ANALYZING TAGUCHI'S EXPERIMENTS USING GLIM WITH INVERSE GAUSSIAN DISTRIBUTION

by

Wong Kwok Keung

A

Thesis

Submitted to

(Division of Statistics)

The Graduate School

of

The Chinese University of Hong Kong

In Partial Fulfillment

of the Requirements of the Degree of

Master of Philosophy

(M. Phil.)

June, 1994

THE CHINESE UNIVERSITY OF HONG KONG

GRADUATE SCHOOL

The undersigned certify that we have read a thesis, entitled "Analyzing Taguchi's experiments using GLIM with Inverse Gaussian distribution" submitted to the Graduate School by Wong Kwok Keung (王國強) in partial fulfillment of the requirement for the degree of Master of Philosophy in Statistics. We recommend that it be accepted.

Dr. N.N. Chan,
Supervisor

Dr. Y. Lam,

Dr. K.H. Li,

Dr. S.K. Tse,
External Examiner.

DECLARATION

No portion of the work referred to in this thesis has been submitted in support of an application for another degree of qualification of this or any other university or other institute of learning.

ACKNOWLEDGMENT

I would like to express my gratitude here to my supervisor, Dr. Nai-Ng Chan for his earnest support and guidance that leads to the completion of this research programme. In addition, it is my pleasure to give thanks to Dr. Kim-Hung Li for his helpful advice.

Abstract

In recent years much attention has been paid to the Taguchi methods whose applications have successfully improved the quality of industrial processes in Japan. The basic concept of the methods is to economically achieve high quality, low variability and consistency of functional performance of a product or a process.

To analyze data from Taguchi's experiments, one of the useful approaches is to use the generalized linear models with varying dispersions. This class of models is usually based on distributions of the exponential families, namely, the normal, the inverse Gaussian and the gamma distributions. In this thesis, we devise a GLIM macro to find the maximum likelihood estimates of such models under the inverse Gaussian distribution assumptions. A simulation study on the analysis of data from these models is conducted. As a preliminary study, properties of the inverse Gaussian distributions are also given.

Table of Contents

1. Introduction

The improvement in the quality of industrial processes in Japan aroused many Western engineers to realize the importance of statistical tools, and in particular, much attention has been paid to the so-called Taguchi methods. Dr. Genechi Taguchi advocated the use of statistically designed experiments to find a best product or process design that is insensitive to any condition.

A powerful class of models for analyzing data from Taguchi's experiments is that of the generalized linear models with varying dispersions (Smyth, 1989). Aitkin (1987) pointed out that such a model can be fitted by maximum likelihood in GLIM using a simple macro. Based on Aitkin's macro, Chan, Li & Sharpe (1992) devised a GLIM macro to analyze Taguchi's experiments based on the class of such models in which the response variables are assumed to have normal distributions.

Analytically, a generalized linear model with varying dispersions has to be founded on distributions in an exponential family. Blaesild & Jensen (1985) showed that the only univariate distributions are those based on the normal, inverse Gaussian and gamma distributions. Therefore, we would mainly consider generalized linear models with varying

dispersions under the assumption of these distributions. The GLIM macro devised by Chan, Li & Sharpe (1992) is based on models under the normality assumption.

In this paper, we make use of the newly available GLIM 4 to devise a macro to analyze data arising from generalized linear models in which the response variables are assumed to have the inverse Gaussian distributions. Section 2 describes Taguchi's philosophy and his approach to the design of experiments. In section 3, we consider properties of the inverse Gaussian distribution. In section 4, we devise a GLIM 4 macro for analyzing data from a generalized linear model with varying dispersions under the inverse Gaussian distribution. We then carry out a simulation study in section 5 to see how well the GLIM 4 macro can help analyzing data from such a model.

2. Taguchi's methodology in design of experiments

Taguchi's fundamental concept rests on the importance of economically achieving high quality, low variability and consistency of functional performance. The objective is to identify the settings of the product or process parameters that maintain the performance characteristic about its target value and reduce the performance variation. In other words, we want to make products that will perform well under the conditions of normal use.

The performance characteristic of a product or a process is a response variable for measuring the quality of that product or process. It is affected by many factors. These factors can be classified into three classes:

1. *Signal factors*. They specify the intended value of the product's response, for example, the speed setting on a table fan. They are selected by the design engineers based on their engineering knowledge.
2. *Noise factors*. They are uncontrollable. They may be the external sources of variation like temperature, humidity, dust, supply voltage and human error, etc. They may also be due to manufacturing imperfections or product deterioration.

