Research Report

KSTS/RR-05/003
Jun. 23, 2005

A generalization of the inverse trinomial

by

Kazuki Aoyama
Kunio Shimizu

Kazuki Aoyama
Keio University
Kunio Shimizu
Keio University

Department of Mathematics
Faculty of Science and Technology
Keio University

3-14-1 Hiyoshi, Kohoku-ku, Yokohama, 223-8522 Japan

KSTS/RR-05/003
June 23, 2005

A GENERALIZATION OF THE INVERSE TRINOMIAL

Kazuki Aoyama[1] and Kunio Shimizu[2]

[1]*School of Fundamental Science and Technology, Keio University, Yokohama 223-8522, Japan, e-mail: mr041101@hc.cc.keio.ac.jp*

[2]*Department of Mathematics, Keio University, Yokohama 223-8522, Japan, e-mail: shimizu@math.keio.ac.jp*

Abstract. The paper provides the generalized inverse trinomial (GIT) distribution as a univariate discrete distribution generated from a modified random walk on the half-plane. The inverse distribution of the GIT with respect to the cumulant generating function is also generated from a modified random walk on the half-plane. The GIT includes twenty-two possible distributions in total. Special cases are the binomial, negative binomial, shifted geometric, inverse binomial, inverse trinomial distributions. A subclass GIT_8 is represented by the independent sum of binomial and negative binomial. Compound or generalized (stopped sum) distributions are studied and some properties of inflated models are given.

Key words and phrases: Binomial, inflated model, inverse binomial, negative binomial, random walk, shifted geometric distribution.

1. Introduction

The univariate inverse trinomial (IT) distribution (Shimizu and Yanagimoto, 1991; Shimizu et al., 1997) is a discrete distribution generated from a modified random walk on the line. Here a particle starts from the origin and steps +1, 0, -1 with probabilities p, q, r $(p, q, r \geq 0; p + q + r = 1)$ respectively until it first reaches the barrier n (positive integer) at the xth step (Fig. 1). A random variable X for the number of steps x has the proper IT, denoted by $\mathrm{IT}(n; p, q, r)$ in this paper, when $p \geq r$. The probability generating function (pgf) of $\mathrm{IT}(n; p, q, r)$ is provided by

$$G_n(t) = E(t^X) = \left[\frac{1 - qt - \sqrt{(1 - qt)^2 - 4prt^2}}{2rt}\right]^n$$

and the probability function (pf) by

$$f_n(x) = \sum_{k=0}^{[(x-n)/2]} \frac{n}{x} \begin{pmatrix} x \\ n+k, x-n-2k, k \end{pmatrix} p^{n+k}\, q^{x-n-2k}\, r^k \tag{1.1}$$

for $x = n, n+1, n+2, \ldots$, where $[a]$ in (1.1) denotes the integral part of the number a and $\begin{pmatrix} x \\ x_1, x_2, x_3 \end{pmatrix} = x!/(x_1!x_2!x_3!)$, the trinomial coefficient under the assumption $x = x_1 + x_2 + x_3$. The $\mathrm{IT}(n; p, q, r)$ reduces to the inverse binomial (Yanagimoto, 1989) or equivalently lost-games distribution (Kemp and Kemp, 1969) if $q = 0$ and to the negative binomial if $r = 0$.

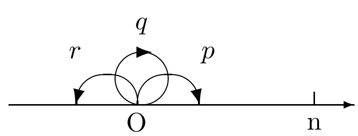


Fig. 1. Random walk for the inverse trinomial on the line

Fig. 2 shows another view of the random walk for the IT on the half-plane $(x \geq 0)$. Here a particle starts from the origin and, for non-negative integer x and integer y $(0 \leq y \leq n-1)$, the particle moves from (x, y) to $(x+1, y+1), (x+1, y), (x+1, y-1)$ with probabilities p, q, r respectively until it first reaches the barrier $y = n$. Notice that the particle moves from (x, y) to $(x+1, y+1)$ and $(x+1, y-1)$ directly with probabilities p and r without calling at $(x+1, y)$. When the particle first reaches the barrier, the coordinate x coincides with the number of steps in Fig. 1, and thus the IT is produced from the random walk pictured in Fig. 2. This readily leads to a generalization of the IT if probabilities from (x, y) to $(x, y+1)$ and $(x, y-1)$ are added to the transition probabilities for the IT. The proposed model with transition probabilities p_1, p_2, p_3, p_4, p_5 ($p_i \geq 0$ for $i = 1, \ldots, 5; \sum_{i=1}^{5} p_i = 1$) and barrier at $y = n$ (positive integer) is shown in Fig. 3. The resulting family of distributions is called the generalized inverse trinomial distribution and denoted by GIT$(n; p_1, p_2, p_3, p_4, p_5)$ or simply GIT.

