Detection of delays at the registration of events of a point process

Streit Franz
Section de Mathématiques de l'Université de Genève
Case Postale 64
1207 Genève, Switzerland
streit@unige.ch

1. Introduction

Due to technical particularities of the recording device it happens frequently that a point process undergoes delayed registration of some or all of its point events. In this contribution we propose statistical tests which may be used to discover modifications of the incoming point process due to such delays. The inferential method to be applied in this context depends on the knowledge of the statistician about the original process input and about the delay mechanism and on the type of data collected. Which technique to implement may for instance be influenced by the choices made among the following contrasting possibilities of interpretation of the task:

- The modification of the incoming point process concerns only some individual point events (local modification) or all its point events (global modification).
- The test to be elaborated is desired to be either locally optimal or else most powerful (this being made possible by working in this case with a prior distribution for the unknown degree of distortion of the incoming point process).
- The realization of the recorded point process is observed during the fixed time period $(0, t]$ or during the random time span up to the registration of the n_0th point event.

The point process is started at time $t = 0$ and we designate by $T(i)$ the time of registration of the ith point event after $t = 0$. The random variables are represented by capital letters and their realized values by the corresponding lower case letters. By $N(t)$ we denote the number of recorded point events in $(0, t]$ and by $G[h] = \exp[\int_0^\infty \log(h(x))dN(x)]$ (defined for $0 \leq h \leq 1$ and $h = 1$ outside a bounded subset of $(0, \infty)$) the generating functional of a point process. $R_{(0,b]}$ stands for the uniform distribution with support $(0, b]\,[b > 0], \varepsilon_\delta$ for the one-point distribution at δ and $N^*(a, g)$ for the normal distribution with mean a and variance g.
In the following typical variants of the problem are treated:

2. Most powerful tests for a fixed period of observation

We choose for the incoming point process a homogeneous Poisson point process with global intensity rate μ. We suppose that any individual point event arriving at $x \in (0, t]$ will be delayed with the same probability δ for a time of length τ which follows the absolutely continuous distribution function F (with $F(0) = 0$); alternatively it is recorded at arrival time with probability $1 - \delta$. In view of the difficulty to specify the exact value of the probability δ, we assume that δ follows a given a priori distribution function F_Δ concentrated on $[0, 1]$. Different choices of F_Δ represent different experimental situations. We consider in particular four cases:

- Case a.) $F_\Delta(\delta) = 2\delta\,[0 < \delta < 1/2]$.
  The point events are more likely to be registered in time than to be delayed.

- Case b.) $F_\Delta(\delta) = \delta\,[0 < \delta \leq 1]$.
  No conceivable positive values of the probability δ are excluded or in any way prefered.
- Case c.) $F_\Delta(\delta) = 0\,[\delta < \delta_0]$ and $F_\Delta(\delta) = 1\,[\delta \geq \delta_0]\,[0 < \delta_0 < 1]$
  Each point event is individually delayed with probability δ_0.
- Case d.) $F_\Delta(\delta) = 0\,[\delta < 1]$ and $F_\Delta(\delta) = 1\,[\delta \geq 1]$.
  Each point event is delayed.