3. *Control factors*. They are to be set to minimize the sensitivity of the product's response to all possible noise factors. The power of the motor of a table fan may be an example of control factors.

Some of these factors affect the mean value of the performance characteristic and these are identified as having location effects. Others might cause variations in the characteristic and they are known as having dispersion effects. Taguchi's approach is to determine levels of control factors at which the effect of the noise factors on the performance characteristic is the least.

Taguchi's methodology is heavily based on statistical methods, especially statistically designed experiments, and concentrates on minimizing the deviation from the target caused by noise factors. He recommends the use of the orthogonal arrays and the 'signal-to-noise' ratio (SN) when conducting experiments. The SN ratio is derived from the quadratic loss function; it is used for measuring sensitivity to noise factors and will take a different form depending on the type of performance characteristic. Three common forms are:

1. $SN_T = 10\log\left(\frac{\bar{y}^2}{s^2}\right)$ which is suitable when we want to reduce variability around a specific target.

2. $SN_L = -10\log\left(\frac{1}{n}\sum_{i=1}^{n}\frac{1}{y_i^2}\right)$ which is suitable when we want the response as large as possible.

3. $SN_S = -10\log\left(\frac{1}{n}\sum_{i=1}^{n}y_i^2\right)$ which is suitable when we want the response as small as possible.

In Taguchi's approach, we would want to consider as many factors as possible in an initial experiment. However, a complete factorial design may involve thousands of treatment combinations. This is obviously expensive, inefficient, time-consuming and may be impossible to carry out. Therefore, Taguchi advocated the use of fractional factorial designs that require far fewer experimental runs.

In designing the experimental trials, Taguchi suggested the use of orthogonal arrays, which allows all combinations of levels to occur an equal number of times in every two columns. As a result of using orthogonal arrays, most of the higher order interaction terms are confounded with the lower order interactions or the main effects, and hence they are ignored as suggested by Taguchi.

There are three stages in a product's (or process's) development; they are (1) system design, (2) parameter design and (3) tolerance design. These stages are described below.

2.1 System design

This stage is usually done by the engineers. They have to use scientific and engineering knowledge to determine the basic configuration, that is, to select the materials, the manufacturing steps and the values or levels of the factors, etc.

System design can play an important role in reducing the sensitivity to noise factors as well as in reducing the manufacturing cost.

2.2 Parameter design

In parameter design, we determine the best settings for the control factors that minimize quality loss, or equivalently the sensitivity of the function of the product or process to all noise factors. During parameter design, we fix the manufacturing cost at a low value, that is, we use low-grade components and materials and consider noise factors with wide tolerance. If at the end of this stage, the quality loss is

within specifications, we have a design with the lowest cost and we need not go to the third stage. However, in practice, the quality loss must be further reduced.

2.3 Tolerance design

Tolerance design is used to determine the best tolerances for the parameters. We have to determine whether to reduce the quality loss due to performance variation or to increase the manufacturing cost for using better materials. Tolerance design should be done only after sensitivity to noise has been minimized through parameter design; otherwise we would have to specify unnecessarily higher-grade materials and components leading to higher manufacturing cost.

3. Inverse Gaussian distribution

3.1 Genesis

Schrödinger (1915) considered the linear Brownian motion of a particle with a positive drift (v) and obtained the first passage time distribution. He showed that the time (X) required for the particle to cover a fixed distance (d) is a random variable with probability density function

$$p_X(x) = \frac{d}{\sqrt{2\pi\beta x^3}} \exp\left\{-(d - vx)^2/(2\beta x)\right\} \qquad (3.1)$$

where β is a diffusion constant.

Alternatively, he found that the distance (D) over which the particle travels in a fixed period of time (x) is a random variable with the normal distribution

$$p_D(d) = \frac{1}{\sqrt{2\pi\beta x}} \exp\left\{-(d - vx)^2/(2\beta x)\right\} \qquad (3.2)$$

While attempting to extend Schrödinger's results, Tweedie (1941) noticed the inverse relationship between the cumulant-generating functions of (3.1) and (3.2), and suggested the name "Inverse Gaussian" for the distribution (3.1). He published a

detailed study of the distribution in 1957 which established many of its important statistical properties.

In 1947, Wald derived this kind of distributions as a limiting form of the sample size in certain sequential probability ratio tests. It is a special case of the inverse Gaussian distribution in which the parameter μ is equal to 1. This distribution is sometimes known as Wald's distribution.

