

Stat 3701 Lecture Notes: Zero-Truncated Poisson Distribution

Charles J. Geyer

April 11, 2017

1 License

This work is licensed under a Creative Commons Attribution-ShareAlike 4.0 International License (<http://creativecommons.org/licenses/by-sa/4.0/>).

2 The Zero-Truncated Poisson Distribution

The zero-truncated Poisson distribution is a Poisson distribution conditioned on being nonzero. It has (exponential family canonical) parameter θ and data x . These satisfy $-\infty < \theta < \infty$ and $x \in \{1, 2, 3, \dots\}$.

Define

$$m = e^\theta$$
$$\mu = \frac{m}{1 - e^{-m}}$$

(μ is the mean of the zero-truncated Poisson distribution and m is the mean of the corresponding untruncated Poisson distribution). Then the log likelihood and derivatives are given by

$$l(\theta) = x\theta - m - \log(1 - e^{-m})$$
$$l'(\theta) = x - \mu$$
$$l''(\theta) = -\mu(1 + m - \mu)$$
$$= -\mu(1 - \mu e^{-m})$$

where, as always in statistics and as in R, \log is the natural logarithm function.

The two formulas for the second derivative have different parameter values for which they suffer catastrophic cancellation. More on this below.

3 Calculation

This is a super hard problem to calculate accurately.

3.1 Large Negative Theta

3.1.1 The Function Itself

First we need to know the behavior of the function near $-\infty$. Naively taking termwise limits gives $(-\infty) - 0 - (-\infty)$, which is undefined. So we apply L'Hospital's rule.

$$\begin{aligned}\lim_{\theta \rightarrow -\infty} e^{l(\theta)} &= \lim_{\theta \rightarrow -\infty} \frac{e^{x\theta} e^{-m}}{1 - e^{-m}} \\ &= \left(\lim_{\theta \rightarrow -\infty} e^{-m} \right) \left(\lim_{\theta \rightarrow -\infty} \frac{e^{x\theta}}{1 - e^{-m}} \right) \\ &= \lim_{\theta \rightarrow -\infty} \frac{e^{x\theta}}{1 - e^{-m}}\end{aligned}$$

because $m \rightarrow 0$ and $e^{-m} \rightarrow 1$ as $\theta \rightarrow -\infty$.

Now we apply L'Hospital's rule to the limit that remains.

$$\lim_{\theta \rightarrow -\infty} \frac{e^{x\theta}}{1 - e^{-m}} = \lim_{\theta \rightarrow -\infty} \frac{x e^{x\theta}}{e^{-m} e^{\theta}} = \begin{cases} 1, & x = 1 \\ 0, & x > 1 \end{cases}$$

Hence (remembering that the limit above is for $e^{l(\theta)}$)

$$\lim_{\theta \rightarrow -\infty} l(\theta) = \begin{cases} 0, & x = 1 \\ -\infty, & x > 1 \end{cases}$$

The problem is to get correct answers, avoiding overflow or underflow, except when the answers are really too large or too small for the computer to represent.

In case $x = 1$, we need to calculate the log of this quantity accurately, avoiding underflow when possible.

$$\log\left(\frac{e^{x\theta}}{1 - e^{-m}}\right) = \log\left(\frac{m}{1 - e^{-m}}\right) = \log(\mu)$$

which we know is near zero when m is near zero. The problem is that, even if we calculate this as

`log(- m / expm1(- m))`

we know that `- expm1(- m)` is very close to m for very small m so we do get very close to $\log(1) = 0$, but in order to avoid catastrophic cancellation, getting exactly zero when we should be getting something nonzero, we need to replace the `log` with `log1p`. But we don't have the expression in the form to do that.

Replace the exponential in the denominator by its Maclaurin series

$$e^{-m} = 1 - m + \frac{m^2}{2} - \frac{m^3}{3!} + \frac{m^4}{4!} - \dots + \frac{(-m)^k}{k!} + \dots$$

so

$$\mu = \frac{m}{1 - \left(1 - m + \frac{m^2}{2} - \frac{m^3}{3!} + \frac{m^4}{4!} - \dots + \frac{(-m)^k}{k!} + \dots\right)} = \frac{1}{1 - \frac{m}{2} + \frac{m^2}{3!} - \frac{m^3}{4!} + \dots + \frac{(-m)^{k-1}}{k!} + \dots}$$

So this gives us a formula for stable computation for large negative θ

$$\begin{aligned}l(\theta) &= x\theta - m - \log(1 - e^{-m}) \\ &= (x - 1)\theta - m + \log(\mu) \\ &= (x - 1)\theta - m - \log\left(1 - \frac{m}{2} + \frac{m^2}{3!} - \frac{m^3}{4!} + \dots + \frac{(-m)^{k-1}}{k!} + \dots\right)\end{aligned}$$

3.1.2 The First Derivative

We already have figured out how to calculate μ stably, but when $x = 1$ and θ is large negative our formula for the first derivative $x - \mu$ will exhibit catastrophic cancellation. We will get zero in computer arithmetic when we should be getting negative numbers less than the machine epsilon in absolute value. So we also need a stable formula for

$$\begin{aligned} 1 - \mu &= 1 - \frac{m}{1 - e^{-m}} \\ &= \frac{1 - m - e^{-m}}{1 - e^{-m}} \\ &= \frac{-\frac{m^2}{2} + \frac{m^3}{3!} - \frac{m^4}{4!} + \dots}{m - \frac{m^2}{2} + \frac{m^3}{3!} - \frac{m^4}{4!} + \dots} \\ &= m \cdot \frac{-\frac{1}{2} + \frac{m}{3!} - \frac{m^2}{4!} + \dots}{1 - \frac{m}{2} + \frac{m^2}{3!} - \frac{m^3}{4!} + \dots} \end{aligned}$$

Thus we get $1 - \mu \approx -m/2$ for $\theta \approx -\infty$. So the first derivative should not underflow until m does, which is at about $\theta = -750$.

