

**RATEMAKING:  
FROM BAILEY AND SIMON (1960)  
TO GENERALISED LINEAR  
REGRESSION MODELS**

*Ben Zehnwirth*

## ABSTRACT

The Bailey and Simon (1960) and Bailey (1963) papers on class and merit rating discuss models and estimation criteria in a non-probabilistic framework. It has been shown, for example, Van Eeghen et al (1983), that the Bailey and Simon criterion of class balance is equivalent to maximum likelihood estimation of a claim frequency model with Poisson distributed claim numbers. It turns out that the Poisson based model is part of a large body of recently developed statistical methodology known as Generalised Linear Regression Modelling.

Indeed, the Bailey and Simon papers provide the motivation for generalised linear regression models. By applying the regression framework some results are developed that relate the various estimation criteria and a number of extensions are given for the case where the condition of class balance is not appropriate as a result of lack of credibility for some of the classes. The regression framework moreover facilitates the consideration of a much wider family of models than that considered by Bailey and Simon. Generalised regression models are also motivated and indeed introduced as an extension to the classical normal based regression models. Many of the benefits afforded by regression modelling are discussed.

## 1. INTRODUCTION AND SUMMARY

The present paper has two important objectives.

- Relate the work of Bailey and Simon [1] and subsequently Bailey [2] to some modern statistical technology that falls under the heading of Generalised Linear Regression Models.
- Demonstrate that Regression modelling is very powerful and accordingly facilitates a number of significant extensions of the works of Bailey and Simon.

There are two principles in insurance that play an important role in ratemaking.

(P1): Each individual risk should pay for its own claims.

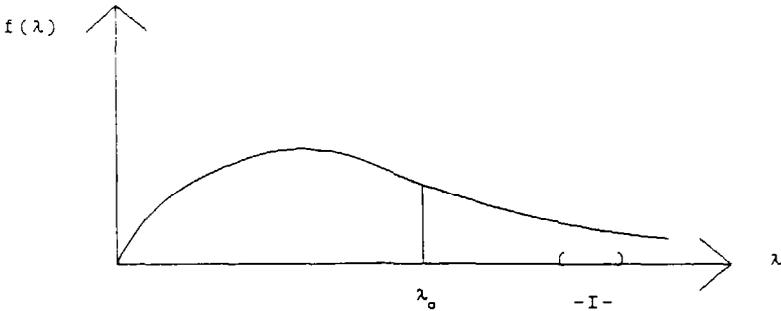
But, on the other hand, the basic quintessential idea of insurance is that

(P2): A given portfolio forms a collective of risks "with equal rights", each of them paying the same pure (collective) premium.

Both principles (P1) and (P2) are in agreement for an homogeneous collective.

The pure premium differentials should be directly related to the 'type and degree' of heterogeneity. How do we measure the heterogeneity in the experience?

Consider the following example. Suppose we have a portfolio of motor car policyholders each having a claim rate  $\lambda$ . The parameter  $\lambda$  is the true claim intensity and is unobservable. Since  $\lambda$  is different for each policyholder we could assume that  $\lambda$  has a distribution  $f(\lambda)$  depicted below.



The mean of  $\lambda$ , viz.,  $\lambda_0 = E[\lambda]$ , is the pure premium for the collective (or portfolio). However, if we consider a sub-population of the collective that has  $\lambda$ -values in the interval  $I$ , then  $\lambda_0$  is biased downwards for the sub-population. The pure premium for the sub-population (or sub-class) is the mean of the  $\lambda$ -values that lie in the interval  $I$ .

Quoting from Bailey [2]:

*"The more we sub-divide the data, the less biased are the resulting rates for each class".*

But we add, that the finer the sub-divisions the "less credible" may be the individual experience for that individual sub-class.

By "less credible" is meant, that an appropriate model, only based on the individual experience, leads to much uncertainty. We discuss a similar example in Section 2 that provides some compelling reasons for regression.

The paper is organised as follows.

Section 2 examines a ratemaking example involving only one risk (tariff) variable. The power and usefulness of a regression model is illustrated, especially in the context of lack of credibility for each level (sub-class) of the

tariff variable. Regression is not a single method. Regression is a modelling approach.

Regression has been used as the scientific method in scientific fields ranging from physics, engineering to psychology. We use regression to extract information from data. We use regression to determine whether the data we have supports our hypotheses.

Section 3 presents well known discrete distributions and their relationships that prove useful in the analysis of categorical data.

In Section 4 we visit a ratemaking example involving two risk variables discussed by Bailey and Simon [1] and Bailey [2].

It is shown that the Bailey and Simon criterion of "class balance" is equivalent to the "criterion of average bias = 0", which in turn is equivalent to "maximum likelihood estimation assuming Poisson counts", which is equivalent to "maximum likelihood estimation for the multinomial distribution".

Moreover, the multiplicative model (or hypothesis) is equivalent to the hypothesis of "independence of the two tariff variables", alternatively, "no interaction" or "no association".

By applying the regression concepts of Section 2 we extend the Bailey and Simon multiplicative model to situations where "class balance" is inappropriate as some classes possess loss experience that is not sufficiently credible for the corresponding class risk parameter.

Brown [3] also introduces a number of models within the Generalised Regression Modelling framework.

Section 5 revisits the well known Normal based regression model as a lead up to the Poisson based regression models of Section 6.

The foregoing sections set up the ground work for the introduction of Generalised Linear Regression Models.

The generalisations from the Normal based regression are in three directions

- \* distributions other than the Normal including Poisson, Binomial and Gamma;
- \* the response variable  $Y$  need to be continuous;
- \* link functions that are not the identity link functions.

Section 8 discusses inference and hypothesis testing in the context of generalised linear models and demonstrates, using an example, that the difference in scaled deviance is the generalisation of the  $F$ -ratio statistic from Normal-based regression models to generalised regression models.

Section 9 provides a summary of conclusions.

The paper is, by and large, pedagogical and is only intended to give the reader a glimpse of the current available statistical technology in the belief that with the advent of fast computers actuaries should be cognisant of it.

There are a number of statistical packages that the reader will find useful for Generalised Linear Regression Modelling, for example, BMDP (Biomedical Computer Programs, UCLA), SAS (SAS Institute, Rayleigh, North Carolina) and GLIM (Numerical Algorithms Group, Oxford, U.K.).

**2. REGRESSION - A RATEMAKING EXAMPLE INVOLVING LACK OF SUB-CLASS BALANCE**

Bailey and Simon [1] argue that a necessary condition to be satisfied by a satisfactory rating system is that for large groups of insureds, (total) premium is approximately equal to (total) observed losses. (This condition is related to the Law of Large numbers). The condition is known as the criterion of class balance.

For the ratemaking example formulated below we have lack of sub-class balance as a result of some sub-classes loss experiences being non-credible. We solve this problem by relating the sub-classes through a regression model.

Suppose  $X$  is a tariff or risk variable, e.g.  $X$  is age.

Aim: Determination (or estimation and uncertainty) of loss size distributions for ten levels (values) of  $X$ , viz.,  $x_1, x_2, \dots, x_{10}$ .

Question: Is there any heterogeneity between the sub-classes  $x_1, x_2, \dots, x_{10}$ ? If the answer is in the affirmative, what is its nature?

Data: Losses  $Y_{11}, \dots, Y_{1n_1}$  corresponding to sub-class (or risk level)  $x_1$ .