3.2 Probability density function

On substituting $\nu = \frac{d}{\mu}$ and $\beta = \frac{d^2}{\lambda}$ into (3.1) we obtain the standard form of the inverse Gaussian distribution as follows:

$$f_X(x|\mu,\lambda) = \sqrt{\frac{\lambda}{2\pi x^3}} \exp\left\{-\frac{\lambda(x-\mu)^2}{2\mu^2 x}\right\}, \qquad x>0;\ \lambda>0;\ \mu>0.$$

The parameter μ is the mean of the distribution, λ is a reciprocal measure of dispersion, and the variance of the distribution is given by μ^3/λ.

There are three equivalent forms obtained by using the relationship $\mu = \lambda/\phi = (2\alpha)^{-1/2}$. They are:

$$f_X(x|\mu,\phi) = \sqrt{\frac{\mu\phi}{2\pi x^3}} \exp\left\{-\frac{\phi x}{2\mu} + \phi - \frac{\mu\phi}{2x}\right\}$$

$$f_X(x|\phi,\lambda) = \sqrt{\frac{\lambda}{2\pi x^3}} \exp\left\{-\frac{\phi^2 x}{2\lambda} + \phi - \frac{\lambda}{2x}\right\}$$

$$f_X(x|\lambda,\alpha) = \sqrt{\frac{\lambda}{2\pi x^3}} \exp\left\{-\alpha\lambda x + \lambda\sqrt{2\alpha} - \frac{\lambda}{2x}\right\}$$

Figures 1 and 2 show the graphs of inverse Gaussian distributions with different values of μ and λ.

Figure 1 Inverse Gaussian densities with $\mu=1$ for six values of λ.

