Department of Statistics

The University of Akron, Department of Statistics
Technical Report Series

Year: 2013 *Paper: 13-01*

Parametric Inference for a Zero-Inflated Poisson Distribution and Its Variants

[1] Santanu Chakraborty
[3] Bashirul A. Polash
[2] Sujay Datta, Corresponding Author

[1] University of Texas Pan American, Edinburg
[2] Fred Hutchinson Cancer Research Center, Seattle
[3] Raytheon Network Centric System

Department of Statistics
and
Center for Statistical Consulting

Buchtel College of Arts and Sciences
Akron, OH 44325-1913
330-972-6886 • 330-972-2028 Fax • http://www.uakron.edu/stat/

Parametric Inference for a Zero-Inflated Poisson Distribution and Its Variants

Santanu Chakraborty[1], Sujay Datta[2*] and Bashirul A. Polash[3]

August 10, 2009

[1] University of Texas Pan American, Edinburg, [2] Fred Hutchinson Cancer Research Center, Seattle, [3] Raytheon Network Centric System, * Corresponding author

1 Introduction

Count data frequently arise in real life and are often modelled by a Poisson distribution. However, it is found sometimes that the number of zeroes in such a dataset is more than what is allowed by a Poisson model. So, in order to accommodate the excess of zeroes, a zero-inflated Poisson distribution (ZIP) was introduced in the literature. See, for example, Cohen (1960), Singh (1963), Katti (1965), Goraski (1977), Kemp (1986) and Heilborn (1989). Some interesting applications of ZIP can also be found in Heilborn and Gibson (1990), Lambert (1992), Gupta et al. (1996), Saei and McGilchrist (1997), Li et al. (1999) and Ghosh et al (1999). There have been attempts to generalize the Poisson distribution in the past, such as the one in Consul and Jain (1973), and these in turn have led to generalized versions of ZIP. Angers and Biswas (2003) presented a Bayesian analysis of such a generalized ZIP model. The Poisson distribution is well-known for the equality of its mean and variance—a feature not suitable for many count datasets whose variances exceed their means. This necessitated the introduction of an overdispersed Poisson (OP) model. See Cox (1983) and Scollnik (1995), among others. Overdispersed Poisson models have been widely used for modeling disease data and one can show that a ZIP model is a special example of OP. Shmueli et al. (2005) revived another variant of the Poisson distribution, called the Conway-Maxwell-Poisson (CMP) distribution, that had been introduced earlier by others and demonstrated its usefulness as a model for discrete data. Later, Kadane et al. (2006) provided a Bayesian analysis of a CMP model using conjugate priors. The CMP distribution can be easily extended to a zero-inflated version.

Here we discuss parametric inference results for a ZIP distribution and its variants mentioned above. We discuss point estimation, hypothesis testing and Bayesian inference. Much of the existing work cited above came as separate developments and are somewhat scattered in the literature. To our knowledge, there is no such unified presentation of inference methodology for this class of distributions. However, this is not meant to be a review article. Most of the results presented here are new (except some basic ones that are quoted from the references cited). Detailed proofs and derivations are provided wherever possible. We believe this article will serve as a starting point for future theoretical and applied research on this interesting family of distributions. The layout of the paper is as follows. In section 2, we provide some background on the ZIP family of distributions. In section 3, we derive the conditional and unconditional distributions of sums of ZIP variables. Section 4 discusses estimation and hypothesis testing for ZIP data. Section 5 provides some distributional derivations for the Conway-Maxwell-Poisson (CMP) distribution, following which, section 6 takes up the zero-inflated version of CMP. Section 7 is devoted to the zero-inflated generalized Poisson distribution and we conclude with a section (section 8) on Bayesian inference for the ZIP model. Some theoretical details are provided in Appendix 1, while Appendix 2 describes some simulation studies related to the methodology outlined in section 4.

2 Preliminaries

The basic idea behind a ZIP distribution with parameters ϕ and λ is to assign a higher probability to zero and lower probabilities to other values as compared to a corresponding Poisson distribution with parameter λ. Thus a ZIP distribution with two parameters is formally defined by:

$$P(X=0) = \phi + (1-\phi)\, e^{-\lambda} \text{ and}$$

$$P(X=x) = (1-\phi)\, e^{-\lambda}\frac{\lambda^x}{x!} \text{ for } x>0,$$

where ϕ is a number between 0 and 1 and $\lambda > 0$. One can also write this as:

$$P(X=x) = \phi\, I(x=0) + (1-\phi)\, e^{-\lambda}\frac{\lambda^x}{x!} \tag{1}$$

for any non-negative integer x, where $I(A)$ is the indicator function of the event A.

The above definition is the most standard way of defining a ZIP distribution. However, there are several other ways of defining it. All of them use two parameters and are based on simple-minded ways of thinking. We shall discuss a few of them here.

The first alternative definition is:

$$P(X=0) = e^{-\lambda} + \gamma \text{ and}$$

$$P(X=x) = \left(1 - \frac{\gamma}{1-e^{-\lambda}}\right) e^{-\lambda}\frac{\lambda^x}{x!} \text{ for } x>0$$

. So, once again, this ZIP distribution has two parameters γ and λ and it is clear that γ is nothing but $\phi(1-e^{-\lambda})$. This definition is guided by the following simple-minded intuition: increase the probability at 0 by a constant number γ. One has to keep in mind that γ should be less than $1 - e^{-\lambda}$.

A second way of defining the ZIP distribution is:

$$P(X=0) = (1+\delta)\, e^{-\lambda} \text{ and}$$

$$P(X=x) = \left(1+\delta - \frac{\delta}{1-e^{-\lambda}}\right) e^{-\lambda}\frac{\lambda^x}{x!} \text{ for } x > 0$$

This time, the two parameters are δ and λ and one can show that δ is nothing but $\frac{\phi(1-e^{-\lambda})}{e^{-\lambda}}$. This definition is guided by the following intuition: increase the probability at 0 by multiplying the original probability by a number bigger than 1. Of course, one has to keep in mind that δ should be less than $e^{\lambda} - 1$.

A third alternative definition of the same distribution is:

$$P(X=0) = 1 - e^{-\mu+\mu\rho} + e^{-\mu} \text{ and}$$

$$P(X=x) = e^{-\mu}\,\frac{(\mu\rho)^k}{k!} \text{ for } k = 1, 2, \ldots,$$

with $0 < r < 1$. Here the two parameters are r and λ. One can easily observe that the two parameters of the original ZIPD, namely, (ϕ, λ) can be recovered from (ρ, μ) as: $\lambda = \mu\rho$, $\phi = 1 - e^{-\mu+\mu\rho}$.

The mean and variance of ZIPD are given by

$$E(X) = (1-\phi)\lambda, \quad V(X) = (1-\phi)\lambda(1+\phi\lambda)$$

The moment generating function $M_X(t)$ and the characteristic function $\Phi_X(t)$ are given by

$$M_X(t) = \phi + (1-\phi)\, e^{-\lambda+\lambda e^t};\ \Phi_X(t) = \phi + (1-\phi)\, e^{-\lambda+\lambda e^{it}}$$

Now it should be clear why a ZIP distribution is an example of an overdispersed Poisson model. Next we look at a weighted version of a Poisson distribution. It is given by

$$P_w(Y^w = y) = \frac{w(y)P(Y=y)}{E[w(Y)]},$$

where Y is a usual Poisson random variable. So, if we want to consider the ZIP distribution as a weighted version of a Poisson distribution,then

$$\frac{w(y)}{E[w(Y)]} = 1 - \phi$$

for $y > 0$ and

$$\frac{w(0)}{E[w(Y)]} = 1 + \delta,$$

where $\delta = \frac{\phi(1-e^{-\lambda})}{e^{-\lambda}}$. As a result, denoting $E[w(Y)]$ by A, we have $w(0) = A(1+\delta)$ and $w(y) = A(1-\phi)$ for $y > 0$. One observes that A can take any positive value. So, it is good enough to choose $w(0) = 1+\delta$ and $w(y) = 1-\phi$ for $y > 0$ where $\delta = \frac{\phi(1-e^{-\lambda})}{e^{-\lambda}}$. Therefore, the ZIP distribution is nothing but a weighted overdispersed Poisson distribution with the weights $w(y)$ chosen as above for each $y \geq 0$.

