Closed-form Maximum Likelihood Estimator for Generalized Linear Models in the case of categorical explanatory variables: Application to insurance loss modeling

–Séminaire de probabilités et statistique du LJK–

Tom Rohmer

in collaboration with:
A. Brouste, V. Désert (Le Mans Univ.)
& C. Dutang (Univ. Paris-Dauphine)

May 2019

Pricing insurance policies

Given an assurance policy, the total claim amount (over a year) is

$$X_{\boldsymbol{\alpha}} = \sum_{k=1}^{M_{\boldsymbol{\alpha}}} Y_{\boldsymbol{\alpha},k}$$

where

- $\boldsymbol{\alpha}$ an individual set of characteristics : warranty type, risk class (for example age, vehicle for motor insurance,...), company size
- $M_{\boldsymbol{\alpha}}$ number of annual claims
- $Y_{\boldsymbol{\alpha},k}$ their amounts.

Pure premium : $\mathbb{E}(X_{\boldsymbol{\alpha}})$

Descriptive statistics

The dataset comes from a private insurer :

- for privacy reason, amounts have been randomly scaled, dates randomly rearranged, variable modalities renamed.
- insurance for private corporates operating in France.
- it consists of 211,739 claims which occurred between 2000 and 2010.

TABLE – Empirical quantiles and moments (in euros) by warranty

	W1	W2	W3	W4	W5	W6	W7
Min.	0	0	0	0	0	0	0
1st Qu.	123	155	235	128	2	1	2
Median	1,253	814	1,955	893	2,977	2	564
3rd Qu.	4,994	2,664	8,246	3,726	39,647	1,560	2,097
Max.	3,882,524	4,529,249	15,315,173	14,272,522	15,688,300	4,888,656	4,670,686
Mean	7,022	4,055	28,429	32,328	110,056	8,388	7,157
Std dev.	39,581	24,620	280,500	273,958	534,337	74,969	60,916
Skewness	49	85	38	22	16	42	35
Kurtosis	3,955	13,620	1,833	738	366	2,399	1,927

Probabilistic model of pricing

Generally claims are decomposed on attritional and atypical part :

$$X_{\boldsymbol{\alpha}} = \sum_{k=1}^{M_{\boldsymbol{\alpha}}^{(1)}} Y_{\boldsymbol{\alpha},k}^{(1)} + \sum_{k=1}^{M_{\boldsymbol{\alpha}}^{(2)}} Y_{\boldsymbol{\alpha},k}^{(2)}$$

where

- $M_{\boldsymbol{\alpha}}^{(1)}$ (resp. $M_{\boldsymbol{\alpha}}^{(2)}$) number of annual attritional (resp. atypical, i.e. greater than a threshold μ) claims for the vector of characteristic $\boldsymbol{\alpha}$.
- $Y_{\boldsymbol{\alpha},k}^{(1)}$ (resp. $Y_{\boldsymbol{\alpha},k}^{(2)}$) their amounts.

Hypothesis :

1. For $i = 1, 2$, $\left(Y_{\boldsymbol{\alpha},k}^{(i)}\right)_k$ are i.i.d. and independent with $M_{\boldsymbol{\alpha}}^{(i)}$
2. $\left(Y_{\boldsymbol{\alpha},k}^{(i)}\right)_{\boldsymbol{\alpha}}$ do not have same distribution !

Premium principles

For a class α, the pure premium is

$$\pi_0^1 = \mathbb{E}(X_{\alpha}) = \sum_{i=1}^{2} \mathbb{E}(M_{\alpha}^{(i)})\mathbb{E}(Y_{\alpha,1}^{(i)})$$

Two considerations for the commercial premium :

1. Expectation principle : $\pi_{\rho_{\alpha}}^1 = (1+\rho_{\alpha})\mathbb{E}(X_{\alpha})$
2. Standard deviation principle : $\pi_{\rho_{\alpha}}^2 = \mathbb{E}(X_{\alpha}) + \rho_{\alpha}\sqrt{\mathrm{var}(X_{\alpha})}$ with

$$\mathrm{var}(X_{\alpha}) = \sum_{i=1}^{2} \mathbb{E}(M_{\alpha}^{(i)})\mathrm{var}(Y_{\alpha,1}^{(i)}) + \mathrm{var}(M_{\alpha}^{(i)})(\mathbb{E}(Y_{\alpha,1}^{(i)}))^2.$$

Estimation of the pure and commercial premiums are major issues for the insurance companies.

ρ_{α} are chosen such that

$$P(\bar{X}_{\alpha} \geq \pi_{\rho_{\alpha}}^{1,2}) \leq 1-\epsilon$$

where $\bar{X}_{\alpha} = \frac{1}{n_{\alpha}}\sum_{i=1}^{n_{\alpha}} X_{\alpha,j}$ is the mean claim amount for a given portfolio

Premium estimation

- Estimation of the 1 and 2 order expectation of number of annual claim number $M_{\boldsymbol{\alpha}}^{(1)}$ and $M_{\boldsymbol{\alpha}}^{(2)}$
- Estimation of the 1 and 2 order expectation of the claim amount for the attritional claim $Y_{\boldsymbol{\alpha}}^{(1)}$: gamma-GLM
- Estimation of the 1 and 2 order expectation of the claim amount for the atypical claim $Y_{\boldsymbol{\alpha}}^{(2)}$: extreme type GLM
- Calibration of the tuning parameter $\rho_{\boldsymbol{\alpha}}$, Monte Carlo procedure

Summary

Summary

parametric assumption : One-parameter exponential family

Consider the sample $\boldsymbol{Y} = (Y_1, \ldots, Y_n)$, composed of independent random variables (but not i.i.d).

- For $i = 1, \ldots, n$, the distribution of Y_i belongs to the one dimensional exponential family with parameter $\lambda_i \in \Lambda \subset \mathbb{R}$:

the log-density or the log p.m.f. of y_i is assumed to be

$$\log L(\lambda_i|\underline{y}) = (\lambda_i y_i - b(\lambda_i))/a(\phi) + c(y_i, \phi), \quad y_i \in \mathbb{Y} \subset \mathbb{R}, \qquad (1)$$

and $-\infty$ if $y_i \notin \mathbb{Y}$, where

- The parameters $\lambda_1, \ldots, \lambda_n$ depend on a finite-dimensional parameter $\boldsymbol{\vartheta} \in \Theta \subset \mathbb{R}^p$ and depend on deterministic exogenous variables $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n \in \mathbb{R}^p$: $\lambda_i = \lambda(\boldsymbol{\vartheta}; \boldsymbol{x}_i)$
- ϕ dispersion parameter

Direct computations lead to

$$b'(\lambda_i) = E_{\boldsymbol{\vartheta}}(Y_i) \quad \text{and} \quad b''(\lambda_i)a(\phi) = Var_{\boldsymbol{\vartheta}}(Y_i).$$

A link function between natural parameter and explanatory variables

Generalized linear models assume :

1. $Y_1, \ldots, Y_n$ are independent observations with density or p.m.f. (1)
2. a linear predictor w.r.t. explanatory variables

$$\langle \boldsymbol{x}_i, \boldsymbol{\vartheta} \rangle = \eta_i = \vartheta_1 + x_i^{(2)} \vartheta_2 + \ldots + \vartheta_p x_i^{(p)}$$

3. a link function g :

$$g(E_{\boldsymbol{\vartheta}}(Y_i)) = g(b'(\lambda_i)) = \eta_i,$$

where g is a twice continuously differentiable and injective on $b'(\Lambda)$.