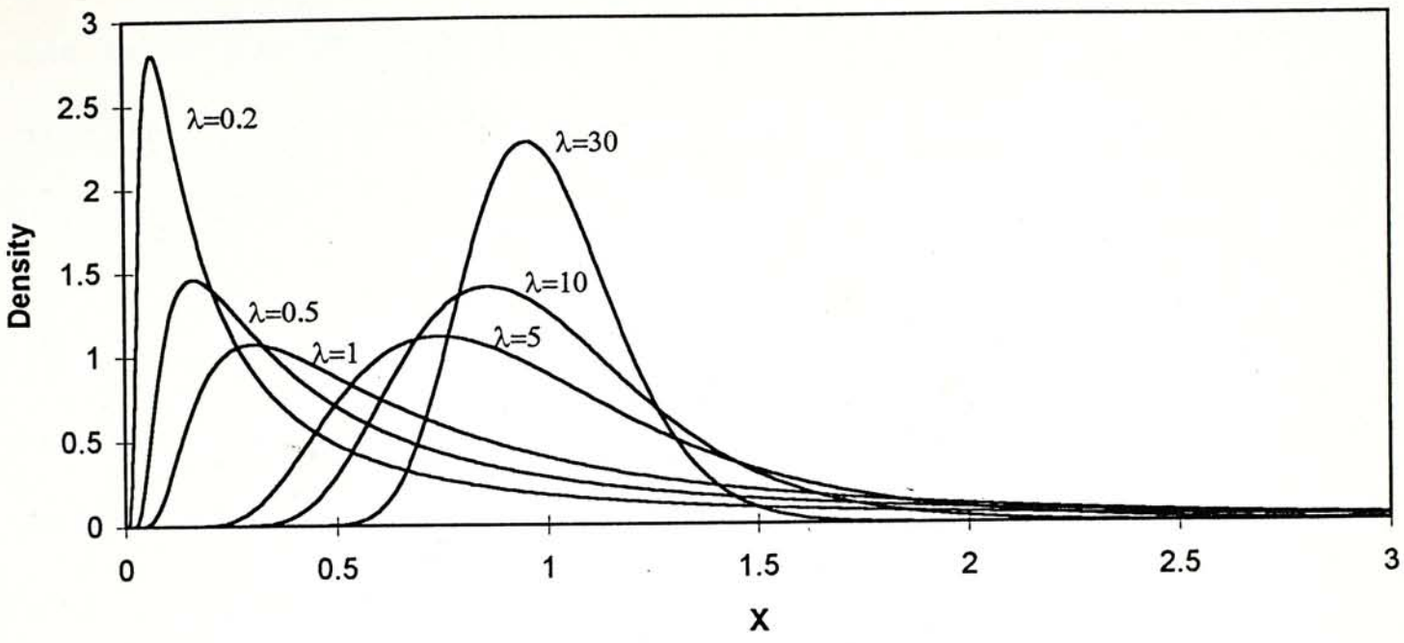
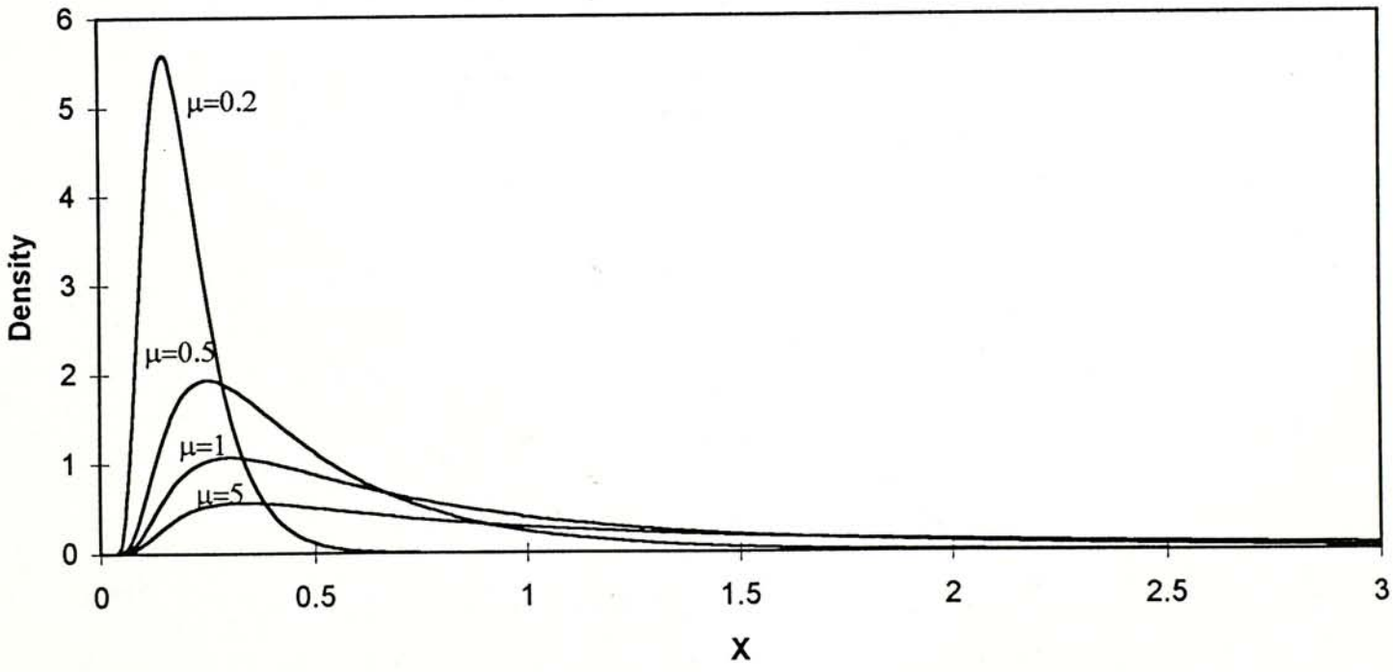


Figure 2 Inverse Gaussian densities with $\lambda=1$ for four values of μ.



In figure 1, we see that for a fixed μ, if λ varies from 0.2 to 30, the distribution will change from a highly skewed one to a symmetrical one. This suggests that the shape of the distribution depends on ϕ only, so ϕ is said to be the shape parameter.

One of the properties of inverse Gaussian distribution is that this family is closed under a change of scale, that is, if X has an inverse Gaussian distribution $IG(\mu,\lambda)$, then for any number $c>0$, cX is inverse Gaussian distributed with parameters $c\mu$ and $c\lambda$.

However, the property of reproducibility does not hold with respect to a change of location. This makes it desirable to introduce an additional threshold parameter for the family of inverse Gaussian distributions. If there is a threshold value θ, a three-parameter inverse Gaussian distribution can be defined by assuming that $(X-\theta)\sim IG(\mu,\lambda)$, and the density function of X is then given by

$$f_X(x|\,\theta,\mu,\lambda)=\left[\frac{\lambda}{2\pi(x-\theta)^3}\right]^{1/2}\exp\left[-\frac{\lambda[(x-\theta)-\mu]^2}{2\mu^2(x-\theta)}\right],\qquad -\infty<\theta<\infty,\ \mu>0,\ \lambda>0,\ x>\theta$$

where the mean is now equal to $\theta+\mu$, while ϕ is still the shape parameter.

