Lecture 6. Poisson regression - Random points in space[1]

Igor Rychlik

Chalmers
Department of Mathematical Sciences

Probability, Statistics and Risk, MVE300 • Chalmers • April 2013

[1] Section 7.1.2 is not included in the course.

Counting number of events N:

Data: Suppose we have observed values of $N_1, \ldots, N_k$, which are equal to $n_1, n_2, \ldots, n_k$, say. The first assumption is that N_i are independent Poisson with constant mean m (the same as N has). Suppose that the test for over-disperion leads to rejection of the hypothesis that N_i are iid Poisson. **Overdispertion can be caused by variable mean of N_i or that Poisson model is wrong.** What can we do?

The first step is to assume that N_i are Poisson but have different expectations m_i.

Little help for predicting future unless one can model variability of m_i!

Check Exposure:

When we compared number of perished in traffic 1998 in US (N_{US}) and Sweden (N_{SE}) the numbers were very different (14500, 500, resp.). This was explained by different numbers of peoples leaving in the countries.

In order to compare risks one tries to find a suitable measure of exposure to the hazard and study the rate of accidents[2]. For example in case of traffic a particularly useful measure of exposure is the total number of kilometers driven during a year. In other situations, e.g. in biology, one may count the number of tree species in a forest, and the rate would be the number of species per square kilometer.

We found that 1998 the rate in US was about 1 person per $100 \cdot 10^6$ km driven while in Sweden, 1 per $125 \cdot 10^6$ km. The rates are close but (is the difference significant or could one assume that rates in both countries are the same?)

[2] Rate is a count of events occurring to a particular unit of observation, divided by some measure of that unit's exposure.

Simple case of two counts N_1 and N_2:

Suppose that $N_i \in \mathrm{Po}(m_i)$, $i = 1, 2$. In general $m_1 \neq m_2$. There are two natural simple models for m_i:

- Model I: $m_1 = m_2 = m$
- Model II: $m_1 = \lambda\, t_1$ and $m_2 = \lambda\, t_2$.
- Model III: Any m_1, m_2 are possible.

Here t_1, t_2 are known exposures for the two counts.

Two hypothesis are of interest: does the data contradicts Model I or Model II? If **yes** then and one draws conclusion that m_1 and m_2 need to be estimated separately.

We will present a test quantity, called **Deviance**, which is a difference between values of the log-likelihoods for the compared models. Let first recall the ML-estimation of parameters in Poisson distribution. **All of this will be done on blackboard!**

- Assumptions: $N_{US} \in \mathrm{Po}(m_{US})$, $N_{SE} \in \mathrm{Po}(m_{SE})$.
- Data: $n_{US} = 41500$ while $n_{SE} = 500$. Exposures were $t_{US} = 4.14 \cdot 10^{12}$, $t_{SE} = 0.0625 \cdot 10^{12}$ [km].
- Model I $m_{US} = m_{SE} = m$, $m^* = (41500 + 500)/2 = 21000$ per year?
- Model II, which we call "simpler model", postulates that $m_{US} = \lambda \cdot t_{US}$ and $m_{SE} = \lambda \cdot t_{SE}$.

$$\lambda^* = \frac{n_{US} + n_{SE}}{t_{US} + t_{SE}} = 0.9994 \cdot 10^{-8}, \quad [\mathrm{km}^{-1}].$$

,

- Model III, called the "complex model", we assume that there is no relation between m_{US} and m_{SE}

Numbers of railway accidents

Authorities are interested in the impact of usage of different track types. Data consists of derailments of passenger trains 1 January 1985 – 1 May 1995. There were $n_1 = 15$ derailments on *welded track with concrete sleepers* and $n_2 = 25$ *welded track with wooden sleepers*. Assume that $N_1 \in \mathrm{Po}(m_{con})$ while $N_2 \in \mathrm{Po}(m_{wod})$ are independent.

- The "simple model" is that $m_{con} = m_{wod} = m$ $m^* = \frac{n_{con}+n_{wod}}{2} = 20$,
- The "complex model" is that the means are different, i.e. $m_{con} \neq m_{wod}$ $m^*_{con} = n_{con} = 15$, $m^*_{wod} = n_{wod} = 25$.

$$\boxed{\mathrm{DEV} = 2 \cdot 15(\ln 15 - \ln 20) + 2 \cdot 25(\ln 25 - \ln 20) = 2.53 < \chi^2_{0.05}(1) = 3.84.}$$

The complex model is not better. **Are both truck equally safe?**[3]

[3] **No**, $t_{con} = 4.21 \cdot 10^8$, $t_{wod} = 0.8 \cdot 10^8$, [km], $\mathrm{DEV} = 40.9$, **Explain it and show details on blackboard.**

Parametric modeling - log-likelihood $l(\theta) = \ln(L(\theta))$:

Suppose $N_i \in \mathrm{Po}(m_i)$, $i = 1, \ldots, k$, are independent and that one has observed $N_i = n_i$. In general m_i can take different values.

