Counting processes and recurrent events beyond the cox model for Poisson process

Vincent Couallier

Institut Mathématique de Bordeaux UMR CNRS 5251, Université de Bordeaux.

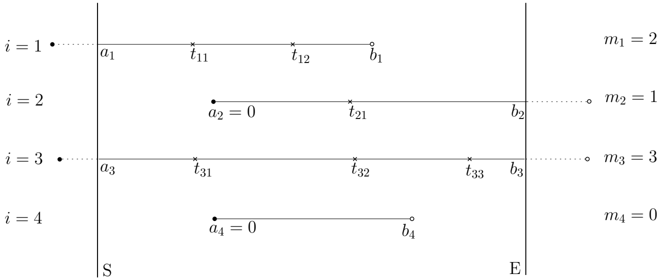
Dynamic predictions for repeated markers and repeated events
Workshop - GS0 2013

Outline

a joint work with
Karim Claudio (PhD LYRE, Suez), Genia Babykina (Post Doc,CRAN), Yves Legat (IRSTEA)

Counting processes and recurrent events

Type of data


$i = 1$
a_1
t_{11}
t_{12}
b_1
$m_1 = 2$
$i = 2$
$a_2 = 0$
t_{21}
b_2
$m_2 = 1$
$i = 3$
a_3
t_{31}
t_{32}
t_{33}
b_3
$m_3 = 3$
$i = 4$
$a_4 = 0$
b_4
$m_4 = 0$
S
E

Main Aim of analysis of recurrent events

- statistical analysis (and modeling) of non-independant occurence times of an event.
  - more than one event per individual.
  - interest in understanding the dependancy between times
- heterogenity and risk factors (covariates observed on individuals)
  - identify the covariates that influence the probability of event.
  - individual prediction of probability of occurrence knowing the characteristics of the individual
- Heterogeneity and frailties
  - does the intensity of events differ from individual to individual because of covariates or past history ?
  - dynamic intensity modeling versus frailty intensity modeling.

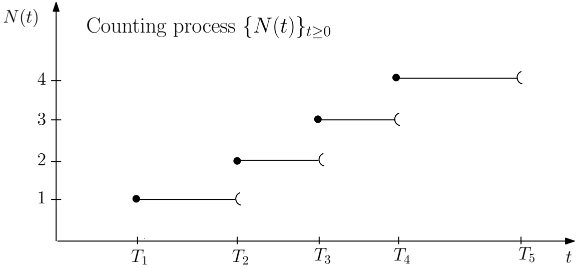
Main Aim of analysis of recurrent events

- statistical analysis (and modeling) of non-independant occurence times of an event.
  - more than one event per individual.
  - interest in understanding the dependancy between times
- heterogenity and risk factors (covariates observed on individuals)
  - identify the covariates that influence the probability of event.
  - individual prediction of probability of occurrence knowing the characteristics of the individual
- Heterogeneity and frailties
  - does the intensity of events differ from individual to individual because of covariates or past history ?
  - dynamic intensity modeling versus frailty intensity modeling.

Main Aim of analysis of recurrent events

- statistical analysis (and modeling) of non-independant occurence times of an event.
  - more than one event per individual.
  - interest in understanding the dependancy between times
- heterogenity and risk factors (covariates observed on individuals)
  - identify the covariates that influence the probability of event.
  - individual prediction of probability of occurrence knowing the characteristics of the individual
- Heterogeneity and frailties
  - does the intensity of events differ from individual to individual because of covariates or past history ?
  - dynamic intensity modeling versus frailty intensity modeling.

The counting process for recurrent events



Definition of the (stochastic) intensity

with respect to a history $\mathcal{H}(t)_{t>0}$(filtration generated by the known history) :

$$\begin{aligned} \lambda(t) &= \lim_{dt\to 0} \frac{1}{dt}\mathbb{P}\left[N(t+dt)-N(t)=1\,|\,\mathcal{H}(t-)\right] \\ \lambda(t)dt &= \mathbb{E}\left[dN(t)\,|\,\mathcal{H}(t-)\right], \end{aligned}$$

The marginal intensity : the rate function

- $\mathcal{H}(t)$ contains jump times and "external covariates" $Z(t)$,
  $\mathcal{H}(t) = \sigma\left(N(t), T_1, ..., T_{N(t)}, Z(t)\right)$

Note : $\lambda(t)$ is a predictable process, it captures the nature of the recurrence of events.