$x_1$	$x_2$	$x_{10}$
$Y_{11}$	$Y_{21}$	$Y_{101}$
.	.	.
.	.	.
$Y_{1n_1}$	.	$Y_{10n_{10}}$
	$Y_{2n_2}$	

Problem: For some sub-classes  $x_i$ , the sample size  $n_i$  is too small to be able to fit or estimate a loss size distribution with reasonable accuracy, (e.g.  $n_i \leq 2$ ) to that sub-class independently.

The sub-classes cannot be analysed separately as for some sub-classes there is insufficient loss experience. Equivalently, some sub-classes are not sufficiently credible. **SO, WE NEED A MODEL.**

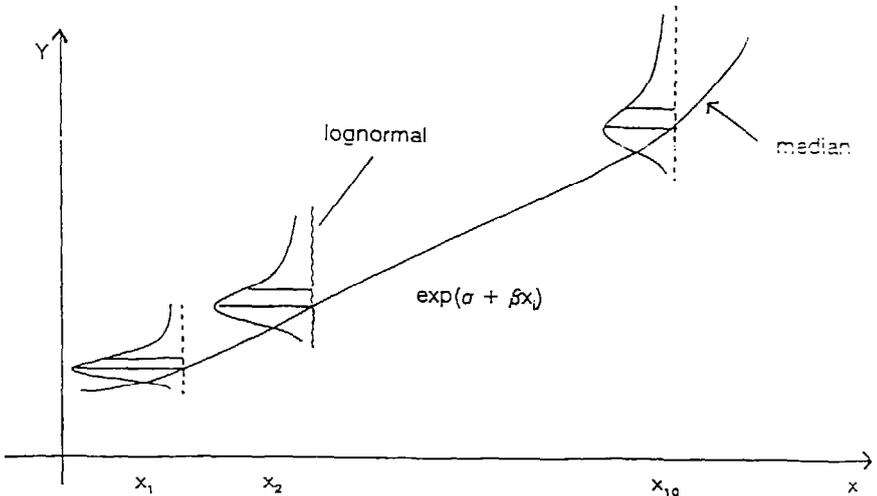
Based on some preliminary diagnostic analyses of the data the following model may suggest itself.

$$Y_{ij} = \exp(\alpha + \beta x_i) \epsilon_i' \quad (2.1)$$

Let  $U_{ij} = \log(Y_{ij})$  and  $\epsilon_i = \log \epsilon_i'$ , then

$$U_{ij} = \alpha + \beta x_i + \epsilon_i \quad (2.2)$$

We also assume that  $\epsilon_i \sim N(0, \sigma^2)$ , so that  $\epsilon_i'$  is lognormally distributed.



. Fig 2.1

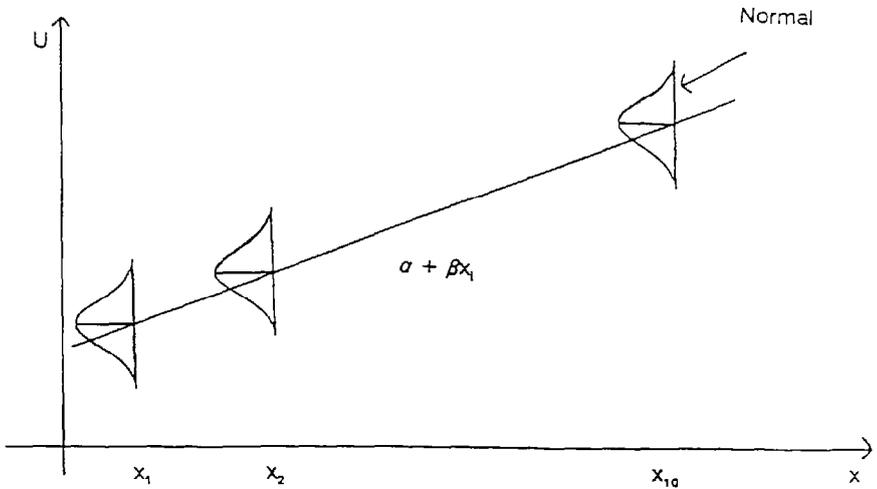


Fig 2.2

Fig 2.1 represents a display of equation (2.1) and Fig 2.2 a display of equation (2.2). Fig 2.1 depicts the regression model on a \$ scale whereas Fig 2.2 depicts the same model on a log scale.

Every regression model contains assumptions or information.

### Assumptions

- (A1): Loss distributions are lognormal (equivalently, normal on a log scale).
- (A2): Constant variance (on log scale) of Normal distributions.
- (A3): Mean of  $U = \log Y$  is linear in  $x$ .

The model relates the loss distributions of the different sub-classes. It is a probabilistic, equivalently, a stochastic model.

The model parameters  $(\alpha, \beta, \sigma^2)$  are estimated using maximum likelihood estimation theory. We emphasise that **THE ASSUMPTIONS (A1), (A2) AND (A3) MUST BE TESTED.**

The assumptions that apply to the model must also apply to the data. Otherwise, all our subsequent computations are meaningless. If none of the assumptions are violated by the data then we have extracted much information. Let's see now. We have an estimate of the loss size distribution for sub-class  $x_j$ , for example, even though we only have two observed losses ( $n_j = 2$ ) corresponding to this sub-class. This information is extracted from the model, since it relates the loss size distributions of the different sub-classes.

Note that the regression model comprises the lognormal distributions whose median is related by  $\exp(\alpha + \beta x_j)$ . So regression modelling is much more than just estimating the curve  $\exp(\alpha + \beta x_j)$ .

### **REGRESSION IS THE ESTIMATION OF DISTRIBUTIONS AND THEIR RELATIONSHIPS.**

Suppose assumptions (A1), (A2) and (A3) are not violated by the data and when we test the hypothesis  $H:\beta=0$  we find that we do not have sufficient evidence to reject it. We can therefore conclude that the sub-classes are indeed homogeneous in terms of loss size distributions.

### **WE ALSO USE REGRESSION MODELLING TO DETERMINE WHICH HYPOTHESES THE DATA SUPPORT.**

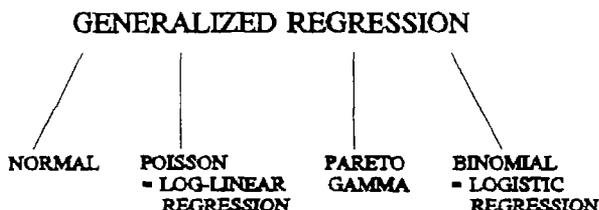
In the case of Normal distributions, maximum likelihood estimation theory is equivalent to least squares estimation theory.

Indeed, the Normal based regression has been the basis of much of statistical modelling until the early 1970's when regression modelling was extended to what is now called:

### **GENERALISED REGRESSION MODELLING**

There are essentially two extensions in the generalisation.

1. From Normal distributions to other distributions including Gamma, Poisson, Pareto and Binomial.
2. From a continuous response variable  $Y$  to categorical response variables.



We will show in subsequent sections that the models and estimation criteria of Bailey and Simon [1] and Bailey [2] are related to log-linear regression.

**REGRESSION IS A VERY POWERFUL TOOL IN RATEMAKING.**

In a way, what we did in the ratemaking example, when faced with insufficient loss experience for each sub-class, is to employ a type of credibility analysis.