3 Conditional and Unconditional Distributions of Sums of ZIP Variables

In order to discuss parametric inference for a ZIP distribution, one has to know the distribution of the convolution of i.i.d. ZIP variables. For this, we first observe that if $X_1, X_2, ..., X_n$ are independent ZIP variables with X_i having parameters (ϕ_i, λ_i) for $i = 1, 2, ..., n$ and if $Z = X_1 + \cdots + X_n$, then using induction on n,

$$P(Z=z) = \prod_{i=1}^{n} \phi_i I(z=0) + \sum_{r=1}^{n} \sum_{(i_1,...,i_r)} \prod_{l=1}^{r} (1-\alpha_{i_l}) \prod_{j=1, j\neq i_1,...,i_r}^{k} \alpha_j e^{-\sum \lambda_{i_l}} \frac{(\sum \lambda_{i_l})^z}{z!}.$$

In case the X_is are i.i.d. with common parameters (ϕ, λ), the convolution distribution reduces to

$$P(Z=z) = \phi^n I(z=0) + \sum_{r=1}^{n} \binom{n}{r} (1-\phi)^r \phi^{(n-r)} e^{-r\lambda} \frac{(r\lambda)^z}{z!}.$$

This can also be written as

$$P(Z=z) = \phi^n I(z=0) + \sum_{r=1}^{n} P(Y_n = r)\, P(Z_r = z), \qquad (2)$$

where Y_n is Binomial with parameters (n, ϕ) and Z_r is Poisson with parameter $r\lambda$. This can also be derived using a different method. One can start out by finding the conditional joint distribution of $X_1, X_2, ..., X_n$ given $n_0 = j$ and the marginal distribution of n_0, where n_0 is the number of zero values among the X_i's. It can be shown that the conditional joint distribution of $X_1, ..., X_n$ given $n_0 = j$ is

$$P(X_1 = x_1, ..., X_n = x_n | n_0 = j) = \binom{n}{j}^{-1} \left(\frac{e^{-\lambda}}{1-e^{-\lambda}}\right)^{n-j} \frac{\lambda^{\sum x_i}}{\prod_{i=1}^{n} x_i!}.$$

Now, for $j = 0$, one observes that the conditional joint distribution of $X_1, X_2, ..., X_n$ given $n_0 = 0$ is the same as the unconditional joint distribution of n i.i.d. truncated Poisson random variables $X_1^*, X_2^*, ..., X_n^*$ (truncated at zero), which is given by

$$P(X_1 = x_1, ..., X_n = x_n | n_0 = 0) = \prod_{i=1}^{n} \frac{e^{-\lambda}\lambda^{x_i}}{(1-e^{-\lambda})\, x_i!}.$$

Notice that this is free from ϕ. From this, we can derive the conditional pmf of $Z = \sum X_i$ given $n_0 = 0$ to be

$$P(\sum X_i = k | n_0 = 0) = \left(\frac{\xi_n(k)}{n^k}\right)(1-e^{-\lambda})^{-n} P(Z_n = k),$$

where $Z_n \sim$ Poisson$(n\lambda)$ and $\xi_n(k)$ is a known, parameter-free function. For $n = 2$, $\xi_2(k) = 2^k - 2$; for $n = 3$, $\xi_3(k) = 3^k - 3.2^k + 3$, and so forth. In general there is a recursive formula for $\xi_n(k)$ as follows:

$$\xi_n(k) = \sum_{l=n-1}^{k-1} \binom{k}{l} \xi_{n-1}(l).$$

Next, it can be shown that the conditional pmf of $Z = \sum_{i=1}^{n} X_i$ given $n_0 = k$ is

$$P(\sum_{i=1}^{n} X_i = a \mid n_0 = k) = \frac{\xi_{n-k}(a)}{(n-k)^a}(1-e^{-\lambda})^{-(n-k)} P(W = a),$$

where W has a usual Poisson distribution with parameter $(n-k)\lambda$. This is exactly the same as $P(\sum_{i=1}^{n-k} X_i = a \mid n_0 = 0)$. Finally, one easily observes that n_0 is Binomial with parameters n and $\phi + (1-\phi)e^{-\lambda}$ and multiplying this with the conditional pmf of $Z = \sum_{i=1}^{n} X_i$ given $n_0 = k$ yields the unconditional distribution of Z mentioned earlier.

We now move on to conditional distributions. We start with the conditional pmf of X_1 given $X_1 + X_2$. A little algebra will reveal that

$$P(X_1 = j | X_1 + X_2 = k) = \frac{a+b}{c+d},$$

where

$$c = 2\phi(1-\phi)e^{-\lambda}\frac{\lambda^k}{k!}, \quad d = (1-\phi)^2 e^{-2\lambda}\frac{(2\lambda)^k}{k!}$$

and either $a = 0$, $b = \binom{k}{j}\left(\frac{1}{2}\right)^k d$ or $a = \frac{c}{2}$, $b = \left(\frac{1}{2}\right)^k d$. In general, the conditional pmf of $\sum_{i=1}^{m} X_i$ given $\sum_{i=1}^{n} X_i$ (for $m \le n$) involves both the parameters and can not be expressed by such a simple formula. This is unlike the ordinary Poisson distribution where the conditional pmf of $\sum_{i=1}^{m} X_i$ given

$\sum_{i=1}^{n} X_i = k$ (for $m \leq n$) is Binomial $(k, \frac{m}{n})$. But, in the case of ZIP variables, if we consider the conditional pmf of $\sum_{i=1}^{m} X_i$ given $\sum_{i=1}^{n} X_i = k$ (for $m \leq n$) and $n_0 = k^*$, then it turns out to be parameter-free, although not binomial. For $k^* = 0$, it is actually symmetric. For example,

$$P(X_1 = 1|\, X_1 + X_2 = 2,\, n_0 = 0) = 1;$$

$$P(X_1 = 1|\, X_1 + X_2 = 3,\, n_0 = 0) = P(X_1 = 2|\, X_1 + X_2 = 3,\, n_0 = 0) = \frac{1}{2};$$

$$P(X_1 = 0|\, X_1+X_2+X_3 = 2,\, n_0 = 1) = \frac{1}{3}, P(X_1 = 1|\, X_1+X_2+X_3 = 2,\, n_0 = 1) = \frac{2}{3};$$

etc. The general formula for $P(X_1 = l|\, X_1 + X_2 = k,\, n_0 = 0)$ is given by

$$P(X_1 = l|\, X_1 + X_2 = k,\, n_0 = 0) = \binom{k}{l} \xi_2^{-1}(k)$$

Similarly,

$$P(X_1 = l|\, X_1 + X_2 + X_3 = k,\, n_0 = 0) = \sum_{m=1}^{k-l-1} \frac{k!}{l!m!(k-l-m)!} \xi_3^{-1}(k)$$

$$P(X_1 + X_2 = l|\, X_1 + X_2 + X_3 = k,\, n_0 = 0) = \sum_{m=1}^{l-1} \frac{k!}{m!(l-m)!(k-l)!} \xi_3^{-1}(k)$$

In general, for $m < n$,

$$P(\sum_{i=1}^{m} X_i = l|\, \sum_{i=1}^{n} X_i = k,\, n_0 = 0) = \sum_{l_1} \cdots \sum_{l_m} \sum_{j_1} \cdots \sum_{j_{n-m}} \frac{k!}{l_1! \cdots l_m! j_1! \cdots j_{n-m}!} \xi_n^{-1}(k)$$

where $\sum_{i=1}^{m} l_i = l$ and $\sum_{i=1}^{n-m} j_i = k - l$.