The rate function (ROCOF in reliability analysis)

$$
\begin{aligned}
r(t) &= \lim_{dt \to 0} \frac{1}{dt} \mathbb{P}\left[N(t+dt) - N(t) = 1 \,|\, Z(t-)\right] \\
r(t)dt &= \mathbb{E}\left[dN(t) \,|\, \mathcal{Z}(t-)\right],
\end{aligned}
$$

Ex : (Chiang, 1968)

$$
\begin{aligned}
\lambda(t) &= \beta_0(t) + \beta_1(t)Z + \beta_2(t)N(t-) \\
r(t) &= \beta_0(t) + \phi(\beta_1, \beta_2)Z + \psi(\beta_0, \beta_2)\beta_2(t),
\end{aligned}
$$

Recall the Cox model for survival analysis

- One event per subject → Survival analysis : $\lambda(t) = h(t)I_{N(t)=0}$

$$h(t) = \lim_{dt \to 0} \frac{1}{dt} P\left(T \in [t, t+dt[| T > t\right)$$

- regression model for covariate → Multiplicative intensity model

$$\lambda(t) = \lambda_0(t) e^{\beta_0 + \beta_1 Z_1 + \cdots + \beta_p Z_p} 1_{N(t-)=0}$$

Remark :

- A dead individual is no more "at risk"
- censoring mechanism may be considered **too** → "At risk" process $Y(t)$.

from Cox regression for lifetimes to Andersen and Gill intensity of counting processes

Remove The ”at risk” indicator $Y(t)$ to remains ”at risk” after the event.

Andersen & Gill, 82

$$\lambda(t)dt = \mathbb{E}\left[dN(t) \,|\, Z_1, \ldots Z_p\right] = \lambda_0(t)e^{\beta_0+\beta_1 Z_1+\cdots+\beta_p Z_p}\, dt$$

- The points of $N(t)_{t\geq 0}$ form a Poisson process conditionally on the values $Z_1, \ldots Z_p$.
- the process intensity does no depend on the past $\rightarrow$ not *dynamic*.

Models for the intensity process in reliability analysis

Some well known models for the intensity process

- dynamic intensity : impact of occurrence of an event on the intensity
  N(t-) or event times themselves ?
- observed "inter-unit" heterogenity : impact of covariates
  multiplicative intensity with "Cox-like" contribution.
- unobserved "inter-unit" heterogeneity : frailty : ? ?
  it may be hard to handle both frailty and dynamic parts.

Rk : in reliability (in this talk) :

- an event = failure + instantaneous maintenance/repair
- the dynamic part explains the maintenance actions.

The Poisson process with covariates

Cox regression model for recurrent events :

$$\lambda(t) = \lambda_0(t)e^{\beta' Z}, \text{ The Z's may be time-dependent}$$

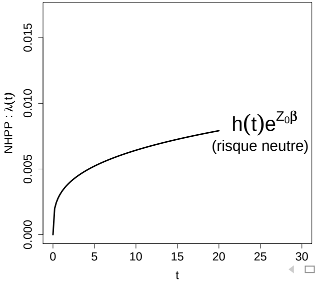
- λ_0 usually increasing, ($\Rightarrow$ ageing).
- $e^{Z'\beta}$: impact of environment and/or individual heterogeneity

The Poisson process with covariates

Cox regression model for recurrent events :

$$\lambda(t) = \lambda_0(t)e^{\beta' Z}, \text{ The Z's may be time-dependent}$$

- λ_0 usually increasing, ($\Rightarrow$ ageing).
- $e^{Z'\beta}$: impact of environment and/or individual heterogeneity