Schema how to define log-likelihood function:

$$l(\theta)\ : \theta \longmapsto (m_1, \ldots, m_k) \longmapsto l(m_1, \ldots, m_k),^4$$

where θ is a parameter (or vector of parameters). Functions $m_i(\theta)$ are models of the expected value variability.

Examples:

- 1) $\theta = m, \quad m_i(\theta) = m;$
- 2) $\theta = (\beta_0, \beta_1), \quad m_i(\theta) = \exp(\beta_0 + \beta_1 \cdot i)$
- 3) $\theta = (m_1, \ldots, m_k), \quad m_i(\theta) = m_i.$

---

[4] $l(m_1, \ldots, m_k) = \sum n_i \ln m_i - \sum m_i - \sum \ln n_i!$

Finding ML-estimates θ^* of parameters:

The ML-estimate θ^* of θ is the solution of equation system $\dot{l}(\theta^*) = 0$.
For example:

- 1) $\theta = m, \quad \theta^* = \sum n_i/k.$
- 2) $\theta = (\beta_0, \beta_1)$, β_0^*, β_1^* has to be solved numerically.
- 3) $\theta = (m_1, \ldots, m_k), \quad m_i^* = n_i.$

Which model should one choose?

Intuitively - likelihood corresponds to odds for parameters: if fraction of likelihood $L(\theta_1)/L(\theta_2) >> 1$ it means that we believe much more in the first values of parameters than in the second ones. Now

$$L(\theta_1)/L(\theta_2) \gg 1 \iff \ln(L(\theta_1)/L(\theta_2)) = l(\theta_1) - l(\theta_2) \gg 0$$

Deviance:

Suppose we wish to choose between a simple model (marked $_{\mathbf{s}}$) and more complex one (marked $_{\mathbf{c}}$) (which include simpler model as a special case).

$$\begin{aligned}
\theta_{\mathbf{c}} &\longmapsto (m_1(\theta_{\mathbf{c}}), \ldots, m_k(\theta_{\mathbf{c}})) \longmapsto l(\theta_{\mathbf{c}}), \\
\theta_{\mathbf{s}} &\longmapsto (m_1(\theta_{\mathbf{s}}), \ldots, m_k(\theta_{\mathbf{s}})) \longmapsto l(\theta_{\mathbf{s}}),
\end{aligned}$$

The maximum likelihood of the parameters are $\theta^*_{\mathbf{s}}$ and $\theta^*_{\mathbf{c}}$ and the values of log likelihoods

$$\begin{aligned}
l(\theta^*_{\mathbf{c}}) &= \sum n_i \ln m_i(\theta^*_{\mathbf{c}}) - \sum n_i - \sum \ln n_i!. \\
l(\theta^*_{\mathbf{s}}) &= \sum n_i \ln m_i(\theta^*_{\mathbf{s}}) - \sum n_i - \sum \ln n_i!
\end{aligned}$$

$$\boxed{\text{If } \mathrm{DEV} = 2\,(l(\theta^*_{\mathbf{c}}) - l(\theta^*_{\mathbf{s}})) = 2\sum n_i\,(\ln m_i(\theta^*_{\mathbf{c}}) - \ln m_i(\theta^*_{\mathbf{s}})) > \chi^2_{\alpha}(f)}$$

then more complex model explains data better than the simpler one does (with approximative significance $1-\alpha$). Here f is the difference between the dimensions of $\theta_{\mathbf{c}}$ and $\theta_{\mathbf{s}}$. In our example simple model has dimension 1 while the complex dimension k hence $f = k - 1$.

A general Poisson process

Let $N(B)$ denote the number of events (or accidents) occurring in a region B. Consider the following list of assumptions:

(A) More than one event can not happen simultaneously.

(B) $N(B_1)$ is independent of $N(B_2)$ if B_1 and B_2 are disjoint.

(C) Events happen in a stationary (time) and homogeneous (space) way, i.e. $N(B)$ cdf depends only on size $|B|$ of B.

The process for which we can motivate that (A–B) are true is called a Poisson point process. It is a stationary (homogenous) process with constant intensity λ if (A–C) holds and $N(B) \in \mathrm{Po}(\lambda|B|)$.