A link function between natural parameter and explanatory variables

Generalized linear models assume :

1. $Y_1, \ldots, Y_n$ are independent observations with density or p.m.f. (1)
2. a linear predictor w.r.t. explanatory variables
$$\langle \boldsymbol{x}_i, \boldsymbol{\vartheta} \rangle = \eta_i = \vartheta_1 + x_i^{(2)} \vartheta_2 + \ldots + \vartheta_p x_i^{(p)}$$
3. a link function g :
$$g(E_{\vartheta}(Y_i)) = g(b'(\lambda_i)) = \eta_i,$$
where g is a twice continuously differentiable and injective on $b'(\Lambda)$.

In other words, $\lambda_i = \lambda(\boldsymbol{\vartheta}, \boldsymbol{y}_i) = \ell(\eta_i) = \ell(\langle \boldsymbol{y}_i, \boldsymbol{\vartheta} \rangle)$ with $\ell = (b')^{-1} \circ g^{-1}$.

	$\langle ., . \rangle$	Linear predictor space		Parameter space
$Y \times \Theta$	$\longrightarrow$	D	$\underset{\ell}{\overset{\ell^{-1}}{\leftrightharpoons}}$	Λ

A link function between natural parameter and explanatory variables

Generalized linear models assume :

1. $Y_1, \ldots, Y_n$ are independent observations with density or p.m.f. (1)
2. a linear predictor w.r.t. explanatory variables

$$\langle \boldsymbol{x}_i, \boldsymbol{\vartheta} \rangle = \eta_i = \vartheta_1 + x_i^{(2)} \vartheta_2 + \ldots + \vartheta_p x_i^{(p)}$$

3. a link function g :

$$g(E_{\vartheta}(Y_i)) = g(b'(\lambda_i)) = \eta_i,$$

where g is a twice continuously differentiable and injective on $b'(\Lambda)$.

In other words, $\lambda_i = \lambda(\boldsymbol{\vartheta}, \boldsymbol{y}_i) = \ell(\eta_i) = \ell(\langle \boldsymbol{y}_i, \boldsymbol{\vartheta} \rangle)$ with $\ell = (b')^{-1} \circ g^{-1}$.

	$\langle .,. \rangle$	Linear predictor space		Parameter space
$Y \times \Theta$	$\longrightarrow$	D	$\overset{\ell^{-1}}{\underset{\ell}{\leftrightharpoons}}$	Λ

- Canonical case : $\ell = id$, i.e. $g = (b')^{-1}$

Score equations for MLE

Let us compute the log-likelihood of $\underline{\boldsymbol{y}} = (y_1, \ldots, y_n)$:

$$\log L(\boldsymbol{\vartheta} \mid \underline{\boldsymbol{y}}) = \sum_{i=1}^{n} \frac{y_i \ell(\eta_i) - b\left(\ell(\eta_i)\right)}{a(\phi)} + \sum_{i=1}^{n} c(y_i, \phi), \tag{2}$$

Here, the vector of the parameters $\boldsymbol{\vartheta}$ is unknown.

If the model is identifiable,

- the sequence of maximum likelihood estimators $(\widehat{\boldsymbol{\vartheta}}_n)_{n\geq 1}$ defined by $\widehat{\boldsymbol{\vartheta}}_n = \arg\max_{\boldsymbol{\vartheta}\in\Theta} L(\boldsymbol{\vartheta} \mid \underline{\boldsymbol{y}})$ asymptotically exists and is consistent [Fahrmeir & Kaufmann 1985].

- The maximum likelihood estimator (MLE) $\widehat{\boldsymbol{\vartheta}}_n$, if it exists, is the solution of the non linear system

$$S_j(\boldsymbol{\vartheta}) = 0 \Leftrightarrow \frac{1}{a(\phi)} \sum_{i=1}^{n} x_i^{(j)} \ell'(\eta_i) \left(y_i - b'\left(\ell(\eta_i)\right)\right) = 0, \quad j = 1, \ldots, p, \tag{3}$$

with $S_j(\boldsymbol{\vartheta})$ are the component of the Score vector.

The typical estimation procedure

Well known procedure :

1. use the iteratively re-weighted least square (IWLS) algorithm to get $\hat{\vartheta}_n$,
2. estimate $\hat{\phi}$ using the sum of square residuals or the log-likelihood,
3. computes predicted values as the expected mean,
4. assess the quality of the fit using residuals.

Summary

Coding categorical variable

In any regression model, categorical explanatory variables have to be coded since their value is a name or a category.

Coding categorical variable

In any regression model, categorical explanatory variables have to be coded since their value is a name or a category.

Consider a single unordered variable with d values : for all $i = 1, \dots, n$,

- $x_i^{(1)} = 1$ is the intercept and $x_i^{(2)}$ takes values in a set of d modalities $\{v_1, \dots, v_d\}$.
- We define the incidence matrix $(x_i^{(2),j})_{i,j}$ where $x_i^{(2),j} = \mathbf{1}_{x_i^{(2)} = v_j}$ is the binary dummy of the jth category.

$x_i^{(2)}$	$\rightarrow$	$\left(x_i^{(2),1}, \dots, x_i^{(2),d}\right)$
v_1		$(1, 0, \dots, 0)$
v_2		$(0, 1, 0, \dots, 0)$
	$\vdots$	
v_d		$(0, \dots, 0, 1)$

Special case : explicit solution

Model :
Consider the following GLM for the explanatory variables $x_i^{(2),1}, \dots, x_i^{(2),d}$:

$$g(E(Y_i)) = \vartheta_{(1)} + \sum_{j=1}^{d} x_i^{(2),j} \vartheta_{(2),j}, \quad i = 1, \dots, n,$$

where $\boldsymbol{\vartheta} = (\vartheta_{(1)}, \vartheta_{(2),1}, \dots, \vartheta_{(2),d})$ is the unknown vector parameters.
At this stage, the model is not identifiable because of the redundancy on the vectors $(x_1^{(2),j}, \dots, x_n^{(2),j})$, $j \in J$ and the ones vector (the design matrix is not of full rank).

We need to impose exactly one linear contrast on $\boldsymbol{\vartheta}$:

$$\boldsymbol{R}\vartheta = 0,$$

with $\boldsymbol{R}$ any non-zero real vector of size $d + 1$.