4 Parametric Inference for a ZIP Distribution

As mentioned earlier, the mean and the variance of a ZIP distribution are given by

$$E(X) = (1-\phi)\lambda, \quad V(X) = (1-\phi)\lambda(1+\phi\lambda)$$

So, if $X_1, X_2, ..., X_n$ is a random sample from a ZIP distribution with parameters (ϕ, λ), then $1 - \frac{\bar{X}}{\lambda}$ is an unbiased estimator for ϕ if λ is known, whereas $\frac{\bar{X}}{1-\phi}$ is an unbiased estimator of λ if ϕ is known. But, in general, both ϕ and λ will be unknown and there is no sample-based unbiased estimator for the parameter-vector (ϕ, λ). However, it is easy to obtain the method of moments (MOM) estimators which are

$$\hat{\lambda}_{MOM} = \frac{\sum X_i^2}{\sum X_i} - 1, \; \hat{\phi}_{MOM} = 1 - \frac{n\bar{X}^2}{\sum X_i^2 - \sum X_i}$$

The sample likelihood function is given by

$$L(\phi, \lambda | x_1, ..., x_n) = (\phi + (1-\phi)e^{-\lambda})^{n_0}((1-\phi)e^{-\lambda})^{n-n_0}\frac{\lambda^{\sum x_i}}{\prod x_i!}$$

where n_0 is once again the number of zeros in the sample. From this, it is easy to see that the maximum likelihood estimators for the parameters can be obtained by solving

$$\hat{\lambda}_{MLE} = \frac{\bar{X}}{1 - \hat{\phi}_{MLE}}, \; \hat{\phi}_{MLE} = \frac{(n_0/n) - e^{-\hat{\lambda}_{MLE}}}{1 - e^{-\hat{\lambda}_{MLE}}}$$

For this pair of equations, no analytic solution exists and it must be solved numerically, which involves the Lambert's W function (see, for example, Corless et al. (1993)). In Appendix 2, we report the results from simulation studies conducted to explore the bias and the variance of the estimators mentioned above as functions of the two parameters. From the likelihood function, it should be clear that the 2×1 vector $(n_0, \sum X_i)$ is jointly sufficient for the two parameters (ϕ, λ). In fact, it can be shown that they are minimally sufficient.

Now suppose that $X_1, \ldots, X_n$ are i.i.d. following a ZIP(ϕ_1, λ_1) distribution, $Y_1, \ldots, Y_n$ are i.i.d. following a ZIP(ϕ_2, λ_2) distribution and we want to test $H_0 : (\phi_1, \lambda_1) = (\phi_2, \lambda_2)$ versus $H_1 :\neq$. Let n_0 and m_0 be the counts of zero values in the two samples. Under H_0, $n_0 + m_0$ has a Binomial$(2n, \phi)$ pmf where ϕ is the common value of ϕ_1 and ϕ_2. If $n_0 + m_0$ is observed to be k and $\sum_{i=1}^{n}(X_i + Y_i)$ is observed to be k^*, we reject H_0 at the preassigned level α provided that the test-statistic $\sum_{i=1}^{n} X_i$ exceeds the $(1 - \frac{\alpha}{2})^{th}$ percentile of, or falls below the $\frac{\alpha}{2}^{th}$ percentile of the conditional pmf of $\sum_{i=1}^{n} X_i$ given $\sum_{i=1}^{n}(X_i + Y_i) = k^*$ and $n_0 + m_0 = k$. Recall from the previous section that this conditional distribution is parameter-free. In order to achieve an exact significance level of α, it may be necessary to randomize this test. The statistics $\sum_{i=1}^{n}(X_i + Y_i)$ and $n_0 + m_0$ together induce a partition of the sample space and here we are conditioning on those partition cells. If the conditional test is performed at level α, the unconditional significance level is α as well, since $P(\text{Type I error} \mid H_0)$ is equal to

$$\sum_{k=0}^{2n}\sum_{k^*=0}^{\infty} P(\text{Type I error} \mid H_0, n_0{+}m_0 = k, \sum_{i=1}^{n}(X_i{+}Y_i) = k^*).P(n_0{+}m_0 = k, \sum_{i=1}^{n}(X_i{+}Y_i) = k^*)$$

This test is, in some sense, a generalization of the one introduced by Przyborowski and Wilenski (1940) for the equality of two Poisson means.

5 The Conway-Maxwell-Poisson Distribution

The CMP distribution which is the abbreviated form of the distribution introduced by Conway and Maxwell (1962) also looks like a variant of the ZIP distribution under certain assumptions. Here the probability mass function is as follows:

$$P(X=x)=\frac{\lambda^x}{(x!)^\zeta\sum_{k=0}^{\infty}\frac{\lambda^k}{(k!)^\zeta}}$$

for $x = 0, 1, 2, ...$ and λ and ζ are positive. For $x = 0$,therefore, the probability is $\left(\sum_{k=0}^{\infty}\frac{\lambda^k}{(k!)^\zeta}\right)^{-1}$. One can observe that for $\zeta = 1$, this probability distribution is nothing but Poisson with parameter λ. For $\zeta > 1$, the probability at 0 is bigger than $e^{-\lambda}$, which may lead one to believe that the CMP distribution with $\zeta > 1$ is an example of a usual ZIP distribution. This, however, is not the case because for values of $x > 0$, the weight function associated with the pmf is dependent on x, unlike the constant weight function of a ZIP distribution. However, we still discuss it here as a generalization of the usual Poisson distribution, whose zero-inflated version will be introduced later. We begin our discussion with the distribution of the convolution of $X_1, X_2, ..., X_n$ which are i.i.d. CMP (λ, ζ). It is given by

$$P(\sum_{i=1}^{n}X_i=k)=\frac{\frac{\lambda^k}{(k!)^\zeta}}{\sum_0^\infty\frac{\lambda^j}{(j!)^\zeta}}\sum_{i_1}\cdots\sum_{i_n}\left(\frac{k!}{i_1!\cdots i_n!}\right)^\zeta$$

for $k = 0, 1, 2, ...$ etc. This can also be proved by induction on n. Calculation of the mean and the variance is complicated in this case and so is the derivation of the method-of-moments (MM) estimators. It may be worth pointing out that for $\zeta > 1$, the CMP distribution is underdispersed, as opposed to the overdispersed ZIP distribution.

Now let us look at the conditional distributions associated with the CMP . The simplest one, namely, $P(X_1 = l|\, X_1 + X_2 = k)$ is given by

$$P(X_1=l|\,X_1+X_2=k)=\frac{\binom{k}{l}^\zeta}{\sum_{i=0}^{k}\binom{k}{i}^\zeta}.$$

Going one step further, we get,

$$P(X_1=l|\,X_1+X_2+X_3=k)=\frac{\sum_{m=0}^{k-l}\left(\frac{k!}{l!m!(k-l-m)!}\right)^\zeta}{\sum_{i_1}\sum_{i_2}\sum_{i_3}\left(\frac{k!}{i_1!i_2!i_3!}\right)^\zeta},$$

where $i_1 + i_2 + i_3 = k$. In general, we have,

$$P(X_1 = l_1 | \sum_{i=1}^{n} X_i = k) = \frac{\sum_{l_2}\sum_{l_3}\cdots\sum_{l_n}\left(\frac{k!}{l_1!l_2!\cdots l_n!}\right)^{\zeta}}{\sum_{x_1}\sum_{x_2}\cdots\sum_{x_n}\left(\frac{k!}{x_1!\cdots x_n!}\right)^{\zeta}},$$

where $\sum_{i=2}^{n} l_i = k - l$ and $\sum_{i=1}^{n} x_i = k$.

Also, if we let $p_{m,n}(l,k) = P(\sum_{i=1}^{m} X_i = l | \sum_{i=1}^{n} X_i = k)$ for $m < n$, then,

$$p_{m,n}(l,k) = \frac{\begin{pmatrix} k \\ l \end{pmatrix}^{\zeta} \sum_{l_1}\cdots\sum_{l_m}\left(\frac{l!}{l_1!\cdots l_m!}\right)^{\zeta}\sum_{j_1}\cdots\sum_{j_{n-m}}\left(\frac{(k-l)!}{j_1!\cdots j_{n-m}!}\right)^{\zeta}}{\sum_{x_1}\cdots\sum_{x_n}\left(\frac{k!}{x_1!\cdots x_n!}\right)^{\zeta}}$$

for $\sum_{i=1}^{m} l_i = l$, $\sum_{i=1}^{n-m} j_i = k - l$, $\sum_{i=1}^{n} x_i = k$.