Another useful property of inverse Gaussian distribution is its additive property. For a linear combination $\sum_i c_i X_i$, $c_i > 0$ where $X_i \sim IG(\mu_i, \lambda_i)$, it follows that if $\frac{\lambda_i}{\mu_i^2 c_i} = \xi$ is a constant for all i, then $\sum_i c_i X_i \sim IG(\sum_i c_i \mu_i , \xi(\sum_i c_i \mu_i)^2)$. In particular, we can obtain the sampling distribution of $\overline{X}$ from this, namely $\overline{X} \sim IG(\mu, n\lambda)$.

3.3 Estimation of parameters

For a random sample $X_1, X_2, \cdots, X_n$ from an inverse Gaussian population $IG(\mu,\lambda)$, we have the likelihood function

$$L = \left(\frac{\lambda}{2\pi}\right)^{n/2} \left(\prod_{i=1}^{n} x_i^{-3/2}\right) \exp\left[-\lambda \sum_{i=1}^{n} \frac{(x_i - \mu)^2}{2\mu^2 x_i}\right] \qquad (3.3)$$

thus

$$\log L = \frac{n}{2}\log\lambda - \frac{n}{2}\log(2\pi) - \frac{3}{2}\sum_{i=1}^{n}\log x_i - \lambda \sum_{i=1}^{n} \frac{(x_i - \mu)^2}{2\mu^2 x_i}. \qquad (3.4)$$

Differentiate (3.4) with respect to μ and λ, and set it to zero, we have

$$\frac{\partial \log L}{\partial \mu} = -\lambda\left(-\frac{\sum x_i}{\mu^3} + \frac{n}{\mu^2}\right) = 0 \qquad (3.5)$$

$$\frac{\partial \log L}{\partial \lambda} = \frac{n}{2\lambda} - \sum \frac{(x_i - \mu)^2}{2\mu^2 x_i} = 0 \qquad (3.6)$$

Solve (3.5), we get $\hat{\mu} = \frac{\sum x_i}{n} = \bar{x}$. Substitute this result into (3.6), we then obtain

$$\frac{1}{\hat{\lambda}} = \frac{2}{n}\sum_{i=1}^{n} \frac{x_i^2 - 2\hat{\mu}x_i + \hat{\mu}^2}{2\hat{\mu}^2 x_i} = \frac{1}{n}\sum_{i=1}^{n} \frac{x_i^2 - 2\bar{x}x_i + \bar{x}^2}{\bar{x}^2 x_i}$$

$$= \frac{1}{n}\sum_{i=1}^{n}\left(\frac{x_i}{\bar{x}^2} - \frac{2}{\bar{x}} + \frac{1}{x_i}\right) = \frac{\bar{x}}{\bar{x}^2} - \frac{2}{\bar{x}} + \frac{1}{n}\sum_{i=1}^{n}\frac{1}{x_i}$$

$$= \frac{1}{n}\sum_{i=1}^{n}\frac{1}{x_i} - \frac{1}{\bar{x}}$$

Therefore, the maximum likelihood estimators of μ and $1/\lambda$ are

$$\hat{\mu} = \bar{X}$$

$$\frac{1}{\hat{\lambda}} = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{1}{X_i} - \frac{1}{\bar{X}}\right) = \frac{1}{\bar{X}_H} - \frac{1}{\bar{X}} \qquad (3.7)$$

i.e. $$\hat{\phi} = \left[\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right)\left(\frac{1}{n}\sum_{i=1}^{n}\frac{1}{X_i}\right) - 1\right]^{-1} = \frac{\bar{X}_H}{\bar{X} - \bar{X}_H}$$

where $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ and $\bar{X}_H = \left[\frac{1}{n}\sum_{i=1}^{n}\frac{1}{X_i}\right]^{-1}$.

It is easily seen that $(\bar{X}, \bar{X}_H)$ is the minimal sufficient statistic for (μ,λ), and because the inverse Gaussian family is an exponential family, $(\bar{X}, \bar{X}_H)$ is complete. As the statistic $\left(\sum X_i, \sum(1/X_i - 1/\bar{X})\right)$ is a one-to-one function of $(\bar{X}, \bar{X}_H)$, it is also

a complete sufficient statistic for (μ,λ). Also, it can be shown that

$$E(\bar{X}) = E\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n}\sum_{i=1}^{n} E(X_i) = \frac{1}{n}\sum_{i=1}^{n} \mu = \mu .$$