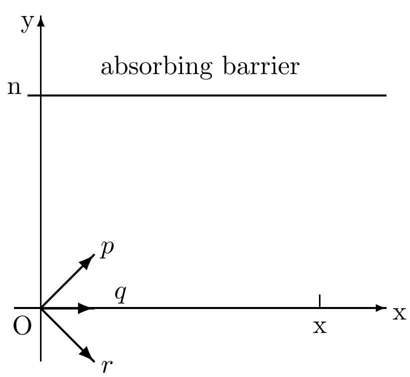


Fig. 2. Random walk for the inverse trinomial on the half-plane

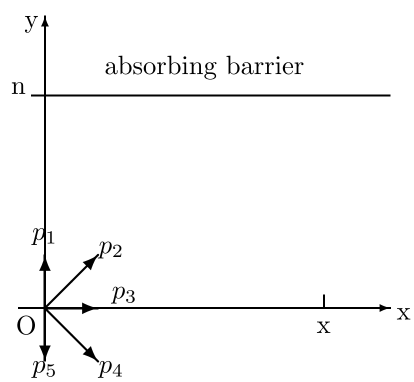


Fig. 3. Random walk for the generalized inverse trinomial

Section 2 provides the pgf and pf of the GIT. The pgf is given by solving the corresponding difference equation with boundary conditions and the pf by expanding the pgf. The proof is lengthy and is placed in Appendix A. The inverse distribution of the GIT with respect to the cumulant generating function is studied in Section 3 with proof in Appendix B. The GIT includes twenty-two possible distributions in total except one-trasition cases; not only the negative binomial, inverse binomial, inverse trinomial but also binomial and shifted geometric distributions. A subclass GIT$(n; p_1, p_2, 0, 0, 0)$ is denoted by $\mathrm{GIT}_1(n;p_1,p_2)$ as an example. The sub number ranges from 1 to 22. Section 4 shows that a distribution $\mathrm{GIT}_8(n;p_1,p_2,p_3)$ of the GIT family is represented as the sum of independent binomial and negative binomial. The reproductive property, moments and approximations of $\mathrm{GIT}_8(n;p_1,p_2,p_3)$ are also studied in Section 4. Finally, Section 5 deals with compound or generalized (stopped sum) distributions (Johnson, Kotz and Kemp, 1992). The inflated-parameter binomial, negative binomial and Poisson distributions by Minkova (2002) are extended in the section.

2. GIT: Generalized inverse trinomial distribution

The concept of a modified random walk on the half-plane to generate the GIT is introduced in Section 1 (Fig. 3). A particle starts from the origin and moves on the lattice of the half-plane as follows. For non-negative integer x and integer y, the particle moves from (x,y) to $(x,y+1)$, $(x+1,y+1)$, $(x+1,y)$, $(x+1,y-1)$, $(x,y-1)$ with probabilities p_1,p_2,p_3,p_4,p_5 ($p_i \geq 0$ for $i=1,\ldots,5; \sum_{i=1}^{5} p_i = 1$) respectively. The process ends once the particle reaches the barrier $y=n$ (positive integer). The $\mathrm{GIT}(n;p_1,p_2,p_3,p_4,p_5)$ is the distribution of a random variable X which represents the coordinate of the horizontal axis when the trials end. The probability function $f_n(x)$ of X satisfies the difference equation

$$f_n(x) = p_1 f_{n-1}(x) + p_2 f_{n-1}(x-1) + p_3 f_n(x-1) + p_4 f_{n+1}(x-1) + p_5 f_{n+1}(x)$$

with boundary conditions $f_0(0)=1,\ f_n(-1)=0,\ n\geq 0,\ f_0(x)=0,\ x\geq 1.$

The pgf $G_n(t)$ of $f_n(x)$ is defined by

$$G_n(t) = \sum_{x=0}^{\infty} f_n(x) t^x,\ n \geq 1,$$

which satisfies the recurrence relation

$$G_n(t) = p_1 G_{n-1}(t) + p_2 t G_{n-1}(t) + p_3 t G_n(t) + p_4 t G_{n+1}(t) + p_5 G_{n+1}(t)$$

or

$$(p_4 t + p_5) G_{n+1}(t) + (-1 + p_3 t) G_n(t) + (p_1 + p_2 t) G_{n-1}(t) = 0, \tag{2.1}$$

with boundary condition $G_0(t) = 1$. If $p_1 + p_2 > 0$, then the solution of (2.1) is provided by

$$\begin{aligned} G_n(t) &= \left[\frac{1 - p_3 t - \sqrt{(1 - p_3 t)^2 - 4(p_1 + p_2 t)(p_4 t + p_5)}}{2(p_4 t + p_5)} \right]^n \\ &= \left[\frac{2(p_1 + p_2 t)}{1 - p_3 t + \sqrt{(1 - p_3 t)^2 - 4(p_1 + p_2 t)(p_4 t + p_5)}} \right]^n \end{aligned} \tag{2.2}$$

for $0 \le t \le 1$. Its proof is in Appendix A. Note that

$$G_n(1) = \begin{cases} 1, & p_1 + p_2 \ge p_4 + p_5, \\ \frac{p_1 + p_2}{p_4 + p_5}, & p_1 + p_2 < p_4 + p_5, \end{cases}$$

from which the condition for which (2.2) gives a proper distribution is $p_1 + p_2 \ge p_4 + p_5$ and then (2.2) is the pgf of the GIT. If $p_1 + p_2 < p_4 + p_5$, (2.2) does not provide a pgf since $\sum_{x=0}^{\infty} f_n(x) = (p_1 + p_2)/(p_4 + p_5) < 1$. In this case the distribution is improper. However, (2.2) with $p_1 + p_2 < p_4 + p_5$ gives a distribution if a probability $1 - (p_1 + p_2)/(p_4 + p_5)$ is added at $x = \infty$.