For a random sample of size n from a CMP population, the likelihood function is given by

$$L(\zeta, \lambda | x_1, \ldots, x_n) = \frac{\lambda^{\sum x_i}}{(\prod_{i=1}^{n} x_i!)^{\zeta}\Psi^n(\zeta,\lambda)},$$

where $\Psi(\zeta,\lambda) = \sum_{k=0}^{\infty}\frac{\lambda^k}{(k!)^{\zeta}}$. Then the vector $(\prod_{i=1}^{n} X_i!, \sum_{i=1}^{n} X_i)$ is jointly minimally sufficient for (ζ,λ). Now let us write $Y = \prod_{i=1}^{n} Y_i = \prod_{i=1}^{n} X_i!$. Then Y_i takes values $0!, 1!, 2!, \ldots$ etc., i.e., $1, 1, 2, 6, 24, \ldots$etc. As a result, if y is the factorial of a nonnegative integer x, then

$$P(Y_i = y) = P(X_i! = x!) = P(X_i = x) = \frac{\lambda^x}{(x!)^{\zeta}\sum_{k=0}^{\infty}\frac{\lambda^k}{(k!)^{\zeta}}}$$

So, we can calculate $P(Y = y)$ where y is the product of n uniquely determined factorials, say, the factorials of $x_1, \ldots, x_n$ (it is not necessary that all $x_1, \ldots, x_n$ are different). Let $y_i = x_i!$ for $i = 1, 2, \ldots, n$. Suppose that the only distinct numbers among $x_1, \ldots, x_n$ are $z_1, \ldots, z_k$. Let n_i be the number of times z_i is present for $i = 1, 2, \ldots, k$, so that $\sum n_i = n$. Then we have

$$P(Y = y) = \frac{n!}{n_1!n_2!\cdots n_k!}\prod_{j=1}^{k}\left(\frac{\lambda^{z_j}}{(z_j!)^{\zeta}\sum_{l=0}^{\infty}\frac{\lambda^l}{(l!)^{\zeta}}}\right)^{n_j}$$

which is the same as

$$P(Y = y) = \frac{n!}{n_1!n_2!\cdots n_k!}\frac{\lambda^{\sum x_i}}{(\prod_{i=1}^{n} x_i!)^{\zeta}\Psi^n(\zeta,\lambda)}$$

6 Zero-inflated Conway-Maxwell-Poisson distribution

The zero-inflated CMP distribution is defined as

$$P(X=0)=\phi+(1-\phi)\frac{1}{\Psi(\zeta,\lambda)}$$

$$\text{and } P(X=k)=(1-\phi)\frac{\frac{\lambda^k}{(k!)^\zeta}}{\Psi(\zeta,\lambda)}$$

for $0<\phi<1$ and $k=1,2,\cdots$, where $\Psi(\zeta,\lambda)$ is as in the previous section. Then the convolution distribution for two i.i.d. ZICMP random variables is given by

$$P(X_1+X_2=0)=\left(\phi+(1-\phi)\frac{1}{\Psi(\zeta,\lambda)}\right)^2$$

$$\text{and } P(X_1+X_2=k)=2\phi(1-\phi)\frac{\frac{\lambda^k}{k!}}{\Psi(\zeta,\lambda)}+(1-\phi)^2\frac{\frac{\lambda^k}{k!}}{(\Psi(\zeta,\lambda))^2}\sum_{l=0}^{k}\binom{k}{l}^\zeta.$$

In general, if we denote $G_{m,n}(k,\zeta)=\sum_{i_{m+1}}\cdots\sum_{i_n}\left(\frac{k!}{i_{m+1}!\cdots i_n!}\right)^\zeta$, then we have

$$P(\sum_{i=1}^n X_i=k)=\sum_{m=0}^{n}\binom{n}{m}\frac{\lambda^k}{k!}\left(\phi+(1-\phi)\frac{1}{\Psi(\zeta,\lambda)}\right)^m\left(\frac{1-\phi}{\Psi(\zeta,\lambda)}\right)^{n-m}G_{m,n}(k,\zeta).$$

If we condition on the event $n_0=0$ where n_0 is once again the number of zero-values, the conditional pmf of $\sum_{i=1}^n X_i$ has a somewhat simpler form. Of course, n_0 itself is binomial with probability of success $p=\phi+(1-\phi)\frac{1}{\Psi(\zeta,\lambda)}$.

The simplest case is

$$P(X_1+X_2=k|\,n_0=0)=\frac{\lambda^k}{(k!)^\zeta}\left(\frac{1}{\Psi(\zeta,\lambda)-1}\right)^2\sum_{l=1}^{k-1}\binom{k}{l}^\zeta.$$

In general,

$$P(\sum_{i=1}^n X_i=k|\,n_0=0)=\frac{\lambda^k}{(k!)^\zeta}\left(\frac{1}{\Psi(\zeta,\lambda)-1}\right)^n\sum_{i_1}\cdots\sum_{i_n}\left(\frac{k!}{i_1!\cdots i_n!}\right)^\zeta,$$

where each i_l is positive and $i_1+\cdots+i_n=k$. Also, as in Section 3, it can be shown that $P(\sum_{i=1}^n X_i=k\mid n_0=j)$ is the same as $P(\sum_{i=1}^{n-j} X_i=k\mid n_0=0)$ As for the conditional joint pmf of $X_1,\cdots,X_n$ given $n_0=0$, it is

$$P(X_1=x_1,\cdots,X_n=x_n|\,n_0=0)=\frac{\lambda^{\sum x_i}}{(\prod x_i!)^\zeta}\left(\frac{1}{\Psi(\zeta,\lambda)-1}\right)^n$$

and, as was the case in Section 3, it turns out to be the same as the unconditional joint pmf of n i.i.d. zero-truncated CMP random variables. Next, as in the case of a usual ZIP distribution, we derive the conditional pmf of X_1 given $\sum_{i=1}^{n} X_i$ which is

$$P(X_1 = l_1 | \sum_{i=1}^{n} X_i = k,\, n_0 = 0) = \frac{\sum_{l_2}\sum_{l_3}\cdots\sum_{l_n}\left(\frac{k!}{l_1!l_2!\cdots l_n!}\right)^{\zeta}}{\sum_{x_1}\sum_{x_2}\cdots\sum_{x_n}\left(\frac{k!}{x_1!\cdots x_n!}\right)^{\zeta}},$$

with $\sum_{i=2}^{n} l_i = k - l$ and $\sum_{i=1}^{n} x_i = k$.

Also, if we denote $p_{m,n}(l,k) = P(\sum_{i=1}^{m} X_i = l | \sum_{i=1}^{n} X_i = k,\, n_0 = 0)$ for $m < n$, then,

$$p_{m,n}(l,k) = \frac{\binom{k}{l}^{\zeta}\sum_{l_1}\cdots\sum_{l_m}\left(\frac{l!}{l_1!\cdots l_m!}\right)^{\zeta}\sum_{j_1}\cdots\sum_{j_{n-m}}\left(\frac{(k-l)!}{j_1!\cdots j_{n-m}!}\right)^{\zeta}}{\sum_{x_1}\cdots\sum_{x_n}\left(\frac{k!}{x_1!\cdots x_n!}\right)^{\zeta}}$$

for $\sum_{i=1}^{m} l_i = l$, $\sum_{i=1}^{n-m} j_i = k - l$, $\sum_{i=1}^{n} x_i = k$. These formulas are exactly same as that of usual COM-Poisson.

