The **J**ournal of
Mathematics and **C**omputer **S**cience

The Generalized Power Series Distributions and their Application

Fereshteh Momeni

Islamic Azad University-Behshahr Branch
momeni@iaubs.ac.ir

Abstract

Necessity of the use of more rich family discrete distributions is practically observable. In this paper, we study about new family of discrete distributions, as Inflated Parameter distributions. Then , we use the application of these distributions for data modeling of Insurance claims and indicate that these distributions have more appropriate approximation then corresponding distributions.

  Keywords: Generalized Power Series Distribution, zero-Inflated distribution.

1. Introduction

The important distributions are extensively used in counting: Binomial, Negative binomial and Poisson distributions. According to be in appropriating classic distributions in modeling of some frequency data or depended data about different areas as Economics , Finance ,Insurance , Ecological and etc ... , Bowers et.al [5], Rolski et.al [6] , Winkelman [6] and Kumer et.al [7] have emphasized the generalized necessity of these distributions. Classic distributions are in doubt about Insurance data modeling. Here, in addition to introducing Inflated-Parameter Power Series distributions, we will indicate the preference of these distributions on corresponding classic distribution by using a real data set related to the number of accident Insurance claims.

2. GPS distributions

Several known univariate discrete probability distributions belong to the class of generalized power series (GPS) distributions. The general from of the probability mass function (pmf) of a random variable (r.v) N of a GPS distribution with parameter $\theta > 0$ is given by

$$P(N = k) = \frac{a(k)\theta^k}{g(\theta)} \quad , \quad k \in S \quad , \quad \theta > 0 \tag{1}$$

Where S is any nonempty enumerable set of none negative integers, $a(k) \geq 0$ and $g(\theta) = \sum_{k \in S} a(k)\theta^k$ is the normalizing constant. Their corresponding probability generating function (pgf) of N has the following from

$$P_N(t) = \frac{g(\theta t)}{g(t)} \quad , \quad t \in [0,1] \tag{2}$$

The Bin(n,π) ,NB(r,π) , Po(λ) and LS(π) distributions belong to this class. This is illustrated in Table (1) below.

Table (1)

	θ	$a(k)$	$g(\theta)$	S	pmf	pgf
$Bin(n,\pi)$	$\dfrac{\pi}{1-\pi}$	$\binom{n}{k}$	$(1+\theta)^n$	$0,1,\dots,n$	$\binom{n}{k}\pi^k(1-\pi)^{n-k}$	$(1-\pi+\pi t)^n$
$Po(\lambda)$	λ	$\dfrac{1}{k!}$	e^θ	$0,1,2,\dots$	$\dfrac{\lambda^k}{k!}e^{-k}$	$e^{\lambda(t-1)}$
$NB(r,\pi)$	$1-\pi$	$\binom{n+k-1}{k}$	$(1-\theta)^{-r}$	$0,1,2,\dots$	$\binom{r+k-1}{k}\pi^r(1-\pi)^k$	$\left(\dfrac{\pi}{1-(1-\pi)t}\right)^r$
$LS(\pi)$	$1-\pi$	$\dfrac{1}{k}$	$-\ln(1-\theta)$	$1,2,3,\dots$	$\dfrac{(1-\pi)^k}{-k\ln\pi}$	$\dfrac{\ln[1-(1-\pi)t]}{\ln\pi}$

Among the elements studied in Risk Theory is the r.v. sum$S_N = X_1 + X_2 + \cdots + X_N$, $S_0 = 0$, called Compound r.v. It describes the aggregate claim amount in the risk model. We assume that N, $X_1, X_2, \dots$ are mutually independent r.v. and say that S_N has a compound distribution. If we suppose that the r.v. N has the GPS distributions. Then the relation (3) of the pmf is useful for deriving the compound distribution.

In Panjer (1981) is shown that the pmf of GPS distributions satisfies the following recursion:

$$p_k = \left(a + \frac{b}{k}\right)p_{k-1} \quad , \quad k = 1,2,\dots \quad , \quad a < 1 \quad , \quad b \geq 0$$

In the particular cases , the constants a and b are given by the following expressions :

$X \sim Bin(n, \pi)$: $\quad a = \dfrac{-\pi}{1 - \pi\pi} \quad , \quad b = \dfrac{(n+1)\pi}{1 - \pi}$

$X \sim Po(\lambda)$: $\quad a = 0 \quad , \quad b = \lambda$

$X \sim NB(r, \pi)$: $\quad a = \pi \quad , \quad b = (r-1)\pi$

3. Zero-Inflated GPS distributions

Gupta et al (1995), Kolev et al (2000) and Minkova (2002) introduced and studied inflated extensions of GPS distributions.