Indeed, regression may be viewed as a type of credibility analysis. See the Zehnwirth notes on credibility presented at the CAS Ratemaking Seminar held in Chicago in 1991.

### 3. PRELIMINARIES ON DISTRIBUTIONS

In the present section we develop some basic distributional theory that will prove useful in relating a number of ratemaking criteria discussed by Bailey and Simon.

#### 3.1 BINOMIAL

We say that  $Y$  has a Binomial distribution with index  $n$  and parameter  $\theta$ , equivalently,  $Y \sim \text{Bin}(n, \theta)$ , if  $Y$  represents the total number of 'successes' in  $n$  independent Bernoulli trials where the probability of success at each trial is  $\theta$ . A Bernoulli trial has only two possible outcomes 'Success' or 'Failure' = 'Not Success'.

#### 3.2 MULTINOMIAL

The Multinomial distribution is an extension of the Binomial where at each trial there are  $k$  possible outcomes, where  $k \geq 2$ .

Prob (jth outcome) =  $\theta_j : \sum \theta_j = 1$ .

Let  $Y_j$  denote the total number of outcomes "j" in  $n$  independent trials.

We write  $(Y_1, \dots, Y_k) \sim \text{Multi}(n; \theta_1, \dots, \theta_k)$ .

Note  $\sum Y_j = n$  and  $\sum \theta_j = 1$ .

$$\text{Prob} [Y_1 = y_1, \dots, Y_k = y_k] = \frac{n!}{y_1! \dots y_k!} \theta_1^{y_1} \dots \theta_k^{y_k} .$$

Note  $Y_j \sim \text{Bin}(n, \theta_j)$  so that

$$E[Y_j] = n\theta_j \text{ and } \text{Var}[Y_j] = n(\theta_j)(1-\theta_j).$$

### 3.3 POISSON

The Poisson can be thought of as a limiting form of the Binomial distribution. It represents the distribution of the number of events in a fixed time interval if it is assumed that (i) each 'small' time interval may be regarded as a Bernoulli trial - either one event occurs in the time interval or it doesn't, (ii) the probability that an event occurs in a small time interval is proportional to the length of the interval and (iii) non-overlapping time intervals are independent.

We say  $Y \sim \text{Po}(\lambda)$  if

$$P(Y=y) = e^{-\lambda} \frac{\lambda^y}{y!}, \quad y=0,1,2\dots$$

We have,

$$E[Y] = \text{Var}[Y] = \lambda.$$

We now state two results and only prove the second.

(R1): If  $Y_1, Y_2, \dots, Y_k$  are independent such that  $Y_j \sim \text{Po}(\lambda_j)$  then

$$\sum Y_j \sim \text{Po}(\sum \lambda_j).$$

(R2): If  $Y_1, Y_2, \dots, Y_k$  are independent such that  $Y_j \sim \text{Po}(\lambda_j)$  then

$$(Y_1, \dots, Y_k \mid \sum Y_j) \sim \text{Multi}(\sum Y_j; \theta_1, \dots, \theta_k)$$

$$\text{where } \theta_j = \frac{\lambda_j}{\sum \lambda_l}.$$

So, if we know the sum of Poisson counts, then the distribution of the individual counts is Multinomial where the probabilities are equal to the relative Poisson means. This result is almost intuitive if we bear in mind that the

Poisson is a limit of Binomial that comprises independent Bernoulli trials.

Proof:  $\Pr(Y_1=y_1, \dots, Y_k=y_k \mid \sum Y_j=n)$

$$= \frac{\Pr((Y_1=y_1, \dots, Y_k=y_k) \cap (\sum Y_j=n))}{\Pr(\sum Y_j=n)}$$

$$= \frac{\Pr((Y_1=y_1, \dots, Y_{k-1}=y_{k-1}, Y_k=n-\sum_1^{k-1} y_j))}{\Pr(\sum Y_j=n)}$$

Bearing in mind that  $\sum Y_j$  is Poisson from (R1), we have the conditional probability

$$\frac{\frac{e^{-\lambda_1} \lambda_1^{y_1}}{y_1!} \dots \frac{e^{-\lambda_{k-1}} \lambda_{k-1}^{y_{k-1}}}{y_{k-1}!} \cdot \frac{e^{-\lambda_k} \lambda_k^{n-\sum_1^{k-1} y_j}}{(n-\sum_1^{k-1} y_j)!}}{e^{-\sum \lambda_j} \frac{(\sum \lambda_j)^n}{n!}}$$

$$= \frac{n!}{y_1! \dots y_k!} \theta_1^{y_1} \dots \theta_k^{y_k} \quad ,$$

where  $y_k = n - \sum_1^{k-1} y_j$  and  $\theta_j = \frac{\lambda_j}{\sum \lambda_j}$ .

The proof is now complete.

## 4. BAILEY [2] - MOTIVATION FOR LOG-LINEAR REGRESSION MODELS

### 4.1 INTRODUCTION

Let us take out an excerpt from Bailey's [2] paper.

*"In making rates for insurance we are faced with the problem that there are many different classes of risks with a different rate for each class, and that no one class by itself has a sufficient volume of premium and losses to give a reliable basis for the rate of that class".*

Is this not the kind of problem we considered in Section 2? And we solved it by using an appropriate regression model. Indeed, Bailey and Simon also solve this problem by indirectly applying regression because class balance is equivalent to Poisson regression. The following cross-classification example is considered by Bailey and Simon [1].

## 4.2 CROSS-CLASSIFICATION EXAMPLE

Consider two categorical tariff variables A and B. Variable A is at I levels whereas B is at J levels.

We can construct the following two-way table.

A	B	1	2	3	j	...	J	
1								$Y_{1\cdot}$
2								$\vdots$
3								$\vdots$
i					$Y_{ij}$			$\vdots$
$\vdots$								$\vdots$
I								$Y_{\cdot I}$
		$Y_{\cdot 1}$						$Y_{\cdot J}$
								$Y_{\cdot\cdot}$

We denote by  $Y_{ij}$  a count representing the total number of claims corresponding to level  $i$  of variable A and level  $j$  of variable B.

$$Y_{i\cdot} = \sum_j Y_{ij} = \text{TOTAL OF ROW } i \\ = \text{TOTAL FOR "CLASS" } i \text{ OF } I$$

$$Y_{\cdot j} = \sum_i Y_{ij} = \text{TOTAL OF COLUMN } j \\ = \text{TOTAL FOR "CLASS" } j \text{ OF } J$$

$$Y_{\cdot\cdot} = \sum_j \sum_i Y_{ij} = \sum_i Y_{i\cdot} = \sum_j Y_{\cdot j} = \text{GRAND TOTAL}$$

Let's assume that the counts  $Y_{ij} \sim \text{Po}(\lambda_{ij})$  and are independent. Without loss of generality, we also assume the exposure  $n_{ij} = 1$ , for all cells  $(i,j)$ .

The parameter  $\lambda_{ij}$  represents the mean claim rate for cell  $(i,j)$ .

Now, most of the cells (sub-classes) are not sufficiently credible in the sense that the individual "loss" experience  $Y_{ij}$  is not a reliable estimate of the corresponding risk parameter  $\lambda_{ij}$ .