7 Generalized Zero-Inflated Poisson Distribution

Angers and Biswas (2003) intoduced the following version of a generalized zero-inflated Poisson (henceforth GZIP) distribution:

$$P(X = 0) = \phi + (1-\phi)e^{-\lambda}$$

$$\text{and } P(X = k) = (1-\phi)\,\frac{(1+\alpha k)^{k-1}}{k!}\,\frac{(\lambda e^{-\alpha\lambda})^k}{e^{\lambda}}$$

for $k = 1, 2, ...$ with the parameters satisfying the restrictions $(1-e^{-\lambda})^{-1} < \phi < 1$, $0 \leq \alpha < \lambda^{-1}$ and $\lambda > 0$. So, a new parameter α comes into the picture. It is easy to see that choosing $\alpha = 0$ brings us back to the original ZIP distribution. See the appendix for a proof of the fact that $\sum_{k=0}^{\infty} P(X = k) = 1$ and that the mean and the variance are given by

$$E(X) = \frac{\lambda(1-\phi)}{1-\alpha\lambda}$$

and

$$V(X) = \frac{\phi(1-\phi)\lambda^2}{(1-\alpha\lambda)^2} + \frac{(1-\phi)\lambda}{(1-\alpha\lambda)^3}$$

The sample likelihood function is given by

$$L(\alpha,\phi,\lambda|x_1,\ldots,x_n) = (\phi+(1-\phi)e^{-\lambda})^{n_0}[(1-\phi)e^{-\lambda}]^{n-n_0}(\lambda e^{-\alpha\lambda})^{\sum x_i}\prod_{i=1}^{n}\frac{(1+\alpha x_i)^{x_i-1}}{x!}$$

The 3×1 vector $(n_0,\ \sum_{i=1}^n X_i,\ \prod_{i=1}^n (1+\alpha x_i)^{x_i-1})$ is jointly minimally sufficient for the three parameters $(\alpha,\ \phi,\ \lambda)$.
Under this set up, the pmf of X_1+X_2 will be

$$P(X_1+X_2=k)=2\phi(1-\phi)P(U_1=k)+(1-\phi)^2P(U_2=k)\left[e^{\alpha\lambda k}\frac{1+\frac{k}{2}\alpha}{1+k\alpha}\right]^{k-1}$$

, where $U_j\sim GP(\alpha,\ j\lambda)$ for $j=1,2$ and $GP(a,b)$ is the pmf given by

$$P(U=0)=e^{-b};\ P(U=k)=\frac{(1+ak)^{k-1}}{k!}\frac{e^{-abk}b^k}{e^b}$$

for positive integer values of k.

An alternative expression for the pmf of X_1+X_2 is

$$P(X_1+X_2=k)=2\phi(1-\phi)P(V_1=k)+(1-\phi)^2P(V_1+V_2=k)$$

where $V_j\sim GP(\alpha,\lambda)$ for $j=1,2$. We can go two steps further and obtain the pmfs of $X_1+X_2+X_3$ and $X_1+X_2+X_3+X_4$ as follows:

$$P(X_1+X_2+X_3=k)=3P(X_1=0)P(X_2+X_3=k)$$
$$+(1-\phi)^3\sum_{k_1,k_2,k_3\neq 0}\prod_{i=1}^{3}\frac{(1+k_i\alpha)^{k_i-1}}{k_i!}e^{-\lambda(3+\alpha k)}\lambda^k,$$

where the sum is over all positive integers k_1,k_2,k_3 adding up to k. This simplifies to

$$P(X_1+X_2+X_3=k)=6\phi(1-\phi)[\phi+(1-\phi)e^{-\lambda}]P(V_1=k)+3\phi(1-\phi)^2P(V_1+V_2=k)$$
$$+(1-\phi)^3P(V_1+V_2+V_3=k).$$

Next,

$$P(X_1+X_2+X_3+X_4=k)=4P(X_1=0)P(X_2+X_3+X_4=k)$$
$$+(1-\phi)^4\sum_{k_1,k_2,k_3,k_4\neq 0}\prod_{i=1}^{4}\frac{(1+k_i\alpha)^{k_i-1}}{k_i!}e^{-\lambda(4+\alpha k)}\lambda^k,$$

where the sum is over all positive integers k_1,k_2,k_3,k_4 adding up to k. This simplifies to

$$P(X_1+X_2+X_3+X_4=k)=24\phi(1-\phi)[\phi+(1-\phi)e^{-\lambda}]^2P(V_1=k)$$
$$+12\phi(1-\phi)^2[\phi+(1-\phi)e^{-\lambda}]P(V_1+V_2=k)$$
$$+4\phi(1-\phi)^3P(V_1+V_2+V_3=k)+(1-\phi)^4P(V_1+V_2+V_3+V_4=k).$$

In general, by induction on n, we can prove that

$$P(X_1+\cdots+X_n=k)=\sum_{j=1}^{n-1}{}^{n}P_{n-j}\phi(1-\phi)^j[\phi+(1-\phi)e^{-\lambda}]^{n-j-1}P(V_1+\cdots+V_j=k)$$

$$+(1-\phi)^nP(V_1+\cdots+V_n=k).$$

Suppose we have proved this for n. Then, we can prove this for $n+1$ as follows:

$$P(X_1+\cdots+X_n+X_{n+1}=k)=(n+1)P(X_1=0)P(X_2+\cdots+X_{n+1}=k)$$

$$+\sum_{k_1,\cdots\cdots,k_{n+1}\neq 0}\prod_{i=1}^{n+1}\frac{(1+k_i\alpha)^{k_i-1}}{k_i!}(1-\phi)^{n+1}e^{-\lambda[(n+1)+k\alpha]}\lambda^k$$

$$=(n+1)[\phi+(1-\phi)e^{-\lambda}][\sum_{j=1}^{n-1}{}^{n}P_{n-j}\,\phi(1-\phi)^j[\phi+(1-\phi)e^{-\lambda}]^{n-j-1}P(V_1+\cdots+V_j=k)$$

$$+(1-\phi)^nP(V_1+\cdots+V_n=k)]+\sum_{k_1,\cdots\cdots,k_{n+1}\neq 0}\prod_{i=1}^{n+1}\frac{(1+k_i\alpha)^{k_i-1}}{k_i!}(1-\phi)^{n+1}e^{-\lambda[(n+1)+k\alpha]}\lambda^k$$

$$=\sum_{j=1}^{n-1}{}^{n+1}P_{n+1-j}\,\phi(1-\phi)^j[\phi+(1-\phi)e^{-\lambda}]^{n-j}P(V_1+\cdots+V_j=k)$$

$$+(n+1)\phi(1-\phi)^nP(V_1+\cdots+V_n=k)$$

$$+(n+1)(1-\phi)^{n+1}e^{-\lambda}P(V_1+\cdots+V_n=k)$$

$$+\sum_{k_1,\cdots\cdots,k_{n+1}\neq 0}\prod_{i=1}^{n+1}\frac{(1+k_i\alpha)^{k_i-1}}{k_i!}(1-\phi)^{n+1}e^{-\lambda[(n+1)+k\alpha]}\lambda^k$$

$$=\sum_{j=1}^{n}{}^{n+1}P_{n+1-j}\,\phi(1-\phi)^j[\phi+(1-\phi)e^{-\lambda}]^{n-j}P(V_1+\cdots+V_j=k)$$

$$+(1-\phi)^{n+1}P(V_1+\cdots+V_n+V_{n+1}=k)$$

8 Bayesian Inference for a ZIP Model

Bayesian analysis of an ordinary Poisson model uses a Gamma density as a conjugate prior on the mean parameter λ. Bayesian analysis of a CMP model has been discussed by Kadane et al. (2006) and that for a generalized Poisson model can be found in Angers and Biswas (2003). Ghosh et al. (2006) introduced a zero-inflated power series (ZIPS) model which includes the ZIP model as a special case and carried out a Bayesian analysis under a generalized linear model setup with covariates. They used a Beta prior on the zero-inflation parameter ϕ and an appropriate prior on λ that is conjugate for the power series model. Here we adopt a different approach. First we describe a Bayesian hierarchical model that leads naturally to a method of two-sample (or multi-sample) comparison. Then we present a conjugate Bayesian analysis for the sum of (conditionally) i.i.d. ZIP random variables.