Interesting contrast vector examples

Typical examples of contrast vectors

No-intercept model	$R = (1, 0, \mathbf{0})$	*i.e.*	$\vartheta_{(1)} = 0;$	(C_0)
Model without factor 1	$R = (0, 1, \mathbf{0})$	*i.e.*	$\vartheta_{(2),1} = 0;$	(C_1)
Zero-sum condition	$R = (0, 1, \mathbf{1})$	*i.e.*	$\sum_{j=1}^{d} \vartheta_{(2),j} = 0.$	(C_Σ)

Closed-form Maximum Likelihood Estimators

Theorem

Let g be an injective link function and Λ its corresponding open parameter space. Suppose that for all $i \in \{1, \dots, n\}$, Y_i takes values in $b'(\Lambda)$. Consider the linear contrast $\boldsymbol{R}$ such that $\boldsymbol{R} = (r_{(1)}, r_{(2),1}, \dots, r_{(2),d})$, with $\sum_{j=1}^{d} r_{(2),j} - r_{(1)} \neq 0$. Then, there exists a unique, consistent and explicit MLE $\widehat{\boldsymbol{\vartheta}}_n = (\widehat{\boldsymbol{\vartheta}}_{n,(1)}, \widehat{\boldsymbol{\vartheta}}_{n,(2),1}, \dots, \widehat{\boldsymbol{\vartheta}}_{n,(2),d})$ of $\boldsymbol{\vartheta}$ given by

$$\widehat{\boldsymbol{\vartheta}}_{n,(1)} = \frac{\sum_{j=1}^{d} r_{(2),j} g(\bar{Y}_n^{(j)})}{\sum_{j=1}^{d} r_{(2),j} - r_{(1)}},$$

$$\widehat{\boldsymbol{\vartheta}}_{n,(2),j} = g(\bar{Y}_n^{(j)}) - \frac{\sum_{j=1}^{d} r_{(2),j} g(\bar{Y}_n^{(j)})}{\sum_{j=1}^{d} r_{(2),j} - r_{(1)}}, \quad j = 1, \dots, d.$$

$\overline{Y}_n^{(j)}$ the mean value of $\underline{\boldsymbol{Y}}$ taking over the jth category :

$$\overline{Y}_n^{(j)} = \frac{1}{m_j} \sum_{i=1}^{n} Y_i x_i^{(2),j}, \quad m_j = \sum_{i=1}^{n} x_i^{(2),j} \quad j \in J.$$

Closed-form estimators for the typical contrast vectors

Typical examples of contrast vectors

No-intercept model	$R = (1, 0, \mathbf{0})$ i.e. $\vartheta_{(1)} = 0;$	(C_0)
Model without factor 1	$R = (0, 1, \mathbf{0})$ i.e. $\vartheta_{(2),1} = 0;$	(C_1)
Zero-sum condition	$R = (0, 1, \mathbf{1})$ i.e. $\sum_{j=1}^{d} \vartheta_{(2),j} = 0.$	(C_Σ)

For which we can compute explicit MLE

- Under (C_0), the unique, consistent and explicit MLE $\widehat{\boldsymbol{\vartheta}}_n$ of $\boldsymbol{\vartheta}$ is

$$\widehat{\vartheta}_{n,(1)} = 0, \quad \widehat{\vartheta}_{n,(2),j} = g\left(\overline{y}_n^{(j)}\right), \quad j \in J.$$

Closed-form estimators for the typical contrast vectors

Typical examples of contrast vectors

No-intercept model	$R = (1, 0, \mathbf{0})$ i.e. $\vartheta_{(1)} = 0;$		(C_0)
Model without factor 1	$R = (0, 1, \mathbf{0})$ i.e. $\vartheta_{(2),1} = 0;$		(C_1)
Zero-sum condition	$R = (0, 1, \mathbf{1})$ i.e. $\sum_{j=1}^{d} \vartheta_{(2),j} = 0.$		(C_Σ)

For which we can compute explicit MLE

- Under (C_1), the unique, consistent and explicit MLE $\widehat{\boldsymbol{\vartheta}}_n$ of $\boldsymbol{\vartheta}$ is

$$\widehat{\vartheta}_{n,(1)} = g\left(\overline{y}_n^{(1)}\right), \quad \widehat{\vartheta}_{n,(2),1} = 0, \quad \widehat{\vartheta}_{n,(2),j} = g\left(\overline{y}_n^{(j)}\right) - \widehat{\vartheta}_{n,1}, \quad j \in J \setminus \{1\}.$$

Closed-form estimators for the typical contrast vectors

Typical examples of contrast vectors

No-intercept model	$R = (1, 0, \mathbf{0})$	i.e. $\vartheta_{(1)} = 0;$	(C_0)
Model without factor 1	$R = (0, 1, \mathbf{0})$	i.e. $\vartheta_{(2),1} = 0;$	(C_1)
Zero-sum condition	$R = (0, 1, \mathbf{1})$	i.e. $\sum_{j=1}^{d} \vartheta_{(2),j} = 0.$	(C_Σ)

For which we can compute explicit MLE

- Under (C_Σ), the unique, consistent and explicit MLE $\widehat{\boldsymbol{\vartheta}}_n$ of $\boldsymbol{\vartheta}$ is

$$\widehat{\vartheta}_{n,(1)} = \frac{1}{d} \sum_{j=1}^{d} g\left(\overline{y}_n^{(j)}\right), \ \widehat{\vartheta}_{n,(2),j} = g\left(\overline{y}_n^{(j)}\right) - \widehat{\vartheta}_{n,1}, \quad j \in J.$$

Sketch of the proof

1. We have to solve the system

$$\begin{cases} S(\vartheta) = 0 \\ \boldsymbol{R}\vartheta = 0. \end{cases} \quad (4)$$

Sketch of the proof

1. We have to solve the system

$$\begin{cases} S(\boldsymbol{\vartheta}) = 0 \\ \boldsymbol{R}\boldsymbol{\vartheta} = 0. \end{cases} \quad (4)$$

2. The system $S(\boldsymbol{\vartheta}) = 0$ has a redundant equation

$$\begin{cases} \sum_{i=1}^{n} \ell'(\eta_i)\left(y_i - b' \circ \ell(\eta_i)\right) = 0 \\ \sum_{i=1}^{n} x_i^{(2),j} \ell'(\eta_i)\left(y_i - b' \circ \ell(\eta_i)\right) = 0, \quad \forall j \in J. \end{cases}$$

Hence $S(\boldsymbol{\vartheta}) = 0 \Leftrightarrow \vartheta_0 + \vartheta_j = g(\bar{y}_n^{(j)}) \ \forall j \in J$

Sketch of the proof

1. We have to solve the system

$$\begin{cases} S(\boldsymbol{\vartheta}) = 0 \\ \boldsymbol{R}\boldsymbol{\vartheta} = 0. \end{cases} \tag{4}$$

2. The system $S(\boldsymbol{\vartheta}) = 0$ has a redundant equation

$$\begin{cases} \sum_{i=1}^{n} \ell'(\eta_i)\left(y_i - b' \circ \ell(\eta_i)\right) = 0 \\ \sum_{i=1}^{n} x_i^{(2),j} \ell'(\eta_i)\left(y_i - b' \circ \ell(\eta_i)\right) = 0, \quad \forall j \in J. \end{cases}$$