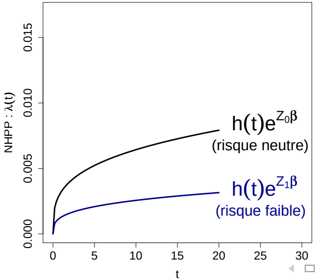


The Poisson process with covariates

Cox regression model for recurrent events :

$$\lambda(t) = \lambda_0(t)e^{\beta' Z}, \text{ The Z's may be time-dependent}$$

- λ_0 usually increasing, ($\Rightarrow$ ageing).
- $e^{Z'\beta}$: impact of environment and/or individual heterogeneity

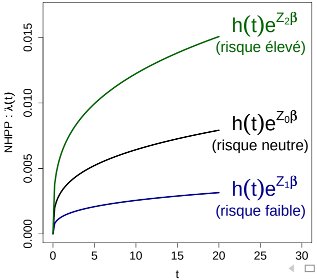


The Poisson process with covariates

Cox regression model for recurrent events :

$$\lambda(t) = \lambda_0(t)e^{\beta' Z}, \text{ The Z's may be time-dependent}$$

- λ_0 usually increasing, ($\Rightarrow$ ageing).
- $e^{Z'\beta}$: impact of environment and/or individual heterogeneity

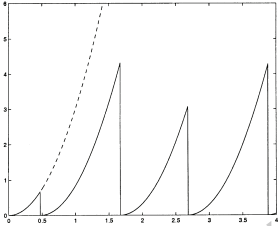


A renewal process for recurrent events

RP process (perfect repairs, AGAN models) :

$$\lambda(t) = \lambda_0(t - T_{N(t)})e^{\beta' Z},$$

- λ_0 usually increasing, $\lambda_0(0) = 0$ ($\Rightarrow$ ageing).
- the intensity uses the *duration* since the last event $t - T_{N(t)}$.
- inter-arrival durations are i.i.d. random variables
- *dynamic* intensity.



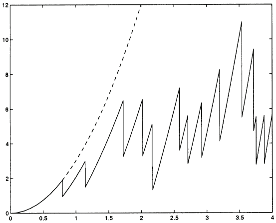
A generalized renewal process for recurrent events

GRP process (imperfect repair, virtual age models) :

$$\lambda(t) = \left(\lambda_0(t) - \rho\lambda_0\left(T_{N(t-)}\right)\right) e^{\beta' Z} \text{ ARI1}$$

$$\lambda(t) = \left(\lambda_0\left(t - \rho T_{N(t-)}\right)\right) e^{\beta' Z} \text{ ARA1}$$

- λ_0 usually increasing, $\lambda_0(0) = 0$ ($\Rightarrow$ ageing).
- ρ captures the *dynamic* of the reduction of intensity

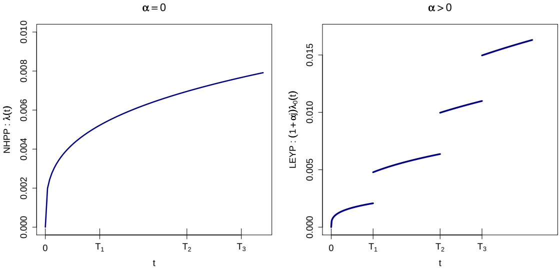


The linear extended Yule process for recurrent events

LEYP process (imperfect repair) :

$$\lambda(t) = (1 + \alpha N(t-))\lambda_0(t)e^{Z(t)'\beta}$$

- a dynamical component ($\alpha > 0$) uses the **number** of previous events.
- a baseline intensity λ_0 (usually parameterized, $\lambda_0(.,\theta)$).
- a Cox-like regression part $e^{Z(t)'\beta}$.



Generalization to Pena & Hollander model

The Pena-Hollander model

$$\lambda(t) = U\rho(N(t-), \alpha)\lambda_0(\epsilon(t))\psi\left(e^{Z(t)'\beta}\right)$$

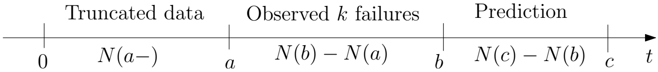
- a frailty component : U (non observable source of heterogeneity).
- a dynamical component $\rho(N(t-), \alpha)$
- a baseline intensity λ_0 (usually parameterized, $\lambda_0(., \theta)$).
- a Cox-like regression part $e^{Z(t)'\beta}$.

