Petr Volf

Institute of Information Theory and Automation
Academy of Sciences of the Czech Republic
Pod vodárenskou věží 4, 182 08 Praha 8

e-mail: volf@utia.cas.cz

Model for Difference of Two Series of Poisson-like Count Data

Abstract

When the discrete count data are analyzed, we are tempted to use the Poisson model. Consequently, we are often facing the problems with insufficient flexibility of Poisson distribution. However, in many instances the variable of the interest is the difference of two count variables. For these cases, so called Skellam distribution is available, derived originally as the difference of two correlated Poissons. The model contains latent variables, which leads quite naturally to the use of Bayes approach and to data augmentation via the Markov Chain Monte Carlo generation. In our contribution we apply the approach to the numbers of serious road accidents. The Skellam distribution is used for the comparison of the situation before and after the introduction of a new "point" system (in 7/2006).

1 Problems with Fit of Poisson Model

The main problem with Poisson model is that it has only one parameter (λ), characterizing both the expectation and variance, so that the flexibility of the model is rather limited. A typical problem encountered here is the "overdispersion", i. e. the case that data shows larger variance than the mean. The case can be solved, explained, modeled, by several methods.

The most common way how to cope with the problem is to consider a random factor as a part of Poisson intensity. Namely, the intensity is now $\lambda = Z \cdot \lambda_0$, where Z is a positive random variable with $EZ = 1$, λ_0 is a baseline intensity. The variable Z is called the heterogeneity, frailty, and represents certain factors which cause the intensity variation and which we are not able to explain more precisely in the present stage of analysis. Thus, if a random variable X has Poisson $(Z \cdot \lambda_0)$ distribution, then $EX = \lambda_0$, $EX^2 = E_z(E(X^2|Z)) = E_z(\lambda_0^2 Z^2 + \lambda_0 Z) = \lambda_0^2 EZ^2 + \lambda_0$, hence

$$\operatorname{var} X = \lambda_0 + \lambda_0^2(EZ^2 - 1) = \lambda_0 + \lambda_0^2 \operatorname{var} Z = \lambda_0(\lambda_0 \operatorname{var} Z + 1).$$

It is seen that the variance is now larger than λ_0 (if var $Z > 0$) and can be adapted to each variation of data. It is also well known that the case with gamma distributed Z leads actually to negative binomial distribution of X.

In frequent cases the data corresponds roughly to the Poisson distribution except at one or several values. In such cases the improvement utilizes the idea of the mixture of distributions,

the Poisson distribution is "inflated" by another discrete one. Then, the data are expected to follow the model

$$P(x) = \pi\, P_0(x) + (1-\pi)\, P_1(x),$$

where $0 \le \pi \le 1$, P_0 is the Poisson probability and P_1 gives more probability to certain points. Thus,

$$\begin{aligned} EX &= \pi\, EX_0 + (1-\pi)\, EX_1, \\ EX^2 &= \pi\, EX_0^2 + (1-\pi)\, EX_1^2, \end{aligned}$$

so that $\operatorname{var} X$ could be either larger or even smaller than EX. For instance, let $P_0 \sim$ Poisson(10), P_1 be concentrated at 10: $P_1(X = 10) = 1$, $\pi = 0,8$. Then $P(X = x) = e^{-10}\, 10^x/(x!) \cdot 0,8$ for $x \neq 10$ and $P(X = 10) = e^{-10}\, 10^{10}/(10!) \cdot 0,8{+}0,2$. EX remains 10, but the variance is $10 \cdot 0,8 = 8$. The same could be expressed by the model $P(x) = C(x) \cdot P_0(x)$, where $C(x)$ is constant at the most of points and supports or weaken the probability of other points.

Convolution of Poisson distribution with another one, which in general leads to the notion of compound point processes or marked point processes. Let us consider here a special case when each event with occurrence given by Poisson(λ) causes (is composed from) a set of other events, their number given (independently) by Poisson(μ). An example – a collision of a particle (event 1) gives a rise to several photons (event 2). The number of photons, X, is then given by $\sum_{i=0}^{Y} Z_i$, where $Y \sim$ Poisson(λ) and $Z_i \sim$ Poisson(μ). Hence, $EX = \lambda \cdot \mu$, while $\operatorname{var} X = \lambda(\mu^2 + \mu) = \lambda\mu(\mu + 1) > EX$.

Quite naturally, in highly heterogeneous cases, i. e. non-proportional, with different development in different groups, the first task is to separate these groups, and after the separation to fit group-specific models. This step of separation can be done for instance by the model-based clustering technique.