Hence $S(\boldsymbol{\vartheta}) = 0 \Leftrightarrow \vartheta_0 + \vartheta_j = g(\bar{y}_n^{(j)}) \ \forall j \in J$

3. The system (4) is

$$\begin{pmatrix} Q \\ \boldsymbol{R} \end{pmatrix} \boldsymbol{\vartheta} = \begin{pmatrix} \boldsymbol{g}(\bar{\boldsymbol{Y}}) \\ 0 \end{pmatrix}, \quad \text{with } Q = \begin{pmatrix} 1 & 1 & 0 & \dots & 0 \\ 1 & 0 & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \dots & 0 \\ 1 & 0 & 0 & \dots & 1 \end{pmatrix}, \boldsymbol{g}(\bar{\boldsymbol{Y}}) = \begin{pmatrix} g(\bar{Y}_n^{(1)}) \\ \vdots \\ g(\bar{Y}_n^{(d)}) \end{pmatrix}.$$

Sketch of the proof

4. Compute the inverse of $M_d = \begin{pmatrix} Q \\ \boldsymbol{R} \end{pmatrix}$

$$M_d^{-1} = \begin{pmatrix} \frac{\boldsymbol{r}}{-r_0 + \boldsymbol{r}\mathbf{1}_d} & \frac{-1}{-r_0 + \boldsymbol{r}\mathbf{1}_d} \\ I_d - \frac{\mathbf{1}_d \boldsymbol{r}}{-r_0 + \boldsymbol{r}\mathbf{1}_d} & \frac{\mathbf{1}_d}{-r_0 + \boldsymbol{r}\mathbf{1}_d} \end{pmatrix}.$$

Sketch of the proof

4. Compute the inverse of $M_d = \begin{pmatrix} Q \\ \boldsymbol{R} \end{pmatrix}$

$$M_d^{-1} = \begin{pmatrix} \frac{\boldsymbol{r}}{-r_0 + \boldsymbol{r}\mathbf{1}_d} & \frac{-1}{-r_0 + \boldsymbol{r}\mathbf{1}_d} \\ I_d - \frac{\mathbf{1}_d \boldsymbol{r}}{-r_0 + \boldsymbol{r}\mathbf{1}_d} & \frac{\mathbf{1}_d}{-r_0 + \boldsymbol{r}\mathbf{1}_d} \end{pmatrix}.$$

5. as long as $r_0 \neq \sum_{j=1}^{d} r_j$

$$\widehat{\boldsymbol{\vartheta}}_n = \begin{pmatrix} \frac{\boldsymbol{r}\boldsymbol{g}(\bar{\boldsymbol{Y}})}{-r_0 + \boldsymbol{r}\mathbf{1}_d} \\ \boldsymbol{g}(\bar{\boldsymbol{X}}) - \mathbf{1}_d \frac{\boldsymbol{r}\boldsymbol{g}(\bar{\boldsymbol{Y}})}{-r_0 + \boldsymbol{r}\mathbf{1}_d} \end{pmatrix}.$$

Some important remarks

1. In Theorem 1, it is worth noting that the value of $\widehat{\boldsymbol{\vartheta}}_n$ does not depend on the distribution of the Y_i.
2. The three different parametrizations depends on the type of application and on the modeler choice.
   1. In statistical software, there is a default choice : the model without the first modality is the default parametrization (see function `glm()` by [?]).
   2. The first option without intercept may be justified when no group can be chosen as the reference group.
3. When g is the identity function, the third option with a zero-sum condition can be interpreted as a generalized analysis of variance (ANOVA) for Y_i with respect to groups defined by the explanatory variable $\boldsymbol{x}^{(2)}$. Even for non-Gaussian random variables, some applications may justify this option.

A first corollary on the optimum likelihood

Corollary

The value of the log-likelihood defined in (2) *(and AIC, BIC) taken on the exact MLE* $\widehat{\vartheta}_n$ *given by* (C_0) *under constraint* (15) *does not depend on the link function* g *:*

$$\forall i \in I, \quad \ell(\hat{\eta}_i) = \tilde{b}(\overline{y}_n^{(j)}) \quad \text{for } j \in J \text{ such that } x_i^{(2),j} = 1,$$

and

$$\log L(\widehat{\vartheta}_n \,|\, \underline{y}) = \frac{1}{a(\phi)} \sum_{j=1}^{d} \sum_{i, x_i^{(2),j}=1} \left(y_i \tilde{b}\left(\overline{y}_n^{(j)}\right) - b\left(\tilde{b}\left(\overline{y}_n^{(j)}\right)\right)\right) + \sum_{i=1}^{n} c(y_i, \phi),$$

with $\tilde{b} = (b')^{-1}$.

A first corollary on the optimum likelihood

Corollary

The value of the log-likelihood defined in (2) *(and AIC, BIC) taken on the exact MLE* $\widehat{\boldsymbol{\vartheta}}_n$ *given by* (C_0) *under constraint* (15) *does not depend on the link function* g *:*

$$\forall i \in I, \quad \ell(\hat{\eta}_i) = \tilde{b}(\overline{y}_n^{(j)}) \quad \textit{for } j \in J \textit{ such that } x_i^{(2),j} = 1,$$

and

$$\log L(\widehat{\boldsymbol{\vartheta}}_n \,|\, \underline{\mathbf{y}}) = \frac{1}{a(\phi)} \sum_{j=1}^{d} \sum_{i, x_i^{(2),j}=1} \left(y_i \tilde{b}\left(\overline{y}_n^{(j)}\right) - b\left(\tilde{b}\left(\overline{y}_n^{(j)}\right)\right)\right) + \sum_{i=1}^{n} c(y_i, \phi),$$

with $\tilde{b} = (b')^{-1}$*.*

$\Rightarrow$ The estimator of ϕ obtained by maximizing $\log L(\widehat{\boldsymbol{\vartheta}}_n \,|\, \underline{\mathbf{y}})$ is also independent of g.

A second corollary on predicted values

The predicted moments for the ith individual is estimated by

$$\widehat{\mathbb{E}Y_i} = b'(\ell(\widehat{\eta}_i)), \quad \widehat{\mathrm{var}\,Y_i} = a(\widehat{\phi})b''(\ell(\widehat{\eta}_i)).$$

A second corollary on predicted values

The predicted moments for the ith individual is estimated by

$$\widehat{\mathbb{E}Y_i} = b'(\ell(\widehat{\eta}_i)), \quad \widehat{\mathrm{var}\, Y_i} = a(\widehat{\phi})b''(\ell(\widehat{\eta}_i)).$$

For i such that $x_i^{(2),j} = 1$, we have $\ell(\hat{\eta}_i) = (b')^{-1}(\overline{y}_n^{(j)})$ and expressions are further simplified

$$\widehat{\mathbb{E}Y_i} = \overline{y}_n^{(j)}, \ \widehat{\mathrm{var}\, Y_i} = a(\widehat{\phi})b'' \circ (b')^{-1}\left(\overline{y}_n^{(j)}\right).$$

Both estimates do not depend on the link function g and the predicted mean does not depend on the function b.

