nu+i)f_i/N$ and $\mathrm{E}[X] = \sum_{i=0}^{n} if_i/N$. Then it is sufficient to study the profile log-likelihood function $L(\hat{r}_\nu,\nu,\hat{\theta}_\nu)$ with ν varying, where $(\hat{r}_\nu,\hat{\theta}_\nu)$ is the solution of the likelihood equations for a given ν. Although this method will take more time than the usual MLE, we can find the maximum point of $L(r,\nu,\theta)$ around the values $\tilde{\nu}$ or $\overline{\nu}$ considered in Section 5.3.1 or 5.3.2.

The Fisher information matrix $I(r,\nu,\theta)$ is seen as $\mathbf{Var}[\log\Gamma(\nu+X), r\psi(\nu+X), X]$. Using the scoring method for solving equations (5.10), given trial values (r_k,ν_k,θ_k), we can update to $(r_{k+1},\nu_{k+1},\theta_{k+1})$ as

$$\begin{pmatrix} r_{k+1} \\ \nu_{k+1} \\ \theta_{k+1} \end{pmatrix} = \begin{pmatrix} r_k \\ \nu_k \\ \theta_k \end{pmatrix} + I(r_k,\nu_k,\theta_k)^{-1} \left.\begin{pmatrix} \sum_{i=0}^{n}\log\Gamma(\nu+i)f_i/N - \mathrm{E}[\log\Gamma(\nu+X)] \\ \sum_{i=0}^{n}\psi(\nu+i)f_i/N - \mathrm{E}[\psi(\nu+X)] \\ \sum_{i=1}^{n} if_i/N - \mathrm{E}[X] \end{pmatrix}\right|_{r=r_k,\nu=\nu_k,\theta=\theta_k}.$$

As the starting point (r_0,ν_0,θ_0), we can choose the estimated parameters introduced in Sections 5.3.1 or 5.3.2.

5.4 Numerical examples

In this section, we give three examples of fittings to real count data sets by the GCOMP distribution and compare them with those by the COMP distribution to illustrate its utility.

The first data set is the length of words in a Hungarian dictionary (Wimmer et al., 1994), which is a under-dispersed data. This data set was used by Shmueli et al. (2005) for illustrating the utility of the COMP distribution. The second data set is the number of spots in southern pine beetle, *Dendroctonus frontalis Zimmermann* (Coleopetra: Scolytidae), in Southeast Texas (Lin, 1985), which is a over-dispersed data. The fitted results and comparisons with the COMP distribution (5.1) are in Tables 5.1 and 5.2. The third data set is the number of roots produced by 270 shoots of the apple cultivar Trajan (Ridout et al., 1998), in which there are two modes. The COMP distribution is not suitable for such a data set and the GCOMP distribution is compared with the zero-inflated COMP (ZICOMP) distribution whose pmf is $\mathrm{P}(X^w = 0) = w + (1 - w)\mathrm{P}(X = 0)$, and $\mathrm{P}(X^w = x) = (1 - w)\mathrm{P}(X = x)$, $x = 1, 2, \ldots$ for $0 < w < 1$, where X is the random variable of the COMP distribution with pmf (5.1). The fitted result is in Table 5.3.

In the first example, the count in the data starts from one and, following Wimmer et al. (1994), the counts in x are treated as in $x - 1$, as if the data were generated from the shifted COMP or GCOMP distributions whose random variable takes positive values. In this case, the sample mean, variance and the third moment about the mean are obtained as $m_1 = 2.3045$, $m_2 = 1.2219$ and $m_3 = 0.6522$, respectively and the estimates by approximating moments introduced in Section 5.3.1 are given by $(\tilde{r}, \tilde{\nu}, \hat{\theta}) = (-0.8735, 0.4673, 4.7195)$. These values are valid as the parameters of the GCOMP distribution, but very rough estimation and therefore, refined by feeding them as initial values into the maximum likelihood estimation.

In the second example, the sample mean, variance and the third moment about the mean are $m_1 = 0.6800$, $m_2 = 2.8117$ and $m_3 = 18.2349$, respectively and we can get the estimates by approximating moments as $(\tilde{r}, \tilde{\nu}, \tilde{\theta}) = (0.8458, 0.5449, 0.8791)$. These are also valid and used for computing the ML estimates.