Suppose we have independent samples from J ZIP populations, say, representing the counts of individuals who respond in a particular way to J different treatments administered independently K times. Let Y_{jk} be the k^{th} replicate count observed under the j^{th} treatment ($j = 1, \ldots, J$ and $k = 1, \ldots, K$). We assume that $Y_{jk} \sim ZIP(p, \lambda_{jk})$. In other words,

$$\mathrm{pr}(Y_{jk} = y) = \phi I(y = 0) + (1 - \phi) P(Y^*_{jk} = y) \tag{3}$$

for some $0 < \phi < 1$, where $Y^*_{jk} \sim \text{Poisson}(\lambda_{jk})$. Then we model $\log(\lambda_{jk})$ as

$$\log(\lambda_{jk}) = \beta_j + \epsilon_{jk} \ , \tag{4}$$

where ϵ_{jk} is the random residual component following $\text{Normal}(0, \sigma^2_\epsilon)$. The use of a residual component in the link-function specification is consistent with the belief that there may be unexplained sources of variation in the data, perhaps due to explanatory variables that were not recorded at first. This is particularly appropriate for Poisson data sets with over-dispersion. The use of residual effects within GLMs is discussed in Sun et al. (2000) and is a special case of the class of generalized linear mixed models (Zeger and Karim, 1991; Breslow and Clayton, 1993). Equation (3) boils down to

$$\log(\lambda_{jk}) \sim \text{Normal}(\beta_j, \sigma^2_\epsilon) \tag{5}$$

where β_j is the effect of the j^{th} treatment. We use conjugate priors in this hierarchical model and center the parameters for efficient MCMC sampling (Gelfand et al.,1995). Let $\mathcal{NIG}$ be the Normal-Inverse Gamma family of conjugate distributions in which, the mean follows a Normal distribution conditionally on the variance and the variance marginally follows an Inverse-Gamma distribution with the hyper-prior parameters u and v having the appropriate subscripts. In other words,

$$\begin{aligned} \theta, \sigma^2 &\sim \mathcal{NIG}(\theta_0, \sigma^2, u, v) \text{ implies that} \\ \theta | \sigma^2 &\sim \mathcal{N}(\theta_0, \sigma^2) \text{ and} \\ \sigma^2 &\sim \mathcal{IG}(u, v) \end{aligned}$$

With this notation in mind, this is how we specify our priors:

$$\begin{aligned} \beta_j, \sigma^2_\beta &\sim \mathcal{NIG}(\mu, \sigma^2_\beta, u_{\beta,\pi}, v_{\beta,\pi}) \\ \mu, \sigma^2_\mu &\sim \mathcal{NIG}(\mu_0, \sigma^2_\mu, u_{\mu,\pi}, v_{\mu,\pi}) \end{aligned}$$

However, the specification of the zero-inflation parameter makes the sampling from the (conditional) posterior distribution extremely difficult. Agarwal et al (2002) and Ghosh et al (2006) cleverly handle the problem by introducing a latent variable. In the present context, denoting the latent variable corresponding to Y_{jk} by Z_{jk}, the complete likelihood of the data is $L(y, z \mid \phi, \lambda) =$

$$\prod_j \prod_k \phi^{z_{jk}} \left\{ (1-\phi) \frac{e^{-\lambda_{jk}} \lambda_{jk}^{y_{jk}}}{y_{jk}!} \right\}^{1-z_{jk}} \tag{6}$$

or, equivalently, $L(y, z \mid \phi, \lambda) =$

$$\phi^{n_0} (1-\phi)^{n-n_0} \prod_{y_{jk}>0} \frac{e^{-\lambda_{jk}} \lambda_{jk}^{y_{jk}}}{y_{jk}!} \prod_{y_{jk}=0} (e^{-\lambda_{jk}})^{1-z_{jk}}, \tag{7}$$

where $n_0 = \sum_j \sum_k z_{jk}$ and $n = JK$.

We assume a Beta(a, b) prior on ϕ and elicit conjugate priors for all the variance parameters. In summary, our hierarchical model is given by:

$$\begin{aligned}
Y_{jk} &\sim ZIP(\phi, \lambda_{jk}) \\
\phi &\sim \text{Beta}(a, b) \\
\log(\lambda_{jk}) &\sim \text{Normal}(\beta_j, \sigma_\epsilon^2) \\
\sigma_\epsilon^2 &\sim \mathcal{IG}(u_{\epsilon,\pi}, v_{\epsilon,\pi}^2) \\
\beta_j, \sigma_\beta^2 &\sim \mathcal{NIG}(\mu, \sigma_\beta^2, u_{\beta,\pi}, v_{\beta,\pi}) \\
\mu, \sigma_\mu^2 &\sim \mathcal{NIG}(\mu_0, \sigma_\mu^2, u_{\mu,\pi}, v_{\mu,\pi})
\end{aligned}$$

Sampling from the posterior distributions can be performed using a block Gibbs sampler. All the conditional distributions except for those of the λ-values and ϕ have conjugate forms. Using latent variables has the advantage that sampling from the conditional distribution of the zero-inflation parameter reduces to sampling from its conjugate distribution. However, a Metropolis-Hastings step is needed for drawing the λ-values and a log-Normal proposal distribution will work. We would prefer using relatively flat priors for all the variance parameters.

Next we move onto a conjugate Bayesian analysis of $Z = \sum_{i=1}^n X_i$ where $X_1, \ldots, X_n$ are conditionally i.i.d. ZIP (λ, ϕ) variables given the parameters. One can assume that

(i) λ has an *a priori* $\Gamma(\mu, \alpha)$ density given by

$$f(\lambda) = \frac{\mu^\alpha}{\Gamma(\alpha)} \lambda^{\alpha-1} e^{-\mu\lambda}$$

for $\lambda > 0$ and a positive integer α;

(ii) ϕ has an *a priori* Beta(a, b) density for some $a > 0$ and $b > 0$ (one can choose $a = b = 1$ yielding a $U(0, 1)$ prior);

(iii) λ and ϕ are independent, in which case the joint prior density denoted by $h(\lambda, \phi)$ will be the product of the above two.

Then, if $\{X_1, X_2, \cdots, X_n\}$ is a sample of size n from the p.m.f. in (1), which implies that $Z = \sum_{i=1}^{n} X_i$, will have the p.m.f. $P(Z = z|\,\lambda, \phi)$ in (2), one does the usual posterior calculation and gets the following. The unconditional or marginal distribution of Z is

$$P(Z = z) = \frac{1}{n+1} I(z = 0) + \frac{1}{n+1} \sum_{r=1}^{n} P(N_r = z),$$

where N_r is negative binomial with parameters $\left(\alpha, \frac{\mu}{\mu+\alpha}\right)$. The posterior joint distribution of λ and ϕ is given by

$$h^*(\lambda, \phi|\,z) = \frac{\left[\phi^n I(z = 0) + \sum_{r=1}^{n} P(Y_n = r)\, P(Z_r = z)\right] \frac{\mu^\alpha}{\Gamma(\alpha)} \lambda^{\alpha-1} e^{-\mu\lambda}}{\frac{1}{n+1} I(z = 0) + \frac{1}{n+1} \sum_{r=1}^{n} P(N_r = z)}$$

APPENDIX

Appendix 1A: Probability mass function of the GZIP distribution

Consider the paper by Consul and Jain (1973). They introduced a generalized Poisson distribution as follows:

$$P(X = x|\,\lambda_1, \lambda_2) = \lambda_1(\lambda_1 + x\lambda_2)^{x-1} e^{-(\lambda_1 + x\lambda_2)}/x!$$

for $x = 0, 1, 2, \cdots$ so that

$$P(X = x|\,\lambda_1, \lambda_2) = 0$$

for $x \geq m$ if $\lambda_1 + m\lambda_2 \leq 0$. But in our case, it is good enough to consider $\lambda \geq 0$ so that $P(X = x|\,\lambda_1, \lambda_2)$ is never zero. One can observe that putting $\lambda_2 = 0$, we can get back the usual Poisson distribution.

