

# Application of GLM Advancements to Non-Life Insurance Pricing

Leonardo Stincone

Università degli Studi di Trieste

18 Maggio 2021



1. Descrizione del problema

2. Dataset

3. Modelli

4. Risultati



**Problema:** prevedere il numero di sinistri ( $N_i$ ) che causerà un assicurato ( $i$ ) a partire dalle informazioni della sua polizza:

$$(x_{i1}, x_{i2}, \dots, x_{ip}) \longmapsto F_{N_i}, E(N_i), Var(N_i)$$

**Soluzione:** stimo un **modello** a partire dai **dati storici**.

**Perché:** prevedere il numero di sinistri è uno degli elementi per **determinare il prezzo** di una polizza assicurativa.

## Origine del Dataset

Portafoglio RCA costituito da polizze di una provincia italiana nel periodo 2014-2019

| Set   | Esposizione<br>(rischi anno) | Numero<br>Sinistri | Frequenza<br>Sinistri |
|-------|------------------------------|--------------------|-----------------------|
| Train | 107 998.4                    | 4 823              | 0.045                 |
| Test  | 26 806.3                     | 1 131              | 0.042                 |
| Tot   | 134 804.7                    | 5 954              | 0.044                 |

| Descrizione                                    | Numero di variabili<br>per categoria |
|--|--------------------------------------|
| Informazioni sul veicolo assicurato            | 12                                   |
| Informazioni generiche sull'assicurato         | 14                                   |
| Informazioni assicurative sull'assicurato      | 9                                    |
| Opzioni della polizza assicurativa             | 11                                   |
| Informazioni sull'assicurato in quanto cliente | 2                                    |
| Dati telematici                                | 4                                    |
| <b>Totale</b>                                  | <b>52</b>                            |

## Ipotesi

- 1  $(Y_1, \dots, Y_n)$  indipendenti con distribuzione appartenente alla medesima Famiglia Esponenziale Lineare
- 2 Predittore lineare

$$\eta_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_p x_{ip} = \mathbf{x}_i^t \boldsymbol{\beta}, \quad i \in \{1, 2, \dots, n\}$$

- 3 Funzione legame

$$g(\mu_i) = \eta_i = \mathbf{x}_i^t \boldsymbol{\beta}, \quad i \in \{1, 2, \dots, n\}$$

## Stima dei parametri

Stima di massima verosimiglianza

$$\hat{\boldsymbol{\beta}} = \arg \max_{\boldsymbol{\beta} \in \mathbb{R}^{p+1}} L(\boldsymbol{\beta}, \phi; \mathbf{y}) = \arg \min_{\boldsymbol{\beta} \in \mathbb{R}^{p+1}} D(\boldsymbol{\beta}, \mathbf{y})$$

## Selezione delle variabili

Algoritmo stepwise basato su un criterio di informazione (AIC)



## Ipotesi

Modello sottostante: GLM

## Stima dei parametri

Stima di Massima Verosimiglianza con Penalizzazione

$$\hat{\beta} = \arg \min_{\beta \in \mathbb{R}^{p+1}} \left\{ D(\beta, \mathbf{y}) + \lambda \sum_{j=1}^p (\alpha |\beta_j| + (1 - \alpha) |\beta_j|^2) \right\}$$

con

- $\lambda \geq 0$  iperparametro di penalizzazione
- $\alpha \in [0, 1]$  iperparametro che determina il peso della penalizzazione LASSO
  - ▶  $\alpha = 0 \implies$  Regressione Ridge
  - ▶  $\alpha = 1 \implies$  Regressione LASSO



## Ipotesi

Predittore lineare

$$\eta_i = \mathbf{x}_i^t \boldsymbol{\beta} + \sum_{l=1}^q f_l(z_{i,l}), \quad i \in \{1, 2, \dots, n\}$$

con  $f_l(\cdot)$  spline cubica

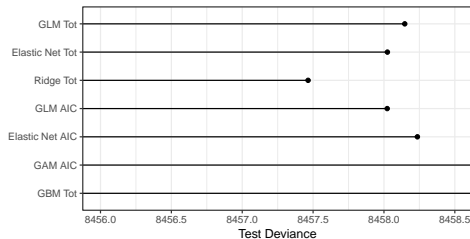
## Stima dei parametri

Stima di Massima Verosimiglianza con Penalizzazione

$$\hat{\mathbf{f}} = \arg \min_{\mathbf{f}} \left\{ D(\mathbf{f}, \mathbf{y}) + \sum_{l=1}^q \lambda_l \int_{a_l}^{b_l} (f_l''(x_l))^2 dx \right\}$$

con  $\lambda_1, \lambda_2, \dots, \lambda_q$  iperparametri di smoothing.





| Id   | Model            | Test Deviance    | Time   | $\alpha$ | $\lambda$ |
|------|------------------|------------------|--------|----------|-----------|
| Mod1 | GLM Tot          | 8 458.147        | 2.7s   | 0        | 0         |
| Mod2 | Elastic Net Tot  | 8 458.024        | 1h 30m | 0.06     | 2.01e-04  |
| Mod3 | <b>Ridge Tot</b> | <b>8 457.465</b> | 1h 30m | 0        | 4.64e-04  |
| Mod4 | GLM AIC          | 8 458.023        | 7h 27m | 0        | 0         |
| Mod5 | Elastic Net AIC  | 8 458.236        | 8h 54m | 0        | 1.63e-05  |
| Mod6 | GAM AIC          | 9 728.570        | 7h 45m | 0        | 0         |
| Mod7 | GBM Tot          | 8 504.178        | 2h 30m |          |           |

Grazie per l'attenzione

