Testing for homogeneity of a Poisson process

Under the direction of Assistant Professor Mrs.Albert, INSA Toulouse

Jonas Meyran - Cassandra Mussard - Lucas Zuliani
5ModIA - GMM department

ENSEEIHT - INSA TOULOUSE

December 2023

Contents

List of Figures

List of Tables

1 Introduction

Between 1980 and 1990, devastating forest fires had a lasting impact on Denmark, leading to terrible economic and environmental consequences. With global warming, the increase in forest fires has become a harsh reality. For insurance companies, predicting these fires has become crucial. By envisioning the likelihood of fires, they can determine appropriate coverage levels and set premiums that reflect the potential risks.

This complex challenge is mathematically expressed through a Poisson process, where a key concern is determining if the process is homogeneous or inhomogeneous. A homogeneous Poisson process is a process with a constant intensity, i.e. events occur at a constant rate over time, while a non-homogeneous process accommodates varying intensities over time.

We aim to understand the timing of forest fires: whether their frequency remains the same over time or changes throughout the year. Firstly, we revisit the basic concepts of Poisson processes, explaining the difference between consistent and varied events. Secondly, we explore the theoretical construction of three statistical tests: Laplace (Ascher (1978)), Weibull (Crow (1974))) and Barlow (Barlow (1972)). Finally, we perform numerical simulations, an essential step in confirming their effectiveness on simulated data.

Finally, we apply these tests to tangible data from Denmark using the R "evir" package. Our goal is to determine if the occurrence of forest fires between 1980 and 1990 remained consistently steady or showed a noticeable tendency to change.

2 Basic concepts of Poisson processes

2.1 Context

To better understand the concepts used throughout this report, we define the most important notions in this section. A Poisson process models random points in time and space. In our case, we want to model the occurence of forest fires (the random points) between the years 1980 and 1990.
To define a Poisson process, we must define a counting process and a point process.

Definition 2.1 (Couting process)**.** *Let N_t (with $t \in \mathbb{R}$) a counting process. It represents the total number of events that occur by time t.*

Definition 2.2 (Point process)**.** *A point process represents the occurrence of events in continuous time, often denoting the times at which events happen. T_i denotes the occurence time of the ith event.*

$$0 \leq T_1 \leq ... \leq T_n$$

There are two types of Poisson processes: homogeneous Poisson processes and inhomogeneous Poisson processes.

Definition 2.3 (Homogeneous Poisson process)**.** *A Poisson process is homogeneous if the intensity $\lambda > 0$ is constant and it satisfies the following properties*

- $N_0 = 0$
- *It has stationary increments (the distribution of the number of events occurring within a specific interval depends only on its length)*
- *It has independent increments (the number of events occurring in non-overlapping time intervals are statistically independent)*
- $\forall t > 0$ *and* $\forall s > 0$,
$$N_{s+t} - N_t \sim \mathcal{P}(\lambda t)$$

Definition 2.4 (Inhomogeneous Poisson process)**.** *A Poisson process is inhomogeneous if the intensity $\lambda \in \mathbf{R}$ varies over time and it satisfies the following properties*

- $N_0 = 0$
- *Independent increments.*
- $\forall t > 0$,
$$N_t \sim \mathcal{P}(\int_0^t \lambda(u)\, du)$$

Now that we have introduced the fundamental concepts of homogeneous and inhomogeneous Poisson processes, we can conduct statistical tests to determine whether the data adheres to a homogeneous or inhomogeneous Poisson process.

3 Construction of various tests

We compare three tests for testing the hypothesis of a constant intensity (H_0) against the alternative hypothesis of an increasing function (H_1) in an inhomogeneous Poisson process.

3.1 Laplace test

We consider the Laplace Poisson process with intensity given by $\lambda(t) = \exp(\beta t)$ for $\beta \geq 0$.
We aim to test

$$\boxed{H_0 : \beta = 0 \quad \text{against} \quad H_1 : \beta > 0}$$

By setting $\beta = 0$ we are in the case of a homogeneous Poisson process and vice versa if $\beta > 0$.
Let $(T_1, \ldots, T_n)$ be the first arrival times of a homogeneous Poisson process and fix T^* and $n \in \mathbb{N}^*$.

$$(T_1, \ldots, T_n) \,|\, \{N = n\} = (U_{(1)}, \ldots, U_{(n)}) \tag{1}$$

where $U_1, \ldots, U_n$ are i.i.d. random variables drawn from the uniform distribution $U([0, T^*])$.