Remark : LEYP $\subset$ Pena & Hollander

Some useful properties for dynamic prediction and statistical estimation for the LEYP model.

Marginal and conditional distributions of N(t)

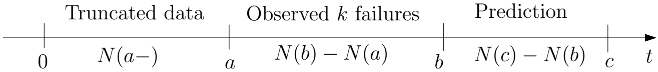
- $\lambda(t) = (1 + \alpha N(t-))\lambda_0(t; Z(t); \delta, \beta)$
- $\Lambda_0(t) = \int_0^t \lambda_0(s; Z(s); \delta, \beta) ds$



$$
\begin{aligned}
N(t) &\sim \mathcal{NB}(\alpha^{-1}, e^{-\alpha\Lambda_0(t)}) \\
\mathbb{E}[N(t)] &= \frac{e^{\alpha\Lambda_0(t)} - 1}{\alpha} \\
\mathbb{V}ar[N(t)] &= \frac{e^{\alpha\Lambda_0(t)}(e^{\alpha\Lambda_0(t)} - 1)}{\alpha}
\end{aligned}
$$

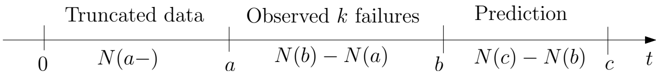
Marginal and conditional distributions of N(t)

- $\lambda(t) = (1 + \alpha N(t-))\lambda_0(t; Z(t); \delta, \beta)$
- $\Lambda_0(t) = \int_0^t \lambda_0(s; Z(s); \delta, \beta) ds$



$$
\begin{aligned}
N(t) &\sim \mathcal{NB}(\alpha^{-1}, e^{-\alpha\Lambda_0(t)}) \\
\mathbb{E}\left[N(t)\right] &= \frac{e^{\alpha\Lambda_0(t)} - 1}{\alpha} \\
\mathbb{V}ar\left[N(t)\right] &= \frac{e^{\alpha\Lambda_0(t)}(e^{\alpha\Lambda_0(t)} - 1)}{\alpha}
\end{aligned}
$$

Marginal and conditional distributions of N(t)



$$[N(b)-N(a)\,|\,N(a-)=k, Z(s), a<s<b] \sim \mathcal{NB}\left(\alpha^{-1}+k, e^{-\alpha[\Lambda_0(b)-\Lambda_0(a)]}\right)$$

$$[N(c)-N(b)\,|\,N(b-)-N(a)=k] \sim \mathcal{NB}\left(\alpha^{-1}+k, \frac{e^{\alpha\Lambda_0(b)}-e^{\alpha\Lambda_0(a)}+1}{e^{\alpha\Lambda_0(c)}-e^{\alpha\Lambda_0(a)}+1}\right)$$

The likelihood

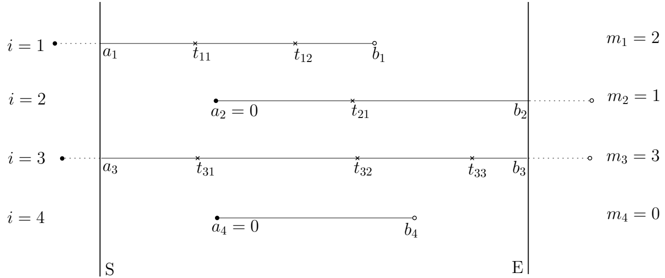
- One individual, observed on $[0, T]$, m events at times t_j ($j = \{1, \ldots m\}$) :

$$L(\theta) = \left(\prod_{j=1}^{m} \lambda(t_j)\right) \times \exp\left(-\sum_{j=0}^{m} \int_{t_j}^{t_{(j+1)}} \lambda(u)\, du\right)$$

Remark : In the following, parametric assumption on the baseline + truncated observation

