Insurance Policy Pricing

Luke Dando

September 12, 2022

Contents

1 Risk

1.1 What is risk?

Risk is the chance that something harmful or unexpected could occur regarding a held policy. This might involve loss, theft, or damage of valuable property and belongings, or it may involve someone being injured. From a statistical point-of-view risk can be defined as:

$$\tau = \mathbb{E}\left[\frac{L}{e}\right]. \tag{1}$$

Here, L is the loss and e is the exposure of the policy. If we now assume that the frequency of a claim is independent from the severity of it, we find:

$$\tau = \mathbb{E}\left[\frac{L}{e}\right], \tag{2}$$

$$= \mathbb{E}\left[\frac{L}{N}\Big|N > 0\right] \cdot \mathbb{E}\left[\frac{N}{e}\right], \tag{3}$$

$$= \mathbb{E}[F] \cdot \mathbb{E}[S]], \tag{4}$$

where N is the number of claims, S is the severity (or size) of the claim and F is the claim frequency.
This forms the basis of how our burn costs are set when defining our technical prices.

2 Linear Models

2.1 Introducing the problem

GLMs are generalized forms of linear models, so in order to understand them it would be ideal to first review classic linear models through an example.
Both these models have the same purpose: to determine the relationship between an observed response variable, Y, and predictor variables X. For example, let there be a total of 4 predictor variables: we then define Y as

$$Y = \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \epsilon. \tag{5}$$

	Urban	Rural
Male	800	500
Female	400	200

Table 1: Covariances for linear model example.

Here, ϵ is an error term, usually Normally distributed with mean zero and variance σ^2 (often written as $\epsilon \sim \mathcal{N}(0, \sigma^2)$). Our covariates will be representing male (X_1), female (X_2), urban (X_3) and rural (X_4). These and their respective values can be seen in table 1.
The problem here is that, with as many parameters as there are combinations of rating factor levels being considered and a linear dependency between the four, the model is not uniquely defined. In order to do so, we remove one of these parameters from our model, leaving us with

$$Y = \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \epsilon. \tag{6}$$

We can express our observations as the following:

$$Y_1 = 800 = \beta_1 + 0 + \beta_3 + \epsilon_1; \tag{7}$$
$$Y_2 = 500 = \beta_1 + 0 + 0 + \epsilon_2; \tag{8}$$
$$Y_3 = 400 = 0 + \beta_2 + \beta_3 + \epsilon_3; \tag{9}$$
$$Y_4 = 200 = \beta_2 + 0 + \epsilon_4. \tag{10}$$

We then want to find the best values for β, which we do so through minimizing the sum of squared errors (SSE):

$$SSE = \epsilon_1^2 + \epsilon_2^2 + \epsilon_3^2 + \epsilon_4^2, \tag{11}$$
$$= (800 - \beta_1 - \beta_3)^2 + (500 - \beta_1)^2 + (400 - \beta_2 - \beta_3)^2 + (200 - \beta_2)^2. \tag{12}$$

We can minimize this by setting the following system of equations

$$\left. \frac{\partial SSE}{\partial \beta_i} \right|_{i=1,2,3} = 0. \tag{13}$$

This is trivial enough and can be solved to derive:

$$\beta_1 = 525, \tag{14}$$
$$\beta_2 = 175, \tag{15}$$
$$\beta_3 = 250. \tag{16}$$

2.2 Vector and Matrix Notation

Let $\mathbf{Y}$ be a column vector with components corresponding to the observed values for the response variable:

$$\underline{Y} = \begin{bmatrix} Y_1 \\ Y_2 \\ Y_3 \\ Y_4 \end{bmatrix} = \begin{bmatrix} 800 \\ 500 \\ 400 \\ 200 \end{bmatrix} \tag{17}$$