Our assumed model currently has  $I \times J$  parameters ( $\lambda_{ij}$ ), too many to be estimated from the 'sparse' data. Let's postulate a model that has fewer parameters.

$$H: \lambda_{ij} = \lambda_i^I \cdot \lambda_j^J \quad .$$

This is called a multiplicative model. It appears that this model has  $I + J$  parameters. Actually, it only has  $I + J - 1$  parameters. This will be explained later.

Now, Bailey and Simon [1] and Bailey [2] argued that the "rate should be balanced in each class and in total".

That is, each class experience is sufficiently credible to estimate the corresponding class parameter.

The class parameter for level  $i$  of variable  $I$  is

$$\begin{aligned} & \sum_j \lambda_i^I \lambda_j^J \\ &= \lambda_i^I \sum_j \lambda_j^J \quad . \end{aligned}$$

The class parameter represents the mean (total) number of claims for the class. Its estimate is the sample total  $Y_{i\cdot}$ , also called the marginal row total.

So, for class i of I we have

$$Y_i = \sum_j \hat{\lambda}_i' \hat{\lambda}_j'' = \hat{\lambda}_i' \sum_j \hat{\lambda}_j'' \quad (4.2.1)$$

Similarly, for class j of J we have

$$Y_j = \sum_i \hat{\lambda}_i' \hat{\lambda}_j'' = \hat{\lambda}_j'' \sum_i \hat{\lambda}_i' \quad (4.2.2)$$

The  $\hat{\phantom{x}}$  above the parameter denotes an estimate.

See Section 4.6 for a discussion of the situation where  $Y_j$  is not credible for the corresponding class mean parameter. That is, we cannot apply "class balance".

Equations (4.2.1) and (4.2.2) are the equations of "class balance". They can be solved iteratively for  $\hat{\lambda}_i'$  and  $\hat{\lambda}_j''$ . It is straightforward to see that a solution is

$$\hat{\lambda}_i' = \frac{Y_i}{\sqrt{Y_{..}}} \quad , \quad \hat{\lambda}_j'' = \frac{Y_{.j}}{\sqrt{Y_{..}}} \quad (4.2.3)$$

Incidentally, if  $\hat{\lambda}_i'$  and  $\hat{\lambda}_j''$  are solutions to (4.2.1) and (4.2.2), then so are  $a\hat{\lambda}_i'$  and  $\hat{\lambda}_j''/a$  for any constant a.

From equations (4.2.1) and (4.2.2) we also have total balance, viz.,

$$Y_{..} = \sum_i \sum_j \hat{\lambda}_i' \hat{\lambda}_j'' \quad (4.2.4)$$

We now re-write the equations of "class balance" as follows:

$$\frac{\sum_i (Y_i - \lambda_i \bar{y}_i)}{\sum_i Y_i} = 0 \quad , \quad (4.2.5)$$

and

$$\frac{\sum_i (Y_i - \lambda_i \bar{y}_i)}{\sum_i Y_i} = 0 \quad . \quad (4.2.6)$$

So, the criterion of "class balance" is equivalent to the criterion of "Average bias = 0".

### 4.3 THE CROSS CLASSIFICATION EXAMPLE AS A RETROSPECTIVE STUDY

Suppose the preceding cross-classification example is regarded as a retrospective study. That is, on our books we have  $Y$  claims that we segment (equivalently, cross-classify) according to the levels of the two risk variables  $A$  and  $B$ .

So if  $Y_j \sim \text{Po}(\lambda_j)$  then  $(Y_{11}, Y_{12}, \dots, Y_{21}, \dots, Y_{\mu}) \mid Y_{\cdot}$  has a Multinomial distribution with cell probabilities  $\theta_{ij} = \frac{\lambda_{ij}}{\sum \sum \lambda_{ij}}$ . See (R2) of Section 3.3.

Assume the hypothesis:

$$H: \theta_{ij} = \theta_i' \cdot \theta_j'' \text{ is valid.}$$

That is, the probability of a claim being in cell  $(i,j)$  is the product of two probabilities.

The hypothesis is equivalent to (i) independence of the two tariff variables  $A$  and  $B$ , (ii) no association between the tariff variables  $A$  and  $B$  and (iii) no interaction between the tariff variables  $A$  and  $B$ .

So, under  $H$

$$\begin{aligned} \theta_{ij} &= \theta_i' \cdot \theta_j'' \\ &: \sum_i \theta_i' = \sum_j \theta_j'' = 1 \quad . \end{aligned} \tag{4.3.1}$$

The Multinomial likelihood function is  $L(Y_{11}, \dots; \theta_{11}, \theta_{12}, \dots)$

$$= \frac{Y_{\cdot}!}{Y_{11}! \dots Y_{\mu}!} \theta_{11}^{Y_{11}} \dots \theta_{\mu}^{Y_{\mu}} \quad , \tag{4.3.2}$$

where  $\theta_{ij} = \theta'_i \theta''_j$

$$\text{and } \sum_i \theta'_i = \sum_j \theta''_j = 1 \quad .$$

So, the log-likelihood  $l$  is given by

$$l = \text{constant} + \sum \sum Y_{ij} \log[\theta'_i \theta''_j] \quad . \quad (4.3.3)$$

The "constant" is a function of the  $Y$ 's but not of the  $\theta$ 's.

To determine the maximum likelihood estimators subject to the constraints, (4.3.1), we set up the Lagrangian.

$$\tau = l - \zeta_1 (\sum \theta'_i - 1) - \zeta_2 (\sum \theta''_j - 1) \quad . \quad (4.3.4)$$

To maximise the Lagrangian, we set the partial derivatives to zero.

$$\frac{\partial \tau}{\partial \theta'_i} = \frac{\partial \tau}{\partial \theta''_j} = \frac{\partial \tau}{\partial \zeta_1} = \frac{\partial \tau}{\partial \zeta_2} = 0 \quad .$$

Now,

$$\frac{\partial \tau}{\partial \theta'_i} = \sum_j Y_{ij} \frac{\theta''_j}{\theta'_i \theta''_j} - \zeta_1 = 0 \quad , \quad \text{whence}$$

$$\hat{\theta}'_i = Y_{i.} / \zeta_1 \quad .$$

$$\text{Similarly,} \quad \hat{\theta}''_j = Y_{.j} / \zeta_2 \quad .$$

$$\text{But} \quad \sum \hat{\theta}'_i = \sum \hat{\theta}''_j = 1 \quad , \quad \text{whence}$$

$$\zeta_1 = \zeta_2 = Y_{..} \quad .$$

$$\text{So, } \hat{\theta}_i = \frac{Y_{i.}}{Y_{..}} , \quad \hat{\theta}'_j = \frac{Y_{.j}}{Y_{..}} . \quad (4.3.5)$$

(almost obvious!)

Equations (4.3.5) can be recast

$$Y_{i.} = Y_{..} \sum_j \hat{\theta}'_j , \quad \hat{\theta}'_j = Y_{..} \hat{\theta}'_j / \sum_i \hat{\theta}'_j , \quad (4.3.6)$$

and,

$$Y_{.j} = Y_{..} \sum_i \hat{\theta}'_i , \quad \hat{\theta}'_i = Y_{..} \hat{\theta}'_i / \sum_j \hat{\theta}'_i . \quad (4.3.7)$$

Equivalently,

$$\hat{\theta}'_i = Y_{i.} / (\sum_j \hat{\theta}'_j) Y_{..} . \quad (4.3.8)$$

$$\hat{\theta}'_j = Y_{.j} / (\sum_i \hat{\theta}'_i) Y_{..} . \quad (4.3.9)$$

Equations (4.3.6) and (4.3.7) are the "equations of class balance". They are equivalent to equations (4.2.3). So, maximum likelihood estimation for the Multinomial is equivalent to the criterion of "class balance" (under the hypothesis of no interaction).