- $\lambda(t) = (1 + \alpha N(t-))\delta t^{\delta-1} e^{Z(t)'\beta} = (1 + \alpha N(t-))\lambda_0(t; \delta, \beta)$
- N individuals, observed on $[a_i, b_i]$, $i = 1 \ldots N$
- $\Lambda_0(t) = \int_0^t \lambda_0(s) ds$

Data for maximum likelihood estimation



And : $Z_1, \ldots, Z_p$, fixed or external time-dependent covariates (no frailty)

The likelihood

- Log-likelihood for truncated data, N individuals

$$\begin{aligned}\ln L(\theta) = \sum_{i=1}^{N} \Big(& m_i \ln \alpha + \ln \Gamma(\alpha^{-1} + m_i) - \ln \Gamma(\alpha^{-1}) \\ - & (\alpha^{-1} + m_i) \ln(e^{\alpha \Lambda_0(b_i;\delta,\beta)} - e^{\alpha \Lambda_0(a_i;\delta,\beta)} + 1) \\ + & \sum_{j=1}^{m_i} (\ln \lambda_0(t_j;\delta,\beta) + \alpha \Lambda_0(t_j;\delta,\beta)) \Big)\end{aligned}$$

Appl. on recurrent failures of water networks

recurrent events of failure-repair on water network pipes

- SEDIF : A public drinking water service in area of Paris.
- stratification : only grey cast iron pipes are considered.
- 21450 pipes to provide drinking water (899km linear).
- failures recorded on 1996-2006 (11 years).
- mean age at inclusion : 35.8 years.
- mean duration of observation : 10 years.
- $\%$ of units with ≥ 1 failure : 12%.
- $\%$ of units with > 1 failures : 3%.

recurrent events of failure-repair on water network pipes

- SEDIF : A public drinking water service in area of Paris.
- stratification : only grey cast iron pipes are considered.
- 21450 pipes to provide drinking water (899km linear).
- failures recorded on 1996-2006 (11 years).
- mean age at inclusion : 35.8 years.
- mean duration of observation : 10 years.
- % of units with ≥ 1 failure : 12%.
- % of units with > 1 failures : 3%.

recurrent events of failure-repair on water network pipes

- SEDIF : A public drinking water service in area of Paris.
- stratification : only grey cast iron pipes are considered.
- 21450 pipes to provide drinking water (899km linear).
- failures recorded on 1996-2006 (11 years).
- mean age at inclusion : 35.8 years.
- mean duration of observation : 10 years.
- % of units with ≥ 1 failure : 12%.
- % of units with > 1 failures : 3%.

recurrent events of failure-repair on water network pipes

- SEDIF : A public drinking water service in area of Paris.
- stratification : only grey cast iron pipes are considered.
- 21450 pipes to provide drinking water (899km linear).
- failures recorded on 1996-2006 (11 years).
- mean age at inclusion : 35.8 years.
- mean duration of observation : 10 years.
- % of units with ≥ 1 failure : 12%.
- % of units with > 1 failures : 3%.

recurrent events of failure-repair on water network pipes

- SEDIF : A public drinking water service in area of Paris.
- stratification : only grey cast iron pipes are considered.
- 21450 pipes to provide drinking water (899km linear).
- failures recorded on 1996-2006 (11 years).
- mean age at inclusion : 35.8 years.
- mean duration of observation : 10 years.
- % of units with ≥ 1 failure : 12%.
- % of units with > 1 failures : 3%.

recurrent events of failure-repair on water network pipes

- SEDIF : A public drinking water service in area of Paris.
- stratification : only grey cast iron pipes are considered.
- 21450 pipes to provide drinking water (899km linear).
- failures recorded on 1996-2006 (11 years).
- mean age at inclusion : 35.8 years.
- mean duration of observation : 10 years.
- % of units with $\geq$ 1 failure : 12%.
- % of units with $>$ 1 failures : 3%.