We deduce

$$\left(\frac{T_1}{T^*}, \ldots, \frac{T_n}{T^*}\right) \,|\, \{N = n\} = (V_{(1)}, \ldots, V_{(n)})$$

where $V_1, \ldots, V_n$ are i.i.d. random variables drawn from the uniform distribution $U([0, 1])$. We define

$$\boxed{L := \textstyle\sum_{i=1}^{n} (\frac{T_i}{T^*})}$$

So L is distributed as the sum of n uniform random variables.

In the case of a homogeneous Poisson process, the T_is are uniformly distributed on $[0, T^*]$, whereas in the inhomogeneous case, the T_is tend to concentrate at the end of the interval. We reject H_0 for large values of L.

Thus, the rejection region is defined as

$$R_\alpha = \{L \geq s_\alpha\}$$

where s_α is defined such that

$$P(L \geq s_\alpha) = \alpha$$

By the Central Limit Theorem (CLT)-type result

$$\left(\frac{T_1}{T^*}, \ldots, \frac{T_n}{T^*}\right)$$

is a sequence of i.i.d. random variables having a moment of order 2.

We have

$$E[\tfrac{T_1}{T^*}] = \tfrac{1}{2} \text{ and } Var(\tfrac{T_1}{T^*}) = \tfrac{1}{12}$$

By the CLT-type result

$$\sqrt{n}\left(\frac{\frac{L}{n} - \frac{1}{2}}{\sqrt{\frac{1}{12}}}\right) \xrightarrow[n\to+\infty]{\text{d}} \mathcal{N}(0,1)$$

This is the pivotal statistic.

We have

$$P_{H_0}(L \geq s_\alpha) = P_{H_0}\left(\sqrt{12n}\left(\frac{L}{n} - \frac{1}{2}\right) \geq \sqrt{12n}\left(\frac{s_\alpha}{n} - \frac{1}{2}\right)\right) = \alpha \xrightarrow[n\to+\infty]{P} (Z \geq z_{1-\alpha}) = \alpha,$$

where Z follows asymptotically a $\mathcal{N}(0,1)$.

We deduce that

$$\sqrt{12n}\left(\frac{s_\alpha}{n} - \frac{1}{2}\right) = z_{1-\alpha}$$

$$\frac{z_{1-\alpha}}{\sqrt{12n}} + \frac{1}{2} = \frac{s_\alpha}{n}$$

$$\sqrt{\frac{n}{12}} \cdot z_{1-\alpha} + \frac{n}{2} = s_\alpha$$

We deduce the following rejection region

$$\boxed{\mathrm{R}_\alpha = \{L \geq \sqrt{\tfrac{n}{12}} \cdot z_{1-\alpha} + \tfrac{n}{2}\}}$$

3.2 Weibull test

Now, we consider the Weibull Poisson Process with intensity given by $\lambda(t) = \beta \cdot t^{\beta-1}$ for $\beta > 0$.
Our aim is to test

$$\boxed{H_0 : \beta = 1 \text{ against } H_1\text{: } \beta > 1}$$

In the case where $\beta = 1$, the intensity is constant and we have a homogeneous Poisson process and vice versa if $\beta > 1$. We fix T^* and $n \in \mathbb{N}^*$.
Let $(T_1, \ldots, T_n)$ be the first arrival times.
We have as in (1)

$$\left(\frac{T_1}{T^*}, \ldots, \frac{T_n}{T^*}\right) \mid \{N = n\} = (V_{(1)}, \ldots, V_{(n)})$$

where $V_1, \ldots, V_n$ are i.i.d. random variables drawn from the uniform distribution $U([0, 1])$.