In the third example, the approximating moment estimates led by the sample mean, variance and the third moment about the mean, $m_1 = 5.0593$, $m_2 = 15.7065$ and $m_3 = 13.5495$, are obtained as $(\tilde{r}, \tilde{\nu}, \tilde{\theta}) = (-0.1592, 96.2370, 28.8972)$. Although this data set is over-dispersed, or $m_1 < m_2$, the estimated value of r is negative. These estimates seem to be invalid because the GCOMP distribution is under-dispersed if $r < 0$. For this data set, we employed the estimation by using the four consecutive probabilities introduced in Section 5.3.2. Utilizing f_x with $x = 0$ instead of P_x in the equations (5.4), (5.8) and (5.9) is a good choice in this case. The estimates are $(\overline{r}, \overline{\nu}, \overline{\theta}) = (0.4132, 0.0011, 2.5988)$ and these are valid and used as a starting point for computing ML estimates .

As seen in Tables 5.1, 5.2 and 5.3, the GCOMP distribution gives good fits. In the first and second examples, the GCOMP distribution gives better fits for almost parts than the COMP distribution. In the third example, the GCOMP distribution gives a better fit for each point than the zero-inflated COMP distribution except $x = 7$ and 13. From this example, we see that we can use the GCOMP distribution for the count data with many zeros without utilizing the special operation like mixture. This is a great advantage of using the GCOMP distribution. Notice that the expected value for $x = 1$ in Table 5.1 and the expected values for $x = 0$ in Tables 5.2 and 5.3 of the GCOMP distribution are always very close to the observed counts. This means the GCOMP distribution is more flexible for zero-valued observations than the COMP distribution.

Table 5.1: The length of words in a Hungarian dictionary (Wimmer et al., 1994)

Count	Observed	COMP	GCOMP
0	1421	1553.17	1416.59
1	12333	12027.81	12319.73
2	20711	20949.95	20954.05
3	15590	15243.58	15023.88
4	5544	5971.32	5949.22
5	1510	1446.93	1497.33
6	289	236.80	261.16
7	60	27.81	33.47
8	1	2.63	3.56
Total	57459	57459.00	57459.00
Log L		−86168.02	−86153.48
χ^2		105.09	70.36
df		4	3
p-value		0.000	0.000

COMP: $\hat{r} = 2.1526$, $\hat{\theta} = 7.7441$.
GCOMP: $\hat{r} = -0.9331$, $\hat{\nu} = 0.5765$, $\hat{\theta} = 5.2019$.

Table 5.2: The number of spots in southern pine beetle, *Dendroctonus frontalis Zimmermann* (Coleopetra: Scolytidae), in Southeast Texas (Lin, 1985)

Count	Observed	COMP	GCOMP
0	1169	927.60	1168.66
1	144	372.48	152.55
2	92	149.57	80.48
3	54	60.06	49.82
4	29	24.12	32.52
5	18	9.68	21.68
6	10	3.89	14.58
7	12	1.56	9.83
8	6	0.63	6.63
9	9	0.25	4.45
10	3	0.10	2.98
11	2	0.04	1.99
12	0	0.02	1.32
13	0	0.01	0.87
14	1	0.00	0.57
15	0	0.00	0.38
16	0	0.00	0.25
17	0	0.00	0.16
18	0	0.00	0.10
19+	1	0.00	0.18
Total	1550	1550.00	1550.00
Log L		−1755.49	−1552.49
χ^2		939.53	10.48
df		8	7
p-value		0.000	0.163

COMP: $\hat{r} = 0.0000$, $\hat{\theta} = 0.4015$.
GCOMP: $\hat{r} = 0.8750$, $\hat{\nu} = 0.1011$, $\hat{\theta} = 0.9699$.

Table 5.3: The number of roots produced by 270 shoots of the apple cultivar Trajan (Ridout et al., 1998)

Count	Observed	WCOMP	GCOMP
0	64	64.00	64.00
1	10	6.48	7.25
2	13	11.93	11.99
3	15	17.61	17.23
4	21	22.21	21.68
5	18	24.79	24.37
6	24	25.06	24.85
7	21	23.28	23.30
8	23	20.10	20.28
9	21	16.28	16.50
10	17	12.45	12.65
11	12	9.03	9.18
12	5	6.25	6.33
13	2	4.14	4.16
14	3	2.63	2.62
15	0	1.61	1.59
16	0	0.95	0.92
17+	1	1.18	1.10
Total	270	270.00	270.00
Log L		−673.19	−672.40
χ^2		11.21	9.45
df		10	10
p-value		0.341	0.490

COMP: $\hat{w} = 0.2281$, $\hat{r} = 0.5463$, $\hat{\theta} = 2.6897$.
GCOMP: $\hat{r} = 0.3826$, $\hat{\nu} = 0.0001$, $\hat{\theta} = 3.3061$.

