Study of electric consumption with data collected from Linky

Kévin Jaunâtre

joint work with
Gilles Durrieu, Ion Grama and Jean-Marie Tricot

Université de Bretagne Sud

Mathematics and Entreprises Days

Plan

SOLENN project

- Directed by ADEME (Agence de l'Environnement et de la Maîtrise de l'Énergie).
- Financed by ENEDIS.
- Taking place in Lorient.
- Gather 12 companies and public entities.
- October 2014 - October 2017.
- Extended until October 2018.

Linky

The SOLENN project was planned as a follow-up of the installation of the electric meter "Linky".

This electric meter is considered a "smart" electric meter as it provides data that were not collected before.

Objectives of the SOLENN project

There are multiple objectives on the SOLENN project, here are the objectives we are working on :

- Follow-up of the electric consumption over the project duration.

Objectives of the SOLENN project

There are multiple objectives on the SOLENN project, here are the objectives we are working on :

- Follow-up of the electric consumption over the project duration.
- Effect of individual visits on the electric consumption.

Objectives of the SOLENN project

There are multiple objectives on the SOLENN project, here are the objectives we are working on :

- Follow-up of the electric consumption over the project duration.
- Effect of individual visits on the electric consumption.
- Test of the electric constraint.

This presentation focus on the last objective.

Data collected

Multiple data have been collected during the project :

- Data from the electric meter.

Data collected

Multiple data have been collected during the project :

- Data from the electric meter.
- Outside temperature from weather stations.

The data are collected on a sample of houses that gave their consent to transfer them.

Plan

I. SOLENN project

II. Electric constraint

III. Model

IV. Application

What is it ?

Everyone with an electric contract subscribe to a maximal electric power. For example, it can be 9 kVA. That means that if the requested electric power exceeds the value of 9 kVA, the breaker cuts off and the house has no electricity.
Putting an electric constraint in the house on the maximal electric power consists of reducing it. In our example, if the house has an electric constraint of 50%, the maximal electric power would be 4.5 kVA.

Why ?

During exceptional events (storms, electric incidents, ...), the only available security on the electric network is the load shedding.
The electric constraint aim to provide an alternative to load shedding.
Example : Instead of having 300 houses with no electricity while 700 have it, applying the electric constraint would mean that the 1000 houses will have electricity to a limited extend.

Electric constraint in SOLENN

In the SOLENN project, the electric constraint on houses is tested with the consent of the habitants. It concerns more than 200 houses with different contracts.

Every winter, 7 to 8 days are chosen (mostly the coldest) for the electric constraint. During these days, we have a period of 2 to 4 hours (for lunch or dinner time) of electric constraint going from 30% to 80%. There is a minimal power to respect to avoid, for example, one house to go from 6 kVA to 1.2 kVA.

A text message is sent one day before the electric constraint in order to inform the inhabitants.

Aim of the approach

- Estimate the probability of exceeding the maximal power during the period of electric constraint.
- Estimate the maximal power during the period of electric constraint such that few houses will be impacted.

Plan

Idea of the model

To estimate the probability of exceeding the maximal power, it would be easy to focus on the associated quantile corresponding to the maximal power. However, one must keep in mind that the data are aggregated every ten minutes. It means that an house can exceed the maximal power between two measurement and open its breaker during the same interval and we will not notice it.

We propose a model which can be viewed as an extension of the regular Cox's proportional hazards model to estimate the probabilities of rare events.

Formulation of the problem

Let be $S(x \mid z) = 1 - P(X \leq x | Z = z)$, $x \geq x_0$, the survival distribution of the random variable X given the covariate vector $z \in [z_{\min}, z_{max}]$, where $x_0 \geq 0$.
We observe the independent duet $\{x_1, z_1\}, \ldots, \{x_n, z_n\}$, where each $\{x_i, z_i\}$ has the survival distribution function $S(x_i \mid z_i)$.