Let ξ be an arbitrary none negative integer-valued r.v. such that $P(\xi = j) = p_j$, $j = 0,1,2, \ldots$ and $\sum_{j=0}^{\infty} p_j = 1$. Also , Let $P_\xi(t) = E(t^\xi)$ be its pgf. An extra proportion of zeros $\rho \in (0 , 1)$ is added to the proportion of zeros from the distribution of the r.v. ξ , while decreasing the remaning proportions in an appropriate way. The zero-Inflated modification η of ξ is defined by

$$P(\eta = j) = \begin{cases} \rho + (1-\rho)p_0 & , j = 0 \\ (1-\rho)p_j & , j = 1, 2 \ldots \end{cases} \tag{3}$$

Then the pgf η is $P_\eta(t) = \rho + (1-\rho) P_\xi(t)$.

If $\rho = 1$, then the corresponding zero-Inflated distributions is degenerated at zero and $\rho = 0$, there is no inflation i.e., $P_\eta(t) = P_\xi(t)$.

Generally, the inflation parameter ρ may also take negative values, provided that $P(\eta = 0) \geq 0$, i.e. $\rho \geq \dfrac{-p_0}{1-p_0}$ and therefore $\max\left\{-1 , -\dfrac{1-\pi}{\pi}\right\} \leq \rho \leq 0$. This case corresponds to the opposite Phenomena i.e excludes a proportion of zeros from the basic discrete distribution if necessary.

In Table (2) below zero-inflated counterparts of known GPS (ZIGPS) distributions with their corresponding pmf and pgf are summarized.

Table (2)

	$P(\eta = 0)$	$P(\eta = 0) , \quad S - \{0\}$	pgf
$ZIBin(n, \pi, \rho)$	$\rho + (1-\rho)(1-\pi)^n$	$(1-\rho)\binom{n}{k}\pi^j(1-\pi)^{n-j}$	$\rho + (1-\rho)(1-\pi+\pi t)^n$
$ZIPo(\lambda, \rho)$	$\rho + (1-\rho)e^{-\lambda}$	$(1-\rho)\dfrac{\lambda^j}{j!}e^{-\lambda}$	$\rho + (1-\rho)e^{\lambda(t-1)}$
$ZINB(r, \pi, \rho)$	$\rho + (1-\rho)\pi$	$(1-\rho)\pi(1-\pi)^j$	$\rho + (1-\rho)\left(\dfrac{\pi}{1-(1-\pi)t}\right)^r$

$ZILS(\pi,\rho)$	ρ	$(1-\rho)\dfrac{(1-\pi)^j}{-j\ln\pi}$	$\rho+(1-\rho)\dfrac{\ln[1-(1-\pi)t]^{-1}}{-\ln\pi}$

4. Inflated-Parameter GPS distributions

Let $\{W_1, W_2, \ldots\}$ be a sequence of ZIBin $(1, \pi, \rho)$ with the inflated parameter $\max\left\{-1, -\frac{1-\pi}{\pi}\right\} \le \rho \le 1$. Also, let r.v. V be the number of trials that we need to achieve the first observed success in the sequence $\{\overline{W_1}, \overline{W_2}, \ldots\}$ of independent binary variables. The pmf of the r.v. V is given by

$$P(V = K) = [(1-\pi)(1-\rho)+\rho]^{k-1} \quad , \quad k = 1,2,3,\ldots$$

Now, define the r.v. S by the following relations

$$P(S = k) = \begin{cases} \pi & , \ k = 0 \\ (1-\pi)[(1-\pi)(1-\rho)+\rho]^{k-1}(1-\rho)\pi & , \ k = 1,2,3,\ldots \end{cases} \qquad (4)$$

Where $\pi \in (0,1)$ and $\rho \in \left(\max\left\{-1, -\frac{1-\pi}{\pi}\right\}, 1\right)$. Then S is called an Inflated-Parameter Geometric distributed r.v. and is denoted by $\sim IGe(\pi, \rho)$. The pgf of $S \sim IGe(\pi, \rho)$ is given by

$$P_S(t) = \frac{\pi(1-t\rho)}{1-t[1-\pi+\rho\pi]} \qquad (5)$$

Which is actually the pgf of $S = X_1 + X_2 + \cdots + X_N$ where $X_1, X_2, \ldots, X_N$ are $Ge(1-\rho)$ distributed i.i.d r.v..'s and are independent of $N \sim Ge(\pi)$. If N is a r.v. distributed according to the $Bin(n,\pi)$, $NB(r,\pi)$, $Po(\lambda)$ and $LS(\pi)$ distribution, then $S = X_1 + X_2 + \cdots + X_N$ is distributed from the $IBin(n,\pi,\rho)$, $INB(r,\pi,\rho)$, $IPo(\lambda,\rho)$ and $ILS(\pi,\rho)$, respectively, where $X_i's$ are $Ge(1-\rho)$ distributed i.i.d r.v.'s and are independent of N.