The pf of the GIT is provided by

$$\begin{aligned} f_n(x) = \sum_{k=0}^{x} \sum_{l=0}^{\infty} \sum_{i=0}^{m} & \frac{n}{n + x - i + k + 2l} \binom{n + x - i + k + 2l}{n - i + k + l, i, x - i - k, k, l} \\ & \times p_1^{n-i+k+l}\, p_2^i\, p_3^{x-i-k}\, p_4^k\, p_5^l, \end{aligned} \tag{2.3}$$

which is obtained from the expansion of (2.2) about t, where $m = \min(n+k+l, x-k)$ and

$\begin{pmatrix} x \\ x_1, x_2, x_3, x_4, x_5 \end{pmatrix} = x!/(x_1!x_2!x_3!x_4!x_5!)$, which is the multinomial coefficient under the assumption $x = x_1 + x_2 + x_3 + x_4 + x_5$.

A model (Fig. 4) is considered by adding a stay probability p_6 $(0 < p_6 < 1)$ from (x, y) to (x, y) to the GIT model above, where $\sum_{i=1}^{6} p_i = 1$. However, this model does not produce an extended family of distributions different from the GIT. The reason is as follows. Apparent new pf $f_n(x)$ and pgf $G_n(t)$ satisfy the difference equation

$$f_n(x) = p_1 f_{n-1}(x) + p_2 f_{n-1}(x-1) + p_3 f_n(x-1) + p_4 f_{n+1}(x-1) + p_5 f_{n+1}(x) + p_6 f_n(x)$$

and recurrence relation

$$(1 - p_6)G_n(t) = p_1 G_{n-1}(t) + p_2 t G_{n-1}(t) + p_3 t G_n(t) + p_4 t G_{n+1}(t) + p_5 G_{n+1}(t)$$

respectively, from which division of both side by $1 - p_6$ leads to

$$G_n(t) = p_1' G_{n-1}(t) + p_2' t G_{n-1}(t) + p_3' t G_n(t) + p_4' t G_{n+1}(t) + p_5' G_{n+1}(t),$$

where $p_i' = p_i/(1 - p_6)$ for $i = 1, \ldots, 5$. This recurrence relation is the same type as (2.1).

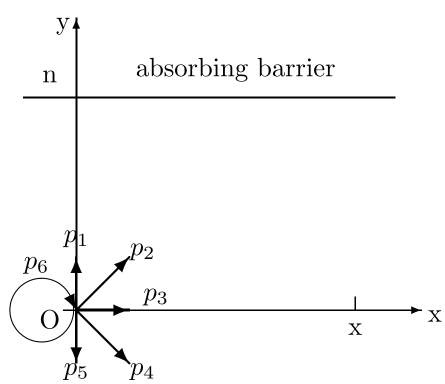


Fig. 4. Random walk for the generalized inverse trinomial with added stay-probability

3. Inverse family of the GIT

When X is distributed as the GIT and its cumulant generating function (cgf) is $C(t) = \log E(e^{-tX})$, the inverse distribution of the GIT is defined by the distribution whose cgf is given by the inverse function $C^{-1}(t)$ of $C(t)$. Since the cgf of GIT$(n; p_1, p_2, p_3, p_4, p_5)$ is

$$C(t) = -n \log \left[\frac{1 - p_3 e^{-t} + \sqrt{(1 - p_3 e^{-t})^2 - 4(p_1 + p_2 e^{-t})(p_4 e^{-t} + p_5)}}{2(p_1 + p_2 e^{-t})} \right], \tag{3.1}$$

its inverse function is provided by

$$C^{-1}(t) = \log \left[\frac{p_2 e^{-2t/n} + p_3 e^{-t/n} + p_4}{-p_1 e^{-2t/n} + e^{-t/n} - p_5} \right]. \tag{3.2}$$

Consider the following modified random walk. A particle starts from the origin and, for intergers x and y $(0 \le y \le m-1)$, moves from (x, y) to $(x+1, y)$, $(x+1, y+1)$, $(x, y+1)$, $(x-1, y+1)$, $(x-1, y)$ with probabilities p_1, p_2, p_3, p_4, p_5 ($p_i \ge 0$ for $i = 1, \ldots, 5$; $\sum_{i=1}^{5} p_i = 1$) respectively until it first reaches the barrier $y = m$ (Fig. 5). If we denote a random variable Z which represents the coordinate x of the horizontal axis when the trials end, then a random variable Z/n in the case $m = 1$ has a distribution whose cgf is $C^{-1}(t)$ in (3.2). See Appendix B for details.

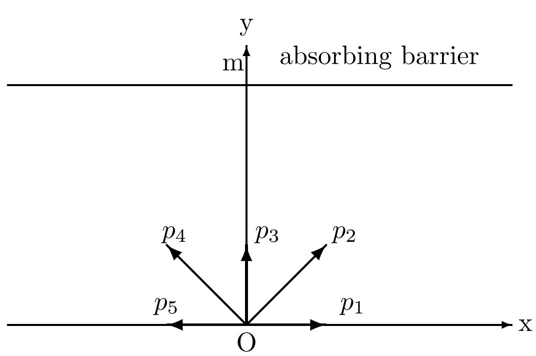


Fig. 5. Random walk for the inverse family of the generalized inverse trinomial

4.1 *Some examples*

There exist twenty-two possible distributions of the GIT in total if one or some of p_1, p_2, p_3, p_4, p_5 are substituted by zero except one-transition cases where $p_1 = 1,\ p_2 = p_3 = p_4 = p_5 = 0$ and $p_2 = 1,\ p_1 = p_3 = p_4 = p_5 = 0$. Table 1 summarizes the possibility of the distributions. Some known classical distributions which belong to the family are below. If $p_3, p_4, p_5 = 0$ for example, the distribuiton is denoted by $\mathrm{GIT}_1(n; p_1, p_2)$.

