

The Bayesian logarithm

Surojit Biswas^{*1}

¹Department of Biomedical Informatics. Harvard Medical School. Boston, MA 02215.
USA.

1 Introduction

When working with count, or more generally, non-negative data it is often the case that one encounters zeros. Furthermore, such data usually demonstrate considerable heteroscedasticity, with variance increasing in mean – a consequence of their governing probabilistic processes (e.g. Poisson, or Negative-Binomial sampling).

In order to render the data more homoscedastic, it is popular to log-transform the data prior to performing data analysis. However, as $\log(0)$ is undefined, applying this transformation across a dataset with zeros is a tricky practice. One often therefore adds a small pseudocount (e.g. 1 when working with count data) prior to applying the log, but this too is a questionable practice. Samples are often unevenly explored, and it can be unclear whether a zero actually suggests zero abundance, or some minuscule abundance that was not within the resolution of sampling.

Using a Poisson-Normal hierarchical model, we propose here a Bayesian version of the logarithm, hereafter “blog”, that extends the ‘pseudocount concept’ in a probabilistically coherent and intuitive manner. Importantly, in the limit of data $\text{blog} = \log$, and in the absence of it, blog returns a prior belief. Furthermore, blog considers the level of confidence in or exploration of a sample, such that a zero for a sample that was well explored is treated differently than a zero for a sample for which we have low confidence. With intelligently chosen priors, which we also discuss, the Bayesian logarithm provides an intuitive and stable alternative for the logarithm for application to count or non-negative data.

2 Methods

2.1 Model

Let $t \in \mathbb{R}_{\geq 0}^n$ be a n -samples long vector of data (e.g. counts). Let $o \in \mathbb{R}_{\geq 0}^n$ be a n -samples long vector of ‘confidences’[†]. For example, in the case where t_i represents the number of times a specific species of animal was observed in a random sampling, o_i would be the size of the random sample.

^{*}Correspondence: surojitbiswas@g.harvard.edu

[†]In the literature on generalized linear models, these are often referred to as ‘offsets’.

We assume that associated with each t_i there is a true, but unmeasured *latent abundance* z_i that t_i is a noisy realization of. Specifically, we assume the following hierarchical model,

$$\begin{aligned} z_i &\sim \mathcal{N}(\mu, \sigma^2) \\ t_i &\sim \text{Continuous-Poisson}(\exp(z_i)o_i) \end{aligned}$$

Here μ and σ^2 denote the mean and variance of z_i . We define the Continuous-Poisson distribution to have the following density function, with support on $x \in [0, \infty)$:

$$f(x|\lambda) = C_\lambda \frac{\lambda^x e^{-\lambda}}{\Gamma(x+1)}$$

where C_λ is a normalization constant that ensures the density integrates to unity. This distribution has the same shape and moments as the Poisson, but has the added advantage of being able to consider all non-negative data – not just counts.

We define the Bayesian logarithm to equal

$$\text{blog}(t_i) = \log(\mathbb{E}[t_i|z_i]) = \begin{cases} z_i + \log(o_i) & \text{if } o_i > 0 \\ z_i & \text{if } o_i = 0 \end{cases}$$

In practice, we will not know z_i and must therefore learn its value.

2.2 Inference

Intuitively, a reasonable estimate of z_i is given by the mode value of its posterior distribution. This objective is given by,

$$\begin{aligned} \hat{z}_i &= \underset{z_i}{\operatorname{argmax}} p(z_i|t_i, \mu, \sigma^2) = \underset{z_i}{\operatorname{argmax}} \log \left[\frac{p(t_i|z_i, \mu, \sigma^2) p(z_i|\mu, \sigma^2)}{\int_{-\infty}^{\infty} p(t_i|x, \mu, \sigma^2) p(x|\mu, \sigma^2) dx} \right] \\ &= \underset{z_i}{\operatorname{argmax}} \log p(t_i|z_i, \mu, \sigma^2) + \log p(z_i|\mu, \sigma^2) \\ &= \underset{z_i}{\operatorname{argmax}} \log \left\{ C_{\exp(z_i)o_i} \frac{[\exp(z_i)o_i]^{t_i} e^{-\exp(z_i)o_i}}{\Gamma(t_i+1)} \right\} \\ &\quad + \log \left\{ \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2}(z_i - \mu)^2\right) \right\} \\ &= \underset{z_i}{\operatorname{argmax}} t_i z_i - \exp(z_i)o_i - \frac{1}{2\sigma^2}(z_i - \mu)^2 \end{aligned}$$

where we have progressively dropped constants that do not depend on z_i .

