4. Fitting a GLM in R

Goals

The goals of this chapter are as follows

1) Overview Non-Life insurance pricing
2) Overview Generalized Linear Models
3) Fit a glm in R (used for Pricing in Non-Life)

Non-Life Insurance Pricing (1/2)

Classical statistical approach for non-life insurance pricing: Let

- R.v. N describing the number of claims,
- r.v. Z_k describing the individual claim sizes,
- R.v. S describing the total claim amount

$$S = Z_1 + \ldots + Z_N = \sum_{k=1}^{N} Z_k$$

Model Assumptions:

(1) N is a discrete r.v. which only takes positive values.

(2) $Z_1, \ldots, Z_N$ are independent and i.i.d.

(3) N and $(Z_1, \ldots, Z_N)$ are independent

Non-Life Insurance Pricing (2/2)

From the model assumptions follows the following decomposition:

$$E[S] = E[N] * E[Z_1],$$

Hence for the modeling of the **pure risk premium E[S]** we can treat the (expected) number of claims E[N] and the (expected) individual claim sized E[Z_1] separately.

Generalized Linear Models (1/4)

- Response Y_i obeying:

$$f(y) = c\left(y, \frac{\varphi}{\nu_i}\right) \exp\left(\frac{y\Theta_i - a(\Theta_i)}{\varphi/\nu_i}\right) \text{ from the EDF}$$

- Linear score (or predictor):

$$score_i = score(x_i) = x_i^T \beta$$

- Expected response $\mu_i = \mu(x_i)$:

$$a'(\Theta_i) = \mu_i$$

$$g(\mu_i) = score_i \,,\ \mu_i = g^{-1}(score_i)$$
$$g \in \{ln, logit, \dots\}$$

- The monotonic link function g is specified by the actuary.

Generalized Linear Models (2/4)

- Hence for every GLM the actuary needs to choose:
  a) The GLM distribution (Normal, Poisson, Gamma, Binomial,…)
  b) The Link function g (how the mean related to the $score_i$)

- The parameter β is estimated by Maximum Likelihood Estimation (MLE).

Generalized Linear Models (3/4)

- Let us look at the Poisson GLM concretely:
  a) Let λ be the expected (claims) frequency λ>0 and a volume v>0. v measures the exposure (for example in yearly units).
  b) The distribution of the response follows a Poisson distribution

  $$N \sim Poi(\lambda v), \text{ i.e. } P[N = k] = \exp(-\lambda v)\frac{(\lambda v)^k}{k!} \text{ and } E[N] = \lambda v$$

  c) The expected frequency λ should allow to incorporate structural differences (heterogeneity) between different risks, i.e. $\boldsymbol{x} = (x_1, \dots, x_d)$, so $\lambda = \lambda(\boldsymbol{x})$.
  d) The Link function g (how the mean related to $score_i$) is log, i.e.

  $$\log(\lambda(x)v) = score(x_i) = x^T \beta$$

Generalized Linear Models (4/4)

- There is much more on GLM’s, see the corresponding literature for further details.
- Let us see how a Poisson GLM can be fitted in R for non-life insurance pricing.