(a) Binomial $\mathrm{GIT}_1(n; p_1, p_2)$, also denoted by $B(n, p_2)$.

$$G_n(t) = (p_1 + p_2 t)^n,$$

$$f_n(x) = \binom{n}{x} p_2^x\ p_1^{n-x}$$

for $x = 0, 1, \ldots, n$.

(b) Negative binomial $\mathrm{GIT}_2(n; p_1, p_3)$, also denoted by $\mathrm{NB}(n, p_3)$.

$$G_n(t) = \left(\frac{p_1}{1 - p_3 t}\right)^n,$$

$$f_n(x) = \binom{n + x - 1}{x} p_3^x\ p_1^n$$

for $x = 0, 1, \ldots$

(c) Shifted geometric $\mathrm{GIT}_3(1; p_2, p_3)$.

$$G_n(t) = \left(\frac{p_2 t}{1 - p_3 t}\right),$$

$$f_n(x) = p_2\ p_3^{x-1}$$

for $x = 1, 2, \ldots$

(d) Inverse binomial $\mathrm{GIT}_6(n; p_2, p_4)$.

$$G_n(t) = \left[\frac{1-\sqrt{1-4p_2p_4t^2}}{2p_4t}\right]^n,$$

$$f_n(x) = \frac{n}{x}\binom{x}{\frac{x-n}{2}} p_2^{\frac{n+x}{2}}\; p_4^{\frac{x-n}{2}}$$

for $x = n, n+2, n+4, \ldots$

(e) Inverse trinomial $\mathrm{GIT}_{11}(n; p_2, p_3, p_4)$.

$$G_n(t) = \left[\frac{1-p_3t-\sqrt{(1-p_3t)^2-4p_2p_4t^2}}{2p_4t}\right]^n,$$

$$f_n(x) = \sum_{k=0}^{[(x-n)/2]} \frac{n}{x}\binom{x}{n+k, x-n-2k, k} p_2^{n+k}\; p_3^{x-n-2k}\; p_4^k$$

for $x = n, n+1, n+2, \ldots$

4.2 *Properties of $GIT_8(n; p_1, p_2, p_3)$*

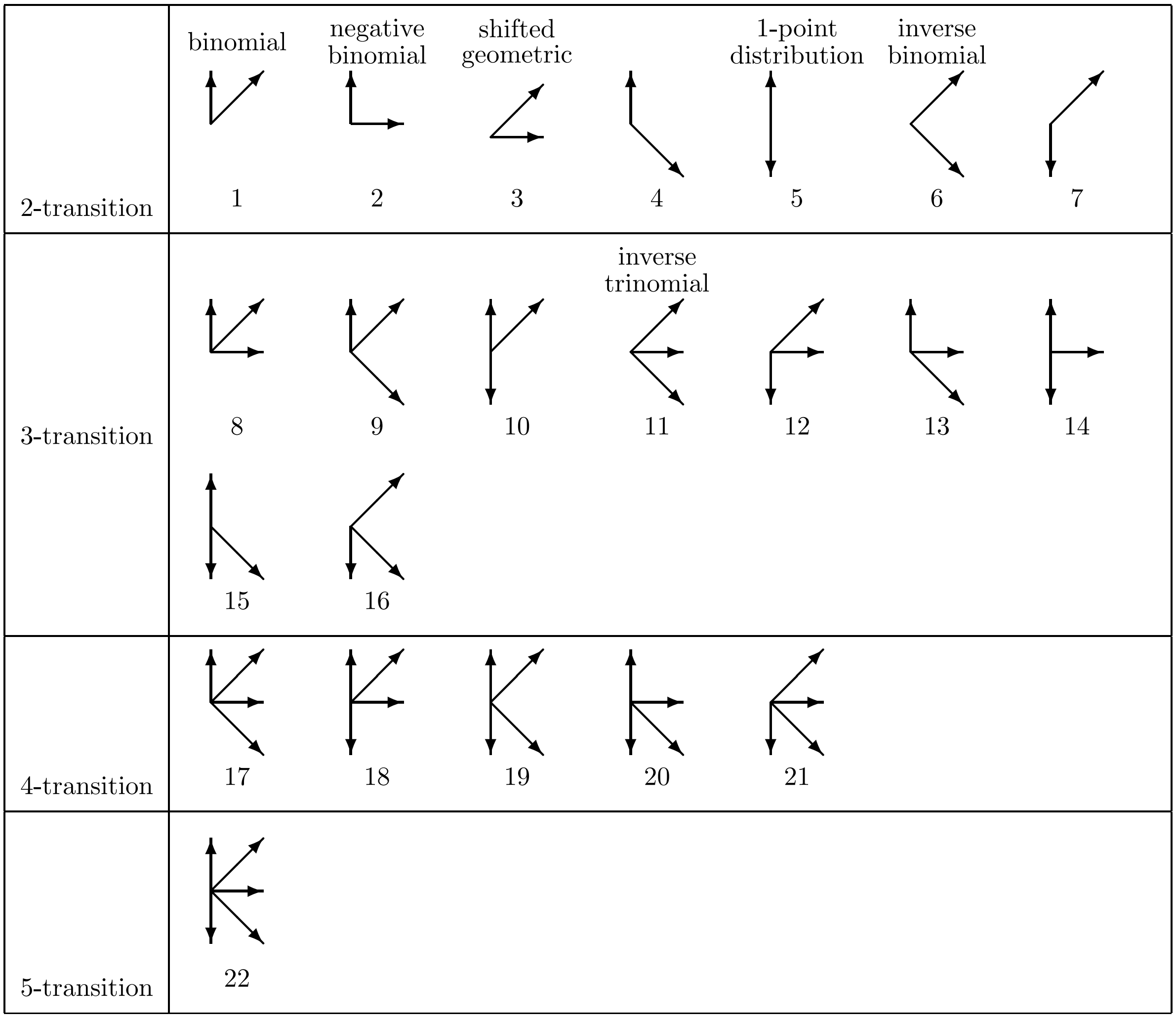
The distribution $\mathrm{GIT}_8(n; p_1, p_2, p_3)$ generalizes the binomial, negative binomial and shifted geometric distributions. Actually the pgf of $\mathrm{GIT}_8(n; p_1, p_2, p_3)$