Consul and Jain referred to Jensen (1902) for proving that $P(X = x|\,\lambda_1, \lambda_2) = 1$. They used one-dimensional Lagrange's formula to prove this:

$$\phi(z) = \phi(0) + \sum_{x=1}^{\infty} \frac{1}{x!} \left(\frac{d^{x-1}}{dz^{x-1}} [f(z)^x \phi'(z)] \right)_{z=0} \left(\frac{z}{f(z)} \right)^x \tag{8}$$

where $\phi(z) = e^{\lambda_1 z}$ and $f(z) = e^{\lambda_2 z}$. Now it can be shown that

$$\left(\frac{d^{x-1}}{dz^{x-1}} [f(z)^x \phi'(z)] \right)_{z=0} = \lambda_1(\lambda_1 + \lambda_2 x)^{x-1}$$

so that (8) becomes

$$e^{\lambda_1 z} = \sum_{x=0}^{\infty} \frac{1}{x!} \lambda_1(\lambda_1 + \lambda_2 x)^{x-1} \left(\frac{z}{f(z)} \right)^x \tag{9}$$

This is true for all z. So, substituting $z = 1$, we get

$$e^{\lambda_1} = \sum_{x=0}^{\infty} \frac{1}{x!} \lambda_1 (\lambda_1 + \lambda_2 x)^{x-1} e^{-\lambda_2 x}$$

which implies

$$1 = \sum_{x=0}^{\infty} \frac{1}{x!} \lambda_1 (\lambda_1 + \lambda_2 x)^{x-1} e^{-(\lambda_1 + \lambda_2 x)}$$

so that the right hand side is nothing but $\sum_{x=0}^{\infty} P(X = x | \lambda_1, \lambda_2)$. Now substituting $\lambda_1 = \lambda$ and $\frac{\lambda_2}{\lambda_1} = \alpha$ in our write up, we get

$$P(X = x | \lambda, \alpha) = \frac{\lambda^x}{x!}(1 + \alpha x)^{x-1} e^{-(1+\alpha x)\lambda}$$

which is nothing but $U \sim GP(\alpha, \lambda)$ in our notations. So, it is proved that

$$\sum_{i=0}^{\infty} P(X = x | \lambda, \alpha) = \sum_{i=0}^{\infty} \frac{\lambda^x}{x!}(1 + \alpha x)^{x-1} e^{-(1+\alpha x)\lambda} = 1$$

Therefore, if we go to zero-inflated generalized Poisson distribution where

$$P(X = 0) = \phi + (1 - \phi)e^{-\lambda};\ P(X = x) = (1 - \phi)[\frac{\lambda^x}{x!}(1 + \alpha x)^{x-1} e^{-(1+\alpha x)\lambda}]$$

for $x > 0$, then we get

$$\sum_{i=0}^{\infty} P(X = x) = \phi + (1 - \phi) \sum_{i=0}^{\infty} \frac{\lambda^x}{x!}(1 + \alpha x)^{x-1} e^{-(1+\alpha x)\lambda} = 1$$

Appendix 1B: Expectation of a GZIP distribution

Here we derive the mean of a zero inflated generalized Poisson distribution. For this, we go back to the paper by Consul and Jain (1973) and work with (9). Consul and Jain have their own way of dealing with this as they are allowing negative values of λ_2. In our case, we can simply differentiate (9) and arrive at the following:

$$\lambda_1 e^{\lambda_1 z} = \sum_{x=1}^{\infty} \frac{\lambda_1 (\lambda_1 + \lambda_2 x)^{x-1}}{(x-1)!} z^{x-1} e^{-\lambda_2 x z} (1 - \lambda_2 z) \tag{10}$$

which is same as

$$\frac{\lambda_1}{1 - \lambda_2 z} = \sum_{x=1}^{\infty} \frac{\lambda_1 (\lambda_1 + \lambda_2 x)^{x-1}}{(x-1)!} z^{x-1} e^{-(\lambda_1 + \lambda_2 x) z}$$

Now putting $z = 1$ once again, we get

$$\frac{\lambda_1}{1-\lambda_2} = \sum_{x=1}^{\infty} \frac{\lambda_1(\lambda_1+\lambda_2 x)^{x-1}}{(x-1)!} e^{-(\lambda_1+\lambda_2 x)}$$

One can observe that the right hand side is the expectation of X so that $E(X) = \frac{\lambda_1}{1-\lambda_2}$. Substituting $\lambda_2 = 0$, one can see that it is the mean of the usual Poisson distribution. Now, coming back to the generalized Poisson distribution $U \sim GP(\alpha, \lambda)$, we get, $E(X) = \sum_{x=1}^{\infty} \frac{\lambda^x (1+\alpha x)^{x-1}}{x!} e^{-(1+\alpha x)\lambda} = \frac{\lambda}{1-\alpha\lambda}$. For zero-inflated generalized Poisson distribution, the expectation is given by

$$E(X) = (1-\phi) \sum_{x=1}^{\infty} \frac{\lambda^x (1+\alpha x)^{x-1}}{x!} e^{-(1+\alpha x)\lambda} = \frac{\lambda(1-\phi)}{1-\alpha\lambda}$$

Substituting $\phi = 0$, we get the mean of the usual generalized Poisson distribution back.

Appendix 1C: Variance of a GZIP distribution

Our next goal is to get the variance of the zero-inflated generalized Poisson distribution. For this, we again go back to the equation (10). We differentiate it once more to get

$$\lambda_1^2 e^{\lambda_1 z} = \sum_{i=2}^{\infty} \frac{\lambda_1(\lambda_1+\lambda_2 x)^{x-1}}{(x-2)!} z^{x-2} e^{-\lambda_2 x z} (1-\lambda_2 z)^2 - \sum_{i=1}^{\infty} \frac{\lambda_1(\lambda_1+\lambda_2 x)^{x-1}}{(x-1)!} z^{x-1} e^{-\lambda_2 x z} \lambda_2 (2-\lambda_2 z)$$

which simplifies to

$$\lambda_1^2 = \sum_{i=2}^{\infty} \frac{\lambda_1(\lambda_1+\lambda_2 x)^{x-1}}{(x-2)!} z^{x-2} e^{-(\lambda_1+\lambda_2 x) z} (1-\lambda_2 z)^2 - \sum_{i=1}^{\infty} \frac{\lambda_1(\lambda_1+\lambda_2 x)^{x-1}}{(x-1)!} z^{x-1} e^{-(\lambda_1+\lambda_2 x) z} \lambda_2 (2-\lambda_2 z)$$

Substituting $z = 1$, we get

$$\lambda_1^2 = \sum_{i=2}^{\infty} \frac{\lambda_1(\lambda_1+\lambda_2 x)^{x-1}}{(x-2)!} e^{-(\lambda_1+\lambda_2 x)} (1-\lambda_2)^2 - \sum_{i=1}^{\infty} \frac{\lambda_1(\lambda_1+\lambda_2 x)^{x-1}}{(x-1)!} e^{-(\lambda_1+\lambda_2 x)} \lambda_2 (2-\lambda_2) \tag{11}$$

Now the second term without the negative sign is just $E(X)$ multiplied by $\lambda_2(2-\lambda_2)$ and hence (11) becomes

$$\lambda_1^2 = \sum_{i=2}^{\infty} \frac{\lambda_1(\lambda_1+\lambda_2 x)^{x-1}}{(x-2)!} e^{-(\lambda_1+\lambda_2 x)} (1-\lambda_2)^2 - \frac{\lambda_1 \lambda_2 (2-\lambda_2)}{1-\lambda_2}$$

From this, one shows that

$$\sum_{i=2}^{\infty} \frac{\lambda_1(\lambda_1+\lambda_2 x)^{x-1}}{(x-2)!} e^{-(\lambda_1+\lambda_2 x)} = \lambda_1 \left[\frac{\lambda_1}{(1-\lambda_2)^2} + \frac{\lambda_2(2-\lambda_2)}{(1-\lambda_2)^3} \right]$$

Now the left hand side is nothing but $E[X(X-1)]$ in Consul and Jain's set up. Adding $E(X)$ to it, we get

$$E(X^2) = \frac{\lambda_1^2}{(1-\lambda_2)^2} + \frac{\lambda_1}{(1-\lambda_1)^3}$$

Finally, subtracting $[E(X)]^2$ from it, we get

$$V(X) = \frac{\lambda_1}{(1-\lambda_2)^3}$$

Substituting $\lambda_2 = 0$, we get back the variance of the usual Poisson distribution. Now, considering the usual generalized Poisson distribution with $U \sim GP(\alpha, \lambda)$, we get

$$E[X(X-1)] = \lambda \left[\frac{\lambda}{(1-\alpha\lambda)^2} + \frac{\alpha\lambda(2-\alpha\lambda)}{(1-\alpha\lambda)^3}\right]$$

so that

$$E(X^2) = \frac{\lambda^2}{(1-\alpha\lambda)^2} + \frac{\lambda}{(1-\alpha\lambda)^3};\ V(X) = \frac{\lambda}{(1-\alpha\lambda)^3}$$

Next, considering the zero-inflated generalized Possion distribution, we get

$$E[X(X-1)] = (1-\phi)\lambda \left[\frac{\lambda}{(1-\alpha\lambda)^2} + \frac{\alpha\lambda(2-\alpha\lambda)}{(1-\alpha\lambda)^3}\right]$$

Adding $E(X) = \frac{(1-\phi)\lambda}{1-\alpha\lambda}$ to both sides, we get

$$E(X^2) = (1-\phi)\lambda \left[\frac{\lambda}{(1-\alpha\lambda)^2} + \frac{1}{(1-\alpha\lambda)^3}\right]$$

Finally, subtracting $(E(X))^2$ from both sides, we get

$$V(X) = \frac{\phi(1-\phi)\lambda^2}{(1-\alpha\lambda)^2} + \frac{(1-\phi)\lambda}{(1-\alpha\lambda)^3}$$

Substituting $\phi = 0$, we get back the variance of the usual generalized Poisson distribution back.

