

# Regression for Randomly Running A Poisson Process

Matthew McGonagle

February 23, 2018

Here we record the calculations for doing a regression on the combination of doing a Bernoulli trial to determine whether to run a Poisson process on a length of time  $T$ . When we don't run the poisson process, we have a resulting count of 0.

- Let  $p$  be the probability of choosing to run the Poisson Process.
- Let  $\lambda$  be the poisson density. That is, for a length of time  $T$ , the poisson rate is  $\lambda T$ .
- Let  $k$  denote the count (i.e. outcome) of a Poisson process.
- The sum  $\sum_{k_i=0}$  will denote a sum over all relevant data samples  $(T_i, k_i)$  where the count  $k_i$  is 0.

Similarily the sum  $\sum_{k_i>0}$  will denote a sum over all samples where the count is non-zero.

We will perform the regression by finding the values of  $p$  and  $\lambda$  that maximize the log-likelihood for a given set of data.

## 1 The Log-Likelihood

First, note that when our Bernoulli trial decides that we don't run a poisson process, we get a count of 0. The probability of this occurring is independent of the poisson density  $\lambda$  and the length of time  $T$ . So we have that:

$$P(0|p, \lambda, T) = P(\text{No Poisson}|p, \lambda) + P(\text{Poisson } k = 0|p, \lambda, T), \quad (1)$$

$$= 1 - p + pe^{-\lambda T}. \quad (2)$$

Next, we have that for counts  $k > 0$  that we must run a Poisson process. So we have that

$$P(k|p, \lambda) = \frac{p(\lambda T)^k}{k!} e^{-\lambda T}. \quad (3)$$

Our data  $X$  consists of time interval lengths  $T_i$  and response counts  $k_i$ . We maximize the log-likelihoods  $L$  for the data  $\{(t_i, k_i)\}$ . The log-likelihoods are given by

$$L(X|p, \lambda) = \sum_{k_i=0} \log(1 - p + pe^{-\lambda T_i}) \quad (4)$$

$$+ \sum_{k_i>0} (\log p + k_i \log \lambda + k_i \log T_i - \lambda T_i - \log k_i!). \quad (5)$$

Note that the last summand in the second sum,  $\log k_i!$  is independent of  $p$  and  $\lambda$ .

## 2 Gradient of Log-Likelihood

Now we compute the gradient of the log-likelihoods.

$$\partial L / \partial \lambda = -p \sum_{k_i=0} \frac{T_i e^{-\lambda T_i}}{1 - p + pe^{-\lambda T_i}} + \sum_{k_i>0} \left( \frac{k_i}{\lambda} - T_i \right). \quad (6)$$

$$\partial L / \partial p = - \sum_{k_i=0} \frac{1 - e^{-\lambda T_i}}{1 - p + pe^{-\lambda T_i}} + \frac{1}{p} \sum_{k_i>0} 1. \quad (7)$$

The equations for the gradient of  $L$  look complicated, but they can be accurately summarized as two equations for the unknown variables  $p$  and  $\lambda$ . Every other term is known from the data. However, the solution of these equations is not of a closed form, and so we will need to resort to numerical methods to solve them. We will use the Newton method to solve for where the gradient vanishes.

## 3 Hessian of Log-likelihood

To use the Newton method, we will need to compute another derivative of  $L$ , i.e. the Hessian of  $L$ . We compute

$$\frac{\partial^2 L}{\partial \lambda^2} = p \sum_{k_i=0} \left( \frac{T_i^2 e^{-\lambda T_i}}{1 - p + pe^{-\lambda T_i}} - \frac{p T_i^2 e^{-2\lambda T_i}}{(1 - p + pe^{-\lambda T_i})^2} \right) - \frac{1}{\lambda^2} \sum_{k_i>0} k_i, \quad (8)$$

$$= p(1 - p) \sum_{k_i=0} \frac{T_i^2 e^{-\lambda T_i}}{(1 - p + pe^{-\lambda T_i})^2} - \frac{1}{\lambda^2} \sum_{k_i>0} k_i. \quad (9)$$

We compute the next second derivative:

$$\frac{\partial^2 L}{\partial p^2} = - \sum_{k_i=0} \frac{(1 - e^{-\lambda T_i})^2}{(1 - p + pe^{-\lambda T_i})^2} - \frac{1}{p^2} \sum_{k_i>0} 1. \quad (10)$$

Next, we compute the mixed second derivatives:

$$\frac{\partial^2 L}{\partial p \partial \lambda} = - \sum_{k_i=0} \left( \frac{T_i e^{-\lambda T_i}}{1-p+pe^{-\lambda T_i}} + \frac{p T_i (1-e^{-\lambda T_i}) e^{-\lambda T_i}}{(1-p+pe^{-\lambda T_i})^2} \right), \quad (11)$$

$$= - \sum_{k_i=0} \frac{T_i e^{-\lambda T_i}}{(1-p+pe^{-\lambda T_i})^2}. \quad (12)$$

## 4 Newton Step

We start with initial guesses  $\lambda_0$  and  $p_0$ . To motivate some convenient guesses, note that for data points where the count is non-zero, we know that a Poisson process occurred. So as a first guess, let's just use the mean count per mean length for the portion of the data set with non-zero counts.

Once we have a guess of the density  $\lambda_0$ , we use the mean length to give an expectation that a Poisson process results in a count of zero. Then we can adjust the proportion of non-zero counts to more accurately reflect the probability of running a poisson process.

Therefore, convenient initial guesses are

$$\lambda_0 = \frac{\sum_{k_i>0} k_i}{\sum_{k_i>0} T_i}, \quad (13)$$

$$p_0 = \frac{\sum_{k_i>0} 1}{(1 - e^{-\lambda_0 \sum_i T_i / \sum_i 1}) \sum_{\text{all data}} 1}. \quad (14)$$

Now let us consider the induction step of Newton's method. We are given the current approximations  $\lambda_i$  and  $p_i$ . We compute new approximations  $\lambda_{i+1}$  and  $p_{i+1}$  in the following manner.

First recall the Hessian

$$H = \begin{pmatrix} \frac{\partial^2 L}{\partial \lambda^2} & \frac{\partial^2 L}{\partial \lambda \partial p} \\ \frac{\partial^2 L}{\partial \lambda \partial p} & \frac{\partial^2 L}{\partial p^2} \end{pmatrix}. \quad (15)$$

We solve for an update vector  $\vec{u}$  given by

$$H \vec{u} = - \begin{pmatrix} \partial L / \partial \lambda \\ \partial L / \partial p \end{pmatrix}. \quad (16)$$

So we get

$$\vec{u} = -H^{-1} \begin{pmatrix} \partial L / \partial \lambda \\ \partial L / \partial p \end{pmatrix}. \quad (17)$$

Note that  $H$  is a  $2 \times 2$  matrix and so its inverse is readily found using a formula.

Then we find our updated approximations from

$$\begin{pmatrix} \lambda_{i+1} \\ p_{i+1} \end{pmatrix} = \vec{u} + \begin{pmatrix} \lambda_i \\ p_i \end{pmatrix}. \quad (18)$$

We repeat until we get enough convergence to approximate values of our parameters. This can be seen by looking at the history of the changes in the parameters every time we run the update step.