$$G_n(t) = \left(\frac{p_1+p_2t}{1-p_3t}\right)^n \tag{4.1}$$

reduces to the binomial pgf if $p_3 = 0$, to the negative binomial pgf if $p_2 = 0$ and to the shifted geometric pgf if $n = 1$, $p_1 = 0$. The pf of $\mathrm{GIT}_8(n; p_1, p_2, p_3)$ is given by

$$f_n(x) = \sum_{i=0}^{\min(n,x)} \frac{n}{n+x-i}\binom{n+x-i}{n-i, i, x-i} p_1^{n-i}\; p_2^i\; p_3^{x-i} \tag{4.2}$$

Table 1. Subclasses of the generalized inverse trinomial



for $x = 0, 1, \ldots$ Some properties of $\mathrm{GIT}_8(n; p_1, p_2, p_3)$ are summarized below except for those of compound distributions, which are studied in Section 5. Here X indicates a random variable having $\mathrm{GIT}_8(n; p_1, p_2, p_3)$.

(a) Reproductive property. If X_1, X_2 are independent random variables and distributed as $\mathrm{GIT}_8(n; p_1, p_2, p_3)$ and $\mathrm{GIT}_8(m; p_1, p_2, p_3)$ respectively, then the sum $X_1 + X_2$ is distributed as $\mathrm{GIT}_8(n + m; p_1, p_2, p_3)$.

(b) The random variable X is expressible as the sum of two independent random variables X_1 and X_2, where X_1 has the binomial $B(n, p_2/(p_1 + p_2))$ and X_2 the negative binomial $\mathrm{NB}(n, p_3)$.

(c) Poisson approximation. The distribution $\mathrm{GIT}_8(n; p_1, p_2, p_3)$ goes to a Poisson distribution with parameter $\lambda_2 + \lambda_3$ as n tends to infinity remaining the relation $np_2 = \lambda_2$ and $np_3 = \lambda_3$.

(d) The rth descending factorial moment of X is

$$E(X(X-1)\cdots(X-r+1)) = \frac{n}{(p_1+p_2)^r}\sum_{i=0}^{r}\left(\begin{array}{c} r \\ \\ i \end{array}\right)\frac{(n+r-i-1)!}{(n-i)!}p_2^i p_3^{r-i}$$

for $r \geq 1$.

(e) The mean and variance of X are obtainable from (d). They are also obtained by using (b). Actually

$$E(X) = E(X_1 + X_2) = n\frac{p_2}{p_1+p_2} + n\frac{p_3}{1-p_3} = n\frac{p_2+p_3}{p_1+p_2},$$

$$V(X) = V(X_1 + X_2) = n\left(\frac{p_1}{p_1+p_2}\right)\left(\frac{p_2}{p_1+p_2}\right) + n\frac{p_3}{(1-p_3)^2} = n\frac{p_1p_2+p_3}{(p_1+p_2)^2},$$

from which the index of dispersion (ID) is

$$\mathrm{ID} = \frac{V(X)}{E(X)} = \frac{p_1p_2+p_3}{(p_1+p_2)(p_2+p_3)}\begin{cases} > 1, \; p_3 > p_2, \\ \\ < 1, \; p_3 < p_2. \end{cases}$$

If $p_2 = p_3$, then ID$= 1$. Note that $\mathrm{GIT}_8(n; 1-2p, p, p)$ with $0 < p < 1/2$ is not a Poisson distribution, but its ID is unity.

(f) Normal approximation. The distribution of $(X - E(X))/\sqrt{V(X)}$ goes to a standard normal distribution as n tends to infinity.

(g) The pf $f_n(x)$ of X satisfies the recurrence relation

$$f_n(x) = \left(a + \frac{b}{x}\right) f_n(x-1) + c\left(1 - \frac{2}{x}\right) f_n(x-2)$$

for $x \geq 2$ with initial conditions $f_n(0) = p_1^n$, $f_n(1) = np_1^{n-1}(p_1p_3+p_2)$, which is an example of Sundt's (1992) recursion, where $a = (p_1p_3 - p_2)/p_1$, $b = (n(p_1p_3+p_2) - (p_1p_3-p_2))/p_1$,

$c = (p_2 p_3)/p_1$ with $p_1 > 0$.