Except of overdispersion, the opposite phenomenon of "underdispersion" can sometimes be encountered, too. It can be caused e. g. by ties (dependencies) in data, also the case when extreme values are not observed may lead to the underdispersion effect.

2 Distribution of Difference of Two Poisson Variables

In many cases of the (discrete time) count data analysis we are interested in the difference of two count variables. Examples are frequent, in medicine, economy, in sport statistics (difference of score of a match). For these cases, so called **Skellam distribution** is available, derived originally as the difference of two independent Poissons (Skellam, 1946).

Let us consider two independent Poisson random variables U, V with parameters λ_1, λ_2, respectively. Then the distribution of the difference $Z = V - U$ is the following:

$$P(Z = z) \;=\; e^{-(\lambda_1+\lambda_2)} \left(\frac{\lambda_2}{\lambda_1}\right)^{z/2} B_{|z|}(2\sqrt{\lambda_1\lambda_2}), \tag{1}$$

where B is the modified Bessel function of the first kind, namely

$$B_r(x) = \left(\frac{x}{2}\right)^r \sum_{k=0}^{\infty} \frac{(x^2/4)^k}{k!\Gamma(k+r+1)}\,.$$

It is also seen that the same distribution is obtained when $Z = Y - X$, where $X = U + W$, $Y = V + W$, U, V are as before and W is an arbitrary discrete distribution, with finite moments, say, in order to overcome formal problems with the moments existence.

Thus, the model offers one of possibilities how to increase the flexibility of Poisson models. If X, Y, Z are observed, then U, V are latent independent Poisson variables, W characterizes the common part of X and Y, so that the distributions of them is rather arbitrary.

In other words, though we observe Z as $Y - X$, we can treat it as $V - U$, of course only in the case that the Skellam distribution fits to observations of Z. This can be tested by the standard chi-square goodness-of-fit test.

3 Estimation of Parameters

Let the random sample $\{Z_i\}, i = 1, ..., n$ be available, the aim is to find corresponding Poisson random variables U, V such that $Z = V - U$, i.e. their parameters or their representation. In such a simple case (of i.i.d. variables), the easiest method just compares the sample moments.

Namely, let $\bar{Z}, s_z^2$ be the sample mean and variance from the data, then, as $\mathrm{E}Z = \lambda_2 - \lambda_1$, $\mathrm{var}(Z) = \lambda_2 + \lambda_1$, a natural estimates are

$$\lambda_1 = \frac{s_z^2 - \bar{Z}}{2},\ \lambda_2 = \frac{s_z^2 + \bar{Z}}{2}$$

(notice that with positive probability such estimates can be negative).

3.1 MCMC Computation

In many applications, however, we deal with the time sequence of Skellam-distributed variables, eventually dependent also on other factors, so that the parameters λ_1, λ_2 follow a regression model and can also develop in time. For such cases we have to find another way of identification of variables U, V. Their nature of latent variables leads to the use of Bayes inference and to the method of data augmentation (i.e. artificial generation of representation of U and V), via the Markov chain Monte Carlo (MCMC) procedure. The main advantage is that while the maximum likelihood estimation (MLE) in the distribution (1) is analytically hardly tractable, the conditional likelihood containing U and V, given Z, is quite simple. Namely, let $z_i = v_i - u_i, i = 1, ..., n$ be random sample of observed variables, u_i, v_i latent Poisson variables, then we have

$$f(u, v|z, \lambda_1, \lambda_2) = \prod_{i=1}^{n} e^{-(\lambda_1+\lambda_2)} \frac{\lambda_1^{u_i} \lambda_2^{v_i}}{u_i! v_i!} \cdot I[z_i = v_i - u_i], \tag{2}$$

where z, u, v denote corresponding vectors $n \times 1$, $I[.]$ is an indicator function. Then the posterior distribution of parameters $\lambda_k, k = 1, 2$ is proportional to (2) times the prior of parameters and the scheme for the Bayes analysis is completed. A natural choice of conjugate priors for λ_k are independent gamma distributions, then the MCMC updating of them uses the Gibbs sampler.

3.2 Updating the Values of U and V

Let us recall here also the Metropolis step used for instance by Karlis and Ntzoufras (2006) for updating the values of latent variables:

Let u_i, v_i be actual values. Then

– if $z_i < 0$, propose v_i^* from Poisson(λ_2), set $u_i^* = v_i^* - z_i$ and accept it with
$p = min\{1, \lambda_1^{(v_i^* - v_i)} \frac{u_i!}{u_i^*!}\}$,

– if $z_i > 0$, propose u_i^* from Poisson(λ_1), set $v_i^* = u_i^* + z_i$ and accept it with
$p = min\{1, \lambda_2^{(u_i^* - u_i)} \frac{v_i!}{v_i^*!}\}$.

