Asymptotic covariance of Poisson regression coefficients

Daniel P. Rice

November 2022

Abstract

The power of Poisson regression detect exponentially growing abundances from count data depends on the variance of the estimated growth rate parameter. Here, I calculate that variance in the limit that you have a lot of data and sample many time points. I find an asymptotic expression for the variance that shows the dependence on the sampling time, the number of samples, the sampling intensity, and the growth parameters.

Let $a(t)$ be the abundance of an exponentially-growing pathogen over time:

$$a(t) = \exp(\beta_0 + \beta_1 t) \tag{1}$$

for $t \in [0, T]$. Suppose we take n measurements at times $\{t_i\}$ with exposures $\{\lambda_i\}$ for $i = 1, \ldots, n-1$. The exposures represent a conversion factor between the abundance and the expected number of counts and lump together the sampling intensity and properties of the sequencing protocol such as taxonomic bias. We model the vector of observed counts $\vec{k}$ as independent Poisson random variables:

$$k_i \sim \text{Poisson}(\lambda_i a(t_i)). \tag{2}$$

We use Poisson regression to estimate the parameters (β_0, β_1). This is equiv-

alent to maximizing the log-likelihood:

$$l(\beta_0, \beta_1; \vec{k}, \vec{\lambda}) = \log \prod_{i=0}^{n-1} \frac{(\lambda_i a(t_i))^{k_i}}{k!} \exp(\lambda_i a(t_i)) \tag{3}$$

$$= \sum_{i=0}^{n-1} \left[k_i(\beta_0 + \beta_1 t_i) - \lambda_i e^{\beta_0 + \beta_1 t_i} \right] + \text{const.} \tag{4}$$

In the limit where we have a lot of data (query what exactly this means in this context), the covariance matrix of the maximum likelihood parameter estimates $(\hat{\beta}_0, \hat{\beta}_1)$ converges to the inverse of the Fisher information matrix,

$$\mathcal{I}_{i,j} = -\mathbb{E}\left[\frac{\partial^2 l}{\partial \beta_i \partial \beta_j} \middle| \beta_0, \beta_1 \right]. \tag{5}$$

Differentiating the log likelihood twice gives us:

$$\frac{\partial^2 l}{\partial \beta_0^2} = -\sum_{i=0}^{n-1} \lambda_i e^{\beta_0 + \beta_1 t} \tag{6}$$

$$\frac{\partial^2 l}{\partial \beta_0 \partial \beta_1} = -\sum_{i=0}^{n-1} \lambda_i t_i e^{\beta_0 + \beta_1 t} \tag{7}$$

$$\frac{\partial^2 l}{\partial \beta_1^2} = -\sum_{i=0}^{n-1} \lambda_i t_i^2 e^{\beta_0 + \beta_1 t}. \tag{8}$$

If we take evenly spaced samples with equal exposure so that $t_i = T(i/n - 1)$ and $\lambda_i = \lambda/n$, with total exposure λ, we have

$$\mathcal{I} = \lambda e^{\beta_0} \begin{pmatrix} S_0 & TS_1 \\ TS_1 & T^2 S_2 \end{pmatrix} \tag{9}$$

where

$$S_j = n^{-1} \sum_{i=0}^{n-1} \left(\frac{i}{n-1} \right)^j \exp\left(\beta_1 T \frac{i}{n-1} \right) \tag{10}$$

Because $\mathcal{I}$ is a 2×2 matrix, we can invert it to get the asymptotic covariance matrix:

$$\mathcal{I}^{-1} = \lambda^{-1} e^{-\beta_0} T^{-2} \frac{1}{S_0 S_2 - S_1^2} \begin{pmatrix} T^2 S_2 & -TS_1 \\ -TS_1 & S_0 \end{pmatrix} \tag{11}$$

For exponential growth detection, we're most interested in

$$\mathrm{Var}(\hat{\beta}_1) = \lambda^{-1} e^{-\beta_0} T^{-2} \frac{S_0}{S_0 S_2 - S_1^2}. \tag{12}$$

We can already learn some important things from Eq. 12:

1. The variance of $\hat{\beta}_1$ is inversely proportional to the total sampling effort, λ, and to the initial abundance, e^{β_0}.