For $x > 1$, we will have $l'(\theta) \approx x - 1$ for $\theta \approx -\infty$ and this is stably calculated by subtraction.

3.1.3 The Second Derivative

Since $m \rightarrow 0$ and $\mu \rightarrow 1$ as $\theta \rightarrow -\infty$, we have $l''(\theta) \rightarrow 0$. Thus our formula for the second derivative will also suffer catastrophic cancellation in this case. We need to rewrite it too.

$$\begin{aligned} l''(\theta) &= -\mu(1 + m - \mu) \\ &= -\mu[(1 - \mu) + m] \end{aligned}$$

but in the preceding section we learned how to calculate $1 - \mu$ stably. Thus we get $l''(\theta) \approx -\mu m/2$ for $\theta \approx -\infty$.

Since this formula does not contain x , it works for all values of x .

3.2 Large Positive Theta

As $\theta \rightarrow \infty$

$$\begin{aligned} m &\rightarrow \infty \\ \mu &\rightarrow \infty \\ l(\theta) &\rightarrow -\infty \\ l'(\theta) &\rightarrow -\infty \\ l''(\theta) &\rightarrow -\infty \end{aligned}$$

and there is nothing that can be done about any of this. The quantities to be calculated are larger than any numbers the computer has. That's what the values `Inf` and `-Inf` were invented for.

However, there are subtractions in the formulas above. We need to be sure that they do not give `Inf - Inf = NaN`. Any formulas that do do that also suffer from catastrophic cancellation for very large θ .

3.2.1 The Function Itself

The formula

$$l(\theta) = x\theta - m - \log(1 - e^{-m})$$

will have no problems (we should use the `log1p` function to calculate the `log`) until $x\theta$ overflows, in which case $m = e^\theta$ has already overflowed, and the computer calculates `Inf - Inf = NaN`, which is wrong. We know the limit is `-Inf`. This is because $x\theta$ goes to infinity much more slowly than e^θ .

So we will have to test for `x * theta == Inf` and return `-Inf` in this case.

3.2.2 The First Derivative

There is no problem with the first derivative formula

$$l'(\theta) = x - \mu$$

When θ is so large that μ overflows, this formula will give the correct result `-Inf`.

3.2.3 The Second Derivative

There is a problem with the second derivative formula

$$l''(\theta) = -\mu(1 + m - \mu)$$

that is good for negative θ or even moderate sized positive θ . For large θ we have $m \approx \mu$ and suffer catastrophic cancellation.

So for large positive θ we need to use our other second derivative formula.

$$l''(\theta) = -\mu(1 - \mu e^{-m})$$

Since $\mu e^{-m} \approx 0$ when $\theta \approx \infty$, this gives the correct answer, approximately $-\mu$ when this does not overflow and `-Inf` when it does overflow.

4 Summary

4.1 Function Itself

There are three cases. For large negative θ , use

$$l(\theta) = (x - 1)\theta - m - \log\left(1 - \frac{m}{2} + \frac{m^2}{3!} - \frac{m^3}{4!} + \dots + \frac{(-m)^{k-1}}{k!} + \dots\right)$$

(since the infinite series is convergent for all θ , this works for all θ but is only necessary for large negative θ).

Otherwise, use

$$l(\theta) = x\theta - m - \log(1 - e^{-m})$$

except when $x\theta$ overflows (is `Inf`), in which case the correct result is `-Inf`.

4.2 Mean

The mean μ is not part of the result but is used in intermediate calculations.

When θ is large negative, use

$$\mu = \frac{1}{1 - \frac{m}{2} + \frac{m^2}{3!} - \frac{m^3}{4!} + \dots + \frac{(-m)^{k-1}}{k!} + \dots}$$

Otherwise, the formula

$$\mu = \frac{m}{1 - e^{-m}}$$

works.

4.3 First Derivative

When $x = 1$ and θ is large negative, use

$$l'(\theta) = m \cdot \frac{-\frac{1}{2} + \frac{m}{3!} - \frac{m^2}{4!} + \dots + \frac{(-1)^{k-1} m^{k-2}}{k!} + \dots}{1 - \frac{m}{2} + \frac{m^2}{3!} - \frac{m^3}{4!} + \dots + \frac{(-m)^{k-1}}{k!} + \dots}$$

In all other cases

$$l'(\theta) = x - \mu$$

gives good answers.

4.4 Second Derivative

When θ is large negative, use

$$l''(\theta) = -\mu(1 + m - \mu)$$

When θ is large positive, use

$$l''(\theta) = -\mu(1 - \mu e^{-m})$$

except when μ overflows (is `Inf`), in which case this formula gives `NaN` but the correct answer is `-Inf`.

For moderate sized θ , positive or negative, either formula works.