Cox model

The Cox model (see Cox [1]) specifies that the hazard function of the random variable X depends on the value of a covariate vector as follow :

$$h(x \mid z) = \exp(\beta \cdot z) h_0(x), \; x \geq x_0,$$

where $x_0 \geq 0$, β is a vector of parameters and $h_0(x) = h(x \mid 0)$ is an unknown baseline hazard function.
Note : The hazard function $h(x \mid z)$ is related to the density function $f(x \mid z)$ and the survival function $S(x \mid z)$ by

$$h(x \mid z) = f(x \mid z) / S(x \mid z),$$

and

$$S(x \mid z) = e^{-\int_{x_0}^{x} h(t \mid z) dt}.$$

Extreme value theory

Assume that $F(.\,|\,z) = 1 - S(.\,|\,z) \in$ DA of the Fréchet law with parameter $1/\theta_z$:

$$\Phi_{1/\theta_z}(x) = \exp\left(-x^{-1/\theta_z}\right), \quad x \geq 0.$$

The following Theorem can be found in Beirlant & Al. [2].

Théorème (Fisher-Tippet-Gnedenko theorem)

$F(.\,|\,z) \in$ *DA of the Fréchet law with parameter* $1/\theta_z$ **iff** *for any* $x \geq 1$,

$$\frac{S(\tau x\,|\,z)}{S(\tau\,|\,z)} \to x^{-1/\theta_z} \quad \textit{as} \quad \tau \to \infty.$$

F-T-G theorem suggests to approximate $\frac{S(\tau x\,|\,z)}{S(\tau\,|\,z)}$ by a Pareto distribution $P_{\theta_z}(x)$ with estimated θ_z for large τ.

Simplification of the problem

Assuming the previous conditions, the problem can be shifted on the baseline survival distribution $S_0(x) = S(x \mid Z = 0)$ as all the survival distributions are linked by the relation

$$S(x \mid z) = S_0(x)^{e^{\beta \cdot z}}.$$

As a consequence, the following semi-parametric model is considered for the baseline survival function :

$$S_{0,\tau,\theta}(x) = \begin{cases} S_0(x) & \text{if } x \in [0, \tau], \\ S_0(\tau) \left(\frac{x}{\tau}\right)^{-1/\theta} & \text{if } x > \tau, \end{cases}$$

where S_0 is fully non-parametric, τ is a threshold parameter and the parametric part is completely described by the Pareto distribution with parameter θ.

Parameters to estimate

We now aim to provide the estimators necessary to estimate the parameters of the previous model. We suppose the regression parameter β known and estimated by the procedure described in Cox [1]. The threshold is considered fixed in the first part, a selection procedure is presented latter. It leaves us the estimation of the non-parametric part S_0 and the parameter of the Pareto distribution θ.

Estimation of θ

Maximizing the log-likelihood function of the semi-parametric model with respect to θ yields the estimator :

Maximum quasi-likelihood estimator of θ with fixed threshold τ :

$$\widehat{\theta}_\tau = \frac{1}{\widehat{n}_\tau} \sum_{i=1}^{n} e^{\beta \cdot z_i} \ln\left(\frac{x_i}{\tau}\right),$$

where $\widehat{n}_\tau = \sum_{i=1}^{n} \mathbb{1}_{\{x_i > \tau\}}$ is the nb. of obs. beyond τ.

Estimation of S_0

Using the well-known Nelson-Aalen estimator ([3] and [4]) which estimate the cumulative hazard function $H_0(x) = \int_{x_0}^{x} h_0(t)dt$ by :

$$\widehat{H}_0(x) = \sum_{x_i \leq x} \widehat{h}_0(x_i).$$

Maximizing the log-likelihood function with respect to $h_0(x_i)$ yields the estimator :

Maximum quasi-likelihood estimator of $h_0(x_i)$:

$$\widehat{h}_0(x_i) = \frac{1}{\sum_{j=1}^{n} e^{\beta \cdot z_j} \mathbb{1}_{\{x_j \geq x_i\}}}.$$