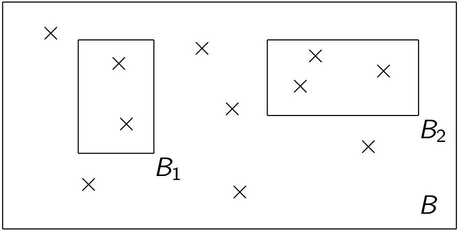


Figure : $N(B) = 11$ while $N(B_1) = 2$, $N(B_2) = 3$.

Some actual history - Bombing raids on London

During the bombing raids on London in World War II, one discussed whether the impacts tended to cluster or if the spatial distribution could be considered random. This was not merely a question of academic interest; one was interested in whether the bombs really targeted (as claimed by Germans) or fell at random[5].

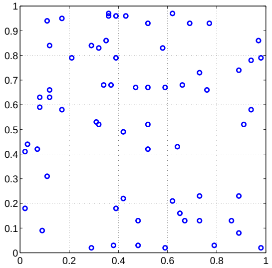
An area in the south of London was divided into 576 small areas of $1/4$ km^2 each; the Poisson distribution was found to be a good model.

---

[5]This problem has even influenced literary texts, as the following excerpt from Pynchon's *Gravity's Rainbow*,

> *Roger has tried to explain to her the V-bomb statistics; the difference between distribution, in angel's-eye view, over the map of England, and their own chances, as seen from down there. She's almost got it, nearly understands his Poisson equation... "...Couldn't there be an equation for us too,..."... "...There is no way, love, not as long as the mean density of the strikes is constant ..."*

Checking for the constant intensity:



The region is 5.7×5.7 m^2, and refered to as one area unit (au). There are 65 pines in the region, i.e. $\lambda^* = 65$ au^{-1}. Divide the region in 25 smaller squares, each of size $0.2 \cdot 0.2 = 0.04$ au^1. It is expected in average $0.04 \cdot 65 = 2.6$ trees in each of smaller regions, since homogeneity of trees locations. The true number differs from the average and their variability is modelled as 25 independent Po(2.6) rv.

It was found 1, 5, 4, 11, 2, 1, 1 regions containing 0, 1, 2, 3, 4, 5, 6 pines. The probability-mass function for Po(2.6) is $p_k = 2.6^k \exp(-2.6)/k!$ and hence one expects to have $25 \cdot p_i$ smaller regions to contain k plants. Since $Q = 8.1 < \chi^2_{0.05}(7-1-1) = 11.07$, the hypothesis about Poisson distribution cannot be rejected.[6]

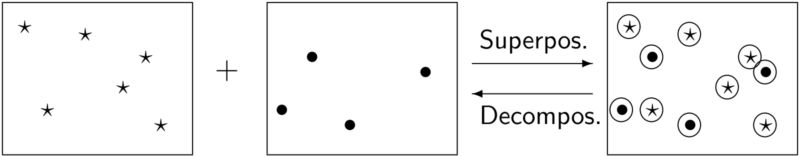
[6] $Q = (1-25p_0)^2/25p_0 + (5-25p_1)^2/25p_1 + \ldots + (1-25p_6)^2/25p_6 = 8.1.$

Decomposition of Poisson process:

Suppose A that is true at point S_i which form PPP with intensity λ.

Let B be a scenario which can be true or false when A occurs. At each point S_i we put a mark "star" if B is true. All remaining S_i (when B is false) are marked by dots ; see Figure.

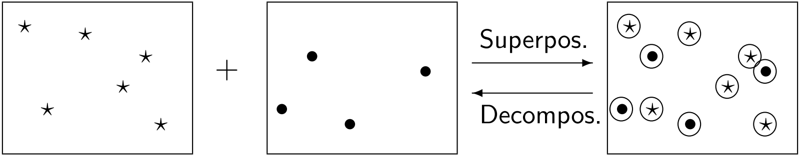
If B is independent of the PPP A, then the **point processes of stars and dots are independent Poisson** having intensities $\mathsf{P}(B)\lambda$, $(1 - \mathsf{P}(B))\lambda$, respectively.



Superposition of Poisson process:

It is not surprising that the reverse operation of superposition of two (or more) independent Poisson processes gives a Poisson process.

Assume that we have two independent Poisson point-processes S_i^I and S_i^{II} with intensities λ^I, λ^{II}, respectively. Consider a point process S_i which is a union of the point processes S_i^I and S_i^{II}. (If S_i^I, S_i^{II} are marked by stars , dots, respectively, replace all symbols with a ring ($\circ$) and let S_i be positions of rings.) The point process of S_i is a superposition of the two processes and is a PPP itself, with intensity $\lambda = \lambda^I + \lambda^{II}$.



In this lecture we met following concepts:

- Deviance, used to select model that best fits data.
- Poisson regression (will be used on Computer lab to model traffic accidents).
- Poisson point process in space.