EdX and its Members use cookies and other tracking technologies for performance, analytics, and marketing purposes. By using this website, you accept this use. Learn more about these technologies in the Privacy Policy.



# 1. Poisson regression

*Instructions: For this problem, whenever a formula box requires you to enter a factorial, enter **fact** to indicate the factorial function. For instance, **fact(10)** denotes $10!$.*

## (a)

5/5 points (graded)

We want to model the rate of infection with an infectious disease depending on the day after outbreak $t$. Denote the recorded number of outbreaks at day $t$ by $k_t$.

We are going to model the distribution of $k_t$ as a Poisson distribution with a time-varying parameter $\lambda_t$, which is a common assumption when handling count data.

First, recall the likelihood of a Poisson distributed random variable $Y$ in terms of the parameter $\lambda$,

$$P(Y=k) = \mathbf{e}^{-\lambda} \frac{\lambda^k}{k!}.$$

Rewrite this in terms of an exponential family. In other words, write it in the form

$$P(Y=k) = h(k) \exp\left[\eta(\lambda) T(k) - B(\lambda)\right].$$

Since this representation is only unique up to re-scaling by constants, take the convention that $T(k) = k$.

$\eta(\lambda) =$ ln(lambda) ✔ **Answer:** ln(lambda)

$B(\lambda) =$ lambda ✔ **Answer:** lambda

$h(k) =$ 1/(fact(k)) ✔ **Answer:** 1/(fact(k))

We can write this in canonical form, e.g. as

$$P(Y = k) = h(k) \exp[k\eta - b(\eta)].$$

What is $b(\eta)$?

$b(\eta) =$ e^eta ✔ **Answer:** exp(eta)

Recall that the mean of a Poisson$(\lambda)$ distribution is $\lambda$. What is the canonical link function $g(\mu)$ associated with this exponential family, where $\mu = \mathbb{E}[Y]$ ? Write your answer in terms of $\lambda$.

$g(\mu) =$ ln(lambda) ✔ **Answer:** ln(lambda)

**Solution:**

We can rewrite the likelihood as

$$\begin{aligned} P(Y = k) &= \mathbf{e}^{-\lambda} \frac{\lambda^k}{k!} \\ &= \frac{1}{k!} \exp[-\lambda + k \ln(\lambda)]. \end{aligned}$$

Hence, given the convention $T(k) = k$ for this specific case, we set

$$\begin{aligned} h(k) &= \frac{1}{k!} \\ B(\lambda) &= \lambda \\ \eta(\lambda) &= \ln(\lambda). \end{aligned}$$

In order to rewrite this in canonical form, solve

$$\ln(\lambda) = \eta \iff \lambda = \mathbf{e}^{\eta},$$

so

$$b(\eta) = \mathbf{e}^{\eta}.$$

The canonical link function is $b'^{-1}$, which is

$$b'^{-1}(\mu) = \ln(\mu)$$

Submit You have used 1 of 3 attempts

Answers are displayed within the problem

## (b)

2/2 points (graded)

What range will the values in $Y$ belong to?

- [ ] $\mathbb{Z} = \{\ldots, -2, -1, 0, 1, 2, \ldots\}$

- $\mathbb{Z}_+ = \{1, 2, 3, \ldots\}$
- (selected) $\mathbb{Z}_{\geq 0} = \{0, 1, 2, 3, \ldots\}$
- $\mathbb{R}$
- $\mathbb{R}_{\geq 0} = \{x \in \mathbb{R} : x \geq 0\}$
- $\mathbb{R}_{>0} = \{x \in \mathbb{R} : x > 0\}$

✔

According to the canonical Generalized Linear Model (your answer from (a)), what is the range of possible predictions for $\lambda$ ?

- $\mathbb{Z} = \{\ldots, -2, -1, 0, 1, 2, \ldots\}$
- $\mathbb{Z}_+ = \{1, 2, 3, \ldots\}$
- $\mathbb{Z}_{\geq 0} = \{0, 1, 2, 3, \ldots\}$
- $\mathbb{R}$
- $\mathbb{R}_{\geq 0} = \{x \in \mathbb{R} : x \geq 0\}$
- (selected) $\mathbb{R}_{>0} = \{x \in \mathbb{R} : x > 0\}$

✔

**Solution:**

$Y$ has the Poisson distribution, so it lives in $\{0, 1, 2, \ldots\}$.

Since the canonical model states $\lambda = e^{\eta}$, the range of $\lambda_t$ is the full range of parameters for a Poisson distribution: $\mathbb{R}_{>0} = \{x \in \mathbb{R} : x > 0\}$.

Submit

You have used 2 of 3 attempts

Answers are displayed within the problem

## (c)

3/3 points (graded)

Return to the original model. We now introduce a Poisson intensity parameter $\lambda_t$ for every time point and denote the parameter ($\eta$) that gives the canonical exponential family representation as above by $\theta_t$ . We choose to employ a linear model connecting the time points $t$ with the canonical parameter $\theta$ of the Poisson distribution above, i.e.,

$$\theta_t = a + bt.$$

In other words, we choose a generalized linear model with Poisson distribution and its canonical link function. That also means that conditioned on $t$, we assume the $Y_t$ to be independent.