While cases a.) - c.) describe local modifications of the incoming point process, case d.) treats a global distortion. It is expected that similar procedures as developed in this section may also be helpful in queueing theory, for instance to check if the recorded point events represent the arrival times or rather the departure times after service for a queue with many servers.
The probability generating functional of the incoming point process observed during $(0,t]$ is
$G_{\varepsilon_0}[h] = \exp[-\int_0^t (1-h(x))\mu dx]$
(see (1),p.225,(7.4.10)). The probability generating functional of the process obtained by retarding all the events of the incoming point process by independent individually F-distributed random delays takes the form
$G_{\varepsilon_1}[h] = \exp[-\int_0^t \int_x^t (1-h(y))dF(y-x)\mu dx] = \exp[-\int_0^t (1-h(y))F(y)\mu dy]$
(see (1), p.241,(8.2.11)). Of course a point event arriving at time $x \in (0,t]$ and subject to a delay τ will not be recorded during the observation period when $x+\tau > t$.
The probability generating functional of the recorded observation in $(0,t]$ when submitting Poisson point events with F_Δ-distributed probability δ to F-distributed delays is thus
$G_{F_\Delta}[h] = \exp[-\int_0^1 \{(1-\delta)\int_0^t (1-h(x))\mu dx + \delta \int_0^t (1-h(y))F(y)\mu dy\} dF_\Delta(\delta)]$
$= \exp[-(1-\int_0^1 \delta dF_\Delta(\delta))\int_0^t (1-h(x))\mu dx - \int_0^1 \delta dF_\Delta(\delta)\int_0^t (1-h(y))F(y)\mu dy]$. Therefore the recorded point process is an (inhomogeneous) Poisson point process with intensity function
$\mu_d(y) = (1-\int_0^1 \delta dF_\Delta(\delta))\mu + (\int_0^1 \delta dF_\Delta(\delta))\mu F(y)$ on $y \in (0,t]$.
Lets derive a most powerful test for the hypotheses $H_0 : G = G_{\varepsilon_0}$ and $H_1 : G = G_{F_\Delta}$. The likelihood function for a typical observation $(n; t_1, \ldots, t_n)$ is given by
$L(G_{F_\Delta} : n; t_1, \ldots, t_n) = \mu^n \prod_{i=1}^n \{(1-\int_0^1 \delta dF_\Delta(\delta)) + \int_0^1 \delta dF_\Delta(\delta)F(t_i)\} \exp[-\mu(1-\int_0^1 \delta dF_\Delta(\delta))t - \mu\int_0^1 \delta dF_\Delta(\delta)\int_0^t F(\tau)d\tau]$.
A statistic yielding a most powerful test of H_0 versus H_1 is thus given by
$\Lambda_{F_\Delta,F} = L(G_{F_\Delta} : N(t); T(1), \ldots, T(N(t)))/L(G_{\varepsilon_0} : N(t); T(1), \ldots, T(N(t)))$
$= \prod_{i=1}^{N(t)} \{(1-\int_0^1 \delta dF_\Delta(\delta)(1-F(T(i)))\} \exp[\mu \int_0^1 \delta dF_\Delta(\delta)(t - \int_0^t F(\tau)d\tau)]$.
An equivalent test statistic is
$\widetilde{\Lambda}_{F_\Delta,F} = \sum_{i=1}^{N(t)} \ln[1-\int_0^1 \delta dF_\Delta(\delta)(1-F(T(i)))]$.
Taking into account that under H_0 $(T(1), \ldots, T(N(t))|N(t) = n$ is the order statistic of a random sample of size n from $T \sim R_{(0,t]}$ we find for the mean and variance:
$E[\widetilde{\Lambda}_{F_\Delta,F} : H_0] = E[N(t)\int_0^t \ln[1-(\int_0^1 \delta dF_\Delta(\delta))(1-F(\tau))]d\tau/t : H_0]$
$= \mu\int_0^t \ln[1-(\int_0^1 \delta dF_\Delta(\delta))(1-F(\tau))]d\tau$ and
$Var[\widetilde{\Lambda}_{F_\Delta,F}; H_0] =$
$E[N(t)Var[\ln[1-\int_0^1 \delta dF_\Delta(\delta)(1-F(T))] : H_0]] + Var[N(t)E[\ln(1-\int_0^1 \delta dF_\Delta(\delta)(1-F(T))] : H_0]] =$
$\mu\int_0^t \ln^2[1-(\int_0^1 \delta dF_\Delta(\delta))(1-F(\tau))]d\tau$.
An application of the central limit theorem leads for $t \to \infty$ and under H_0 to
$\widetilde{\Lambda}_{F_\Delta,F} \sim$
$N^*(\mu\int_0^t \ln[1-(\int_0^1 \delta dF_\Delta(\delta))(1-F(\tau))]d\tau, \mu\int_0^t \ln^2[1-(\int_0^1 \delta dF_\Delta(\delta))(1-F(\tau))]d\tau)$.
The critical region of a test of size α for large t is thus given by
$[\mu\int_0^t \ln^2[1-\int_0^1 \delta dF_\Delta(\delta)(1-F(\tau))]d\tau]^{-1/2}(\tilde{\lambda}_{F_\Delta,F} - \mu\int_0^t \ln(1-\int_0^1 \delta dF_\Delta(\delta)(1-F(\tau))d\tau) > z(1-\alpha)$,
where $\Phi(z(1-\alpha)) = 1-\alpha$ and Φ denotes the distribution function of the standard normal distribution.