recurrent events of failure-repair on water network pipes

- SEDIF : A public drinking water service in area of Paris.
- stratification : only grey cast iron pipes are considered.
- 21450 pipes to provide drinking water (899km linear).
- failures recorded on 1996-2006 (11 years).
- mean age at inclusion : 35.8 years.
- mean duration of observation : 10 years.
- % of units with ≥ 1 failure : 12%.
- % of units with > 1 failures : 3%.

recurrent events of failure-repair on water network pipes

- SEDIF : A public drinking water service in area of Paris.
- stratification : only grey cast iron pipes are considered.
- 21450 pipes to provide drinking water (899km linear).
- failures recorded on 1996-2006 (11 years).
- mean age at inclusion : 35.8 years.
- mean duration of observation : 10 years.
- % of units with $\geq$ 1 failure : 12%.
- % of units with $>$ 1 failures : 3%.

recurrent events of failure-repair on water network pipes

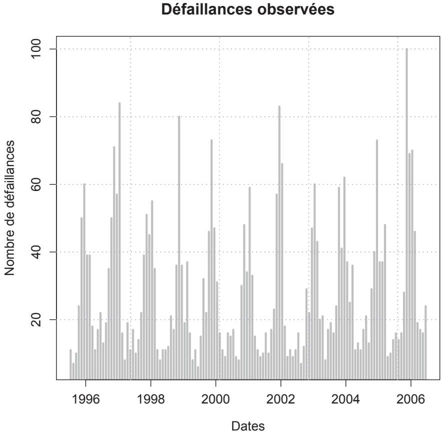
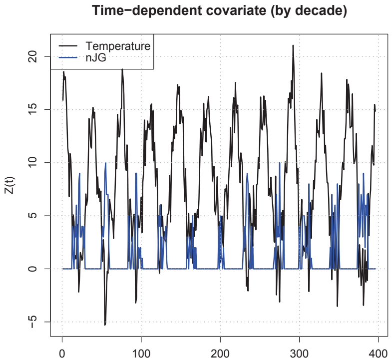


FIGURE : observed monthly number of failures

recurrent events of failure-repair on water network pipes

- the length L of the pipe influences the intensity
- the climate influences the intensity (time dependent)
  $\rightarrow X_2$: air temperature average over 10-days periods.

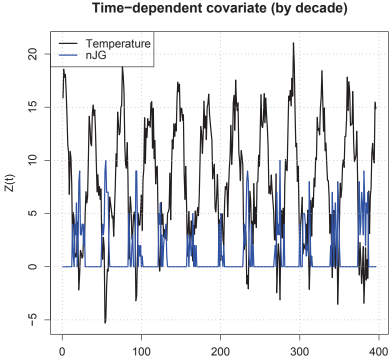
$$\lambda(t) = (1 + \alpha N(t-))\delta t^{\delta-1} e^{\beta_0+\beta_1 \ln(L)+\beta_2 X_2(t)}$$



recurrent events of failure-repair on water network pipes

- the length L of the pipe influences the intensity
- the climate influences the intensity (time dependent)
  $\rightarrow X_2$: air temperature average over 10-days periods.

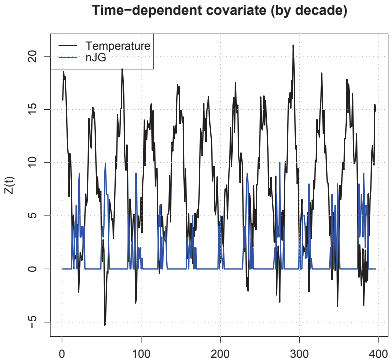
$$\lambda(t) = (1 + \alpha N(t-))\delta t^{\delta-1} e^{\beta_0 + \beta_1 \ln(L) + \beta_2 X_2(t)}$$



recurrent events of failure-repair on water network pipes

- the length L of the pipe influences the intensity
- the climate influences the intensity (time dependent)
  $\rightarrow X_2$: air temperature average over 10-days periods.