Next, let $\underline{X}_1$, $\underline{X}_2$, $\underline{X}_3$ denote the column vectors with components equal to the observed values for the respective indicator variables (eg the ith element of $\underline{X}_1$ is 1 when the ith observation is male, and 0 if female):

$$\underline{X}_1 = \begin{bmatrix} 1 \\ 1 \\ 0 \\ 0 \end{bmatrix}, \qquad \underline{X}_2 = \begin{bmatrix} 0 \\ 0 \\ 1 \\ 1 \end{bmatrix}, \qquad \underline{X}_3 = \begin{bmatrix} 1 \\ 0 \\ 1 \\ 0 \end{bmatrix}. \tag{18}$$

We can then finally denote $\underline{\beta}$ and $\underline{\epsilon}$ as

$$\underline{\beta} = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \end{bmatrix}, \qquad \underline{\epsilon} = \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \epsilon_3 \\ \epsilon_4 \end{bmatrix}. \tag{19}$$

The system of equations can now take the form:

$$\underline{Y} = \beta_1 \underline{X}_1 + \beta_2 \underline{X}_2 + \beta_3 \underline{X}_3 + \underline{\epsilon}. \tag{20}$$

We can further simplify this by aggregating $\underline{X}_1$, $\underline{X}_2$, $\underline{X}_3$ into a single matrix $\mathbf{X}$. This is called the **design matrix** and would (in this example) be defined as:

$$\mathbf{X} = \begin{bmatrix} 1 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 1 \\ 0 & 1 & 0 \end{bmatrix}. \tag{21}$$

We now find that the system of equations takes the form

$$\underline{Y} = \mathbf{X} \cdot \underline{\beta} + \underline{\epsilon}. \tag{22}$$

2.3 Assumptions

From the vectorized forms, we can determine the two elements required for a linear model

1. a set of assumptions about the relationship between $\underline{Y}$ and the predictor variables,

2. an objective function which is to be optimized in order to solve the problem. It can be shown outside of this material that the parameters which minimize the sum of squared error (SSE) also maximize the likelihood.

Finally, we now make three explicit assumptions about the model:

- **(LM1)** *Random component*: Each component of $\underline{Y}$ is independent and is Normally distributed. The mean, μ_i, of each component is allowed to differ, but they all have common variance σ^2;

- **(LM2)** *Systematic component*: The p covariates are combined to give the linear predictor $\underline{\eta}$:

$$\underline{\eta} = \mathbf{X} \cdot \underline{\beta}; \tag{23}$$

- **(LM3)** *Link function*: The relationship between the random and systematic components is specified via a link function. In the linear model the link function is equal to the identity function so that:

$$\mathbb{E}[\underline{Y}] \equiv \underline{\mu} = \underline{\eta}. \tag{24}$$

3 Generalized Linear Model

A GLM allows for a non-linear dependence through what's called a link function, g:

$$\mathbb{E}[\mathbf{Y}] = g^{-1}(\mathbf{X} \cdot \beta). \tag{25}$$

Here, Y is the dependent variable, X the independent variables and β the parameters that are fitted through regression.

3.1 Assumptions

There are three main assumptions that will form the foundation of our model building:

- **Policy independence:** For n considered policies, $X_1, \ldots, X_n$ are independent where X_i denotes the response for policy i.
  There are some situations that obviously don't abide by this assumption (two Hastings Direct customers claiming against one another), but the effect of neglecting this should be small.

- **Time independence:** Consider n disjoint time intervals. For any response type, let X_i denote the response in time interval i. Then $X_1 \ldots, X_n$ are independent.
  This essentially states that any amount of claims made in one time interval will not directly influence the number made in another. Although not entirely true, again it's a reasonable enough assumption to make in order to help simplify the calculations needed to create the statistical model.

- **Homogeneity:** Consider any two policies with the same exposure and the same

3.2 Simple examples

We will solve simple examples using the same observations previously discussed in sec. (2.2). The general procedure for solving a GLM involves the following:

1. Specify the matrix $\mathbf{X}$ and the vector of parameters $\underline{\beta}$;

2. Choose the error structure and link function;

3. Identify the log-likelihood function;

4. Take the logarithm to convert the product of many terms into a sum;

5. Maximize the logarithm of the likelihood function;

6. Compute the predicted values.

We will follow the following models:

- Normal error structure with an identity link function.