3.3 Modified EM Algorithm

EM algorithm is a standard method used in cases of missing (i.e. also latent) data. In our case, the M step (MLE of parameters provided latent data are available) is straightforward:

$$\lambda_1 = \bar{U},\ \lambda_2 = \bar{V}.$$

However, the E step computing the expectation of latent variables, given observed data and parameters, from distribution (2), is rather difficult. Therefore, we consider the variant combining the M-step (as above) with Monte Carlo updating of latent values following the method described in the preceding part.

In the sequel we shall employ either the MCMC procedure or also the variant using the direct computation of parameters via the M-step. In both cases the result consists of the samples of generated values of U and V (representing the distributions of latent variables) and samples of model parameters representing their posterior distributions. The cases with time or regressor-dependent variables and parameters can be incorporated quite easily.

3.4 Prediction of Future Values

Once the model is evaluated, we can use it for the prediction of new data (i.e. under non-changed conditions). In the Bayes scheme it means to construct the predictive distribution

$$p(x_{new}|\boldsymbol{x}) = \int_\Theta p_m(x|\theta) g_a(\theta|\boldsymbol{x}) d\theta,$$

where p_m describes the model, θ its parameters and g_a the aposterior distribution of them, which is based on observed data $\boldsymbol{x}$. After MCMC solution, instead of g_a the sample $\{\theta^{(i)}\}$ representing it is available. Hence, instead of integration, the averaging is used. In some cases the average of $p_m(x|\theta^{(i)})$ can be obtained directly, in a tractable form, for each possible value x. However, the sampling approach is preferred usually. Namely, the sample representing the predictive distribution is obtained in such a way that from each $p_m(x|\theta^{(i)})$ one value (or fixed number of values) is generated.

4 Artificial Example

First, let us demonstrate the procedure of solution on simple artificially generated data. We generated Poisson data U, V, n=100, with parameters $\lambda_1 = 5, \lambda_2 = 10$, and set $Z = V - U$. Estimated mean and variance of Z were: 4.7600, 15.1027, hence the moment method estimates of λ_1, λ_2 were 5.1714, 9.9314,

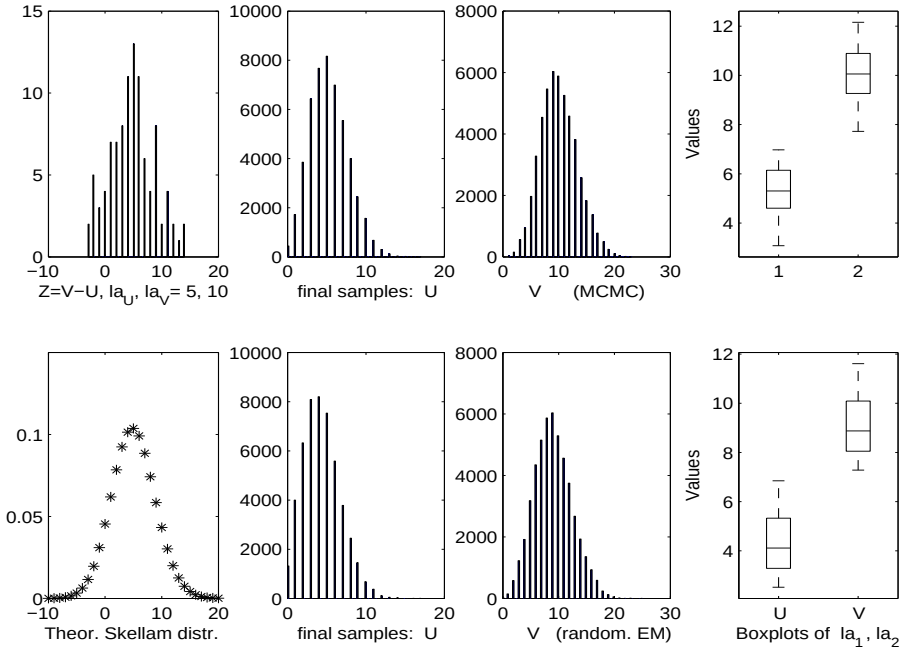

Figure 1: Results of artificial example: samples of latent variables, boxplots of posterior samples of parameters λ_1, λ_2

Then, the MCMC method was used. We selected the same (and rather wide) Gamma priors for both λ-s with parameters $a_0 = 1, b_0 = 10$, i.e. with E=10, var=100. The generation started from randomly uniformly selected integers between 1 and 20 for u and v. 1000 sweeps (iterations of the procedure) were performed, results computed from last 500 of them are displayed in Figure 1. The samples had the following characteristics:

(mean U, var U)= (5.2845, 5.9704), (mean V, var V)= (10.0445, 10.9520).

