EBCT 2

Let $Y_1, \ldots, Y_n$ denote the number of claims or aggregate claims for a particular risk. We need to estimate Y_{n+1}.

For each Y_i, we associate a risk volume P_i, which represents the amount of business in year i, e.g. number of Policies sold.

We know P_{n+1}.

Define $X_j = \frac{Y_j}{P_j}$, the Y_j standardised by risk volume.

Assumptions on the X_j

1. The distribution of the X_j depends on and an unknown parameter θ.
2. Given θ, the X_j's are independent and identically distributed.
3. Therefore, $E[X_j|\theta]$ does not depend on j.
4. We assume $P_j \operatorname{Var}(X_j|\theta)$ does not depend on θ.

Define $m(\theta) = E(X_j|\theta)$

$$s^2(\theta) = P_j \cdot Var(X_j|\theta)$$

Example

Suppose the claims from P_j policies are independent, where P_j is the number of policies in year j.

Suppose for fixed θ, each ~~X_j~~ claim has mean $m(\theta)$ and variance $s^2(\theta)$.

Y_j is the aggregate claims from all P_j policies in year j.

$$\Rightarrow E(Y_j|\theta) = P_j m(\theta)$$

$$Var(Y_j|\theta) \; P_j \cdot s^2(\theta)$$

$$E(X_j|\theta) = m(\theta)$$

$$Var(X_j|\theta) = \frac{1}{P_j^2} Var(Y_j|\theta)$$

$$= \frac{1}{P_j^2} P_j \cdot s^2(\theta) = \frac{1}{P_j} s^2(\theta)$$

$$\Rightarrow s^2(\theta) = P_j \, Var(X_j|\theta)$$

Define $X_{i,j} = \dfrac{Y_{i,j}}{P_{i,j}}$

The credibility estimate of

$$Y_{1,n+1} = P_{1,n+1} X_{1,n+1}$$

is

$$P_{1,n+1}\left[Z_1 \bar{X}_1 + (1 - Z_1) E[m(\theta)] \right]$$

(note that Z now depends on i)

where $E[m(\theta)]$ is estimated by

$$\bar{X} = \frac{\sum_{i=1}^{N} \sum_{j=1}^{n} Y_{i,j}}{\sum_{i=1}^{N} \sum_{j=1}^{n} P_{i,j}}$$

$$\bar{X}_1 = \frac{\sum_{j=1}^{n} Y_{1,j}}{\sum_{j=1}^{n} P_{1,j}}$$

The estimate of Y_{n+1} given $Y_1, \dots, Y_n$ is

$$P_{n+1}\left[Z\bar{X} + (1-Z)\, E[m(\theta)] \right]$$

where

$$\bar{X} = \frac{\sum_{j=1}^{n} P_j X_j}{\sum_{j=1}^{n} P_j} = \frac{\sum_{j=1}^{n} Y_j}{\sum_{j=1}^{n} P_j}$$

and

$$Z = \frac{\sum_{j=1}^{n} P_j}{\sum_{j=1}^{n} P_j + \frac{E[s^2(\theta)]}{Var[m(\theta)]}}$$

As we did for EBCT1, we have a matrix $Y_{i,j}$, but now we have a second matrix $P_{i,j}$, where i is the number of the risk j is the year.

$$Z_1 = \frac{\sum_{j=1}^{n} P_{1,j}}{\sum_{j=1}^{n} P_{1,j} + \frac{E[s^2(\theta)]}{Var[m(\theta)]}}$$

where

$$\frac{1}{N(n-1)} \sum_{i=1}^{N} \sum_{j=1}^{n} P_{i,j} \left(X_{i,j} - \bar{X}_i\right)^2$$

estimates $E[s^2(\theta)]$

where $$\bar{X}_i = \frac{\sum_{j=1}^{n} \cancel{Y}_{i,j}}{\sum_{j=1}^{n} P_{i,j}}$$

and

$$\frac{1}{P^*} \left(\frac{1}{Nn-1} \sum_{i=1}^{N} \sum_{j=1}^{n} P_{i,j} \left(X_{i,j} - \bar{X}\right)^2 - \text{estimator for } E[s^2(\theta)] \right)$$

estimates $Var[m(\theta)]$

where $$P^* = \frac{1}{Nn-1} \sum_{i=1}^{N} \bar{P}_i \left(1 - \frac{\bar{P}_i}{\bar{P}}\right)$$

where $\bar{P}_i = \sum_{j=1}^{n} P_{i,j}$ $\quad \bar{P} = \sum_{i=1}^{N} \bar{P}_i = \sum_{i=1}^{N} \sum_{j=1}^{n} P_{i,j}$

Generalised Linear Models

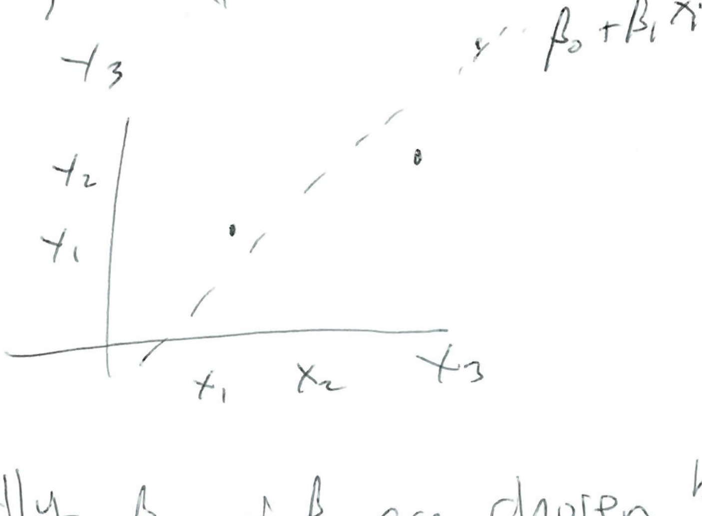
Generalised Linear Models (GLM) is an extension of linear regression.

A linear regression takes the form

$$Y_i = \beta_0 + \beta_1 X_i + \epsilon_i$$

where X_i is a variable
Y_i is a covariate
$\epsilon_i \sim N(0, \sigma^2)$ i.i.d.

Given Y_i, we are trying to estimate β_0 and β_1.



Typically β_0 and β_1 are chosen by the method of least squares.

$\mu_i = E(Y_i) = \beta_0 + \beta_1 X_i$

μ_i is linear in β_0, β_1

$\mu_i = \beta_0^2 + \ln \beta_1 X_i$ is not linear in β_0, β_1.

The differences between GLM and Linear Regression are

- The Y_i do not have to be normal.
- We can have more than one variable.
- instead of setting $\mu_i = E(Y_i)$ equal to a linear predictor, we can set a function of μ_i equal to a linear ~~preto~~ predictor.

Above $g(\mu_i) = \beta_0 + \beta_1 x_i$

when $g(\mu_i) = \mu_i$

but now g can be some other function.

A GLM consists of three components

1. Distribution of Y_i
2. Linear predictor or estimator (~~the~~ function of the variables which is linear in the coefficients)
3. Link function g, so that

$$g(EY_i) = \text{linear prector}$$

The distribution of Y_i should come from the exponential family of distributions.
It includes: normal, Poisson, binomial, gamma, but not uniform.

Definition

A distribution belongs to the exponential family if it's p.m.f. or p.d.f. has the form

$$f_Y(y;\theta,\phi) = \exp\left[\frac{y\theta - b(\theta)}{a(\phi)} + c(y,\phi)\right]$$

does not depend on θ

Remark:
1. θ is important for predicting y
2. If a distribution only has one parameter we take $\phi = 1$.