2. The standard error is inversely proportional to the total time spanned by the sampling, T.

3. The variance depends on the true growth rate through the dimensionless parameter $\beta_1 T$, which represents the log-fold increase in abundance over T.

Next, we want to clarify the dependence of $\mathrm{Var}(\hat{\beta}_1)$ on the number of sampled timepoints, n, and on the log fold-increase $\beta_1 T$, which are opaque in Eq. 12. First, we will get the exact equation for the situation with no growth ($\beta_1 = 0$). Then, we'll look at the general case and study the large n behavior with an asymptotic series in n^{-1}.

When $\beta_1 = 0$, the sums become simple:

$$S_0 = 1 \tag{13}$$

$$S_1 = \frac{n-1}{2} \tag{14}$$

$$S_2 = \frac{(n-1)(2n-1)}{6} \tag{15}$$

so that we have

$$\mathrm{Var}(\hat{\beta}_1) = \lambda^{-1} e^{-\beta_0} \frac{12}{T^2} \left(\frac{n-1}{n+1} \right) \tag{16}$$

$$\sim \lambda^{-1} e^{-\beta_0} \frac{12}{T^2} \left(1 - \frac{2}{n} + \frac{2}{n^2} + \cdots \right) \tag{17}$$

as $n \to \infty$.

Note that the variance converges to a constant as the number of samples increases (holding the total exposure constant). This is intuitive because we expect diminishing returns to ever-finer temporal resolution. Also note that

the error of the first order approximation in n^{-1} is already only 2% once you have 10 sampled times, so for the full case, we won't bother with higher-order approximations.

For the case of arbitrary β_1, we can still compute the sums S_j in closed form, but they become significantly more complicated:

$$S_0 = \frac{e^{\beta_1 T \frac{n}{n-1}} - 1}{n\left(e^{\beta_1 T \frac{1}{n-1}} - 1\right)} \tag{18}$$

$$S_1 = \frac{(n-1)e^{\beta_1 T \frac{n+1}{n-1}} - ne^{\beta_1 T \frac{n}{n-1}} + e^{\beta_1 T \frac{1}{n-1}}}{n(n-1)\left(e^{\beta_1 T \frac{1}{n-1}} - 1\right)^2} \tag{19}$$

$$S_2 = \frac{(n-1)^2 e^{\beta_1 T \frac{n+2}{n-1}} + (-2n^2+2n+1)e^{\beta_1 T \frac{n+1}{n-1}} + n^2 e^{\beta_1 T \frac{n}{n-1}} - e^{\beta_1 T \frac{2}{n-1}} - e^{\beta_1 T \frac{1}{n-1}}}{n(n-1)^2\left(e^{\beta_1 T \frac{1}{n-1}} - 1\right)^3} \tag{20}$$

To get an asymptotic series for the variance, we can either do pages and pages of tedious algebra, or we can turn to Mathematica. Here's the answer we get from the later:

$$\begin{aligned}\mathrm{Var}(\hat{\beta}_1) = \lambda^{-1} e^{-\beta_0} T^{-2} \Bigg[& \frac{(\beta_1 T)^3 \left(e^{\beta_1 T} - 1\right)}{e^{2\beta_1 T} - ((\beta_1 T)^2 + 2)e^{\beta_1 T} + 1} \\ & + n^{-1} \frac{(\beta_1 T)^3 \left((\beta_1 T - 2)\, e^{\beta_1 T} + \beta_1 T + 2\right) \left(e^{2\beta_1 T} + ((\beta_1 T)^2 - 2)e^{\beta_1 T} + 1\right)}{2\left(e^{2\beta_1 T} - ((\beta_1 T)^2 + 2)e^{\beta_1 T} + 1\right)^2} \\ & + \mathcal{O}(n^{-2}) \Bigg]\end{aligned} \tag{21}$$

TODO:

1. Plots
2. Include Mathematica code
3. Interpretation of final equation
4. P-values under Gaussian approximation to null distribution