A variant with direct estimation of parameters (randomized EM algorithm) yielded the following:

(mean U, var U)= (4.3520, 5.6449), (mean V, var V)= (9.1120, 11.4121).

Some other features of solution were noticed, for instance: 1000 iterations quite sufficed – increase of number of iterations did not change results significantly. Increase of n led to narrower posterior distributions of parameters λ, simultaneously they were closer to real values (the consistency).

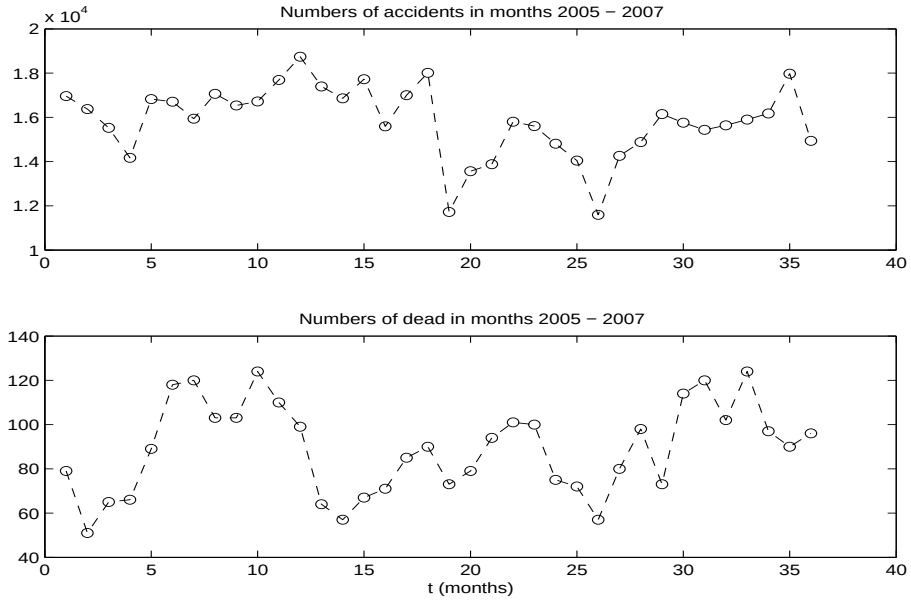

Figure 2: Numbers of accidents (above) and deaths (below), in months 2005–2007

5 Real Data Example

Figure 2 shows the development of monthly numbers of (reported) car accidents in Czech Republic in years 2005–2007, in the upper subplot, and the numbers of their fatal consequences (dead people), below. A new point system together with significant increase of fines and other sanctions was introduced from July, 2006 (here month 19). It is seen that there was certain decline, especially in the numbers of accidents, but the values reached the former level very

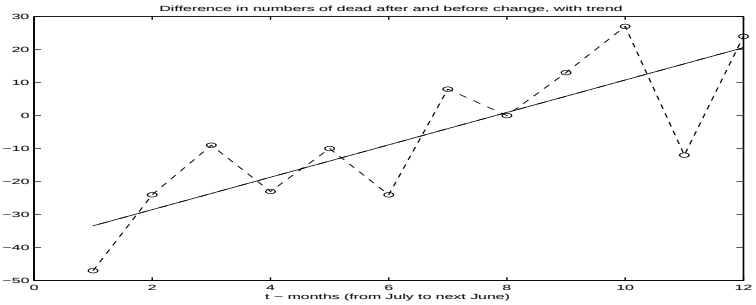

Figure 3: Differences of numbers of deaths, from July 2006–July 2005 to June 2007–June 2006, and fitted linear trend

quickly. The system had just temporal effect and now the situation, especially regarding the serious consequences of accidents, is worse than before.