Particular cases:
Statistics to be used:
Case a.): $\widetilde{\Lambda}_{R_{(0,1/2)},F} = \sum_{i=1}^{N(t)} \ln(3/4 + F(T(i))/4)$
Case b.): $\widetilde{\Lambda}_{R_{(0,1]},F} = \sum_{i=1}^{N(t)} \ln(1/2 + F(T(i))/2)$
Case c.): $\widetilde{\Lambda}_{\varepsilon_{\delta_0},F} = \sum_{i=1}^{N(t)} \ln(1 - \delta_0 + \delta_0 F(T(i)))$
Case d.): $\widetilde{\Lambda}_{\varepsilon_1,F} = \sum_{i=1}^{N(t)} \ln(F(T(i)))$.
Using the equation $\ln(3/4+F(T(i))/4) = \ln(3/4)+\ln(1+F(T(i))/3)$ in case a.) and similar relations in cases b.) and c.) we obtain for the mean value and the variance of $\widetilde{\Lambda}$ in case a.)
$E[\widetilde{\Lambda}_{R_{(0,1/2)},F} : H_0] = \mu(t\ln(3/4) + \int_0^t \ln(1 + F(u)/3)du)$ and
$Var[\widetilde{\Lambda}_{R_{(0,1/2)},F} : H_0] = \mu(t\ln^2(3/4) + 2\ln(3/4)\int_0^t \ln(1 + F(u)/3)du + \int_0^t \ln^2(1 + F(u)/3)du)$ and analogous results in the cases b.) –d.). The tests statistics and the critical regions for the case a.) (resp. for the case b.)) are the same as in case c.) with $\delta_0 = 1/4$ (resp. $= 1/2$).
When choosing $F = R_{(0,t]}$ (uniformly distributed delays) we find with $F(\tau) = \tau/t\,[0 \le \tau \le t]$:

Case	$E[\widetilde{\Lambda} : H_0]$	$Var[\widetilde{\Lambda} : H_0]$
a.)	$(3\ln(4/3) - 1)\mu t = -0,137\mu t$	$(-3\ln^2(4/3) - 6\ln(4/3) + 2)\mu t = 0,026\mu t$.
b.)	$(\ln(2) - 1)\mu t = -0,307\mu t$	$(-\ln^2(2) - 2\ln(2) + 2)\mu t = 0,133\mu t$
c.)	$((1 - 1/\delta_0)\ln(1 - \delta_0) - 1)\mu t$	$(1 - 1/\delta_0)\ln^2(1 - \delta_0) - 2(1 - 1/\delta_0)\ln(1 - \delta_0) + 2)\mu t$
d.)	$-\mu t$	$2\mu t$.