(h) The rth moment, $\mu_r' = E(X^r)$, of X about zero satisfies the recurrence relation

$$\mu_r' = \sum_{j=0}^{r-1} \begin{pmatrix} r-1 \\ j \end{pmatrix} \{(a + 2^{r-j-1}c)\mu_{j+1}' + (a+b)\mu_j'\}$$

for $r \geq 1$ with initial condition $\mu_0' = 1$ and the understanding that $0! = 1$, where a, b and c are given in (g). This is proved by using the recurrence relation of the pf in (g) as follows

$$\begin{aligned}
\mu_r' &= \sum_{x=1}^{\infty} x^r f(x) \\
&= f(1) + \sum_{x=0}^{\infty} (x+2)^r f(x+2) \\
&= \left(a + \frac{b}{1}\right) f(0) + \sum_{x=0}^{\infty} (x+2)^r \left(a + \frac{b}{x+2}\right) f(x+1) + \sum_{x=0}^{\infty} (x+2)^r c \left(1 - \frac{2}{x+2}\right) f(x) \\
&= \sum_{x=0}^{\infty} (x+1)^r \left(a + \frac{b}{x+1}\right) f(x) + \sum_{x=0}^{\infty} (x+2)^r c \left(1 - \frac{2}{x+2}\right) f(x) \\
&= \sum_{x=0}^{\infty} (x+1)^r \left(\frac{ax + (a+b)}{x+1}\right) f(x) + \sum_{x=0}^{\infty} (x+2)^r c \left(\frac{x}{x+2}\right) f(x) \\
&= \sum_{j=0}^{r-1} \sum_{x=0}^{\infty} \begin{pmatrix} r-1 \\ j \end{pmatrix} \{a x^{j+1} + (a+b) x^j\} f(x) + \sum_{j=0}^{r-1} \sum_{x=0}^{\infty} \begin{pmatrix} r-1 \\ j \end{pmatrix} 2^{r-1-j} c x^{j+1} f(x).
\end{aligned}$$

5. Compound distributions

5.1 *Inflated model*

This section studies the distribution of the random variable $S = X_1 + \cdots + X_N$ with the understanding that $S = 0$ when $N = 0$, where $X_1, X_2, \ldots$ are independent and identically distributed (iid) as $\text{GIT}_8(1; q_1, q_2, q_3)$, N as $\text{GIT}(n; p_1, p_2, p_3, p_4, p_5)$, and N is independent of $X_1, X_2, \ldots$ Let $G_X(t)$ and $G_N(t)$ denote the pgf's of X_i $(i = 1, 2, \ldots)$ and

N respectively. Then, from (4.1) and (2.2), the pgf $G_S(t)$ of S is provided by

$$\begin{aligned}G_S(t) &= G_N\left(G_X(t)\right) \\ &= \left[\frac{1-\alpha_3 t+\sqrt{(1-\alpha_3 t)^2-4(\alpha_1+\alpha_2 t)(\alpha_4 t+\alpha_5)}}{2(\alpha_1+\alpha_2 t)}\right]^{-n},\end{aligned} \tag{5.1}$$

which is the pgf of GIT$(n;\alpha_1,\alpha_2,\alpha_3,\alpha_4,\alpha_5)$ for $0\le t\le 1$, where $\alpha_1=(p_1+p_2q_1)/(1-p_3q_1)$, $\alpha_2=(-p_1q_3+p_2q_2)/(1-p_3q_1)$, $\alpha_3=(p_3q_2+q_3)/(1-p_3q_1)$, $\alpha_4=(p_4q_2-p_5q_3)/(1-p_3q_1)$, $\alpha_5=(p_4q_1+p_5)/(1-p_3q_1)$. Note that the range of parameters is extended to $\alpha_1+\alpha_2+\alpha_3+\alpha_4+\alpha_5=1$, $-1\le\alpha_2,\alpha_4\le 1$, $0\le\alpha_1,\alpha_3,\alpha_5\le 1$, $\alpha_1\alpha_3+\alpha_2\ge 0$, $\alpha_4+\alpha_3\alpha_5\ge 0$, $\alpha_1+\alpha_2>0$ and $\alpha_1+\alpha_2\ge\alpha_4+\alpha_5$, whereas $p_i\ge 0$ $(i=1,\ldots,5)$, $\sum_{i=1}^5 p_i=1$, $p_1+p_2>0$, $p_1+p_2\ge p_4+p_5$, and $q_j\ge 0$ $(j=1,2,3)$, $\sum_{j=1}^3 q_j=1$. Thus (5.1) is an inflated model of the GIT. As a particular case, if $X_1,X_2,\ldots$ are iid as GIT$_8(1;q_1,q_2,q_3)$ and N as GIT$_8(n;p_1,p_2,p_3)$, and N is independent of $X_1,X_2,\ldots$, then $S=X_1+\cdots+X_N$ has an inflated GIT$_8(n;\gamma_1,\gamma_2,\gamma_3)$, where $\gamma_1=(p_1+p_2q_1)/(1-p_3q_1)$, $\gamma_2=(-p_1q_3+p_2q_2)/(1-p_3q_1)$, $\gamma_3=(p_3q_2+q_3)/(1-p_3q_1)$, $\gamma_1+\gamma_2+\gamma_3=1$, $-1\le\gamma_2\le 1$, $0\le\gamma_1,\gamma_3\le 1$. This shows that even the inflated GIT$_8$ is closed under the generalization by the inflated GIT$_8$.