Here, we analyze the data displayed in Figure 3, namely the differences in numbers of dead, after and before the change of punishment system, in corresponding months. Namely, July 06 – July 05, August 06 – August 05, ... , till June 07 – June 06, so that graph contains 12 such differences. Their linear trend is evident, its estimate (described further) is displayed, too. As such data forms a rather short time series, with no additional factors, we consider the following simple model: Observed data, $Z(t), t = 1, .., 12$, are Skellam variables. It means that they may be expressed as differences $Z(t) = V(t) - U(t)$, where $V(t), U(t)$ are Poisson. Further, we assume that they both have parameters with a linear trend, namely $\lambda_u(t) = a_u + b_u t$, $\lambda_v(t) = a_v + b_v t$, Then, the task is to estimate those four trend parameters. It also means that the mean and variance of $Z(t)$ develop linearly, too, with parameters given by the difference and sum, respectively, of parameters of $V(t)$ and $U(t)$.

As regards the computations, we preferred the randomized EM procedure. We started from $U(t), V(t)$ equal to observed numbers of dead in corresponding months. They evidently have a more complex structure, for instance they contain a seasonal component. It is assumed that this is common for both, and is substracted away. Then, parameters of linear trend, in a framework of Poisson model, were fitted. It means to find maximum likelihood estimates of a and b, i.e. the maximizer of log-likelihood, which is (for variables u, for instance):

$$L = \sum_{t=1}^{12} \left[-a - bt + u(t) \cdot log(a + bt) \right].$$

It was solved with the aid of several iterations of Newton–Raphson type. Then, from estimated parameters, new values of $u(t), v(t)$ were generated by the step described in part 3.2. Such a loop (one sweep) was repeated 1000 times The results displayed further were obtained from last 500 sweeps. Figure 4 shows histograms of samples of all four linear trend parameters. The means of these samples were (in order $\hat{a_u}, \hat{b_u}, \hat{a_v}, \hat{b_v}$) : $105.0160, 2.3203, 66.6882, 7.2287$.

Then the means of trend parameters for variables $Z(t)$ are the differences, namely $\hat{a_z} = -38.3278, \hat{b_z} = 4.9084$, with Bayes credibility intervals, given by 2,5% and 97,5% quantile of

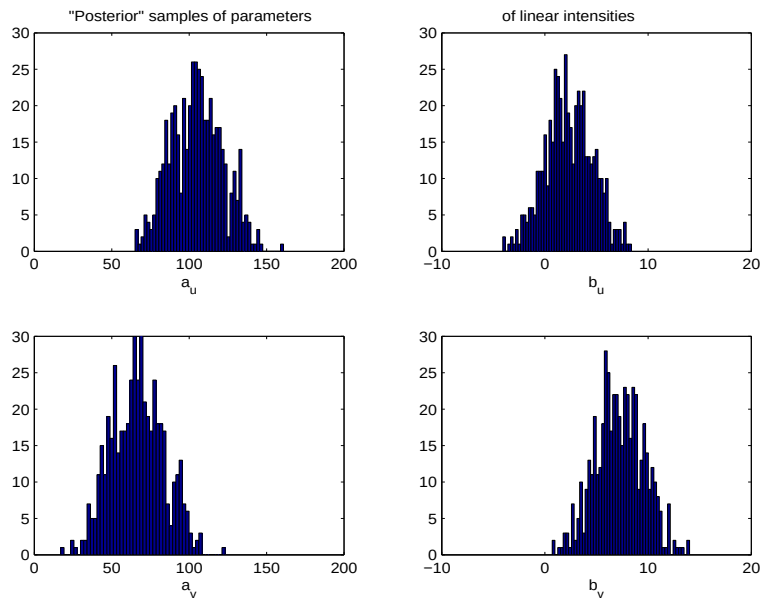

Figure 4: Histograms of approximate posterior samples of trend parameters

posterior sample: (-39.8290, -36.6471) for a_z and (4.6508, 5.1404) for b_z. Hence, Figure 3 contains the trend line with parameters $\hat{a_z}, \hat{b_z}$.

6 Conclusion

We have introduced the distribution for difference of two independent Poisson variables and presented corresponding methods of statistical analysis. What remains to be done is the testing the model fit, at least by the test whether resulting samples of latent components of U, V correspond to Poisson distribution.

Acknowledgement:
The research was supported by the grant of GACR No 402/07/1113.

References

[1] Karlis, D., Ntzoufras, I.: Bayesian analysis of the differences of count data. Statistics in Medicine 25, 1885-1905 (2006).

[2] Skellam, J.G.: The frequency distribution of the difference between two Poisson variates belonging to different populations. Journal of the Royal Statistical Society, Series A, 109, p.296 (1946).