The r.v. S belongs to the family of Inflated-Parameter GPS distributions, if its pmf can be represented as

$$P(S = k) = \frac{1}{g(\theta)} \sum_{k_1, k_2, \ldots} c(k) [\theta(1-\rho)]^{k_1+k_2+\cdots} \rho^{k_2+2k_3+\cdots} \qquad (7)$$

With $k = 0,1,2,3,\ldots$ and $\rho \in [0,1)$, $\theta > 0$ and the summation is on the set of all nonnegative integers $k_1, k_2, \ldots$ such that $k_2 + 2k_3 + \cdots = k$. If $X \sim ILS(\pi, \rho)$, then its realizations begin from 1 and the summation is over the nonnegative integers such that $k_2 + 2k_3 + \cdots = k + 1$.

Also, the pgf of S has the following from

$$P_S(t) = \frac{g(\theta P_X(t))}{g(\theta)} \quad , \quad t \in [0,1] \tag{8}$$

Where $P_X(t) = \frac{t(1-\rho)}{1-t\rho}$, $\quad t \in [0,1]$ is the pgf $Ge(1-\rho)$.

The scond from of pmf , which can be found by Taylor expansion of the pgf , is given by

The corresponding θ , $g(\theta)$, $c(k)$, $a(k)$, pgf P_S, H , $P(S = 0)$, mean and variance for specified distributions are given in Table (3) .

The pmf of Inflate – arameter Power Series distributions satisfies the following recursion:

$$p_k = \left(a + \frac{b}{k}\right) p_{k-1} + \left(1 - \frac{2}{k}\right) c \, p_{k-2} \quad , \quad k = 1,2, \dots \; , p_{-1} = 0$$

the particular cases , the constants a , b and c are given by the following expressions :

$S \sim IBin(n, \pi, \rho)$: $\; a = 2\rho - \pi(1-\rho) \; , \; b = (n+1)\pi(1-\rho) - 2\rho \, , c$
$\qquad\qquad\qquad = -\rho\big(\rho - \pi(1-\rho)\big)$

$S \sim IPo(\lambda, \rho)$: $\; a = 2\rho \; , \; b = \lambda(1-\rho) - 2\rho \, , c = -\rho^2$

$S \sim INB(r, \pi, \rho)$: $\; a = (1-\rho)(1-\pi) + 2\rho \; , \; b = (r-1)(1-\rho)(1-\pi) - 2\rho \; ,$
$\qquad\qquad\qquad c = -\rho\left(1 - \pi(1-\rho)\right)$

5. Application

In this chapter, we approximate the classic poisson distribution and Inflated-Parameter poisson distribution to a real data group that is related to the number of accident Insurance claims and it has prepared on the basis of 16760 policies (number of policyholders) in Mazandaran province, we indicate that Inflated-Parameter poisson distribution is strongly appropriate. The Table of frequency data is following:

k	0	1	2	3	≥ 4
o_k	13772	2631	318	34	5

And then, $\bar{X} = 2.02 \quad , S^2 = 0.21502$.

5.1 Fit of Classic Poisson distribution

Assumed that p_k is corresponding probability of class k_{th} , and e_k is expected values of class k_{th} . In this case , with estimating parameter $\hat{\lambda} = \bar{X} = 2.02$, statistics goodness of fit χ^2 for poisson distribution is calculated on Table (3) .

Then , the assumption of data poisson in level 0.05 and even in significance level 0.005 is strongly rejected , because the observed value $\chi^2 = 40.4$ is bigger than critical point $\chi^2_{0.05,3} = 7.815$.

5.2 Fit of Inflated-Parameter poisson distributeon

According to the mean and variance distribution $IPo(\lambda, \rho)$ in Table (2), we obtain the following moment estimates for the parameters , ρ :

$$\hat{\rho} = \frac{S^2 - \bar{X}}{S^2 + \bar{X}} \quad \text{and} \quad \hat{\lambda} = \frac{2\bar{X}^2}{S^2 + \bar{X}}$$

Then , we have $\hat{\lambda} = 0.1959$,$\hat{\rho} = 0.031$, and therefore , the value calculation Table of goodness of fit statistic χ^2 , in this manner will obtained as Table (5) . Then , against before manner , since $\chi^2 = 1.407$, $\chi^2{}_{0.05,2} = 5.991$, Null hypothesis based on fit of Inflated-Parameter poisson distribution to data is strongly accepted.

