

# Quasi-Poisson regression

M. Rule

March 12, 2014

In neuroscience, we often model the number of spikes observed from a neuron as a Poisson distribution with rate  $\lambda$ :

$$y \sim \text{Poisson}(\lambda) \tag{1}$$

$$\Pr(y|\lambda) = \frac{1}{\Gamma(y+1)} \lambda^y e^{-\lambda} \tag{2}$$

In a typical log-linear Poisson Generalized-Linear Point-Process (PP-GLM) regression, we estimate the logarithm of  $\lambda$  as a function of other, measured covariances “ $\mathbf{x}$ ”, that is  $\ln \lambda = \mathbf{w}^\top \mathbf{x}$ . However, for these notes we will simply assume that  $\lambda$  is already available, and we’d like to model  $\Pr(y|\lambda)$ .

**Over- and Under-dispersion** One limitation of the Poisson distribution is that it has only one parameter,  $\lambda$ , which controls both the mean  $\mu = \lambda$  and the variance  $\sigma^2 = \lambda$ . However, spike-count data are often over- or under-dispersed in practice, meaning that the variance might be larger or smaller than  $\lambda$ , respectively.

There are many way to model over-dispersed data. For example, using the Negative Binomial distribution is common. It is also possible to use the “zero inflated Poisson distribution”. In this case, the probability of observing any spikes at all is controlled by another Bernoulli ( $s = \text{Bernoulli}(p)$ ) variable, and the distribution of spike-counts given  $s = 1$  is then Poisson distributed. Both of these approaches can only increase the variance, so they can only handle over-dispersion.

**Quasi-Poisson observation model** Another approach to handling over/under dispersion is the so-called “quasipoisson” regression. This uses a function for  $\Pr(y|\lambda)$  which is not a probability density function. Nevertheless, it provides a quick and useful hack to control the dispersion without changing much in the Poisson regression.

To adjust the dispersion in quasipoisson regression, one can multiply both the observed spike counts  $y$  and the firing rate  $\lambda$  by a dispersion-control parameter “ $\kappa$ ”.

$$Q(y|\lambda) \approx \frac{\kappa}{\Gamma(\kappa y+1)} (\kappa \lambda)^{\kappa y} e^{-\kappa \lambda} \tag{3}$$

On the surface, this corresponds to a new Poisson distribution with rate  $\lambda' = \kappa \lambda$ , so its mean and variance are  $\mu' = \sigma'^2 = \kappa \lambda$ . However, we evaluate it at  $\tilde{y} = \kappa y$  rather than  $y$ . This makes this an improper distribution for discrete, integer-value count data.

**Adding a scale parameter to an existing distribution** Recall: If  $x \sim P(x)$  is a continuous, one-dimensional random variable with known distribution “ $Q(x)$ ”, then the distribution of  $\tilde{x} = 2x$  is narrower. In particular

$$\tilde{x} \sim 2 \cdot P(\tilde{x} = 2x) \tag{4}$$

The multiplication by 2 adjusts the normalization factor to account for the fact that a function that is half as wide also has half as much area, so we need to multiply by 2 to ensure that the density still integrates to 1. If  $x$  has known mean  $\mu$  and standard deviation  $\sigma$ , then the rescaled  $\tilde{x}$  will have mean  $\mu/2$  and  $\sigma/2$ , respectively.

**Scaling of observation error variance in the quasi-Poisson model** So, when we multiply our rate  $\lambda$  by  $\kappa$ , we move the mean and variance to  $\mu' = \sigma'^2 = \kappa\lambda$ . But, by evaluating this density at  $\tilde{y} = \kappa y$ , we then divide the mean and standard deviation by  $\kappa$ . The net transformation leads to

$$\tilde{\mu} = (\kappa\lambda)/\kappa = \mu = \lambda \tag{5}$$

$$\tilde{\sigma}^2 = (\kappa\lambda)/\kappa^2 = \lambda/\kappa \tag{6}$$

We see then that  $\kappa$  controls the over/under dispersion.  $\kappa = 1$  is the original Poisson distribution.  $\kappa > 1$  reduces the variance.  $\kappa < 1$  increases it.

**Log-likelihoods** The log-likelihood for the quasi-Poisson model is

$$\ln Q(y|\lambda) = \ln(\kappa) - \ln \Gamma(\kappa y + 1) + (\kappa y) \ln(\kappa\lambda) - \kappa\lambda \tag{7}$$

If  $k$  is known and  $y$  is fixed, we can instead optimize

$$\ln Q(y|\lambda) = \kappa[y \ln(\lambda) - \lambda] + \text{constant} \tag{8}$$

In this case quasi-Poisson regression can be thought of as Poisson regression, but with an extra “fudge” factor  $\kappa$  that re-weights the importance of our evidence.

**Take care** Care must be taken when mixing the quasi-Poisson observation model with Bayesian methods. For the most part, things will work. However, since the quasi-Poisson model is not a distribution, it is unclear how it should be normalized. In particular, estimating  $k$  rigorously is not well-posed. I suppose one might attempt to e.g. adjust  $k$  to maximize and evidence lower-bound, retaining then the terms  $\ln(\kappa) - \ln \Gamma(\kappa y + 1) + y\kappa \ln(\kappa)$  from the quasi-Poisson likelihood that depend on  $k$ . I have not tried this, but perhaps it usually works out ok.