Reformulation of the previous case

Let define $Q = (A_0, A_1)$, with A_0 is the ones vector of size d, A_1 the identity matrix of size d and

$$\boldsymbol{g}(\bar{\boldsymbol{Y}}) = \begin{pmatrix} g(\bar{Y}_n^{(1)}) \\ \vdots \\ g(\bar{Y}_n^{(d)}) \end{pmatrix}.$$

The closed form of the MLE $\widehat{\boldsymbol{\vartheta}}_n$ can be reformulated as

$$\widehat{\boldsymbol{\vartheta}}_n = \begin{pmatrix} Q \\ \boldsymbol{R} \end{pmatrix}^{-1} \begin{pmatrix} \boldsymbol{g}(\bar{\boldsymbol{Y}}) \\ 0 \end{pmatrix}.$$

We have

$$\sum_{j=1}^{d} r_{(2),j} - r_{(1)} \neq 0 \Leftrightarrow \text{rank} \begin{pmatrix} Q \\ \boldsymbol{R} \end{pmatrix} = d + 1.$$

With this formulation, the estimator of $\boldsymbol{\vartheta}$ is

$$\widehat{\boldsymbol{\vartheta}}_n = (Q'Q + R'R)^{-1} Q' \boldsymbol{g}(\bar{\boldsymbol{Y}}).$$

Notation for two explanatory variables

Consider two unordered variables with d values : for all $i = 1, \ldots, n$,

- $x_i^{(1)} = 1$ is the intercept and $x_i^{(2)}$ takes values in a set of d modalities $\{v_1, \ldots, v_d\}$.
- Set $d_{2,3}^\star = \# KL^\star$, for $l \in L$, $K_l^\star = \{(k,l) \in K^\star \times \{l\}; m_{kl} > 0\}$, $d_{(3),l}^\star = \# K_l^\star$ and for $k \in K^\star$, $L_k^\star = \{(k,l) \in \{k\} \times L^\star; m_{kl} > 0\}$, $d_{(2),k}^\star = \# L_k^\star$.
- We define

Dummy	Frequency	Mean	Index
$x_i^{(2),k} = \mathbf{1}_{x_i^{(2)} = v_{2k}}$	$m_k^{(2)} = \sum_{i=1}^n x_i^{(2),k}$	$\overline{y}_n^{(2),k} = \frac{1}{m_k^{(2)}} \sum_{i=1}^n y_i x_i^{(2),k}$	$k \in K$
$x_i^{(3),l} = 1_{x_i^{(3)} = v_{3l}}$	$m_l^{(3)} = \sum_{i=1}^n x_i^{(3),l}$	$\overline{y}_n^{(3),l} = \frac{1}{m_l^{(3)}} \sum_{i=1}^n y_i x_i^{(3),l}$	$l \in L$
$x_i^{(k,l)} = x_i^{(2),k} x_i^{(3),l}$	$m_{kl} = \sum_{i=1}^n x_i^{(k,l)}$	$\overline{y}_n^{(k,l)} = \frac{1}{m_{kl}} \sum_{i=1}^n y_i x_i^{(k,l)}$	$(k,l) \in K \times L$

where $\overline{y}_n^{(k,l)}$ is computer over $KL^\star = (K \times L) \setminus \{(k,l) \in K \times L; m_{kl} = 0\}$.

Explicit solutions

Consider the following GLM for explanatory variables $x_i^{(1)}, x_i^{(2),j}, x_i^{(3),j}$

$$g(\mathbb{E}Y_i) = \vartheta_1 + \sum_{k=1}^{d_2} x_i^{(2),k}\vartheta_{(2),k} + \sum_{l=1}^{d_3} x_i^{(3),l}\vartheta_{(3),l} + \sum_{(k,l)\in KL^\star} x_i^{(k,l)}\vartheta_{kl}, \qquad (5)$$

where $\vartheta_{(1)}, (\vartheta_{(2),k})_{k\in K}, (\vartheta_{(3),l})_{l\in L}, (\vartheta_{kl})_{(k,l)\in KL^\star}$ are the $d_2 + d_3 + d_{2,3}^\star + 1$ unknown parameters.

As previously, we need to impose $q \geq 1 + d_2 + d_3$ linear constraints on the vector parameters $\boldsymbol{\vartheta}$

$$R\boldsymbol{\vartheta} = \mathbf{0}_q, \qquad (6)$$

where R is a $q \times (1 + d_2 + d_3 + d_{2,3}^\star)$ real matrix of linear contrasts, with $rank(R) = 1 + d_2 + d_3$ and $\mathbf{0}_q$ the zeros vector of size q.

Explicit solutions

Let define $Q \in \mathbb{R}^{d^\star_{2,3} \times (1+d_2+d_3+d^\star_{2,3})}$ by

$$Q = (A_0, A_1, A_2, A_{12})$$

with $A_0 = \mathbf{1}_{d^\star_{2,3}}$ the $d^\star_{2,3} \times 1$ ones matrix ; $A_1 = (\mathrm{diag}(\mathbf{1}_{d^\star_{(2),k}}))_{k \in K}$, the $d^\star_{2,3} \times K$ diagonal block matrix of ones vector of size $d^\star_{(2),k}$; $A_2 = (I^{\star,k}_{d_3})_{k \in K}$, the $d^\star_{2,3} \times L$ matrix where $I^{\star,k}_{d_3}$ is the identity matrix of size d_3 without rows l for which $m_{kl} = 0$; $A_{12} = I_{d^\star_{2,3}}$ the $d^\star_{2,3} \times d^\star_{2,3}$ identity matrix.

Explicit solutions

Let define $Q \in \mathbb{R}^{d^\star_{2,3} \times (1+d_2+d_3+d^\star_{2,3})}$ by

$$Q = (A_0, A_1, A_2, A_{12})$$

with $A_0 = \mathbf{1}_{d^\star_{2,3}}$ the $d^\star_{2,3} \times 1$ ones matrix ; $A_1 = (\mathrm{diag}(\mathbf{1}_{d^\star_{(2),k}}))_{k \in K}$, the $d^\star_{2,3} \times K$ diagonal block matrix of ones vector of size $d^\star_{(2),k}$; $A_2 = (I^{\star,k}_{d_3})_{k \in K}$, the $d^\star_{2,3} \times L$ matrix where $I^{\star,k}_{d_3}$ is the identity matrix of size d_3 without rows l for which $m_{kl} = 0$; $A_{12} = I_{d^\star_{2,3}}$ the $d^\star_{2,3} \times d^\star_{2,3}$ identity matrix.