Appendix 2: Simulation results for biases and variances of estimators

For the maximum likelihood and method-of-moments estimators of λ and ϕ that we mentioned in Section 4, here we report the results of some simulation studies regarding their bias and variance. For a fixed value of $\phi \in (0,1)$, we generated $M = 1000$ random samples of size $n = 100$ from a ZIP(ϕ,λ) distribution with $\lambda = 2, 3, \ldots, 10$ and each time computed the bias in the method-of-moments estimator for λ. For each λ, we then averaged the 1000 bias values and plotted this average bias against λ. Also, for each λ, we computed the sample variance of the 1000 $\hat{\lambda}_{MOM}$ values and plotted it against λ. We repeated this exercise for 9 different values of ϕ and the resulting 9 'λ vs. average bias' graphs, represented by different colors, were superimposed. Similarly, the 9 'λ vs. var($\hat{\lambda}_{MOM}$)' graphs, represented by different colors, were superimposed.

Next we implemented the same procedure with the roles of λ and ϕ switched. In other words, we now fixed λ at an integer value between 2 and 10 and in each case, created a 'ϕ vs. average bias' graph and a 'ϕ vs. var($\hat{\phi}_{MOM}$)' graph based on $M = 1000$ random samples of size $n = 100$ drawn from a ZIP(ϕ,λ) distribution with $\phi = 0.01, 0.02, \ldots, 0.09$. Once again, each of the two resulting sets of 9 graphs (represented by different colors) were superimposed.

Subsequently, the same procedure was repeated for the MLEs of λ and ϕ, resulting in 4 sets of graphs which are λ vs. avg($\hat{\lambda}_{MLE} - \lambda$), λ vs. var($\hat{\lambda}_{MLE}$), ϕ vs. avg($\hat{\phi}_{MLE} - \phi$) and ϕ vs. var($\hat{\phi}_{MLE}$).

Finally, all of the above were carried out again with $M = 10000$. It is clear that the MLEs of both parameters are asymptotically unbiased. For $\hat{\lambda}_{MLE}$ the bias is quite close to zero even for moderate values of λ such as 6 or 7. In the case of $\hat{\phi}_{MLE}$, the bias is negligible irrespective of ϕ for λ upwards of 5. The variance of $\hat{\lambda}_{MLE}$, for the value-range of λ considered here, are below 0.4 except for some extreme values of ϕ. Also, it shows a slight increasing trend with λ. As for the variance of $\hat{\phi}_{MLE}$, the only one that stands out is the case $\lambda = 2$. Whether this aberrant behavior was caused by a computational error needs to be investigated. Moving on to the bias and the variance of $\hat{\lambda}_{MOM}$, the bias shows a lot more fluctuations in this case than for the MLE, but it is close to zero except perhaps some extremely small values of ϕ. The variance of $\hat{\lambda}_{MOM}$ once again shows a slight increasing trend with λ and is below 0.4 over this value-range of λ except for some extremely small values of ϕ. The bias in $\hat{\phi}_{MOM}$ also shows more fluctuations compared to that in the MLE and the variance of $\hat{\phi}_{MOM}$ shows a concave quadratic trend with the maximum occurring around the mid-range of ϕ (except for some small values of λ).

Acknowledgment

Research of the second author (Datta) was partially supported by the Department of Statistics, Texas A & M University, through a National Cancer Institute grant (CA90301).

References

[1] Angers, Jean-Francois, Biswas, Atanu, 2003: A Bayesian analysis of zero-inflated generalized Poisson model. *Computational Statistics & Data Analysis*, **42**, p.37-46.

[2] Castillo, J. del, Perez-Casany, Marta, 2005: Overdispersed and underdispersed Poisson generalizations. *Journal of Statistical Planning and Inference*, **134**, p.486-500.

[3] Cohen, A.C.,1960: Estimating the parameters of a modified Poisson distribution. *J. Amer. Statist. Assoc.*, **55**, p.139-143.

[4] Consul, P.C., Jain, G.C., 1973: A Generalization of the Poisson distribution. *Technometrics*, **15(4)**, p.791-799.

[5] Cox, D.R., 1983: Some remarks on overdispersion. *Biometrika*, **70(1)**, p.269-274.

[6] Dahiya, R.C., Gross, A.J., 1973: Estimating the zero class from a truncated Poisson sample. *J. Amer. Statist. Assoc.*, **68**, p.731-733.

[7] Ghosh, S.K., Mukhopadhyay, P., Lu, J.C., 2006: Bayesian analysis of zero-inflated regression models. *J. Statist. Plan. Inf.*, **136**, p.1360-1375.

[8] Gupta, P.L., Gupta, R.C., Tripathi, R.C., 1996: Analysis of zero-zdjusted count data. *Comput. Statist. Data Anal.*, **23**, p.531-547.

[9] Heilborn, D.C., 1989: Generalized linear models for altered zero probabilities and over dispersion in count data. *Technical Report*, Department of Epidemiology and Biostatistics, University of California, San Francisco.

[10] Heilborn, D.C., Gibson, D.R., 1990: Shared needle use and health beliefs concerning AIDS: regression modeling of zero-heavy count data. Poster session. *Proceedings of the Sixth International Conference on AIDS*, San Francisco, CA.

[11] Kadane, J.B., Shmueli, G., Minka, T.P., Borle, S. and Boatwright, P., 2006: Conjugate analysis of the Conway-Maxwell-Poisson distribution. *Bayesian Analysis*, **1(2)**, p.363-374.

[12] Kemp, A.W., 1986: Weighted discrepancies and maximum likelihood estimation for discrete distribution. *Commun. Statist. A - Theory and Methods*, **15**, p.783-803.

[13] Kokonendji, C.C., Mizere, D., Balakrishnan, N., 2008: Connections of the Poisson weight function to overdispersion and underdispersion. *J. Statist. PLan. Inf.*, **138(5)**, p.1287-1296.

[14] Lambert, D., 1992: Zero inflated Poisson regression with an application to defects in manufacturing. *Technometrics*, **34**, p.1-14.

[15] Li, C.S., Lu, J.C., Park, J., Kim, K.M., Brinkley, P.A., Peterson, J., 1999: A multivariate zero-inflated Poisson distribution and its inference. *Technometrics*, **41**, p.29-38.

[16] Martin, D.C., Katti, S.K., 1965: Fitting of some contagious distributions to some available data by the maximum likelihood method. *Biometrics*, **21**, p.34-48.

[17] Przyborowski, J., Wilenski, H., 1940: Homogeneity of results in testing samples from Poisson Series. *Biometrika*, **31**, p.313-323.

[18] Saei, A., McGilchrist, C.A., 1997: Random threshold models applied to zero class data. *Austral. J. Statist.*, **39**, p.5-16.

[19] Scollnik, D.P.M., 1995: Bayesian analysis of two overdispersed Poisson models. *Biometrics*, **51(33)**, p.1117-1126.

[20] 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. *J. Roy. Statist. Soc. Series C*, **54(1)**, p.127-142.

[21] Singh, S.N., 1963: A note on inflated Poisson distribution. *J. Indian. Statist. Assoc.*, **1**, p.140-144.