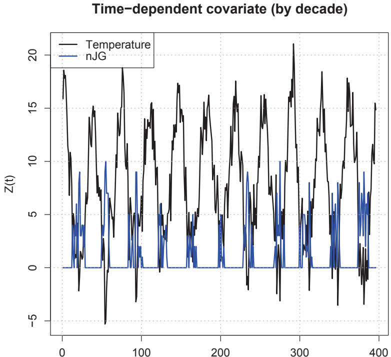
$$\lambda(t) = (1 + \alpha N(t-))\delta t^{\delta-1} e^{\beta_0 + \beta_1 \ln(L) + \beta_2 X_2(t)}$$



recurrent events of failure-repair on water network pipes

- the length L of the pipe influences the intensity
- the climate influences the intensity (time dependent)
  $\rightarrow X_2$: air temperature average over 10-days periods.

$$\lambda(t) = (1 + \alpha N(t-))\delta t^{\delta-1} e^{\beta_0+\beta_1 \ln(L)+\beta_2 X_2(t)}$$



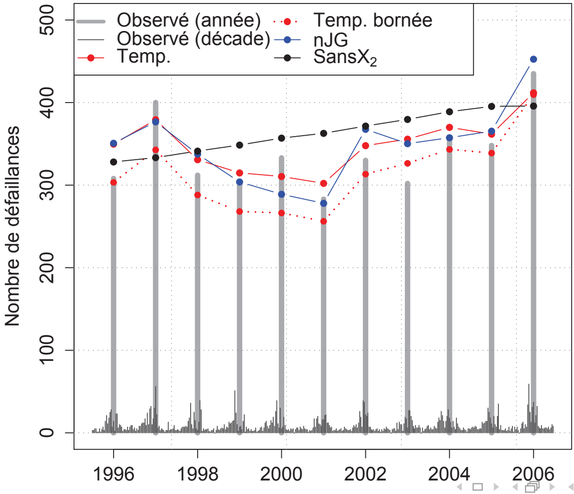
recurrent events of failure-repair on water network pipes

The results (maximum likelihood estimation)

Parameter	Model	Estimate	Standard Dev.	95% CI Estimate ± 1.96 std
α	with X_2	0.96	0.078	$[0.81, 1.12]$
	without X_2	0.98	0.079	$[0.82, 1.13]$
δ	with X_2	1.14	0.094	$[0.96, 1.32]$
	without X_2	1.11	0.094	$[0.93, 1.30]$
β_0	with X_2	-7.03	0.45	$[-7.92, -6.14]$
	without X_2	-7.52	0.45	$[-8.41, -6.63]$
β_1	with X_2	0.67	0.020	$[0.63, 0.71]$
	without X_2	0.65	0.020	$[0.61, 0.69]$
β_2	with X_2	-0.10	0.0.003	$[-0.11, -0.09]$

recurrent events of failure-repair on water network pipes

Global prediction of failures on the water network



recurrent events of failure-repair on water network pipes

individual prediction of failure on [U,V], at time T $(\leq U < V)$?

- at time T, use the adjusted model, the known history of the unit (age, covariate, number of past failures) to compute the Negative Binomial distribution for $N_i(V) - N_i(U)$.
- ranking of the sample by ordering $P[N_i(V) - N_i(U)]_{i=1\ldots n}$.
- use this ranking to provide a *preventive maintenance policy*.
- or do a graph of the Lift Curve to validate the prediction efficiency of the model.

recurrent events of failure-repair on water network pipes

individual prediction of failure on [U,V], at time T $(\leq U < V)$?

- at time T, use the adjusted model, the known history of the unit (age, covariate, number of past failures) to compute the Negative Binomial distribution for $N_i(V) - N_i(U)$.
- ranking of the sample by ordering $P[N_i(V) - N_i(U)]_{i=1\ldots n}$.
- use this ranking to provide a *preventive maintenance policy*.
- or do a graph of the Lift Curve to validate the prediction efficiency of the model.

recurrent events of failure-repair on water network pipes

individual prediction of failure on [U,V], at time T $(\leq U < V)$?