Theorem

Suppose that for all $i \in \{1, \dots, n\}$, Y_i takes values in $b'(\Lambda)$. Under constraint (6) and if R such that $(Q'\ R')$ is of rank $d^\star_{2,3}$, there exists a unique, consistent and explicit MLE $\widehat{\boldsymbol{\vartheta}}_n$ of $\boldsymbol{\vartheta}$ given by

$$\widehat{\boldsymbol{\vartheta}}_n = (Q'Q + R'R)^{-1} Q' \mathbf{g}(\bar{\mathbf{Y}}), \tag{7}$$

where the vector $\mathbf{g}(\bar{\mathbf{Y}})$ is $((g(\bar{Y}_n^{(k,l)}))_{l \in L^\star_k})_{k \in K}$.

Classical situation # 1 – No-intercept model

The model with no intercept and no single-variable dummy is

$$\vartheta_1 = 0,\ \vartheta_{(2),k} = \vartheta_{(3),l} = 0,$$

$\forall k \in K\, \forall l \in L$.

Therefore, the unique, consistent and explicit MLE $\widehat{\boldsymbol{\vartheta}}_n$ of $\boldsymbol{\vartheta}$ is

$$\widehat{\vartheta}_{n,kl} = g\left(\overline{Y}_n^{(k,l)}\right), \quad (k,l) \in KL^{\star}.$$

Classical situation # 2 – Zero-sum condition

The model with zero-sum conditions assumes

$$\sum_{k\in K} m_k^{(2)}\vartheta_{(2),k} = \sum_{l\in L} m_l^{(3)}\vartheta_{(3),l} = 0,\ \forall l \in L, \sum_{k\in K_l^\star} m_{kl}\vartheta_{kl} = 0,\ \forall k\in K, \sum_{l\in L_k^\star} m_{kl}\vartheta_{kl} = 0.$$

Therefore, the unique, consistent and explicit MLE $\widehat{\boldsymbol{\vartheta}}_n$ of $\boldsymbol{\vartheta}$ is

$$\begin{cases} \widehat{\vartheta}_{n,(1)} = \frac{1}{n} \sum_{(k,l)\in KL^\star} m_{kl}\, g\left(\overline{Y}_n^{(k,l)}\right) \\ \widehat{\vartheta}_{n,(2),k} = \frac{1}{m_k^{(2)}} \sum_{l\in L_k^\star} m_{kl}\, g\left(\overline{Y}_n^{(k,l)}\right) - \widehat{\vartheta}_{n,1}, \quad k\in K \\ \widehat{\vartheta}_{n,(3),l} = \frac{1}{m_l^{(3)}} \sum_{k\in K_l^\star} m_{kl}\, g\left(\overline{Y}_n^{(k,l)}\right) - \widehat{\vartheta}_{n,1}, \quad l\in L \\ \widehat{\vartheta}_{n,kl} = g\left(\overline{Y}_n^{(k,l)}\right) - \widehat{\vartheta}_{n,(2),k} - \widehat{\vartheta}_{n,(3),l} - \widehat{\vartheta}_{n,1}, \quad (k,l)\in KL^\star. \end{cases}$$

Concluding remarks on explicit solutions

- so if we use only GLMs to predict, there is no reason to select a particular distribution.

- this approach can be further expanded to a high number of explanatory variables.

- results work for any distribution of the one-parameter exponential family : Poisson, Gaussian, gamma.

Summary

Pareto 1 distribution – setting

The density with scale and shape parameter μ and $\lambda_i(\boldsymbol{\vartheta})$, $i \in I$ is

$$f(x) = \lambda_i(\boldsymbol{\vartheta}) \frac{\mu^{\lambda_i(\boldsymbol{\vartheta})}}{x^{\lambda_i(\boldsymbol{\vartheta})+1}}, \quad x \in \mathbb{X} = [\mu, \infty),$$

where μ is fixed and known.

- For Pareto 1 distribution, we recall that

$$E(X_i) = \frac{\lambda_i(\boldsymbol{\vartheta})\mu}{\lambda_i(\boldsymbol{\vartheta}) - 1} < +\infty \quad \text{iff } \lambda_i(\boldsymbol{\vartheta}) > 1$$
$$\text{and } E(X_i^2) = \frac{\lambda_i(\boldsymbol{\vartheta})\mu^2}{\lambda_i(\boldsymbol{\vartheta}) - 2} < +\infty \quad \text{iff } \lambda_i(\boldsymbol{\vartheta}) > 2.$$

- Unlike the known parameter μ, the parameter $\boldsymbol{\vartheta}$ is to be estimated.

- $Z_i = T(X_i) = -\log(X_i/\mu)$ belongs to the exponential family, with

$$\alpha(\phi) = 1, \quad \varphi(\lambda) = -\log(\lambda), \quad \text{and} \quad c(z, \phi) = 0, \quad z \in T(\mathbb{X}) = \mathbb{R}^-, \quad \lambda \in \Lambda.$$

TABLE – Typical link functions for Pareto I

Names	$\ell(\eta_i)$	$g^{-1}(t)$	$g(t)$	Λ	$\varphi'(\Lambda)$
canonical	η_i	$-\frac{1}{t}$	$-\frac{1}{t}$	$(0, +\infty)$	$(-\infty, 0)$
log-inv	e^{η_i}	$-e^{-t}$	$\log(-\frac{1}{t})$	$(0, +\infty)$	$(-\infty, 0)$
shifted log-inv	$e^{\eta_i} + 1$	$-\frac{1}{e^t + 1}$	$\log(-\frac{1}{t} - 1)$	$(1, +\infty)$	$(-1, 0)$

Pareto 1 distribution – explicit MLE 1

We consider the model (contrast (C_0))

$$g(E(Z_i)) = y_i^{(2),1}\vartheta_{(2),1} + \cdots + y_i^{(2),d}\vartheta_{(2),d}, \quad i \in I, \tag{8}$$

where $\boldsymbol{\vartheta} = (\vartheta_{(2),1}, \ldots, \vartheta_{(2),d})$ is the vector of unknown parameter.

Example 1 – canonical link :
We have $\overline{z}_n^{(j)} \in \varphi'(\Lambda) = (-\infty, 0)$ and using Theorem 1, the MLE is

$$\widehat{\vartheta}_{n,(2),j} = -m_j \left(\sum_{i=1}^{n} y_i^{(2),j} z_i \right)^{-1} = -\frac{1}{\overline{z}_n^{(j)}}, \quad j \in J.$$

- Hence, $\widehat{\vartheta}_{n,(2),j}$ follows an Inverse Gamma distribution with shape parameter m_j and rate parameter $m_j\vartheta_{(2),j}$.

Pareto 1 distribution – explicit MLE 1

We consider the model (contrast (C_0))

$$g(E(Z_i)) = y_i^{(2),1}\vartheta_{(2),1} + \cdots + y_i^{(2),d}\vartheta_{(2),d}, \quad i \in I, \tag{8}$$

where $\boldsymbol{\vartheta} = (\vartheta_{(2),1}, \ldots, \vartheta_{(2),d})$ is the vector of unknown parameter.