From Johnson & Kotz (1970), we have the first negative moment of the inverse Gaussian distribution $IG(\mu,\lambda)$ equal to

$$E\left(\frac{1}{X}\right) = \frac{1}{\mu} + \frac{1}{\lambda} .$$

Moreover, from section 3.2, we know that $\bar{X} \sim IG(\mu, n\lambda)$, hence we have

$$E\left(\frac{1}{\bar{X}}\right) = \frac{1}{\mu} + \frac{1}{n\lambda}$$

and

$$E\left(\frac{1}{\hat{\lambda}}\right) = E\left(\frac{1}{n}\sum_{i=1}^{n}\left(\frac{1}{X_i} - \frac{1}{\bar{X}}\right)\right) = \frac{1}{n}\sum_{i=1}^{n} E\left(\frac{1}{X_i} - \frac{1}{\bar{X}}\right)$$

$$= \frac{1}{n}\sum_{i=1}^{n}\left(\frac{1}{\mu} + \frac{1}{\lambda} - \frac{1}{\mu} - \frac{1}{n\lambda}\right) = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{n-1}{n\lambda}\right)$$

$$= \left(\frac{n-1}{n}\right)\frac{1}{\lambda}$$

Therefore, the uniform minimum variance unbiased estimators (UMVUEs) of μ and $1/\lambda$ are $\bar{X}$ and $\frac{1}{n-1}\sum_{i=1}^{n}\left(\frac{1}{X_i} - \frac{1}{\bar{X}}\right)$ respectively.

To find the moment estimators, we consider

$$\frac{1}{n}\sum_{i=1}^{n} X_i = E(X) = \mu$$

$$\frac{1}{n}\sum_{i=1}^{n} X_i^2 = E(X^2) = \frac{\mu^3}{\lambda} + \mu^2$$

and solve this system of equations, we will get

$$\tilde{\mu} = \bar{X}$$

$$\frac{1}{\tilde{\lambda}} = \frac{\sum (X_i - \bar{X})^2}{n\bar{X}^3} \qquad (3.8)$$

i.e. $$\tilde{\phi} = \frac{n\bar{X}^2}{\sum (X_i - \bar{X})^2} = (\text{sample coefficient of variation})^{-2}$$

To compare the maximum likelihood estimators with the moment estimators, we must consider their variances; some of the results are shown below.

Because $\hat{\mu}$ is the same as $\tilde{\mu}$,

$$Var(\hat{\mu}) = Var(\tilde{\mu}) = Var(\bar{X}) = \frac{\mu^3}{n\lambda}.$$

For n large,

$$n\,Var(\hat{\phi}) \approx 2\phi^2 + \phi$$

and

$$n\,Var(\tilde{\phi}) \approx 10\phi^2 + 19\phi\,.$$

The variance of $\hat{\phi}$ is smaller than that of $\tilde{\phi}$, showing that the maximum likelihood estimator of ϕ is more efficient than the moment estimator. This is of no surprise because the maximum likelihood estimator $(\hat{\mu},\hat{\phi})$ is a complete sufficient statistic for (μ,ϕ).

When μ is known, if we change the scale, we may take μ as one, and we only have to estimate a single parameter ϕ of a standard Wald distribution. The maximum likelihood estimator of ϕ is

$$\hat{\phi}' = \left[\frac{1}{n}\sum_{i=1}^{n}\left(X_i + \frac{1}{X_i}\right) - 2\right]^{-1} \tag{3.9}$$

and the moment estimator is

$$\tilde{\phi}' = \frac{n}{\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2}. \tag{3.10}$$

For n large,

$$n\,Var(\hat{\phi}') \approx 2\phi^2$$

and

$$n\,Var(\tilde{\phi}') \approx 2\phi^2 + 15\phi .$$

We note that the efficiency of the moment estimator is also less than that of the maximum likelihood estimator.

However, both estimators have smaller variances than those when μ is unknown.

Since ϕ is exactly equal to λ when μ is equal to one, and we are also interested in the variance of $1/\lambda$, it is of interest to look at the variances of the estimators of ϕ^{-1}. In fact,

$$Var\left(\frac{1}{\hat{\phi}'}\right)=\frac{2}{\phi^2 n}.$$

and

$$Var\left(\frac{1}{\tilde{\phi}'}\right)=\frac{1}{n}\left(\frac{2}{\phi^2}+\frac{15}{\phi^3}\right).$$

In case μ is known but not equal to one, the results are different merely by a scale multiple.