- at time T, use the adjusted model, the known history of the unit (age, covariate, number of past failures) to compute the Negative Binomial distribution for $N_i(V) - N_i(U)$.
- ranking of the sample by ordering $P[N_i(V) - N_i(U)]_{i=1\ldots n}$.
- use this ranking to provide a *preventive maintenance policy*.
- or do a graph of the Lift Curve to validate the prediction efficiency of the model.

recurrent events of failure-repair on water network pipes

individual prediction of failure on [U,V], at time T $(\leq U < V)$?

- at time T, use the adjusted model, the known history of the unit (age, covariate, number of past failures) to compute the Negative Binomial distribution for $N_i(V) - N_i(U)$.
- ranking of the sample by ordering $P[N_i(V) - N_i(U)]_{i=1\ldots n}$.
- use this ranking to provide a *preventive maintenance policy*.
- or do a graph of the Lift Curve to validate the prediction efficiency of the model.

recurrent events of failure-repair on water network pipes

individual prediction of failure on [U,V], at time T $(\leq U < V)$?

- at time T, use the adjusted model, the known history of the unit (age, covariate, number of past failures) to compute the Negative Binomial distribution for $N_i(V) - N_i(U)$.
- ranking of the sample by ordering $P[N_i(V) - N_i(U)]_{i=1\ldots n}$.
- use this ranking to provide a *preventive maintenance policy*.
- or do a graph of the Lift Curve to validate the prediction efficiency of the model.

Conclusion - Bilan

Conclusion

- many models to handle dynamic intensity of recurrent events.
- the LEYP is one of them, with useful properties
- semiparametric framework has not been investigated (yet)
- recurrent events problems : epidemiology / biostatistics / reliability analysis : bridges exist.
- dynamic versus frailty models : a risk to badly identify dynamic components (in fact due to frailty component) : The negative Binomial distribution is also the marginal for Gamma mixed Poisson processes (Poisson with Gamma frailty)

Conclusion

- many models to handle dynamic intensity of recurrent events.
- the LEYP is one of them, with useful properties
- semiparametric framework has not been investigated (yet)
- recurrent events problems : epidemiology / biostatistics / reliability analysis : bridges exist.
- dynamic versus frailty models : a risk to badly identify dynamic components (in fact due to frailty component) : The negative Binomial distribution is also the marginal for Gamma mixed Poisson processes (Poisson with Gamma frailty)

Conclusion

- many models to handle dynamic intensity of recurrent events.
- the LEYP is one of them, with useful properties
- semiparametric framework has not been investigated (yet)
- recurrent events problems : epidemiology / biostatistics / reliability analysis : bridges exist.
- dynamic versus frailty models : a risk to badly identify dynamic components (in fact due to frailty component) : The negative Binomial distribution is also the marginal for Gamma mixed Poisson processes (Poisson with Gamma frailty)

Conclusion

- many models to handle dynamic intensity of recurrent events.
- the LEYP is one of them, with useful properties
- semiparametric framework has not been investigated (yet)
- recurrent events problems : epidemiology / biostatistics / reliability analysis : bridges exist.
- dynamic versus frailty models : a risk to badly identify dynamic components (in fact due to frailty component) : The negative Binomial distribution is also the marginal for Gamma mixed Poisson processes (Poisson with Gamma frailty)

Conclusion

- many models to handle dynamic intensity of recurrent events.
- the LEYP is one of them, with useful properties
- semiparametric framework has not been investigated (yet)
- recurrent events problems : epidemiology / biostatistics / reliability analysis : bridges exist.
- dynamic versus frailty models : a risk to badly identify dynamic components (in fact due to frailty component) : The negative Binomial distribution is also the marginal for Gamma mixed Poisson processes (Poisson with Gamma frailty)

Conclusion

- many models to handle dynamic intensity of recurrent events.
- the LEYP is one of them, with useful properties
- semiparametric framework has not been investigated (yet)
- recurrent events problems : epidemiology / biostatistics / reliability analysis : bridges exist.
- dynamic versus frailty models : a risk to badly identify dynamic components (in fact due to frailty component) : The negative Binomial distribution is also the marginal for Gamma mixed Poisson processes (Poisson with Gamma frailty)

Thank you for your attention.