**4.4 THE CROSS-CLASSIFICATION EXAMPLE AS A PROSPECTIVE (POISSON) MODEL**

Consider the multiplicative model (Bailey and Simon [1]) hypothesis again, viz.,

$$H: \lambda_{ij} = \lambda_i' \cdot \lambda_j''.$$

The study here is regarded as an observational study where we begin recording the claims at the beginning of the year and conduct the analysis at the end of the year. Accordingly, we do not condition on the total count  $Y_{..}$ .

The likelihood function is given by

$$L(Y_{ij}; \lambda_i', \lambda_j'', i=1, \dots, I; j=1, \dots, J) \\ = \prod_{ij} e^{-\lambda_{ij}} \frac{(\lambda_i' \lambda_j'')^{Y_{ij}}}{Y_{ij}!} .$$

Therefore, the log-likelihood  $l$  is given by

$$l = \text{constant} + \sum_i \sum_j Y_{ij} [\log \lambda_i' \lambda_j''] - \sum_i \sum_j \lambda_i' \lambda_j'' .$$

So,  $\frac{\partial l}{\partial \lambda_i'} = \sum_j Y_{ij} / \lambda_i' - \sum_j \lambda_j'' = 0$ , which implies that

$$Y_{i.} = \hat{\lambda}_i' \sum_j \hat{\lambda}_j'' . \tag{4.4.1}$$

Similarly, we obtain

$$Y_{.j} = \hat{\lambda}_j'' \sum_i \hat{\lambda}_i' . \tag{4.4.2}$$

Observe that (4.4.1) and (4.4.2) are equivalent to (4.2.1) and (4.2.2) respectively.

So, maximum likelihood estimation for the multiplicative Poisson prospective model is also equivalent to the Bailey and Simon [1] criterion of class balance.

The maximum likelihood estimation criterion is also equivalent to what is sometimes termed "the method of marginal totals".

It is also important to recognise that maximum likelihood for the Multinomial retrospective multiplicative model yields the same results as maximum likelihood for the Poisson prospective multiplicative model. This is because both likelihoods only depend on the observed cell frequencies and the expected cell frequencies. One conclusion here is the interpretation of the Bailey and Simon [1] multiplicative model:

$$H: \lambda_{ij} = \lambda_i^I \cdot \lambda_j^{II}.$$

The assumption is equivalent to the assumption of "no interaction between rating variables A and B"  $\equiv$  "independence of rating variables A and B"  $\equiv$  "no association between variables A and B".

In the more recent statistical literature, the method of estimation based on equations (4.4.1) and (4.4.2) is called *iterative proportional fitting* procedure.

The *iterative proportional fitting* procedure involves adjusting the expected cell frequencies until they add up to the required marginal totals (at least to within some specified accuracy). We begin with some initial estimates of

$$\hat{\lambda}_1^{II}, \dots, \hat{\lambda}_J^{II}, \quad \text{say, adjust the estimates of } \hat{\lambda}_1^I, \dots, \hat{\lambda}_I^I \text{ using equation (4.4.1)}$$

so that the row margins are the sum of the cell frequencies in the corresponding row. Then using equation (4.4.2), adjust the estimates of

$$\hat{\lambda}_1^{II}, \dots, \hat{\lambda}_J^{II}, \quad \text{so that the column margins are the sum of the cell frequencies}$$

in the corresponding column, and so on.

### Summary

#### ITERATIVE PROPORTIONAL FITTING

- ≡ MAXIMUM LIKELIHOOD PROSPECTIVE MULTIPLICATIVE POISSON
- ≡ MAXIMUM LIKELIHOOD RETROSPECTIVE MULTINOMIAL
- ≡ CRITERION OF CLASS BALANCE
- ≡ AVERAGE BIAS = 0

AND the multiplicative hypothesis is

- ≡ INDEPENDENCE (OF THE TWO RATING VARIABLES)
- ≡ NO INTERACTION
- ≡ NO ASSOCIATION

## 4.5 NON-CONSTANT EXPOSURES

Suppose that for the cross-classification of the two categorical rating variables A and B, the exposures  $n_{ij}$  corresponding to cell (i,j) are not constant.

We have

$$Y_{ij} = \text{COUNT IN CELL } (i,j)$$

$$n_{ij} = \text{EXPOSURE (e.g. number of policyholders)}$$

$$\begin{aligned} \lambda_{ij} &= \text{CLAIM RATE} \\ &= \text{MEAN CLAIM/EXPOSURE UNIT.} \end{aligned}$$

The count  $Y_{ij} \sim \text{Po}(n_{ij}\lambda_{ij})$ , where  $n_{ij}$  is known and we also assume the hypothesis  $H: \lambda_{ij} = \lambda_i' \cdot \lambda_j''$ .

Equations (4.4.1) and (4.4.2) are now

$$Y_i = \hat{\lambda}_i' \sum_j n_j \hat{\lambda}_j'' \quad (4.5.1)$$

and

$$Y_j = \hat{\lambda}_j'' \sum_i n_i \hat{\lambda}_i' \quad (4.5.2)$$

So, these equations provide the maximum likelihood solutions for  $\lambda_i'$  and  $\lambda_j''$ . Here, there is not a close solution for the estimates  $\hat{\lambda}_i'$  and  $\hat{\lambda}_j''$ .

The summary of results given in the foregoing section still apply.

#### 4.6 EXTENSIONS TO SITUATIONS WHERE CLASS BALANCE IS NOT APPROPRIATE

Hitherto we have assumed, as Bailey and Simon do, the condition of class balance. That is, the loss experience of a class is sufficiently credible for the corresponding risk parameter for that class.

Suppose instead that some of the classes  $j$  of variable B are not credible in the sense that  $Y_j$  is not a reliable estimator of the corresponding class pure premium,

$$\lambda_j'' \sum_i \lambda_i' \quad .$$

Furthermore, let us assume that variable B (e.g. age) can also be regarded as a continuous variable. In place of having an "independent" parameter  $\lambda_j''$  for each class  $j$ , we could relate the parameters thus:

$$\log \lambda_j'' = \alpha + \beta x_j \quad (4.6.1)$$

Our multiplicative Poisson model now contains only  $I - 1 + 2$  parameters.

Of course, all the assumptions of this model would have to be tested to ensure they are supported by the data.

By way of summary, we have used regression to circumvent the problem of lack of balance.

We make the additional remark the even when class balance is appropriate, the model involving the relation (4.6.1) is better than the model involving  $J$  "independent"  $\lambda_j''$  parameters, provided it is not violated by the data. Parsimony is an important principle in modelling.

## 5. NORMAL BASED REGRESSION MODEL

The model:

$$\underline{Y} = X\underline{\beta} + \underline{\epsilon} : \underline{\epsilon} \sim N(0, \sigma^2 I), \quad (5.1)$$

has formed the basis of most statistical analyses of continuous data. Models ranging from "straight line" regression to analysis of variance and analysis of covariance possess the formulation (5.1).