- Poisson error structure with a log link function.

- Gamma error structure with an inverse link function.

Note that for these examples, the prior weights ($\underline{\omega}$) and the offset term ($\underline{\xi}$) are set to 1 and 0 respectively in order to simplify these calculations.

3.2.1 Normal error structure with identity link function

This is a good introductory example as it also shows where the link is between the SSE and the likelihood function falls with regards to the classic linearization model.
The predicted values in this example will take the following form:

$$\mathbb{E}[\underline{Y}] = g^{-1}(X \cdot \underline{\beta}) = \begin{bmatrix} g^{-1}(\beta_1 + \beta_3) \\ g^{-1}(\beta_1) \\ g^{-1}(\beta_2 + \beta_3) \\ g^{-1}(\beta_2) \end{bmatrix} = \begin{bmatrix} \beta_1 + \beta_3 \\ \beta_1 \\ \beta_2 + \beta_3 \\ \beta_2 \end{bmatrix}. \tag{26}$$

Next, we can define our density function f using our prior assumptions that the mean is μ and our variance is σ^2:

$$f(y; \mu, \sigma^2) = \exp\left\{-\frac{(y_i - \mu_i)^2}{2\sigma^2} - \frac{1}{2}\ln\left(2\pi\sigma^2\right)\right\}. \tag{27}$$

Our likelihood function is then

$$L(y; \mu, \sigma^2) = \prod_{i=1}^{n} \exp\left\{-\frac{(y_i - \mu_i)^2}{2\sigma^2} - \frac{1}{2}\ln\left(2\pi\sigma^2\right)\right\}. \tag{28}$$

We know that maximizing the likelihood function is equivalent to maximizing the log-likelihood function:

$$\log(L(y; \mu, \sigma^2)) = l(y; \mu, \sigma^2) = \log\left(\prod_{i=1}^{n} \exp\left\{-\frac{(y_i - \mu_i)^2}{2\sigma^2} - \frac{1}{2}\ln\left(2\pi\sigma^2\right)\right\}\right). \tag{29}$$

We can use our fundamental laws of indices and products/summations to adapt

$$\prod_{i=1}^{n} e^{x_i} = e^{x_1} \cdot e^{x_2} \cdot \ldots \cdot e^{x_n} = e^{x_1 + x_2 + \ldots x_n} = e^{\sum_{i=1}^{n} x_i}, \tag{30}$$

and apply it to our log-likelihood found in eq. (29). This gives us

$$l(y; \mu, \sigma^2) = \sum_{i=1}^{n} -\frac{(y_i - \mu_i)^2}{2\sigma^2} - \frac{1}{2}\ln\left(2\pi\sigma^2\right). \tag{31}$$

Here, we insert our link function (identity link function in this case), $\mu_{ij} = \sum_j X_{ij}\beta_j$, to find

$$l(y; \mu, \sigma^2) = \sum_{i=1}^{n} -\frac{\left(y_i - \sum_j X_{ij}\beta_j\right)^2}{2\sigma^2} - \frac{1}{2}\ln\left(2\pi\sigma^2\right). \tag{32}$$

We now expand the right hand side in order to observe

$$l(y; \mu, \sigma^2) = -\frac{(800 - (\beta_1 + \beta_3))^2}{2\sigma^2} - \frac{(500 - \beta_1)^2}{2\sigma^2} - \frac{(400 - (\beta_2 + \beta_3))^2}{2\sigma^2} - \frac{(200 - \beta_2)^2}{2\sigma^2} - 2\ln(2\pi\sigma^2), \tag{33}$$

$$l^*(y; \mu, \sigma^2) = -\frac{(800 - (\beta_1 + \beta_3))^2}{2\sigma^2} - \frac{(500 - \beta_1)^2}{2\sigma^2} - \frac{(400 - (\beta_2 + \beta_3))^2}{2\sigma^2} - \frac{(200 - \beta_2)^2}{2\sigma^2}, \tag{34}$$

which is up to that constant we "removed" $(2\ln(2\sigma^2))$.