Estimators of μ and λ and their asymptotic variances were given in Patel (1965) for both the one-sided and two-sided truncated inverse Gaussian distributions. Johnson and Kotz (1970) well summarized his results.

3.4 Applications

Inverse Gaussian distribution is usually used in studying reliability and life testing problems in lifetime models. Being the first passage time distribution for the Wiener process, it is particularly appropriate for failure or reaction time data analysis.

Besides, the inverse Gaussian distribution has been used in a wide range of applications. Most of these applications are based on the idea of first passage time for an underlying process. In chapter 10 of Chhikara & Folks (1989), there are examples from many diverse fields such as cardiology, hydrology, demography, linguistics, employment service, labour disputes, and finance. Though it seems farfetched to postulate the existence of an underlying Brownian motion in studying human behavior, this is not at all unreasonable. The use of the inverse Gaussian distribution upon an underlying Wiener process is not critical; we may use inverse Gaussian distributions when it is justified by the goodness-of-fit test. This situation is similar to that of the normal distribution. Normal distributions were primarily used to describe a law of errors, but gradually it became acceptable to use them to describe all sorts of data.

When dealing with skewed distributions, we usually resort to a transformation, such as Box-Cox transformation, to normalize the data, but it is sometimes difficult to interpret the result based on the transformed variable. Therefore, if possible, it is desirable to analyze data using statistical methods based on skewed distributions. An inverse Gaussian distribution can represent a highly skewed to an almost normal distribution; so it is one of the choices. Moreover, a statistical model involving inverse Gaussian distributions allows for quite a wide range of statistical methods such as the t-test for μ, the chi-square test for λ and analysis of residuals. This is also similar to the case of the normal distribution.

4. Iterative procedures and Derivation of the GLIM 4 macros

4.1 Generalized linear models with varying dispersions

In a classical linear model, we usually assume that the response variables $Y_1, Y_2, \ldots, Y_n$ are independently and normally distributed with means $\mu = (\mu_1, \mu_2, \ldots, \mu_n)$ and constant variance σ^2, and that the systematic part of the model is given by

$$\mu = \mathbf{X}\beta$$

where $\mathbf{X}$ is the design matrix and β is the vector of unknown parameters. This model relies on strong assumptions of independence and constant variance of the errors, but, in many cases, they are not easily fulfilled.

Generalized linear models (Nelder & Wedderburn, 1972) are an extension of the classical linear models. The generalization allows analysis on non-normal data such as probit analysis and contingency tables. A generalized linear model has three components:

1. The *random component*: The response variables $Y_1, Y_2, \ldots, Y_n$ are assumed to share the same distribution from the exponential family with means μ.

2. The *systematic component*: the covariates $\mathbf{t}_1, \mathbf{t}_2, \ldots, \mathbf{t}_p$ (the column vectors in the design matrix $\mathbf{X}$) and the unknown parameters $\beta_1, \beta_2, \ldots, \beta_p$ (in $\boldsymbol{\beta}$) produce a linear predictor $\boldsymbol{\eta}$ given by

$$\boldsymbol{\eta} = \mathbf{X}\boldsymbol{\beta} = \sum_{j=1}^{p} \mathbf{t}_j \beta_j \, .$$

3. The *link* function between the random and systematic components: it is a monotonic differentiable function $g(\cdot)$ that relates the i^{th} component η_i of the linear predictor $\boldsymbol{\eta}$ to the expected value of a datum, namely,

$$\eta_i = g(\mu_i) \, .$$

When we consider the ordinary linear models with normal errors, the random component specifies $Y_i \overset{i.i.d.}{\sim} N(\mu_i, \sigma^2)$ and the link function will be an identity function. Thus the means are equivalent to the linear predictors, so that $Y_i \overset{i.i.d.}{\sim} N(\mathbf{x}_i'\boldsymbol{\beta}, \sigma^2)$ where $(\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_n) = X'$.

In general, for a generalized linear model with **varying dispersions**, we assume that each component of Y has a distribution in the exponential family, taking the form

$$f_Y(y|\,\theta, \phi) = \exp\{(y\theta - b(\theta))/d(\phi) + c(y, \phi)\} \, ,$$

for some specific functions b, c and d (McCullagh and Nelder, 1989). In this specification, θ is called the canonical parameter, ϕ is called the dispersion parameter, and $b(\theta)$ is called the cumulant function. The function $d(\phi)$ is generally of the form $d(\phi)=\phi/w$, where w is a known prior weight that varies from observation to observation. If ϕ is unknown, c will have the form

$$c(y,\phi)=-w\phi^{-1}a(y)-\frac{1}{2}s(-w\phi^{-1})+t(y)$$

for some functions a, s and t. Thus, the overall log-likelihood function becomes

$$l(\theta_i,\phi_i|\,y_i)=w_i\phi_i^{-1}\left[y_i\theta_i-b(\theta_i)-a(y_i)\right]-\frac{1}{2}s(-w_i\phi_i^{-1})+t(y_i) \qquad (4.1)$$

