



www.sce.unimore.it  
Scienze della Comunicazione  
e dell'Economia

# A Bayesian test for Poissonness

La Rocca, L.\* Consonni, G.\*\* Dossou-Gbete, S.\*\*\*

\* University of Modena and Reggio Emilia, Italy

\*\* University of Pavia, Italy

\*\*\* University of Pau and Pays de l'Adour, France



UNIVERSITÀ DEGLI STUDI  
DI MODENA E REGGIO EMILIA

## Introduction

The Poisson family constitutes a benchmark model for count data, which can be interpreted in terms of **rare events**.

We give a Bayesian procedure to test the adequacy of this model, against a recently revived alternative model.

## The CMP family

The Conway-Maxwell-Poisson family is a two-parameter model defined by

$$\frac{\Pr(Y = y | \lambda, \nu)}{\Pr(Y = y - 1 | \lambda, \nu)} = \frac{\lambda}{y^\nu},$$

$y = 1, 2, \dots$  ( $\lambda > 0, \nu > 0$ ); it bridges (Shmueli *et al.*, 2005) three well-known sampling models:

- geometric ( $\nu \downarrow 0, \lambda < 1$ );
- Poisson ( $\nu = 1$ );
- Bernoulli ( $\nu \uparrow \infty$ ).

Both **overdispersion** ( $\nu < 1$ ) and **underdispersion** ( $\nu > 1$ ) can be modelled.

## Conjugate analysis

The CMP likelihood, having observed the counts  $y_1, \dots, y_n$ , can be written as

$$L(\lambda, \nu | s_1, s_2) = \frac{\lambda^{s_1} e^{-\nu s_2}}{Z(\lambda, \nu)^n},$$

where  $s_1 = \sum_{i=1}^n y_i$ ,  $s_2 = \sum_{i=1}^n \log(y_i!)$ , and

$$Z(\lambda, \nu) = \sum_{j=0}^{\infty} \frac{\lambda^j}{(j!)^\nu}.$$

Hence, the CMP model constitutes an **exponential family** and is amenable to conjugate analysis (Kadane *et al.*, 2006).

## Bayes factor

We obtain a BF against Poissonness as

$$\text{BF}_{H, H_0}(s_1, s_2) = \frac{m_H(s_1, s_2)}{m_{H_0}(s_1)},$$

where  $m_H$  ( $m_{H_0}$ ) is the **marginal likelihood** under the CMP (Poisson) model; notice that  $H_0$  is nested in  $H$ .

## Prior choice

Under  $H$  we let  $a$ ,  $b$  and  $c$  in

$$p_H(\lambda, \nu) \propto \frac{\lambda^a e^{-\nu b}}{Z(\lambda, \nu)^c}$$

be the sufficient statistics and sample size of a **training sample**  $x_1, \dots, x_c$ .

For a fair comparison of  $H$  against  $H_0$  we centre  $p_H(\lambda, \nu)$  on  $H_0$  by using a “perfectly Poissonian” training sample:

$$\#\{k : x_k = x\} \simeq m \frac{\hat{\lambda}^x}{x!} e^{-\hat{\lambda}},$$

$x = 0, 1, 2, \dots$ , where  $m$  is a tentative value for  $c$  and  $\hat{\lambda} = \frac{s_1}{n}$  (data mean).

Under  $H_0$  we let  $p_{H_0}(\lambda) = p_H(\lambda | \nu = 1)$ , and get a gamma density with shape  $a$  and rate  $c$  (conjugate analysis).

## Computation

While  $m_{H_0}(s_1)$  is available in closed form, non-trivial **numerical evaluation** of

$$k(a, b, c)^{-1} = \int_0^\infty \int_0^\infty \frac{\lambda^a e^{-\nu b}}{Z(\lambda, \nu)^c} d\lambda d\nu$$

is needed to compute  $m_H(s_1, s_2) = k(a, b, c) / k(a + s_1, b + s_2, c + n)$ .

## An example

Letting  $\Pr(H_0) = \frac{1}{2}$ , we study the **sensitivity** of  $\Pr(H_0 | s_1, s_2)$  to  $c$ , as a fraction of  $n$ , for the chromosome interchange data reported by Kadane *et al.* (2006):  $n = 2566$ ,  $s_1 = 303$ ,  $s_2 = 10.4$ .

$c/n$	$\Pr(H_0   s_1, s_2)$	$c/n$	$\Pr(H_0   s_1, s_2)$
0.03	0.74	0.61	0.57
0.23	0.64	0.81	0.59
0.42	0.59	1.00	0.57

## Future plans

We plan to carry out a more structured objective Bayesian analysis using **intrinsic prior** methodology; see Pericchi (2005) for a review, and Consonni and La Rocca (2007) for an application.

## Acknowledgements

This work was started when the first two authors were visiting the third author at the **University of Pau and Pays de l'Adour**. Financial support was given by MIUR, Rome, PRIN 2005132307, and the **University of Pavia**.

## References

- Consonni, G., and La Rocca, L. (2007). Intrinsic tests for the equality of two correlated proportions. **Preprint arXiv:0707.3877v1 [math.ST]**.
- 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**, 363–374.
- Pericchi, L. R. (2005). Model selection and hypothesis testing based on objective probabilities and Bayes factors. In: **Handbook of Statistics**, vol. 25, Bayesian Thinking, Modeling and Computation (Dey, D. and Rao, C.R., Eds.) North Holland.
- Shmueli, G., Minka, T. P., Kadane, J. B., Borle, S., and Boatwright P. (2005). A useful distribution for fitting discrete data: Revival of the COM-Poisson. **Journal of the Royal Statistical Society, Series C**, **54**, 127–142.