3. Locally most powerful tests for point processes observed up to the occurrence of the n_0th recorded event

Suppose now that the incoming point process is globally modified by delays with probability δ and assume that the gap lengths between successive point events are independent and identically distributed with density function f_T if no delay occurs and with density function f_{T+D} in the presence of delay. The likelihood of the typical observation $(t_1, \ldots, t_{n_0})$ is of the mixed ordinary renewal point process type specified by
$L(\delta : t_1, \ldots, t_{n_0}) =$
$(1 - \delta)f_T(t_1)f_T(t_2 - t_1)\cdots f_T(t_{n_0} - t_{n_0-1}) + \delta f_{T+D}(t_1)f_{T+D}(t_2 - t_1)\cdots f_{T+D}(t_{n_0} - t_{n_0-1})$.
Testing $H_0 : \delta = 0$ versus $H_1 : \delta > 0$ for small δ leads to the score statistic given by
$T = [f_T(U(1))\ldots f_T(U(n_0))]^{-1}[f_{T+D}(U(1))\ldots f_{T+D}(U(n_0))] - 1$
with $U(\nu) := T(\nu) - T(\nu - 1)$ and $P(T(0) = 0) = 1$ which is equivalent to
$\tilde{T} = \sum_{i=1}^{n_0} \ln[f_{T+D}(U(i))/f_T(U(i))]$.
Since the $U(i)$'s and also the additive terms of $\tilde{T}$ are independent identically distributed, $\tilde{T}$ has for $n_0 \to \infty$ according to the standard central limit theorem an asymptotical normal distibution provided that the mean and variance are finite and the variance is positive. In particular we get under H_0 that
$E[\tilde{T} : \delta = 0] = n_0 \int_0^\infty \ln[f_{T+D}(u)/f_T(u)]f_T(u)du$
and
$Var[\tilde{T} : \delta = 0] = n_0(\int_0^\infty \ln^2[f_{T+D}(u)/f_T(u)]f_T(u)du - E[\tilde{T} : \delta = 0]^2)$.
Particular case:
If the incoming point process is again homogeneous Poisson with global rate
μ and if the gap length density function under H_1 is $f_{T+D}(u) = \mu^2 u e^{-\mu u}\,[u > 0]$ we obtain
$E[\tilde{T} : \delta = 0] = n_0 \int_0^\infty \ln(z)e^{-z}dz = -Cn_0 = -0,557n_0$ and
$Var[\tilde{T} : \delta = 0] = n_0(\int_0^\infty \ln^2(z)e^{-z}dz - C^2) = \pi^2 n_0/6 = 1,645n_0$
(see (2), pp. 573, 574, 1080; C designates Euler's constant). The critical region of a locally most powerful test of size α for large n_0 can thus be taken as
$(\sum_{i=1}^{n_0} \ln(\mu U(i)) + 0,577n_0)/\sqrt{1,645n_0} > z(1 - \alpha)$.
On the other hand, if the incoming point process is only locally modified by delaying point events

individually with probability δ, then we work with the score statistic
$T^* = \sum_{i=1}^{n_0}([f_{T+D}(U(i))/f_T(U(i)] - 1)$
It is easy to proof that
$E[T^* : \delta = 0] = 0$ and
$Var[T^* : \delta = 0] = n_0 \int_0^\infty \{(f_{T+D}(u) - f_T(u))^2/f_T(u)\}du.$
In the above described special case we get in particular
$T^* = \sum_{i=1}^{n_0}(\mu U(i) - 1).$
This statistic follows asymptotically for $n_0 \to \infty$ and H_0 the $N^*(0, n_0)$-distribution, which permits to set up in the usual way a test for large n_0. However also an exact locally most powerful test is available due to the fact that the equivalent test statistic $\tilde{T}^* = \sum_{i=1}^{n_0}(\mu U(i))$ is Gamma-(and $\chi^2/2$)-distributed under H_0 with density function $f_{\tilde{T}^*}(u) = \Gamma^{-1}(n_0)u^{n_0-1}e^{-u}\ [u > 0]$.

REFERENCES

(1) Daley,D.J. & Vere-Jones,D. An introduction to the theory of point processes.Springer, New York, 1988.
(2) Gradshteyn,I.S. & Ryzhik,I.M. Table of integrals, series and products. Academic Press, New York, 1965.

ABSTRACT

When observing the realization of a point process the question naturally arises whether this output represents a homogeneous Poisson point process or a more complicated stochastic model. Very often technical imperfections of the recording device lead to delayed registration of some or all of its points events. It is the aim of this contribution to develop statistical tests helpful to discover the presence of such delay mechanisms. Most powerful and locally most powerful tests are proposed which are useful in these regards, in particular if the period of observation of the point process is long. The fact that in many experimental set-ups the observed process of possibly delayed point events of a Poisson point process turns out to be again a Poisson point process facilitates the elaboration of such tests.

RÉSUMÉ

En observant la réalisation d'un processus ponctuel on est amené à se poser la question si ces données sont engendrées par un processus homogène de Poisson ou par un modèle stochastique plus compliqué. Assez souvent des inperfections techniques de l'instrument d'enregistrement a comme effet de saisir quelques-uns ou tous les événements ponctuels trop tard. Le but de cette contribution est de proposer des tests statistiques utiles à découvrir la présence d'un tel méchanisme de retardement. Des tests localement les plus puissants et des tests les plus puissants sont indiqués pour traiter ce problème; ils sont particulièrement performants quand la période d'observation du processus ponctuel est longue. Le fait que dans bien des situations expérimentales le processus des événements potentiellement retardés d'un processus ponctuel de Poisson est de nouveau un processus de Poisson facilite l'élaboration de tels tests.