Chapter 6

Conclusion

The binomial distribution is one of the under-dispersed distributions and the negative binomial distribution is one of the over-dispersed distributions. These distributions are classically utilized for analyzing count data and a lot of their generalizations are considered for extending their applications. This thesis proposed the distributions which admits both under-dispersion and over-dispersion and showed that each distribution can be adapted to the various types of count data.

In Chapter 2, some properties of general Lagrangian distributions are considered and the LNNBD_1 introduced in Chapter 3 is generated as a member of general Lagrangian distributions. The LNNBD_1 includes the Charlier series and non-central negative binomial distributions, which are generalizations of the binomial and negative binomial distribution, respectively, and has the properties similar to these particular cases. Moreover, the LNNBD_1 includes some important distributions which are well studied in Lagrangian distributions and thus, has an advantage of the use as a general model since this allows a particular case to be selected. We showed in the theorem of Chapter 2 that any Lagrangian distribution can be generated through different sets of functions and, from this theorem, we can get the various expressions of pmf and recursive formulas about the LNNBD_1. The theorem in Chapter 2 has an application to the tandem queueing system with a single moving server where customers at the first counter have priority over customers at the second counter. There are several applications of the tandem queueing system in robotic systems, network systems and telecommunication systems and therefore, we can see the potential applicability of Lagrangian distributions to such systems.

The $\text{EGIT}_{3,1}$ introduced in Chapter 4 is generated from a convolution of binomial and negative binomial variables. This distribution plays the role of a continuous bridge between under- and over-dispersion and includes the Charlier series and non-central negative binomial distributions as limiting distributions. The $\text{EGIT}_{3,1}$ can be derived from various stochastic processes and, in this thesis, we consider the three stochastic processes which lead to the $\text{EGIT}_{3,1}$, a three-dimensional random walk, a birth, death and immigration process and a thinned stochastic process. The generations from various types of stochastic processes lead the application of the $\text{EGIT}_{3,1}$ to various phenomena.

The GCOMP distribution introduced in Chapter 5 is a generalization of Conway–Maxwell–Poisson distribution which includes the negative binomial distribution. The distribution also plays the role of a continuous bridge between under- and over-dispersion with simple pmf and has the flexibility for zero-valued observations. Moreover, the GCOMP distribution can become a bimodal distribution with one mode at zero and this fact leads to the use for the count data with excess zeros without utilizing the special operation like mixture. From these facts, the proposed distribution is more flexible distribution than the COMP distribution.

Acknowledgement

I would like to express my sincere gratitude to my supervisor, Professor Kunio Shimizu, for providing me with continuing support and direction during my studies. Besides supervisor, I would like to thank Professor Ritei Shibata, Professor Mihoko Minami and Professor Shun Shimomura for various insightful comments and suggestions on the paper. My sincere gratitude also goes to Professor Seng Huat Ong for his inspiration and wonderful advice during my research into discrete distributions.

Bibliography

[1] Aoyama, K., Shimizu, K., and Ong, S. H. (2008). A first–passage time random walk distribution with five transition probabilities: a generalization of the shifted inverse trinomial. *Annals of the Institute of Statistical Mathematics*, **60**, 1–20.

[2] Benes, V. E. (1957). On queues with Poisson arrivals. *The Annals of Mathematical Statistics*, **28**, 670–677.

[3] Blumer, M. G. (1974). On the fitting the Poisson-lognormal distribution to species abundance data. *Biometrics*, **30**, 101–110.

[4] Castillo, J., and Pérez-Casany, M. (1998). Weighted Poisson distributions for overdispersion and underdispersion situtations. *Annals of the Institute of Statistical Mathematics*, **50**, 567–585.

[5] Castillo, J., and Pérez-Casany, M. (2005). Overdispersed and underdispersed Poisson generalizations. *Journal of Statistical Planning and Inference*, **134**, 486–500.