Estimation of $S_{0,\tau,\theta}$

The estimator of the survival function $S_{0,\tau,\theta}$ is then given by :

$$\widehat{S}_{0,\tau,\widehat{\theta}_\tau}(x) = \begin{cases} \widehat{S}_0(x) & \text{if } x \in [0,\tau], \\ \widehat{S}_0(\tau)\left(\frac{x}{\tau}\right)^{-1/\widehat{\theta}_\tau} & \text{if } x > \tau, \end{cases}$$

where the non-parametric estimator of S_0 is defined by $\widehat{S}_0(x) = e^{-\widehat{H}_0(x)}$

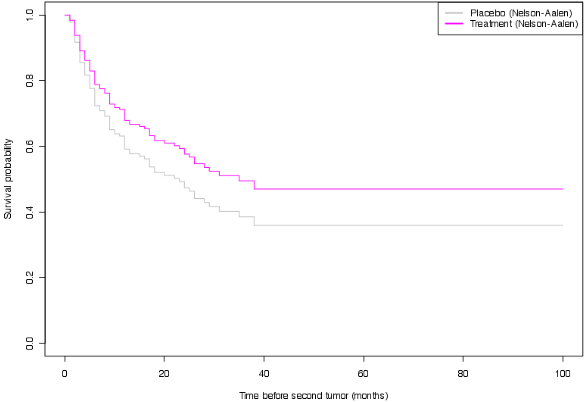
Selection procedure of the threshold τ

We propose a procedure to choose the threshold τ in two steps :

- Propagation step :
  given a sequence of embedded Pareto models, it consists in determining the largest one to fit the data.
- Selection step :
  determine the fitted model by penalized maximum likelihood.

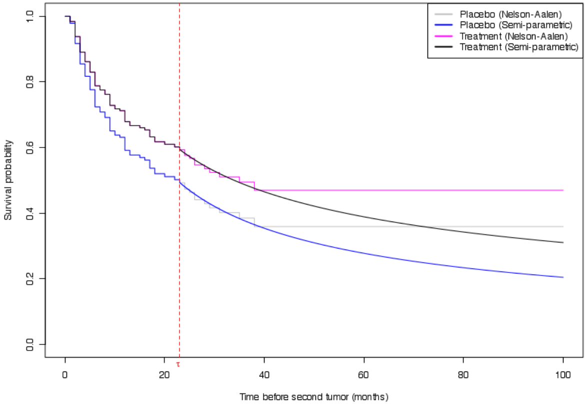
Illustration

Cox model with Nelson-Aalen estimator on the data set : "bladder" ("R" package "survival").



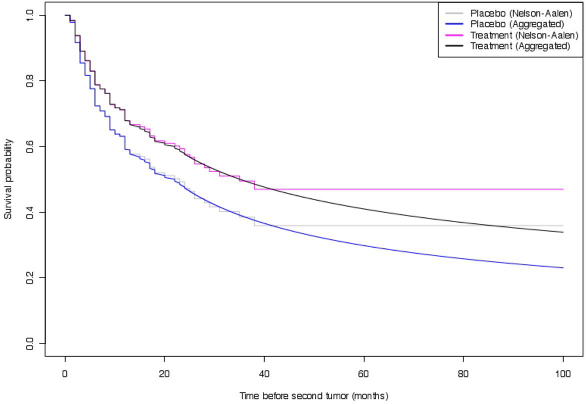
Illustration

Cox model with our approach on the data set : "bladder" ("R" package "survival").



Illustration

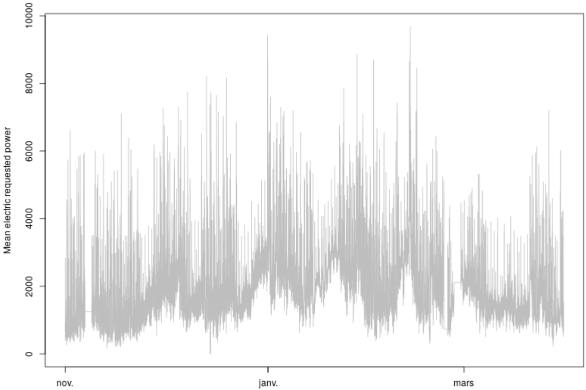
Cox model with an aggregation of our approach on the data set : "bladder" ("R" package "survival").