The model can be written as follows:

$\underline{Y} = (Y_1, \dots, Y_n)$ ,  $\underline{\beta} = (\beta_1, \dots, \beta_p)'$  and  $\underline{\epsilon} = (\epsilon_1, \dots, \epsilon_n)'$ . The zero mean error terms  $\epsilon_1, \dots, \epsilon_n$  are independent from  $N(0, \sigma^2)$ . The Normal distribution is an integral part of the model. See Section 2.

We can write

$$\begin{aligned} Y_i &= \mu_i + \epsilon_i, \text{ where } \mu_i = E[Y_i], \text{ and} \\ \mu_i &= \underline{x}'_i \underline{\beta}. \end{aligned}$$

The vector  $\underline{x}'_i = (x_{i1}, \dots, x_{ip})$  is the "design" corresponding to  $Y_i$ . The design matrix  $X$  is given by

$$X = \begin{pmatrix} x_{11} & \dots & x_{1p} \\ \cdot & & \cdot \\ x_{n1} & \dots & x_{np} \end{pmatrix} = \begin{pmatrix} \underline{x}'_1 \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ \underline{x}'_n \end{pmatrix}.$$

The mean of  $Y_i$ ,  $\mu_i$  is linear in the parameters  $\beta_1, \dots, \beta_p$ , hence the term "linear" in linear regression.

### Example

Consider the example of Section 2.

$$U_{ij} = \alpha + \beta x_i + \epsilon_{ij} : \quad \epsilon_{ij} \sim N(0, \sigma^2)$$

$$i = 1, \dots, 10$$

$$j = 1, \dots, n_i$$

Using the notation of the current section,

$$\underline{Y} = (U_{11}, \dots, U_{1n_1}, U_{21}, \dots, U_{10,1}, \dots, U_{10n_{10}})'$$

$$\underline{\beta} = (\alpha, \beta)'$$
 and the design  $\underline{x}_{ij}'$  corresponding to  $U_{ij}$  is
$$\underline{x}_{ij}' = (1, x_i).$$

In order to find the least squares estimates of  $\alpha$  and  $\beta$  and other statistics using a statistical package the user typically has to specify the design matrix  $X$ .

Just a brief comment about the link function.

We can write  $g(\mu_i) = \underline{x}_i' \underline{\beta}$  where  $g$  a function of the mean  $\mu_i$  of  $Y_i$ , that links  $\mu_i$  with a linear function of parameters,  $\underline{x}_i' \underline{\beta}$ . Here  $g$  is the identity function. So, the link function for the Normal based regression is the identity function.

6. LOG-LINEAR (REGRESSION) MODEL

Consider the model

$$Y_{ij} \sim \text{Po}(\lambda_{ij}) \text{ with}$$

$$H: \lambda_{ij} = \lambda_i' \lambda_j'' \quad (6.1)$$

This suggests the log as the natural link function.

The link function relates  $\lambda_{ij}$  ( $= E[Y_{ij}]$ ), the mean of  $Y_{ij}$ , to a linear combination of parameters.

The above hypothesis regarding multiplicativity may be re-cast,

$$\log \lambda_{ij} = \mu + \alpha_i + \beta_j \quad (6.2)$$

$$= \tilde{x}'_{ij} \tilde{\beta} \quad (6.3)$$

So, the log of the mean is related linearly to a set of parameters  $\mu, \alpha_i$  and  $\beta_j$ .

The parameter vector  $\tilde{\beta} = (\mu, \alpha_1, \alpha_2, \dots, \alpha_i, \beta_1, \beta_2, \dots, \beta_j)'$ , and the design

$$\tilde{x}'_{ij} = (1, \underbrace{0, \dots, 0}_{i-1}, 1, \underbrace{0, \dots, 0}_{j-1}, \underbrace{0, \dots, 0}_{j-1}, \underbrace{0, \dots, 0}_{j-1})$$

The Poisson based regression model is called log-linear regression because the link function is the logarithm function.

We return to this log-linear regression model in Section 7.

The log-linear regression model of Section 4.6 is formulated as follows:

$$Y_{ij} \sim \text{Po}(\lambda_{ij}) \quad (6.4)$$

with

$$\log \lambda_{ij} = \mu + \alpha_i + \beta x_j \quad (6.5)$$

The design corresponding to  $Y_{ij}$  is

$$\tilde{x}'_{ij} = (1, \overbrace{0, \dots, 0}^{i-1}, 1, \overbrace{0, \dots, 0}^{I-1}, x_j)$$

and the parameter vector  $\underline{\beta}$  is

$$\underline{\beta} = (\mu, \alpha_1, \dots, \alpha_v, \beta)'$$

For those readers who have not used a statistical package to conduct log-linear analysis it is important to recognise that the design matrix is specified indirectly by just specifying the variables.

## 7. GENERALISED LINEAR REGRESSION MODELS

### 7.1 EXPONENTIAL FAMILY OF DISTRIBUTIONS

Recent advances in statistical theory and computer software allows us to use methods other for those developed for Normal based linear regression models (5.1).

There have been a number of generalisations or extensions:

- (1) response variable  $Y$  has a distribution other than the Normal;
- (2) response variable  $Y$  can be categorical;
- (3) the link function is not necessarily the identity function.

It turns out that some of the 'nice' properties of the normal distribution are shared by a wider class of distributions called the exponential family of distributions.

Consider a random variable  $Y$  that is discrete or continuous whose distribution depends on a single parameter of interest  $\theta$ .

The distribution belongs to the exponential family if it is of the form

$$f(y;\theta) = s(y) \tau(\theta) \exp[a(y)b(\theta)], \quad (7.1.1)$$

where  $a$ ,  $b$ ,  $s$  and  $\tau$  are known functions.

So,  $f(y;\theta)$  can be written

$$f(y;\theta) = \exp [a(y)b(\theta) + c(\theta) + d(y)], \quad (7.1.2)$$

where  $s(y) = \exp [d(y)]$  and  $\tau(\theta) = \exp [c(\theta)]$ .

If in equation (7.1.2)  $a(y) = y$ , then (7.1.2) is said to be the canonical form and  $b(\theta)$  is called the natural parameter.

Any other parameters apart from  $\theta$  are regarded as nuisance parameters.

Example 1: Poisson

$$\text{Suppose } f(y; \theta) = \frac{e^{-\theta} \theta^y}{y!}; y = 0, 1, 2, \dots$$

We can express  $f$  in exponential form as follows:

$$f(y; \theta) = \exp(y \log \theta - \theta - \log y!).$$

So, here  $a(y) = y$ ,  $b(\theta) = \log \theta$ ,  $c(\theta) = -\theta$  and  $d(y) = -\log y!$ . Since  $f$  is expressed in canonical form,  $\log \theta$  is the natural parameter for the Poisson distribution.

Example 2: Binomial

$$\text{Suppose } f(y; \theta) = \binom{n}{\theta} \theta^y (1-\theta)^{n-y}; y=0, 1, \dots, n.$$

Therefore,

$$\begin{aligned} f(y; \theta) &= \exp\left(y \log \theta + (n-y) \log(1-\theta) + \log \binom{n}{\theta}\right) \\ &= \exp\left(y \log \left(\frac{\theta}{1-\theta}\right) + n \log(1-\theta) + \log \binom{n}{\theta}\right) \end{aligned}$$

So, here  $a(y) = y$ ,  $b(\theta) = \log \frac{\theta}{1-\theta}$ ,  $c(\theta) = n \log(1-\theta)$  and  $\alpha(y) = \log \binom{n}{\theta}$ .