[6] Charalambides, Ch. A. (1986). On some properties of the linear function Poisson, binomial and negative binomial distributions. *Metrika*, **33**, 203–216.

[7] Chihara, T. S. (1978). *An Introduction to Orthogonal Polynomials.* New York, Gordon and Breach.

[8] Consul, P. C. (1990). Geeta distribution and its properties. *Communications in Statistics–Theory and Methods*, **19(8)**, 3051–3068.

[9] Consul, P. C., and Famoye, F. (2006). *Lagrange Probability Distributions.* New York: Birkhäuser.

[10] Consul, P. C., and Jain, G. C. (1973) A generalization of the Poisson distribution. *Technometrics*, **15**, 791–799.

[11] Consul, P. C., and Shenton, L. R. (1972). Use of Lagrange expansion for generating discrete generalized probability distributions. *SIAM Journal of Applied Mathematics*, **23**, 239–248.

[12] Consul, P. C., and Shenton, L. R. (1973). Some interesting properties of Lagrangian distributions. *Communications in Statistics–Simulation and Computation*, **2**, 263–272.

[13] Consul, P. C., and Shenton, L. R. (1975). On the probabilistic structure and properties of discrete Lagrangian distributions. In G. P. Patil, S. Kotz, and J. K. Ord, editors, *Statistical Distributions in Scientific Work, Vol. 1: Models and Structures*, 41–48. New York: North–Holland.

[14] Conway, R. W., and Maxwell, W. L. (1962). A queueing model with state dependent service rates. *Journal of Industrial Engineering*, **12**, 132–136.

[15] Delaporte, P. (1959). Quelques problems de statistique mathematique poses par l'assurance automible et le bonus pour non sinister. *Bulletin Trimestriel de l'Institut des Actuaires Francais*, **227**, 121–135.

[16] Doray, L. G., and Luong, A. (1995). Quadratic distance estimators for the zeta family. *Insurance: Mathematics and Economics* , **16**, 255–260.

[17] Doray, L. G., and Luong, A. (1997). Efficient estimators for the Good family. *Communications in Statistics–Simulation and Computation*, **26**, 1075–1088.

[18] Erdélyi, A., Magnus, W., Oberhettinger, F., and Tricomi, G. F. (1953). *Higher Transcendental Functions*, Vol. 1. New York, McGraw–Hill.

[19] Erikson, R., and Åberg, R. (1987). *Welfare in Transition.* Oxford: Claredon Press.

[20] Good, I. J. (1949). The number of individuals in a cascade process. *Proceedings of the Cambridge Philosophical Society*, **45**, 360–363.

[21] Gupta, R. C., and Ong, S. H. (2005). Analysis of long-tailed count data by Poisson mixtures. *Communications in Statistics–Theory and Methods*, **34**, 557–573.

[22] Haight, F. A., and Breuer M. A. (1960). The Borel–Tanner Distribution. *Biometrika Trust*, **47**, 143–150.

[23] Imoto, T. (to appear, a) Convolution of binomial and negative binomial variables. *Communications in Statistics–Theory and Methods.*

[24] Imoto, T. (to appear, b) Properties of Lagrangian distributions. *Communications in Statistics–Theory and Methods.*

[25] Imoto, T. (submitted) A generalized Conway–Maxwell–Poisson distribution which includes the negative binomial distribution.

[26] Imoto, T., and Ong, S. H., (2013). A Lagrangian non central negative binomial distribution of the first kind. *Communications in Statistics–Theory and Methods*, **42**, 466–477.

[27] Jain, G. C., and Consul, P. C. (1971). A generalized negative binomial distribution. *SIAM Journal of Applied Mathematics*, **21**, 501–513.

[28] Janardan, K. G. (1997). A wider class of Lagrange distributions of the second kind. *Communications in Statistics–Theory and Methods*, **26**, 2087–2091.

[29] Janardan, K. G., and Rao, B. R. (1983). Lagrange distributions of the second kind and weighted distributions. *SIAM Journal of Applied Mathematics*, **43**, 302–313.

[30] Johnson, N. L., Kemp, A. W., and Kotz, S. (2005). *Univariate Discrete Distributions. Third Edition.* New Jersey: Wiley.

[31] Katayama, N. (1992). Performance analysis and optimization of a cyclic-service tandem queueing system with multi-class customers. *Computers & Mathematics with Applications*, **24**, 25–33.