For $i \in [1, n]$, $\frac{Ti}{T^*} \sim \mathcal{U}([0, 1])$.
We denote $Z = log(\frac{T_i}{T^*}) = -log(\frac{T^*}{T_i})$.
Show that $Z \sim \mathcal{E}(1)$

$$\begin{aligned}
p(Z \leq z) &= p\left(-\log\left(\frac{T_i}{T^*}\right) \leq z\right) \\
&= p\left(\log\left(\frac{T_i}{T^*}\right) \geq -z\right) \\
&= p\left(\frac{T_i}{T^*} \geq \exp(-z)\right) \\
&= 1 - p\left(\frac{T_i}{T^*} \leq \exp(-z)\right) \\
&= 1 - \exp(-z)
\end{aligned}$$

We recognize the cumulative distribution function of an exponential distribution with the parameter $\lambda = 1$.

Then, we have $Z = -log(\frac{T^*}{T_i}) = log(\frac{T_i}{T^*}) \sim \mathcal{E}(1)$. For any $i \in [1, n]$, $(log(\frac{T_1}{T^*}), ..., log(\frac{T_i}{T^*}))$ are i.i.d. random variables drawn from the exponential distribution with $\lambda = 1$.
So, under H_0, for any $n \in \mathbb{N}^*$,

$$\begin{aligned}
\sum_{i=1}^{n} \log\left(\frac{T^*}{T_i}\right) &\sim \Gamma(n, 1) \\
2\sum_{i=1}^{n} \log\left(\frac{T^*}{T_i}\right) &\underset{H_0}{\sim} \Gamma(n, \frac{1}{2})
\end{aligned}$$

$$\boxed{Z = 2\sum_{i=1}^{n} \log\left(\frac{T^*}{T_i}\right) \sim \chi^2(2n)}$$

H_0 is rejected for small values of Z for the same reason as the first test (here we have $\frac{T^*}{T_i}$ instead of $\frac{T_i}{T^*}$).

The rejection zone is then defined as follows $R_\alpha = \{Z \leq s_\alpha\} = \alpha$, with s_α the α-quantile of a $\chi^2(2n)$.

3.3 Barlow Test

The test statistic is given by the following formula:

$$\boxed{F = \frac{(n-d)T_d}{d(T_n - T_d)}}$$

We recall that, under the H_1 hypothesis, we consider an inhomogeneous Poisson process. We start by analysing F under this assumption. In this case, $T_n - T_d$ is small and T_d is large, resulting in high value for F. So, under H_1, F will tend to take larger values than under H_0, then we will tend to reject H_0 when F is large enough. With these observations, we deduce the form of the reject zone $R_\alpha = \{F \geq s_\alpha\}$.

Under H_0, the law of the test statistic F is given by a Snedecor distribution with $2d$ and $2(n-d)$ degrees of freedom. $F \underset{H_0}{\sim} \mathcal{F}_{(2d, 2(n-d))}$.

Demonstration:

Under H_0, we can rewrite as $\frac{\sum_{j=1}^{d} Y_j}{\sum_{j=d+1}^{n} Y_j}$ where $Y_j \overset{iid}{\sim} \mathcal{E}(\lambda)$ Then we have the following results

$$\sum_{j=1}^{d} Y_j \sim \Gamma(d, \lambda)$$
$$2\lambda \sum_{j=1}^{d} Y_j \sim \Gamma(d, \frac{1}{2}) \overset{(d)}{=} \chi^2(2d)$$

In the same way, we obtain that $2\lambda \sum_{j=d+1}^{n} Y_j \sim \Gamma(n-d, \frac{1}{2}) \overset{(d)}{=} \chi^2(2(n-d))$.

Then, we calculate the ratio of these two variables divided by their degree of freedom to reveal the Snedecor distribution:

$$\begin{aligned}\frac{(2\lambda \sum_{j=1}^{d} Y_j)/2d}{(2\lambda \sum_{j=d+1}^{n} Y_j)/(2(n-d))} &= \frac{\sum_{j=1}^{d} Y_j/d}{\sum_{j=d+1}^{n} Y_j/(n-d)} \\ &= \frac{\sum_{j=1}^{d} Y_j}{\sum_{j=d+1}^{n} Y_j} \quad \text{because } d = \frac{n}{2}\end{aligned}$$

4 Power comparisons

In order to confirm the theoretical results, Bain (1985) computes the power of each test.