The quantity  $\log \frac{\theta}{1-\theta}$  is called  $\text{logit}\theta$  and is the natural parameter for the Binomial.  $\text{Logit}\theta$  is the log "odds ratio". In Binomial regression the link function is the logit so that we have

$$\text{logit}\theta_i = \underline{x}'_i \underline{\beta} \quad \text{where}$$

$$Y_i \sim \text{Bin}(n_i, \theta_i),$$

$\underline{\beta}$  is a vector of parameters and  $\underline{x}'_i$  is the "design" corresponding to  $Y_i$ . Binomial regression is more popularly known as logistic regression.

### Example 3: Normal

$$\begin{aligned} f(y; \mu, \sigma) &= \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{1}{2}\left(\frac{y-\mu}{\sigma}\right)^2\right); \quad -\infty < y < \infty \\ &= \exp\left(-\frac{1}{2}\frac{y^2}{\sigma^2} + \frac{y\mu}{\sigma^2} + \frac{1}{2}\frac{\mu^2}{\sigma^2} - \log\sigma\sqrt{2\pi}\right). \end{aligned}$$

The parameter of interest is  $\mu$ , so that  $\sigma$  is a nuisance parameter. The exponential expression shows that  $\mu$  is the natural parameter. That's the principal reason why in Normal based regression the link function is the identity function. See Sections 5 and 6.

## 7.2 GENERALISED LINEAR REGRESSION MODELS

We are now in a position to generalise the Normal based linear regression model.

1.  $Y_i$  has probability density (mass) function  $f(y_i; \theta)$  such that  $f(y_i; \theta) = \exp [y_i b(\theta) + c(\theta) + d(y_i)]$
2.  $E[Y_i] = \mu_i$
3.  $g(\mu_i) = \underline{x}'_i \underline{\beta}$

$g$  is called the link function;  
 $\underline{x}'_i$  is a vector of explanatory variables;  
 $\underline{\beta}$  is a vector of parameters.

Let's reconsider the example of Section 6.

$$Y_{ij} \sim \text{Po}(\lambda_{ij}), \text{ where} \\ \log \lambda_{ij} = \mu + \alpha_i + \beta_j .$$

This model is equivalent to the multiplicative model. We also require the side conditions  $\sum_i \alpha_i = \sum_j \beta_j = 0$ .

In Sections 4.3 and 4.4 we developed the maximum likelihood estimators of  $\lambda_i''$  and  $\lambda_i'$  (also  $\theta_i'$  and  $\theta_i''$ ).

Let the total number of counts  $Y_{..}$  be denoted by  $n$ .

$$\text{We had } \hat{\theta}_i' = \frac{Y_{i.}}{n} \text{ and } \hat{\theta}_i'' = \frac{Y_{.i}}{n} .$$

So, the fitted (expected) cell frequency  $\hat{\lambda}_i \hat{\lambda}_j'' = \frac{Y_i Y_j}{n}$  .

But,  $\log \hat{\lambda}_i \hat{\lambda}_j'' = \hat{\mu} + \hat{\alpha}_i + \hat{\beta}_j$  .

So,  $\log Y_i + \log Y_j - \log n = \hat{\mu} + \hat{\alpha}_i + \hat{\beta}_j$  . (7.2.1)

Summing (7.2.1) over i and j and setting  $\sum_i \hat{\alpha}_i = \sum_j \hat{\beta}_j = 0$  , we have

$$\hat{\mu} = \frac{1}{I} \sum_i \log Y_i + \frac{1}{J} \sum_j \log Y_j - \log n \quad . \quad (7.2.2)$$

Summing (7.2.1) only over j and using (7.2.2) we have

$$\hat{\alpha}_i = \log Y_i - \frac{1}{I} \sum_i \log Y_i \quad . \quad (7.2.3)$$

Similarly,  $\hat{\beta}_j = \log Y_j - \frac{1}{J} \sum_j \log Y_j$  . (7.2.4)

For this 'simple' generalised regression model we are able to compute the maximum likelihood estimates by 'hand'. A statistical package facilitates the computation of maximum likelihood estimates and associated statistics for any generalised linear regression model.

We remark again that the model

$$\log \lambda_{ij} = \mu + \alpha_i + \beta_j \quad , \quad (7.2.5)$$

is equivalent to

$$\lambda_{ij} = \lambda_i' \lambda_j'' \quad . \quad (7.2.6)$$

The parameter  $\mu$  represents the overall mean effect. The parameter  $\alpha_i$

represents the differential effect of row  $i$  beyond the average  $\mu$  and similarly the parameter  $\beta_j$  represents the differential effect of column  $j$  beyond the average effect  $\mu$ .

As with analysis of variance models, model (7.2.5) has too many

parameters, so that the constraints  $\sum_i \alpha_i = \sum_j \beta_j = 0$  are needed.

## 8. INFERENCE BASED ON GENERALISED LINEAR REGRESSION MODELS

Inferences involving Normal based regression models are based on F-ratio statistics (and T-ratio statistics).

In the present section we define the scaled deviance and show that it represents the generalisation of the F-ratio statistic to generalised regression models.

Statistical modelling involves an iterative cycle containing three or four steps.

- \* Preliminary diagnostic analysis
  - \* Specifying a model
  - \* Testing assumptions
  - \* MAKING INFERENCES
- 

If any of the assumptions possessed by the model are not supported by the data (equivalently, are violated by the data), then a new model has to be specified and tested.

The identified probabilistic (stochastic) model is not intended to represent the generating process of the data. Rather, it explains the salient features in the data and the residual variation resides in the error term (or distribution).

The observed data may be regarded as a sample (path) from the identified model.

In order to draw inferences based on the identified (or assumed) model, we can proceed in two alternative ways.

For the generalised linear regression model it can be shown that

$$\hat{\beta}_{MLE} - \beta \sim N(0, \mathfrak{I}^{-1}),$$

where  $\mathfrak{I}$  is the Fisher information matrix. So we can use the sampling distribution of the maximum likelihood estimator  $\hat{\beta}_{MLE}$  of  $\beta$ . Alternatively, we can test hypotheses about  $\beta$  by specifying each hypothesis in terms of the corresponding model and comparing goodness of fit.

Consider the "maximal" model where the number of parameters = number of observations;

$$\beta_M = [\beta_1, \dots, \beta_N]'$$

Suppose for the hypothesised model  $\beta = [\beta_1, \dots, \beta_P]'$ ,  $P < N$ .

The parameter  $\beta$  would be regarded as a "poor" description if and only if

$$L(\beta; y) \ll L(\beta_M; y)$$

That is, if the likelihood at  $\beta$  is 'much' less than the likelihood at  $\beta_M$ .

Equivalently,  $L(\hat{\beta}; y) \ll L(\hat{\beta}_M; y)$ , where  $\hat{\beta}$  denotes maximum likelihood estimator of  $\beta$  and  $\hat{\beta}_M$  denotes the maximum likelihood estimator of  $\beta_M$ .

$$\text{Let } \Lambda = \frac{L(\hat{\beta}_M; y)}{L(\hat{\beta}; y)}$$

Large values of  $\log \Lambda$  (log likelihood ratio) provide evidence that the model is poor. It turns out that if the hypothesis assumption about  $\beta$  is valid then

$$D = 2 \log \Lambda \sim \chi_{N-P}^2,$$

where  $\chi_{N-P}^2$  is a chi-square distribution with  $N-P$  degrees of freedom.