Example 1 – canonical link :
We have $\overline{z}_n^{(j)} \in \varphi'(\Lambda) = (-\infty, 0)$ and using Theorem 1, the MLE is

$$\widehat{\vartheta}_{n,(2),j} = -m_j \left(\sum_{i=1}^{n} y_i^{(2),j} z_i \right)^{-1} = -\frac{1}{\overline{z}_n^{(j)}}, \quad j \in J.$$

- We deduce that for $j \in J$, $m_j > 2$,

$$E(\widehat{\vartheta}_{n,(2),j}) = \frac{m_j}{m_j - 1}\vartheta_{(2),j}, \quad \text{and} \quad Var(\widehat{\vartheta}_{n,(2),j}) = \frac{m_j^2}{(m_j-1)^2(m_j-2)}\vartheta_{(2),j}^2.$$

Pareto 1 distribution – explicit MLE 1

We consider the model (contrast (C_0))

$$g(E(Z_i)) = y_i^{(2),1}\vartheta_{(2),1} + \cdots + y_i^{(2),d}\vartheta_{(2),d}, \quad i \in I, \tag{8}$$

where $\boldsymbol{\vartheta} = (\vartheta_{(2),1}, \ldots, \vartheta_{(2),d})$ is the vector of unknown parameter.

Example 1 – canonical link :
We have $\overline{z}_n^{(j)} \in \varphi'(\Lambda) = (-\infty, 0)$ and using Theorem 1, the MLE is

$$\widehat{\vartheta}_{n,(2),j} = -m_j \left(\sum_{i=1}^{n} y_i^{(2),j} z_i \right)^{-1} = -\frac{1}{\overline{z}_n^{(j)}}, \quad j \in J.$$

- An unbiased estimator of $\theta_{(2),j}$ is then $\widehat{\vartheta}^{\star}_{n,(2),j} = \frac{m_j - 1}{m_j}\widehat{\vartheta}_{n,(2),j}$ which has a lower variance

$$Var(\widehat{\vartheta}^{\star}_{n,(2),j}) = \frac{\vartheta^2_{(2),j}}{m_j - 2} \leq Var(\widehat{\vartheta}_{n,(2),j}), \quad j \in J.$$

Pareto 1 distribution – explicit MLE 2

Example 2 – log-inv link :
We also have $\bar{z}_n^{(j)} \in \varphi'(\Lambda) = (-\infty, 0)$ for all $j \in J$ and using Theorem 1, the MLE is

$$\widehat{\vartheta}_{n,(2),j} = -\log\left(-\bar{z}_n^{(j)}\right), \quad j \in J.$$

- the distribution of $-\widehat{\vartheta}_{n,(2),j}$ is the distribution of the log of a gamma distributed r.v.

Pareto 1 distribution – explicit MLE 2

Example 2 – log-inv link :
We also have $\bar{z}_n^{(j)} \in \varphi'(\Lambda) = (-\infty, 0)$ for all $j \in J$ and using Theorem 1, the MLE is

$$\widehat{\vartheta}_{n,(2),j} = -\log\left(-\bar{z}_n^{(j)}\right), \quad j \in J.$$

- Some calculus lead to

$$E(\widehat{\vartheta}_{n,(2),j}) = \vartheta_{(2),j} + \log m_j - \psi(m_j) \quad \text{and} \quad Var(\widehat{\vartheta}_{n,(2),j}) = \psi'(m_j), \qquad j \in J.,$$

where ψ et ψ' the digamma and trigamma functions.

Pareto 1 distribution – explicit MLE 2

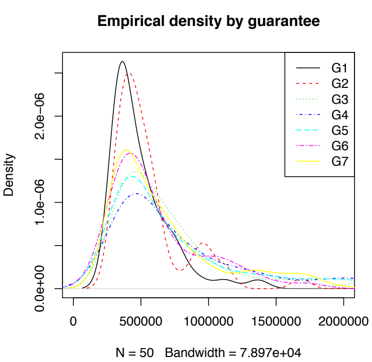
Example 2 – log-inv link :
We also have $\bar{z}_n^{(j)} \in \varphi'(\Lambda) = (-\infty, 0)$ for all $j \in J$ and using Theorem 1, the MLE is

$$\widehat{\vartheta}_{n,(2),j} = -\log\left(-\bar{z}_n^{(j)}\right), \quad j \in J.$$

- Hence $\widehat{\vartheta}_{n,(2),j}$ is asymptotically unbiased, and an unbiased estimator of $\theta_{(2),j}$ is

$$\widehat{\vartheta}^{\star}_{n,(2),j} = \widehat{\vartheta}_{n,(2),j} - (\log(m_j) - \psi(m_j)), \quad j \in J.$$

back to application : Graphics for excesses over μ =300,000



Fitted parameters

TABLE – Fitted parameters with the guarantee variable

	Pareto 1			Log-normal
Link	canonical	log-inv	shifted log-inv	canonical
(Intercept)	1.980	0.683	-0.020	11.675
G2	-0.265	-0.144	-0.315	0.330
G3	-0.830	-0.543	-1.875	0.880
G4	-0.970	-0.673	-4.543	1.024
G5	-0.839	-0.551	-1.936	0.797
G6	-0.527	-0.310	-0.773	0.386
G7	-0.526	-0.309	-0.769	0.275
log lik.	-950.76	-950.76	-950.76	-2168.18

- For Pareto 1, the lower the coefficient value, the heavier the tail is.
- For log-normal, the higher the coefficient value, the heavier the tail is.
- G4 has (unsurprisingly) the heavier tail.
- the log-likelihood value depends only on the distribution.

Summary

Enhanced standard errors – Pareto 1 with canonical link

For each n with $p = 3$

- we simulate samples of Pareto 1 distribution of size n.
- we estimate coefficient by explicit formula and IWLS.
- we compute the asymptotic confidence interval and the true one.

Enhanced standard errors – Pareto 1 with canonical link

For each n with $p = 3$

- we simulate samples of Pareto 1 distribution of size n.
- we estimate coefficient by explicit formula and IWLS.
- we compute the asymptotic confidence interval and the true one.

We observe that the theoretical confidence interval is always better than the asymptotic ones.

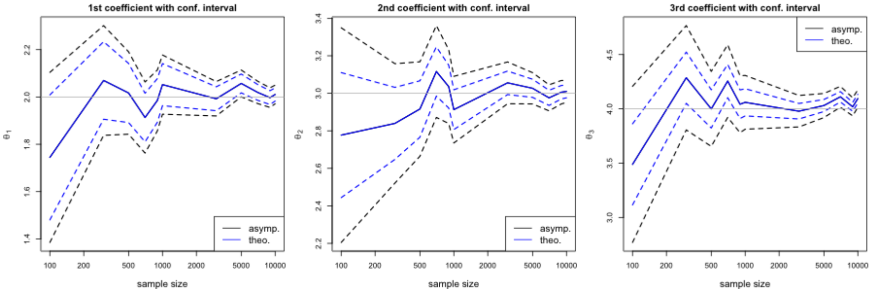


FIGURE – Floating point operation number given the size of the dataset

Floating point operation numbers

We compute the floating point operation numbers given the size of the input dataset.