The power of a test is the probability that the test correctly rejects the null hypothesis (H_0) when a specific alternative hypothesis (H_1) is true. We reject H_0 if the p-value is less than or equal to the confidence level α.

The p-value is the highest level α allowing the acceptance of H_0 in view of the data.

4.1 Laplace test

We have shown previously that $\sqrt{12n}\left(\frac{L}{n}-\frac{1}{2}\right) \xrightarrow[n\to+\infty]{\mathrm{d}} \mathcal{N}(0,1)$

Let $Z_{\text{obs}} = \sqrt{12n}\left(\frac{L_{\text{obs}}}{n}-\frac{1}{2}\right)$ with $L_{\text{obs}} = \sum_{i=1}^{n}\left(\frac{t_i}{T^*}\right)$.

We deduce the following expression for the p-value.

$$\hat{\alpha} = P_{H_0}(Z \geq Z_{\text{obs}}) \approx 1 - F_{\mathcal{N}(0,1)}(Z_{\text{obs}})$$

where Z asymptotically follows $\mathcal{N}(0,1)$.

4.2 Weibull test

We have shown previously that $2\sum_{i=1}^{n}\log\left(\frac{T^*}{T_i}\right) \sim \chi^2(2n)$

We deduce the following expression for the p-value:

$$\hat{\alpha} = P_{H_0}(Z \leq Z^{\text{obs}}) \approx F_{\chi^2(2n)}(Z^{\text{obs}})$$

where $Z \sim \chi^2(2n)$.

4.3 Barlow test

We have shown previously that $F = \frac{(n-d)T_d}{d(T_n-T_d)} \underset{H_0}{\sim} \mathcal{F}_{(2d,2(n-d))}$

We deduce the following expression for the p-value:

$$\hat{\alpha} = P_{H_0}(F \geq F^{\text{obs}}) \approx 1 - \mathcal{F}_{(2d,2(n-d))}(F^{\text{obs}})$$

Table 1: Theoretical simulation on Laplace and Weibull test for different values of T^* and n with the exponential intensity

Test with exponential intensity			
T^*	1	2	4
n	1.71	6.38	53.59
L (Laplace)	0.2467	0.3877	1
Z (Weibull)	0.2643	0.3703	1

Table 2: Theoretical simulation on Laplace and Weibull test for different values of T^* and n with the logarithmic intensity

Test with logarithmic intensity			
T^*	10	15	25
n	15.37	28.36	58.71
L (Laplace)	0.4061	0.5478	0.7543
Z (Weibull)	0.4791	0.674	0.8812

Table 3: Theoretical simulation on Laplace and Weibull test for different values of θ and n with the Weibull intensity

Test with Weibull intensity			
θ	1	2	4
n	2	4	16
L (Laplace)	0.175	0.2912	0.9991
Z (Weibull)	0.177	0.3048	0.9992

The goal of this part is to identify if there is a test which is more powerfull than the other ones in order to apply it on our dataset. The previous results show that Laplace and Weibull tests are equivalent.

5 Application using the Danish dataset

In the previous section, we simulated our own data. In this section, we consider the Danish dataset of R evir package. This dataset represents the costs (in millions of Danish Crowns) due to forest fires between 1980 and 1990. Our goal is to determine if there was a constant trend in the occurrence of those forest fires or not. It can be modeled by a Poisson process because it is a rare and random event. Furthermore, one forest fire generally does not trigger another, which validates the assumption of stationary increments. Ultimately, our goal is to determine whether the stationary hypothesis is confirmed or not, by examining whether the occurrence of forest fires remains constant or varies over time. To proceed, we must use the tests mentioned earlier. However, since we lack knowledge about the intensity function λ representing the average frequency of forest fires, we are unable to favor one test over the others. It is necessary to evaluate all three tests in this context.

Firstly, we establish T_i as the arrival time of a forest fire at the ith instance. Constructing T_i within the dataset involves normalizing the data within the range of 0 to 11. For example, a value between 0 and 1 means that the corresponding event occurred in 1980.