5.2 *A comment*

Minkova (2002) studied the family of inflated-parameter power series distributions or a shifted geometric distribution generalized by the generalizing power series distribution. If X has a shifted geometric distribution and N the Poisson, binomial, negative binomial, logarithmic series as a member of the power series distributions, then the distribution of $S=X_1+\cdots+X_N$ is called the inflated-parameter Poisson (IPo), binomial (IBi), negative binomial (INB), logarithmic series respectively. The inflated GIT$_8$ in

Section 5.1 extends the IPo, IBi, INB because the GIT_8 includes the shifted geometric distribution $\mathrm{GIT}_3(1; q_2, q_3)$ as well as the binomial $\mathrm{GIT}_1(1; p_1, p_2)$ and negative binomial $\mathrm{GIT}_2(1; p_1, p_3)$. The inflated GIT in Section 5.1 is more extended.

Acknowledgements

The work was supported in part by The Ministry of Education, Culture, Sports, Science and Technology in Japan under a Grant-in-Aid of the 21st Century Center of Excellence for Integrative Mathematical Sciences: Progress in Mathematics Motivated by Social and Natural Sciences.

Appendix A: The probability function of the GIT

The pf of the GIT is found from the difference equation given by (2.1)

$$(p_4t + p_5)G_{n+1}(t) + (-1 + p_3t)G_n(t) + (p_1 + p_2t)G_{n-1}(t) = 0$$

with boundary condition $G_0(t) = 1$. We look for particular solutions $G_n(t)$ of the form $G_n(t) = \{\lambda(t)\}^n$. Substitution of this expression into (2.1) gives the quadratic equation

$$(p_4t + p_5)\lambda^2(t) + (-1 + p_3t)\lambda(t) + (p_1 + p_2t) = 0, \tag{A.1}$$

which has the two roots

$$\lambda_1(t) = \frac{1 - p_3t - \sqrt{(1 - p_3t)^2 - 4(p_4t + p_5)(p_1 + p_2t)}}{2(p_4t + p_5)},$$

$$\lambda_2(t) = \frac{1 - p_3t + \sqrt{(1 - p_3t)^2 - 4(p_4t + p_5)(p_1 + p_2t)}}{2(p_4t + p_5)}.$$

The range of t for which $(1-p_3t)^2 - 4(p_4t+p_5)(p_1+p_2t) \ge 0$ is $0 \le t \le (b-\sqrt{b^2 - ac})/a$, with $a = p_3^2 - 4p_2p_4$, $b = p_2 + 2p_2p_5 - 4p_1p_4$, $c = 1 - 4p_1p_5$ and $(b-\sqrt{b^2 - ac})/a > 1$. Since

$0 < \lambda_1(t) < 1$ and $\lambda_2(t) > 1$ for $0 < t < 1$, $\{\lambda_2(t)\}^n$ is inappropriate for a solution and $G_n(t) = B(t)\{\lambda_1(t)\}^n$ is a solution to (A.1). From the boundary condition $G_0(t) = 1$, we obtain $B(t) = 1$. Hence the solution is

$$G_n(t) = \{\lambda_1(t)\}^n. \tag{A.2}$$

The pf is provided by expanding (A.2) about t. From the formula (Abramowitz and Stegun, 1972 [15.1.13])

$$ {}_2F_1\left(a, \frac{1}{2} + a; 1 + 2a; z\right) = 2^{2a}\{1 + (1 - z)^{1/2}\}^{-2a}$$

for $|z| < 1$ where ${}_2F_1$ stands for the Gauss hypergeometric function, (A.2) is transformed into

$$\begin{aligned}
G_n(t) &= \left[\frac{1 - p_3t + \sqrt{(1 - p_3t)^2 - 4(p_4t + p_5)(p_1 + p_2t)}}{2(p_1 + p_2t)}\right]^{-n} \\
&= \left(\frac{1 - p_3t}{2(p_1 + p_2t)}\right)^{-n}\left[1 + \sqrt{1 - \frac{4(p_1 + p_2t)(p_4t + p_5)}{(1 - p_3t)^2}}\right]^{-n} \\
&= \left(\frac{1 - p_3t}{2(p_1 + p_2t)}\right)^{-n} {}_2F_1\left[\frac{n}{2}, \frac{n + 1}{2}; n + 1; \frac{4(p_1 + p_2t)(p_4t + p_5)}{(1 - p_3t)^2}\right] \\
&= \left(\frac{1 - p_3t}{2(p_1 + p_2t)}\right)^{-n} \sum_{k=0}^{\infty} \frac{\left(\frac{n}{2}\right)_k \left(\frac{n+1}{2}\right)_k}{(n + 1)_k\, k!} \left(\frac{4(p_4t + p_5)(p_1 + p_2t)}{(1 - p_3t)^2}\right)^k,
\end{aligned}$$

where $(x)_i = x(x + 1)\cdots(x + i - 1) = \Gamma(x + i)/\Gamma(x)$. From the duplication formula for the gamma function

$$\Gamma(2z) = \frac{1}{(2\pi)^{1/2}} 2^{2z-1/2}\Gamma(z)\Gamma(z + 1/2),$$

we obtain

$$G_n(t) = \sum_{k=0}^{\infty} \frac{n}{n + k} \binom{n + 2k - 1}{k} (p_1 + p_2t)^{n+k}(p_4 + p_5)^k \left(\frac{1}{1 - p_3t}\right)^{n+2k}$$

$$= \sum_{j}^{\infty}\sum_{k}^{\infty}\sum_{l}^{k}\sum_{i}^{n+k} \frac{n}{n+2k+j} \begin{pmatrix} n+2k+j \\ n+k-i, i, j, k-l, l \end{pmatrix} p_1^{n+k-i} p_2^i p_3^j p_4^{k-l} p_5^l t^{i+j+k-l},$$

which leads to

$$\sum_{x}^{\infty}\sum_{k'}^{\infty}\sum_{l}^{\infty}\sum_{i}^{n+k'+l} \frac{n}{n+2k'+2l+j} \begin{pmatrix} n+2k'+2l+j \\ n+k'+l-i, i, j, k', l \end{pmatrix} p_1^{n+k'+l-i} p_2^i p_3^j p_4^{k'} p_5^l t^{i+j+k'}$$

after the replacement $k - l = k'$. The coefficient of t^x, where $x = i + j + k'$, in $G_n(t)$ is (2.3), the pf of the GIT.