Floating point operation numbers

We compute the floating point operation numbers given the size of the input dataset.

We observe that the explicit solution is far less computer intensive (4000 times faster) than the IWLS algorithm which takes 5 or 6 iterations to get the solution.

IWLS algorithm				exact method				
n	*p*	iter. nb.	flop x1000	*n*	*p*	iter. nb.	flop x1000	rel. gain
200	5	6	1206.2	200	5	1	1.6	753.9
400	5	5	4010.4	400	5	1	3.2	1253.2
600	5	6	10818.6	600	5	1	4.8	2253.9
800	5	5	16020.8	800	5	1	6.4	2503.2
1000	5	5	25026	1000	5	1	8	3128.2
1200	5	5	36031.2	1200	5	1	9.6	3753.2
1400	5	5	49036.4	1400	5	1	11.2	4378.2
1600	5	5	64041.6	1600	5	1	12.8	5003.2
1800	5	5	81046.8	1800	5	1	14.4	5628.2
200	7	6	1686.2	200	7	1	2	843.1
400	7	6	6732.4	400	7	1	4	1683.1
600	7	6	15138.6	600	7	1	6	2523.1
800	7	6	26904.8	800	7	1	8	3363.1
1000	7	6	42031	1000	7	1	10	4203.1
1200	7	6	60517.2	1200	7	1	12	5043.1
1400	7	6	82363.4	1400	7	1	14	5883.1
1600	7	6	107569.6	1600	7	1	16	6723.1
1800	7	6	136135.8	1800	7	1	18	7563.1

TABLE – Floating point operation number given the size of the dataset

References I

J. Beirlant, Y. Goegebeur, J. Teugels, and J. Segers.
Statistics of extremes : Theory and applications.
Wiley & Sons, 2004.

A. Brouste, C. Dutang, and T. Rohmer.
Closed form Maximum Likelihood Estimator for Generalized Linear Models in the case of categorical explanatory variables : Application to insurance loss modelling.
working paper or preprint, 2019.

A. Brouste, A. Matoussi, T. Rohmer, C. Dutang, V. Désert, E. Gales, P. Golhen, B. Milleville, and W. Lekeufack.
Solvency tuned premium for a composite loss distribution.
working paper or preprint, September 2018.

H. Bühlmann and A. Gisler.
A course in credibility theory and its applications.
Springer Science & Business Media, 2006.

A.C. Davison and R.L. Smith.
Models for exceedances over high thresholds.
Journal of the Royal Statistical Society. Series B, 52(3) :393–442, 1990.

References II

Ludwig Fahrmeir and Heinz Kaufmann.
Consistency and asymptotic normality of the maximum likelihood estimator in generalized linear models.
The Annals of Statistics, pages 342–368, 1985.

P. McCullagh and J. A. Nelder.
Generalized linear models, volume 37.
CRC press, 1989.

Esbjorn Ohlsson and Bjorn Johansson.
Non-Life Insurance Pricing with Generalized Linear Models.
Springer, 2010.

Expectation of the annual number of claim $\mathbb{E}(M_\alpha^{(i)})$, $i = 1, 2$.

Problematic : some policies are not observed during the whole year !
Instead of $M_\alpha^{(i,j)}$ we observe $(d_\alpha^{(i,j)}, N_\alpha^{(i,j)})_{j=1,\dots,n_\alpha}$, where

- $d_\alpha^{(i,j)}$ Contract duration (in days)
- $N_\alpha^{(i,j)}$ the number of claim during the period d_{ij}.

In our model, we assume that the number of claims follows an homogeneous Poisson process with intensity $\lambda_\alpha^{(i)}$ Since the annual aggregate claim is considered,

$$\mathbb{E}(M_{\alpha_i}^{(i)}) = \mathrm{var}(M_{\alpha_i}^{(i)}) = \lambda_\alpha^{(i)}.$$

The Poisson process structure gives a natural estimation of λ_{α_i} with

$$\widehat{\lambda}_{\alpha_i}^{(i)} = \frac{\sum_{j=1}^{n_{\alpha_i}^{(i)}} N_{\alpha_i,j}^{(i)}}{\sum_{j=1}^{n_{\alpha_i}^{(i)}} d_\alpha^{(i,j)}}.$$

- Non-parametric right-censored based estimation ?

Calibration of the tunning parameter

α	$\widehat{\pi}^1_{\alpha,0}$	Attr (%)	Atyp (%)
1	3704	16	84
2	5751	15	85
3	1285	17	83
4	1792	14	86
5	6568	9	91
6	6413	5	95
7	2265	12	88
8	4756	12	88
9	4143	17	83
10	24631	15	85
11	935	23	77

α	$\widehat{\pi}^1_{\alpha,0}$	$\widehat{\pi}^1_{\alpha,\rho'}$	$\widehat{\pi}^2_{\alpha,\rho^*}$
1	3704	7860	9548
2	5751	12205	11363
3	1285	2727	4283
4	1792	3803	5893
5	6568	13938	15610
6	6413	13610	15749
7	2265	4806	6988
8	4756	10093	11305
9	4143	8793	9619
10	24631	52269	38350
11	935	1984	3377

TABLE – Respective contributions of the attritional and atypical claims in the pure premium for the different risk class in the GLM-GPD model (on the left). Estimation of the different premium principles ; for the expectation principle with a global safety loading calibrated to $\rho' = 1.122$ and for the standard deviation principle with a global safety loading calibrated to $\rho^* = 0.048$.

Model diagnostics – residuals

- Pareto 1 distribution :
  Using $-Z_i = \log(X_i/\mu) \sim \mathcal{E}(\ell(\eta_i))$, we define the residuals

$$R_i = -\ell(\eta_i) Z_i, \quad i \in I.$$

  Hence $R_1, \ldots, R_n$ are i.i.d. and have an exponential distribution $\mathcal{E}(1)$.

- Lognormal distribution :
  Using $Z_i = \log(X_i - \mu)$, we define the residuals

$$R_i = \frac{Z_i - \ell(\eta_i)}{\sqrt{\phi}}, \quad i \in I.$$

  Hence $R_1, \ldots, R_n$ are i.i.d. and have a normal distribution $\mathcal{N}(0, 1)$.

Log of a gamma distribution

Let $L = \log(G)$ when G is gamma distributed with shape parameter $a > 0$ and rate parameter $\lambda > 0$. We have by elementary manipulations the moment generating function of L :

$$M_L(t) = \mathbb{E}e^{tL} = \frac{\Gamma(a+t)}{\Gamma(a)}\lambda^{-t}, \quad t > -a,$$

where Γ denotes the usual gamma function. Therefore by differentiating and evaluating at 0, we deduce that the expectation and the variance of L are

$$\mathbb{E}L = M_Y'(0) = \psi(a) - \log\lambda$$

and

$$\mathrm{var} L = M_Y''(0) - M_Y'(0)^2 = \psi'(a),$$

where the functions ψ and ψ' are the digamma and trigamma function, see e.g. [**?**].