[32] Kemp, A. W. (1979). Convolutions involving binomial pseudo-variables. *Sankhyā: Series A*, **41**, 232–243.

[33] Kemp, A. W., and Kemp, C. W. (1968). On a distribution associated with certain stochastic process. *Journal of the Royal Statistical Society: Series B*, **30**, 160–163.

[34] Kocherlakota, S., and Kocherlakota, K. (1992). *Bivariate Discrete Distributions.* New York, Marcel Dekker.

[35] Kulasekera, K. B., and Tonkyn, D. W. (1992). A new discrete distribution, with applications to survival, dispersal and dispersion. *Communications in Statistics–Simulation and Computation*, **21**, 499–518.

[36] Lin, S-K. (1985). Characterization of lightning as a disturbance to the forest ecosystem in East Texas. M.Sc. thesis, Texas A&M University, College Station.

[37] Luke, G. (1969). *The Special Functions and Their Approximations*, Vol. I. New York, Academic Press.

[38] McKenzie, E. (1986). Autoregressive moving-average processes with negative-binomial and geometric marginal distributions. *Advances in Applied Probability*, **18**, 679–705.

[39] Menon, M. V. (1966). Characterization theorems for some univariate probability distributions. *Journal of the Royal Statistical Society: Series B*, **28**, 143–145.

[40] Nair, S. S. (1971). A single server tandem queue. *Journal of Applied Probability*, **8**, 95–109.

[41] Nelson, R. T. (1968). Dual-resource constrained series service systems. *Operations Research*, **16**, 324–341.

[42] Ong, S. H. (1988). A discrete Charlier series distribution. *Biometrical Journal*, **30**, 1003–1009.

[43] Ong, S. H. (1995). Some stochastic models leading to the convolution of two binomial variables. *Statistics & Probability Letters*, **22**, 161–166.

[44] Ong, S. H., and Lee, P. A. (1979). The non-central negative binomial distribution. *Biometrical Journal*, **21**, 611–627.

[45] Ong, S. H., Chakraborty, S., Imoto, T. and Shimizu, K. (2012). Generalizations of non-central negative binomial, Charlier series distributions and their extensions by Lagrange expansion. *Communications in Statistics–Theory and Methods*, **41**, 571–590.

[46] Otter, R. (1949). The multiplicative process. *The Annals of Mathematical Statistics*, **45**, 360–363.

[47] Panaretos, J. (1989). On the evolution of surnames. *International Statistical Review*, **57**, 161–179.

[48] Ridout, M., Demetrio, C. G. B., and Hinde, J. (1998). Models for count data with many zeros. International Biometric Conference, Cape Town, December.

[49] Riordan, J. (1968). *Combinatorial Identities.* New York: Wiley.

[50] Shimizu, K. and Yanagimoto, T. (1991). The inverse trinomial distribution. *Japanese Journal of Applied Statistics*, **20**, 89–96. In Japanese.

[51] Shmueli, G., Minka, T. P., Kadane, J. B., Borle, S., and Boatwright, P. (2005). A useful distribution for fitting discrete data: revival of the Conway-Maxwell-Poisson distribution. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, **54**, 127–142.

[52] Sibuya, M., Miyawaki, N., and Sumita, U. (1994). Aspects of Lagrangian probability distributions. *Journal of Applied Probability*, **31**, 185–197.

[53] Sundt, B. (1992). On some extensions of Panjer's class of counting distributions. *Astin Bulletin*, **22**, 61–80.

[54] Takács, L. (1967). *Combinatorial Methods in the Theory of Stochastic Process.* New York: Wiley.

[55] Van Oyen, M. P., and Teneketzis, D. (1994). Optimal stochastic scheduling of forest networks with switching penalties. *Advances in Applied Probability*, **26**, 474–497.

[56] Wimmer, G., Kohler, R., Grotjahn, R., and Altmann, G. (1994). Toward a theory of word length distributions. *Journal of Quantitative Linguistics*, **1**, 98–106.

[57] Yanagimoto, T. (1989). The inverse binomial distribution as a statistical model. *Communications in Statistics–Theory and Methods*, **18**, 3625–3633.

[58] Zörnig, P., and Altmann, G. (1995). Unified representation of Zipf distributions. *Computational Statistics & Data Analysis*, **19**, 461–473.