We remind that here the positiveness of estimate $\hat{\rho} = 0.031$ is indicated that the observance of zero are more than this that it is forcasted by classic poisson distribution.

Table (3)

	$IBin(n,\pi,\rho)$	$IPo(\lambda,\rho)$	$INB(r,\pi,\rho)$	$ILS(\pi,\rho)$
θ	$\dfrac{\pi}{1-\pi}$	λ	$1-\pi$	$1-\pi$
$g(\theta)$	$(1+\theta)^n$	e^θ	$(1-\theta)^{-r}$	$-\ln(1-\theta)$
$c(k)$	$\begin{pmatrix} n \\ k_0{}^*, k_1, k_2, \ldots \end{pmatrix}$	$\dfrac{1}{k_1! k_2! \ldots}$	$\begin{pmatrix} k_1 + k_2 + \cdots + (r-1) \\ k_1, k_2, \ldots, (r+1) \end{pmatrix}$	$\dfrac{(-1 + k_1 + k_2 + \cdots)!}{k_1! k_2! \ldots}$
$a(i)$	$\binom{n}{i}$	$\dfrac{1}{i!}$	$\binom{n+i-1}{i}$	$\dfrac{1}{i}$
pgf	$\left[1 - \dfrac{\pi(1-t)}{1-t\rho}\right]^n$	$e^{\frac{\lambda(t-1)}{1-t\rho}}$	$\left[\dfrac{\pi(1-t\rho)}{1-t[1-\pi+\pi\rho]}\right]^r$	$\dfrac{1}{-\ln\pi} \ln\left[\dfrac{1-\rho t}{1-t(1-\pi+\pi\rho)}\right]$
H	$\min\{n,k\}$	k	k	k
$P(S=0)$	$(1-\pi)^n$	$e^{-\lambda}$	π^r	0
Mean	$\dfrac{n\pi}{1-\rho}$	$\dfrac{\lambda}{1-\rho}$	$\dfrac{r(1-\pi)}{\pi(1-\rho)}$	$\dfrac{-(1-\pi)}{\pi(1-\rho)\ln\pi}$
Variance	$\dfrac{n\pi(1-\pi+\rho)}{(1-\rho)^2}$	$\dfrac{\lambda(1+\rho)}{(1-\rho)^2}$	$\dfrac{r(1-\pi)(1+\pi\rho)}{\pi^2(1-\rho)^2}$	$\dfrac{-(1-\pi)[\ln\pi(1+\pi\rho) + 1 - \pi]}{\pi^2(1-\rho)^2(\ln\pi)^2}$

Table (4)

K	o_k	π_k	e_k	$\dfrac{(e_k - o_k)^2}{e_k}$
0	13772	0.81693	13691.77	0.47
1	2631	0/16518	2768.47	6.82
2	318	0.01671	279.99	5.16
3	34	0.00112	18.86	12.15
≥ 4	5	6.02×10^{-5}	1.01	15.76
sum	16760	1	16760	$\chi^2 = 40.4$

Table (5)

K	o_k	π_k	e_k	$\dfrac{(e_k - o_k)^2}{e_k}$
0	13772	0.822	13777.2	0.002
1	2631	0/1561	2616.35	0.080
2	318	0.0196	329.53	0.403
3	34	0.002	33.64	0.004
≥ 4	5	1.95×10^{-4}	3.268	0.918
sum	16760	1	16760	$\chi^2 = 1.407$

References

[1] Jonson, N. L., Kotz, S. and Kemp, A. W. "Univariate Discrete Distributions". Wiley Series in Probability and Mathematical Statistics.2 nd edition, (1992).

[2] Minkova, L.,"A Generalization of the Classical Discrete Distributions", Compt. Randue Bulg. Acad,Aci , (2002)54,2.

[3] Kolve, N, Minkova, L. and Neytchev, P., "Inflated – Parameter Family of Generalized Power Series Distributions and their applications in Analysis of Overdispesed Insurance Data". ARCH, Research Clearing House, 2, (2000).

 [4]Gapta, L.P. and Gupta, R. C, Inflatrd Modified Power Series Distributions with applications. Communications in Statistics;Theory and Methods, (1995)24(9),2355-2374

[5]Bowers, N. L, Gerber, H. U., Hickman, J. C., Jones, D. A. and Nesbitt, C. J. Actuarial Mathematics,2nd ed . Schaumburg, III.:Society of Actuaries(1997).

[6] Rolski, T. Schmidli, H., Schmidt, V. and Teugels, J. Stochastic Processes for Insurance and Finance, John Wiley&Sons, Chichester (1999).

[7]Winkelmann, R. Econometric Analysis of Count Data ,3rd Edition . Springer-Verlag: Berlin (2000).