Plan

Example on one house

The data set is the mean electric power requested every 10 minutes by one house with a contract of 9 kVA during the winter from November 2016 until March 2017 and participating to the electric constraints.

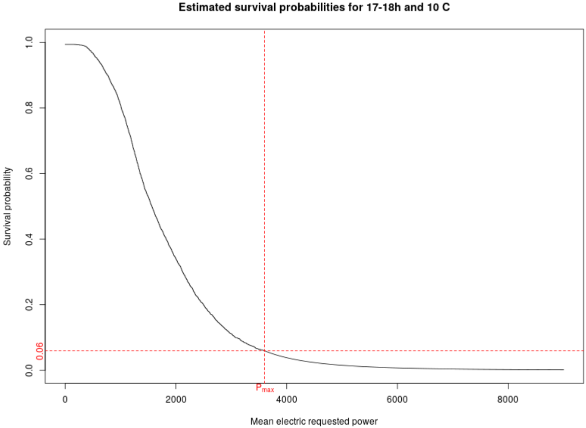


Preparation of the data set

- Creation of the covariate
- Verification of the hazards proportionality.

First electric constraint

The first electric constraint during the winter 2016-2017 was from 17h until 21h the 22nd of November, 2016. The electric constraint was of 60%, the max. power was 3.6 kVA.



Estimated probabilities of exceeding the maximal power

By applying this method for every houses, we observe three behaviours :

- Quick learners : Cutting of the breaker on the first electric constraint period only.
- Slow learners : Cutting of the breaker for almost every electric constraint period.
- Adapted behaviour : No cutting of the breaker during the electric constraint period.

Estimated probabilities of exceeding the maximal power

Example :

TABLE 1 – Estimated probabilities of exceeding the maximal power during the electric constraint period.

House	22 Nov.	14 Dec.	5 Jan.	17 Jan.	1 Feb.	28 Feb.	30 Mar.
House 1	0.547	0.566	0.042	0.761	0.645	0.866	0.701
House 2	0.38	0.827	0.477	0.865	0.5	0.661	0.932
House 3	0.229	0.888	0.003	0.892	0.824	0.767	0.429

Estimated maximal power

We are now interested in the estimation of the maximal power such that a small amount of house will have their breaker cut.
On a four hours constraint, the probability to have one measurement exceeding the maximal power is $1/24 \approx 0.04$ and on a two hours constraint, this probability is $1/12 \approx 0.08$.

TABLE 2 – Average estimated maximal power with probability of having one measurement exceeding it. Houses with a 9 kVA subscription.

	22 Nov.	14 Dec.	5 Jan.	17 Jan.	1 Feb.	28 Feb.	30 Mar.
Estimated P_{max} (kVA)	5.287	4.572	4.725	4.112	3.656	4.290	3.133
Estimated reduction	41%	49%	48%	54%	59%	52%	65%

Conclusion

- The model presented allows to estimate the probability of exceeding the maximal power during the electric constraint. This estimation is necessary to classify the house into categories which depends on its behaviour during the constraint.
- It is interesting to focus on the house perspective and to adjust the inconvenience by estimating the maximal power which may reduce the number of cut breakers.

Thank you for your attention !

D. R. Cox.
Regression models and life-tables.
Journal of the Royal Statistical Society. Series B (Methodological), 34(2) :187–220, 1972.

J. Beirlant, Y. Goegebeur, J. Teugels, and J. Segers.
Statistics of Extremes : Theory and Applications.
John Wiley & Sons, 2004.

W. Nelson.
Theory and applications of hazard plotting for censored failure data.
Technometrics, 14(4) :945–966, 1972.

O. Aalen.
Nonparametric inference for a family of counting processes.
The Annals of Statistics, pages 701–726, 1978.