That is, if model is good under

$$H: \underline{\beta} = (\beta_1, \dots, \beta_p)' ,$$

then  $D \sim \chi^2_{N-p}$ .  $D$  is called the scaled deviance. So if the hypothesised model describes the data as well as the maximal model then  $D \sim \chi^2_{N-p}$  and so it is unlikely that we would observe "large" values of  $D$ .

Consider the hypothesis testing situation:

$$H_0: \underline{\beta} = (\beta_1, \dots, \beta_q)' = \underline{\beta}_0$$

versus the alternative

$$H_1: \underline{\beta} = (\beta_1, \dots, \beta_p)' = \underline{\beta}_1 \text{ such that } Q < P < N.$$

If  $D_0$  is the scaled deviance under  $H_0$  and  $D_1$  is the called deviance under  $H_1$ , then if both models give a "good" description of the data then

$$D = D_0 - D_1 \sim \chi^2_{P-Q} .$$

If the value of  $D$  is not in the tail of the  $\chi^2_{P-Q}$  distribution then we prefer the model corresponding to  $H_0$  since it is more parsimonious.

### Example

Consider two random samples:

$$Y_{11} \dots Y_{1n_1} \stackrel{\text{iid}}{\sim} N(\mu_1, \sigma^2) ;$$

$$Y_{21} \dots Y_{2n_2} \stackrel{\text{iid}}{\sim} N(\mu_2, \sigma^2) .$$

We want to test the null hypothesis;

$H_0: \mu_1 = \mu_2$  versus the alternative,

$H_1: \mu_1 \neq \mu_2$ .

	$H_0$	$H_1$
MAXIMUM LIKELIHOOD ESTIMATES	$\hat{\mu} = \bar{Y}_{..} = \frac{\sum_{j=1}^2 \sum_{i=1}^{n_j} Y_{ij}}{n_1 + n_2}$	$\hat{\mu}_i = \bar{Y}_{i.} = \frac{\sum_{j=1}^{n_j} Y_{ij}}{n_i}$ <p style="text-align: right;">(<math>i=1,2</math>).</p>

Now, the log (likelihood) for  $\hat{\beta}_0$  is

$$l(\hat{\beta}_0; y) = -(n_1 + n_2) \log \sigma \sqrt{2\pi} - \frac{1}{2} \sum_{i=1}^2 \sum_{j=1}^{n_i} \frac{(Y_{ij} - \bar{Y}_{..})^2}{\sigma^2} .$$

The log (likelihood) for  $\hat{\beta}_1$  is

$$l(\hat{\beta}_1; y) = -(n_1 + n_2) \log \sigma \sqrt{2\pi} - \frac{1}{2} \sum_{i=1}^2 \sum_{j=1}^{n_i} \frac{(Y_{ij} - \bar{Y}_{i.})^2}{\sigma^2} .$$

So,

$$\begin{aligned} \sigma^2 (D_0 - D_1) &= \sum_{i=1}^2 \sum_{j=1}^{n_i} (Y_{ij} - \bar{Y}_{..})^2 - \sum_{i=1}^2 \sum_{j=1}^{n_i} (Y_{ij} - \bar{Y}_{i.})^2 \\ &= \text{"Total sum of squares"} - \text{"Within group sum of squares"} \\ &= \text{"Between group sum of squares"} \\ &= n_1(\bar{Y}_{1.} - \bar{Y}_{..})^2 + n_2(\bar{Y}_{2.} - \bar{Y}_{..})^2 . \end{aligned}$$

Straightforward algebra shows that

$$\sigma^2(D_0 - D_1) = \frac{n_1 n_2}{n_1 + n_2} (\bar{Y}_1 - \bar{Y}_2)^2$$

Now under  $H_0$

$$D_0 - D_1 = \frac{n_1 n_2}{(n_1 + n_2)} (\bar{Y}_1 - \bar{Y}_2)^2 / \sigma^2$$

is  $\sim \chi^2_1$ .

We cannot test  $H_0$  directly since the quantity  $\sigma^2$  is unknown. We can estimate  $\sigma^2$  by the pooled sample variance

$$S_p^2 = \frac{(n_1 - 1) S_1^2 + (n_2 - 1) S_2^2}{n_1 + n_2 - 2}$$

The statistic  $(n_1 + n_2 - 2) S_p^2 / \sigma^2 \sim \chi^2_{n_1 + n_2 - 2}$

So, if  $H_0$  is correct

$$f = \frac{(D_0 - D_1) / 1}{\frac{S_p^2}{\sigma^2}} \sim F_{1, n_1 + n_2 - 2}$$

where  $F_{1, n_1 + n_2 - 2}$  is the F-distribution.

The quantity  $f$  is the F-ratio statistic for testing equality of population means.

$$\text{Now, } f = \frac{\frac{n_1 n_2}{n_1 + n_2} (\bar{Y}_1 - \bar{Y}_2)^2}{S_p^2}$$

$$= \left( \frac{\bar{Y}_1 - \bar{Y}_2}{S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \right)^2$$

=  $t^2$  where  $t$  is the T-ratio statistic with  $n_1 + n_2 - 2$  degrees of freedom.

The one-way ANOVA model can be cast in regression form as follows.

$$Y_{ij} = \mu_i + \epsilon_{ij} \quad ; \quad \epsilon_{ij} \sim N(0, \sigma^2) \quad , \quad \text{and}$$

$$\mu_i = \underline{x}'_i \underline{\beta} \quad .$$

The link function is the identity function. The parameter vector  $\underline{\beta}$  is given by  $\underline{\beta} = (\mu_1, \mu_2)$  and the corresponding design is

$$\underline{x}'_i = (1, 0) \quad \text{if } i=1$$

and  $\underline{x}'_i = (0, 1) \quad \text{if } i=2.$

So, for Normal based regression models hypothesis testing based on differences of scaled deviances is equivalent to hypothesis testing based on F-ratio statistics.

Scaled deviances are generalisations of F-ratio statistics to generalised regression models.

## 9. CONCLUSIONS

We discussed a number of relationships between the seminal works of Bailey and Simon [1] and Bailey [2] and the modern statistical body of methodology known as Generalised Linear Regression Modelling.

The relationships facilitate the consideration of a broad family of ratemaking models in a probabilistic framework. In particular, it was shown how to apply regression concepts in the case where the condition of class balance may be inappropriate as a result of the class experience being insufficiently *credible for the corresponding class risk parameter*.

Regression models are employed to extract maximum information from the data and to draw inferences from the data. They also afford the principle advantage of being testable.

The advent of statistical software packages that run on PC's has made interactive regression modelling possible.

## REFERENCES

- [1] R.A. Bailey and L. Simon, "Two studies in Automobile Insurance Ratemaking," *ASTIN Bulletin*, 1960, Vol.1, pp 192-217.
- [2] R.A. Bailey, "Insurance rates with minimum bias", PCAS L, 1963, pp 4-11.
- [3] R.L. Brown, "Minimum Bias with Generalised Linear Models", PCAS, 1986.
- [4] J.V. Eeghen, E.K. Greup and J.A. Nijssen, "Surreys of Actuarial Studies: Ratemaking", *Nationale - Nederlanden, N.V.*, 1980.