To maximize l^*, we take derivatives with respect to β_1, β_2 and β_3 and then setting them to zero.

$$\frac{\partial l^*}{\partial \beta_1} = 0 \implies \beta_1 + \beta_3 + \beta_1 = 800 + 500 + 1300; \tag{35}$$

$$\frac{\partial l^*}{\partial \beta_2} = 0 \implies \beta_2 + \beta_3 + \beta_2 = 400 + 200 = 600 \tag{36}$$

$$\frac{\partial l^*}{\partial \beta_3} = 0 \implies \beta_1 + \beta_3 + \beta_2 + \beta_3 = 800 + 400 = 1200. \tag{37}$$

You can now see how these equations are identical to those that would be found in the simple linear model. These again are solved to derive:

$$\beta_1 = 525; \tag{38}$$

$$\beta_2 = 175; \tag{39}$$

$$\beta_3 = 250. \tag{40}$$

These are the same as the solutions found in eq. (15-16). This in turn produces the following predicted values:

	Urban	Rural
Male	775	525
Female	425	175

Table 2: Predicted values using the Normal error structure with identity link function.

3.2.2 The Poisson error structure with a logarithmic link function

For our Poisson model, we begin with our predicted values:

$$\mathbb{E}[\underline{Y}] = g^{-1}(X \cdot \underline{\beta}) = \begin{bmatrix} g^{-1}(\beta_1 + \beta_3) \\ g^{-1}(\beta_1) \\ g^{-1}(\beta_2 + \beta_3) \\ g^{-1}(\beta_2) \end{bmatrix} = \begin{bmatrix} e^{\beta_1+\beta_3} \\ e^{\beta_1} \\ e^{\beta_2+\beta_3} \\ e^{\beta_2} \end{bmatrix}. \tag{41}$$

Using the Poisson distribution density function $(f(y; \mu) = e^{-\mu}\mu^y/y!)$, we can introduce the log-likelihood function as

$$l(y; \mu) = \sum_{i=1}^{n} \ln f(y; \mu) = \sum_{i=1}^{n} -\mu_i + y_i \ln \mu_i - \ln(y_i!). \tag{42}$$

We then introduce the link function which in this case is the logarithmic link function $\mu_i = \exp\left(\sum_j X_{ij}\beta_j\right)$. This reduces the log-likelihood function to

$$l(y; e^{X\beta}) = \sum_{i=1}^{n} -\exp\left(\sum_j X_{ij}\beta_j\right) + y_i \sum_j X_{ij}\beta_j - \ln(y_i!). \tag{43}$$

We can (however tedious) expand this into

$$l(y; \mu) = -e^{(\beta_1+\beta_3)} + 800 \cdot (\beta_1 + \beta_3) - \ln 800! - e^{\beta_1} + 500 \cdot \beta_1 - \ln 500!$$
$$- e^{(\beta_2+\beta_3)} + 400 \cdot (\beta_2 + \beta_3) - \ln 400! - e^{\beta_2} + 200 \cdot \beta_2 - \ln 200!. \tag{44}$$

Ignoring the constant, the following function is to be maximized:

$$l^*(y; \mu) = -e^{(\beta_1+\beta_3)} + 800 \cdot (\beta_1 + \beta_3) - e^{\beta_1} + 500 \cdot \beta_1 - e^{(\beta_2+\beta_3)} + 400 \cdot (\beta_2 + \beta_3) - e^{\beta_2} + 200 \cdot \beta_2. \tag{45}$$