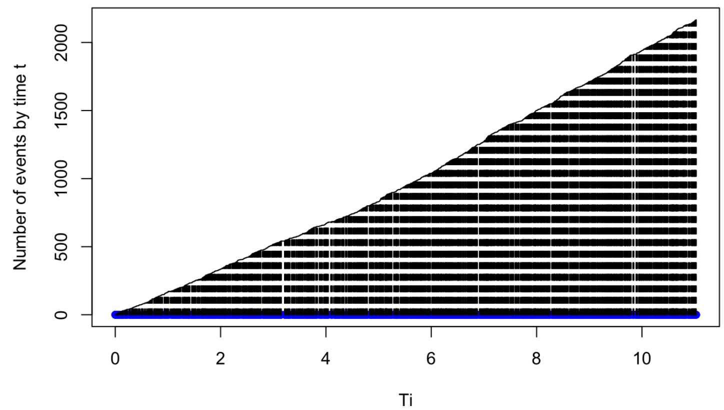

Figure 1: Linear variation of the number of forest fire from 1980-01-03 to 1990-12-31

Figure 1 depicts the temporal progression of the event count. As illustrated, the progression appears to be linear, and it is anticipated that the data will conform to a homogeneous Poisson process.

Secondly, we apply Laplace, Weibull and Barlow test to these data. The p-value for each test was computed, with the number of forest fires (n) set to 2197 and T^* representing the time period over which we observe the Poisson process set to $11 + \frac{3}{365}$ (such that $0 < T_1 < T_2 < .. < T_n < T$ and because there were 3 leap years).

Table 4: Estimated p-value of Laplace, Weibull and Barlow tests on the Danish dataset with $T^* = 11 + \frac{3}{365}$, $n = 2197$

	Laplace test	Weibull test	Barlow test
p-value	$9.35 * 10^{-9}$	$4.30 * 10^{-7}$	$3.97 * 10^{-11}$

By setting α to 0.05, with alpha representing the probability threshold below which we reject H_0, we observe in Table 4 that the p-value is lower than α. This implies that we reject H_0 at a 5% risk level. In summary, contrary to the appearance of a linear trend in Figure 1 suggesting an increase in the number of forest fires, the data arises from a Poisson process that is inhomogeneous. We deduce that the occurrence of forest fires varied between 1980 and 1990.

6 Conclusion

In conclusion, our objective was to investigate whether the occurrence of forest fires was constant or varied between 1980 and 1990.

To explore this, we employed three statistical tests, namely Laplace, Weibull and Barlow. The computed p-value for the Laplace test was 9.35×10^{-9}, while the Weibull test generated a p-value of 4.30×10^{-7}. Finally, we observed a p-value of $3.97 * 10^{-11}$ for the Barlow test. Since the p-values from those three tests are below the significance level of $\alpha = 0.05$, we reject the hypothesis that the occurrence of forest fires remained constant between 1980 and 1990 at a 5% risk level.

One of the main limitations is that we only studied the occurence of forest fires between 1980 and 1990. It might be more informative to consider a longer timeframe, especially including recent years, to understand the impact of global warming. Additionally, our focus on Denmark is another limitation. Exploring forest fire occurrences across Europe would be beneficial due to the diverse climates present in different regions.

7 References

References

Ascher, H. E., and Feingold, H. (1978), "Application of Laplace's Test to Repairable System Reliability", in *Proceedings of the International Conference on Reliability and Maintainability* (Societe Pour la Diffusion des Sciences et des Arts, France), 219-225.

Crow, L. H. (1974), "Reliability and Analysis for Complex Repairable Systems," in *Reliability and Biometry, eds. F. Proschan and R. J. Serfling, Philadelphia*: Society for Industrial and Applied Mathematics, pp. 379-410.

Bain, L. J., Engelhardt, M. and Wright F. T. (1985), "Tests for an Increasing Trend in the Intensity of a Poisson Process: A Power Study," in *Journal of the American Statistical Association, Vol. 80, No. 390,* pp. 419-422

Barlow, R. E., Bartholomew, D. J., Bremner, J. M., and Brunk, H. D. (1972), Statistical Inferences Under Order Restrictions, New York: John Wiley.