Imagine we observe the following data:

$t_1 = 1$ 1 outbreaks

$t_2 = 2$ 3 outbreaks

$t_3 = 4$ 10 outbreaks

We want to produce a maximum likelihood estimator for $(a, b)$. To this end, write down the log likelihood $\ell(a, b)$ of the model for the provided three observations at $t_1$, $t_2$, and $t_3$ (plug in their values).

$\ell(a,b) =$

14*a+47*b-e^(a+b)-e^(a+2*b)-e^(a+4*b)-16.89617

**Answer:** -ln(6)-ln(fact(10))-exp(a+b)-exp(a+2*b)-exp(a+4*b)+(14*a)+(47*b)

$$14 \cdot a + 47 \cdot b - e^{a+b} - e^{a+2\cdot b} - e^{a+4\cdot b} - 16.89617$$

What is its gradient? Enter your answer as a pair of derivatives.

$\partial_a \ell(a,b) =$

14-e^(a+b)-e^(a+2*b)-e^(a+4*b)

**Answer:** -exp(a+b)-exp(a+2*b)-exp(a+4*b)+14

$$14 - e^{a+b} - e^{a+2\cdot b} - e^{a+4\cdot b}$$

$\partial_b \ell(a,b) =$

47-e^(a+b)-2*e^(a+2*b)-4*e^(a+4*b)

**Answer:** -exp(a+b)-2*exp(a+2*b)-4*exp(a+4*b)+47

$$47 - e^{a+b} - 2 \cdot e^{a+2\cdot b} - 4 \cdot e^{a+4\cdot b}$$

**Solution:**

The likelihood for one observation is given by

$$P(Y_t = k_t) = \frac{1}{k_t!} \exp\left[-\exp(a+bt) + k_t(a+bt)\right].$$

That means the log likelihood for the model for n observations is

$$\ell(a,b) = \sum_{i=1}^{n} \left[-\ln(k_t!) - \exp(a+bt_i) + k_{t_i}(a+bt_i)\right].$$

Plugging in the provided values, we get

$$\begin{aligned}\ell(a,b) = \quad & -\ln(1!) - \ln(3!) - \ln(10!) \\ & -\exp(a+b) - \exp(a+2b) - \exp(a+4b) \\ & +14a + 47b.\end{aligned}$$

Its derivative with respect to $a$ is

$$\partial_a \ell(a,b) = \quad -\exp(a+b) - \exp(a+2b) - \exp(a+4b) + 14.$$

Its derivative with respect to $b$ is

$$\partial_b \ell(a,b) = \quad -\exp(a+b) - 2\exp(a+2b) - 4\exp(a+4b) + 47.$$

Submit You have used 3 of 3 attempts

Answers are displayed within the problem

## (d)

1/1 point (graded)

In order to find the maximum likelihood estimator, we have to solve the nonlinear equation

$$\nabla\ell(a,b) = 0,$$

which in general does not have a closed solution.

Assume that we can reasonably estimate the likelihood estimator using numerical methods, and we obtain

$$\widehat{a} \approx -0.43, \quad \hat{b} \approx 0.69.$$

Given these results, what would be the predicted expected number of outbreaks for $t = 3$? Round your answer to the nearest $0.001$.

5.15517

✔ **Answer:** 5.1551695

**Solution:**

We obtain the expected number of outbreaks as

$$\mathbb{E}[Y_t|t] = \lambda_t,$$

since the expectation of a Poisson random variable is equal to its rate parameter. With this and the relation $\lambda_t = \exp(a + bt)$, we obtain the prediction

$$\lambda_3 = \exp(\widehat{a} + \hat{b}t) \approx 5.1551695.$$

Submit You have used 1 of 3 attempts

Answers are displayed within the problem

## Discussion

Hide Discussion

**Topic:** Unit 7 Generalized Linear Models:Homework 11 / 1. Poisson regression

Add a Post

Show all posts ▼ by recent activity ▼

? [Staff] Help / Request for Review of Calculations 6
I am posting my notes on this problem both as a request for someone to review my calculations and to help individuals still confused get some footing. Please let me know if ...

Python Code 7
For this problem the value of $\hat{a}, \hat{b}$ is given. Here is a Python code for estimation of $\hat{a}, \hat{b}$ : import numpy as np x = np.zeros((2,1)) #initial value of x x[0,0] = 0 x[1,0] = 0 f = np.zer...

? [Staff] Calculating theta 4

? [STAFF] Factorial required 3
Hi, isn't a factorial required for part a)? The system doesn't let me put it as an input.

? [STAFF] Log Likelihood Function in (c) 8
Hi, I calculated the log likelihood function using MLE method and got the gradient of a and b correct, but the log likelihood was incorrect. I checked my calculation and believe...

I screwed a quite dumb point here!!! 1
Just two tips: 1. Pay attention to what happens to $Y!$ in c. 2. You can use $\exp\left(f\infty\right)$ function in the last exercise in this page. Fortunately, I only missed one of them. But almo...

? [Staff] One more attempt? 2
My calculus has been rusty, plus having not enough exercise in the lecture, I kindly request one more attempt in this homework set. Thanks

© All Rights Reserved