Differentiating we get,

$$\frac{\partial}{\partial z_i} t_i z_i - \exp(z_i)o_i - \frac{1}{2\sigma^2}(z_i - \mu)^2 = t_i - \exp(z_i)o_i - \frac{z_i - \mu}{\sigma^2}.$$

We cannot solve for this gradient analytically, so instead we rely on Newton-Raphson to optimize the gradient numerically. The required Hessian is given by,

$$\frac{\partial}{\partial z_i} t_i - \exp(z_i)o_i - \frac{z_i - \mu}{\sigma^2} = -\exp(z_i)o_i - \frac{1}{\sigma^2} < 0.$$

Note that the Hessian is negative-definite, which implies there is a unique maximum for our objective.

Therefore, we have

$$\text{blog}(t_i) = \log(\mathbb{E}[t_i|\hat{z}_i]) = \begin{cases} \hat{z}_i + \log(o_i) & \text{if } o_i > 0 \\ \hat{z}_i & \text{if } o_i = 0 \end{cases}$$

2.3 Properties

We now note some important properties of the Bayesian logarithm. Note that given z_i , $t_i \rightarrow \infty$ as $o_i \rightarrow \infty$ since $\mathbb{E}[t_i|z_i] = \exp(z_i)o_i$. Thus in the limit of data we have,

$$\lim_{t_i \rightarrow \infty, o_i \rightarrow \infty} \underset{z_i}{\operatorname{argmax}} t_i z_i - \exp(z_i)o_i - \frac{1}{2\sigma^2}(z_i - \mu)^2 = \underset{z_i}{\operatorname{argmax}} t_i z_i - \exp(z_i)o_i$$

which is simply an optimization of the Poisson likelihood. With one data point the maximum likelihood estimator of a Poisson mean is just the value of the datum itself. Consequently,

$$\lim_{t_i \rightarrow \infty, o_i \rightarrow \infty} \text{blog}(t_i) = \log(\mathbb{E}(t_i|z_i)) = \log(t_i).$$

In contrast, consider the case where we have limited data (e.g. $t_i = 0$ and $o_i = 0$). Here we have,

$$\lim_{t_i \rightarrow 0, o_i \rightarrow 0} \underset{z_i}{\operatorname{argmax}} t_i z_i - \exp(z_i)o_i - \frac{1}{2\sigma^2}(z_i - \mu)^2 = \underset{z_i}{\operatorname{argmax}} -\frac{1}{2\sigma^2}(z_i - \mu)^2$$

which is simply the Normal likelihood, for which $z_i = \mu$ is the maximizer. Thus,

$$\lim_{t_i \rightarrow 0, o_i \rightarrow 0} \text{blog}(t_i) = \log(\mathbb{E}(t_i|z_i)) = \mu.$$

3 Results

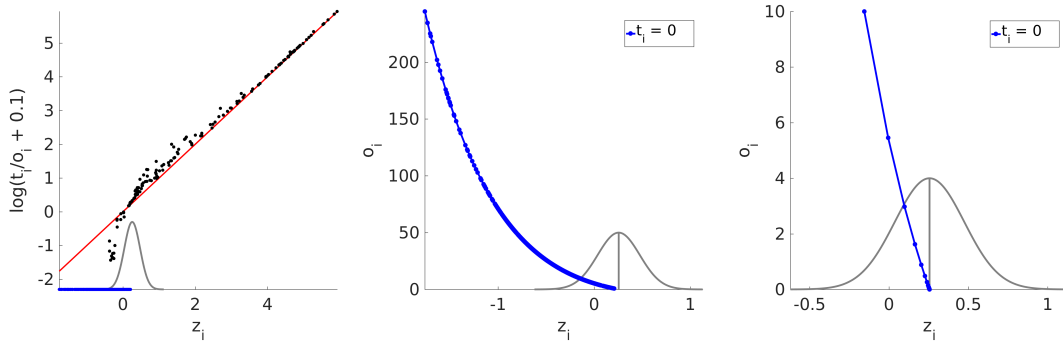


Figure 1