Appendix B: The inverse distribution of the GIT

We consider the distribution derived from the modified random walk defined in Section 3 (Fig. 5). The difference equation of the pf is

$$h_m(x) = p_1 h_m(x-1) + p_2 h_{m-1}(x-1) + p_3 h_{m-1}(x) + p_4 h_{m-1}(x+1) + p_5 h_m(x+1)$$

with initial condition $h_0(x) = 1$ if $x = 0$, 0 if $x \neq 0$, and the recurrence relation of the corresponding pgf is

$$H_m(t) = p_1 t H_m(t) + p_2 t H_{m-1}(t) + p_3 H_{m-1}(t) + p_4 t^{-1} H_{m-1}(t) + p_5 t^{-1} H_m(t)$$

with initial condition $H_0(t) = 1$. From this we obtain the pgf

$$H_m(t) = \left(\frac{p_2 t^2 + p_3 t + p_4}{-p_1 t^2 + t - p_5} \right)^m .$$

On the other hand to get the inverse of (3.1), we set $\log E(e^{-tX}) = s$. Then

$$\frac{1 - p_3 e^{-t} - \sqrt{(1 - p_3 e^{-t})^2 - 4(p_1 + p_2 e^{-t})(p_4 e^{-t} + p_5)}}{2(p_1 + p_2 e^{-t})} = e^{-s/n}.$$

If we set $e^{-t} = T$ and $e^{-s/n} = S$, then we have

(B.1) $0 = S^2(p_1^2 + 2p_1p_2T + p_2^2T^2) + S(p_2p_3T^2 + p_1p_3T - p_2T - p_1)$

$$
\begin{aligned}
&+(p_2p_4T^2+p_1p_4T+p_2p_5T+p_1p_5)\\
&= p_2^2\left(S^2+\frac{p_3}{p_2}S+\frac{p_4}{p_2}\right)T^2+p_1p_2\left\{2S^2+\left(\frac{p_3}{p_2}-\frac{1}{p_1}\right)S+\left(\frac{p_4}{p_2}+\frac{p_5}{p_1}\right)\right\}T\\
&+p_1^2\left\{S^2-\frac{1}{p_1}S+\frac{p_5}{p_1}\right\}\\
&= p_2^2(S^2+aS+c)T^2+p_1p_2\left\{2S^2+(a-b)S+(c+d)\right\}T+p_1^2(S^2-bS+d)\\
&= AT^2+BT+C,
\end{aligned}
$$

where $a=p_3/p_2,\ b=1/p_1,\ c=p_4/p_2,\ d=p_5/p_1,\ A=p_2^2(S^2+aS+c),\ B=p_1p_2(2S^2+(s-b)S+(c+d)),\ C=p_1^2(S^2-bS+d)$. The solution of (B.1) is

$$
T=\frac{-B\pm\sqrt{B^2-4AC}}{2A}
$$

with $B^2-4AC=p_1^2p_2^2\{(a+b)S+(c-d)\}^2\geq 0$. We see that

$$
\begin{aligned}
T &= \frac{-p_1p_2\{2S^2+(a-b)S+(c+d)\}\pm p_1p_2\{(a+b)S+(c-d)\}}{2(p_2^2S^2+p_2p_3S+p_2p_4)}\\
&= \frac{-p_1S^2+S-p_5}{p_2S^2+p_3S+p_4},\ \frac{-p_1p_2S^2-p_1p_3S-p_1p_4}{p_2^2S^2+p_2p_3S+p_2p_4}
\end{aligned}
$$

and the second solution is inappropriate since it is negative. Thus we obtain (3.2).

References

Abramowitz, M. and Stegun, I. A. Eds. (1972). *Handbook of Mathematical Functions*, Dover Publications, Inc., New York.

Johnson, N. L., Kotz, S. and Kemp, A. W. (1992). *Univariate Discrete Distributions*, Wiley, New York.

Kemp, A. W. and Kemp, C. D. (1969). Branching and clustering models associated with the 'lost-games' distribution, *Journal of Applied Probability*, **6**, 700–703.

Minkova, L. D. (2002). A generalization of the classical discrete distributions, *Communications in Statistics-Theory and Methods*, **31**(6), 871–888.

Shimizu, K. and Yanagimoto, T. (1991). The inverse trinomial distribution, *Japanese Journal of Applied Statistics*, **20**(2), 89–96. In Japanese.

Shimizu, K., Nishii, N. and Minami, M. (1997). The multivariate inverse trinomial distribution as a Lagrangian probability model, *Communications in Statistics-Theory and Methods*, **26**(7), 1585–1598.

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

Yanagimoto, T. (1989). The inverse binomial distribution as a statistical model, *Communications in Statistics-Theory and Methods*, **18**(10), 3625–3633.