To maximize l^*, the derivatives with respect to $\beta_{1,2,3}$ are set to zero and the following three equations are derived:

$$\frac{\partial l^*}{\partial \beta_1} = 0 \implies e^{\beta_1} \cdot (e^{\beta_3} + 1) = 1300, \tag{46}$$

$$\frac{\partial l^*}{\partial \beta_2} = 0 \implies e^{\beta_2} \cdot (e^{\beta_3} + 1) = 600, \tag{47}$$

$$\frac{\partial l^*}{\partial \beta_3} = 0 \implies e^{\beta_3} \cdot (e^{\beta_1} + e^{\beta_2}) = 1200. \tag{48}$$

We can solve this system of equations to derive the following parameter estimates:

$$\beta_1 = 6.1716, \tag{49}$$

$$\beta_2 = 5.3984, \tag{50}$$

$$\beta_3 = 0.5390, \tag{51}$$

which in turn produces the following predicted values:

	Urban	Rural
Male	821.1	479.0
Female	378.9	221.1

Table 3: Predicted values using the Poisson error structure with logarithmic link function.

3.2.3 The Gamma error structure with an inverse link function

A Gamma distribution is substantially more difficult that the prior two examples, so feel free to skip it if it is out of the scope of your prior knowledge.

For the Gamma error structure, the predicted values take the form:

$$\mathbb{E}[\underline{Y}] = g^{-1}(X \cdot \underline{\beta}) = \begin{bmatrix} g^{-1}(\beta_1 + \beta_3) \\ g^{-1}(\beta_1) \\ g^{-1}(\beta_2 + \beta_3) \\ g^{-1}(\beta_2) \end{bmatrix} = \begin{bmatrix} (\beta_1 + \beta_3)^{-1} \\ (\beta_1)^{-1} \\ (\beta_2 + \beta_3)^{-1} \\ (\beta_2)^{-1} \end{bmatrix}. \tag{52}$$

The gamma error structure has the following density function:

$$f(x; \mu, \phi) = \frac{x^{-1}}{\Gamma(1/\phi)} \left(\frac{x}{\mu\phi}\right)^{1/\phi} e^{\left(-\frac{x}{\mu\phi}\right)}. \tag{53}$$

It's log-likelihood function is derived as

$$l(x; \mu, \phi) = \sum_{i=1}^{n} \frac{1}{\phi}\left(\ln\frac{x_i}{\mu_i} - \frac{x_i}{\mu_i}\right) - \ln x_i - \frac{\ln\phi}{\phi} - \ln\Gamma\left(\frac{1}{\phi}\right). \tag{54}$$

We can then introduce our inverse link function, $\mu_i = 1/(\sum_j X_{ij}\beta_j)$, to reduce the log-likelihood function to

$$l(x; 1/X\beta, \phi) = \sum_{i=1}^{n} \frac{1}{\phi}\left(\ln\left(x_i \cdot \sum_{j=1}^{p} X_{ij}\beta_j\right) - x_i \cdot \sum_{j=1}^{p} X_{ij}\beta_j\right) - \ln x_i - \frac{\ln\phi}{\phi} - \ln\Gamma\left(\frac{1}{\phi}\right). \tag{55}$$

This is then expanded to

$$\begin{aligned} l(x; \mu) = & \frac{1}{\phi}\left(\ln\left(800 \cdot (\beta_1 + \beta_3)\right) - 800 \cdot (\beta_1 + \beta_3)\right) - \ln 800 - \frac{\ln\phi}{\phi} - \ln\Gamma\left(\frac{1}{\phi}\right) \\ & + \frac{1}{\phi}\left(\ln\left(500 \cdot (\beta_1)\right) - 500 \cdot (\beta_1)\right) - \ln 500 - \frac{\ln\phi}{\phi} - \ln\Gamma\left(\frac{1}{\phi}\right) \\ & + \frac{1}{\phi}\left(\ln\left(400 \cdot (\beta_2 + \beta_3)\right) - 400 \cdot (\beta_2 + \beta_3)\right) - \ln 400 - \frac{\ln\phi}{\phi} - \ln\Gamma\left(\frac{1}{\phi}\right) \\ & + \frac{1}{\phi}\left(\ln\left(200 \cdot (\beta_2)\right) - 200 \cdot (\beta_2)\right) - \ln 200 - \frac{\ln\phi}{\phi} - \ln\Gamma\left(\frac{1}{\phi}\right). \end{aligned} \tag{56}$$

If we again ignore the constant terms (after multiplying by ϕ), we then maximize the following function:

$$\begin{aligned} l^*(x; \mu) = & \ln\left(800 \cdot (\beta_1 + \beta_3)\right) - 800 \cdot (\beta_1 + \beta_3) + \ln\left(500 \cdot \beta_1\right) - 500 \cdot \beta_1 \\ & + \ln\left(400 \cdot (\beta_2 + \beta_3)\right) - 400 \cdot (\beta_2 + \beta_3) + \ln\left(200 \cdot \beta_2\right) - 200 \cdot \beta_2. \end{aligned} \tag{57}$$

To maximize l^*, the derivatives with respect to $\beta_{1,2,3}$ are set to zero and the following three equations are derived:

$$\frac{\partial l^*}{\partial \beta_1} = 0 \implies \frac{1}{\beta_1 + \beta_3} + \beta_1 = 1300, \tag{58}$$

$$\frac{\partial l^*}{\partial \beta_2} = 0 \implies \frac{1}{\beta_2 + \beta_3} + \frac{1}{\beta_2} = 600, \tag{59}$$

$$\frac{\partial l^*}{\partial \beta_3} = 0 \implies \frac{1}{\beta_1 + \beta_3} + \frac{1}{\beta_2 + \beta_3} = 1200. \tag{60}$$

Solving these gives the following:

$$\beta_1 = 0.00223804, \tag{61}$$
$$\beta_2 = 0.00394964, \tag{62}$$
$$\beta_3 = -0.00106601, \tag{63}$$

which results in the following predicted values:

	Urban	Rural
Male	853.2	446.8
Female	346.8	253.2

Table 4: Predicted values using the Gamma error structure with inverse link function.

3.3 The Tweedie Distribution

Modelling premiums can often be problematic, with one main problem being the high frequency of no-claims causing a large spike at zero and a wide range of amounts following on from this. May other more commonly taught distributions don't accurately account for this. In order to understand how this distribution is derived, we must break down what these distributions are attempting to model.

We essentially want to model the number of claims received, N, against the size of these claims, Z_i. Respectively, these can be accurately modelled through a Poisson distribution with mean $\lambda_i w_i$ ($N \sim P(\lambda_i w_i)$) and a gamma distribution with mean $\tau_i = \alpha \beta_i$ and shape parameter α (β represents a rate parameter of our distribution). Therefore, we define our random variable Z as the following:

$$Z = \begin{cases} 0, & \text{if } N = 0, \\ \sum_{i=1}^{N} Z_i, & \text{if } N > 0. \end{cases} \tag{64}$$

This will evidently take into account the fact that if $N = 0$ (No policy claims made), then the claim amount will *also* be zero.

3.4 Large Datasets and Numerical Techniques

As the number of observations being modelled increases, the practicality of finding values of $\underline{\beta}$ which maximize likelihood using an explicit technique becomes less and less. Although there is a general case for solving the GLM problem using an assumed exponential distribution, we will now turn to numerical techniques to aid us in approaching our optimum.

References

[1] Duncan Anderson et al. "A Practitioners Guide to Generalized Linear Models". In: Towers Watson. 2008.

[2] A. Géron. *Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems*. O'Reilly Media, 2019. ISBN: 9781492032618. URL: `https://books.google.co.uk/books?id=HHetDwAAQBAJ`.

[3] Samuel Tober. *Basics of Insurance Pricing*. 2020. URL: `https://towardsdatascience.com/basics-of-insurance-pricing-47243c2630b9`.