It follows from differentiating (4.1) that the means and variances are given by $E(y_i)=\mu_i=\dot{b}(\theta_i)$ and $Var(y_i)=w_i^{-1}\phi_i\ddot{b}(\theta_i)$ (Smyth, 1989). Barndorff-Nielsen and Blaesild (1983) showed that $\dot{a}$ must be the inverse of $\dot{b}$, and hence $\theta_i=\dot{b}^{-1}(\mu_i)=\dot{a}(\mu_i)$.

Equation (4.1) may be considered as the log-likelihood for a generalized linear model with weights $w_i\phi_i^{-1}$ and dispersion parameter 1 if we hold all ϕ_i fixed. We call this the *mean submodel*.

On the other hand, if we assume all μ_i are fixed and let the deviance components be $d_i(y_i,\mu_i) = -2w_i[y_i\theta_i - b(\theta_i) - a(y_i)]$, the log-likelihood becomes

$$l(\xi_i | y_i,\theta_i) = \frac{1}{2}[\xi_i d_i - s(w_i\xi_i)] + t(y_i) \qquad (4.2)$$

This has the form of a generalized linear model with observations d_i and canonical parameters $\xi_i = -\phi_i^{-1}$, and we call it the *dispersion submodel*. This dispersion submodel itself has a dispersion parameter 2 and weight 1. Differentiating (4.2), we may get $E(d_i) = \delta_i = w_i\dot{s}(w_i\xi_i)$ and $Var(d_i) = 2w_i^2\ddot{s}(w_i\xi_i)$.

The fact that the mean and dispersion of a generalized linear model are orthogonal allows us to estimate the unknown parameters β and γ one at a time. The Fisher scoring algorithm for the simultaneous maximum likelihood estimation of β and γ then reduces to two separate algorithms for β and γ.

The scoring iteration in the mean submodel can be written as a weighted least squares calculation

$$\beta^{k+1} = F_1(\beta^k,\gamma) = (X'WX)^{-1}X'W\mathbf{r}$$

where X is the matrix of covariates affecting the mean, W is the weight (diagonal) matrix

$$W = \mathrm{diag}\left(\dot{g}(\mu_i)^2 Var(Y_i)\right)^{-1} \qquad (4.3)$$

and **r** is the working vector with components

$$r_i = \dot{g}(\mu_i)(y_i - \mu_i) + g(\mu_i) .$$

For the dispersion submodel, we assume the link function $f(\cdot)$ to be

$$f(\delta_i) = \mathbf{z}_i'\gamma .$$

Then the scoring iteration would be given by

$$\gamma^{k+1} = F_2(\beta, \gamma^k) = (Z'VZ)^{-1} Z'V\mathbf{s}$$

where Z is the matrix of covariates affecting the dispersion, V is the weight matrix

$$V = \mathrm{diag}\left(\dot{f}(\delta_i)^2 Var(d_i)\right)^{-1} \qquad (4.4)$$

and **s** is the working vector with components

$$s_i = \dot{f}(\delta_i)(d_i - \delta_i) + f(\delta_i) .$$

Smyth (1989) suggested several iteration schemes and studied their convergence rates.

4.2 Mean and dispersion models for inverse Gaussian distribution

The log-likelihood function of a random sample from $IG(\mu_i, \lambda_i)$ is given by

$$l = \sum_i \left[-\lambda_i \left(\frac{y_i}{2\mu_i^2} - \frac{1}{\mu_i} + \frac{1}{2y_i} \right) + \frac{1}{2}\log\lambda_i - \frac{1}{2}\log(2\pi y_i^3) \right]. \qquad (4.5)$$

We consider this as the likelihood of the mean submodel and compare this to (4.1). Then we have $w_i = 1$,

$$\theta_i = -\frac{1}{2\mu_i^2},$$

$$\phi_i = \frac{1}{\lambda_i},$$

$$b(\theta_i) = -\frac{1}{\mu_i} = -(-2\theta_i)^{1/2},$$

$$a(y_i) = \frac{1}{2y_i},$$

$$s(\xi_i) = -\log\lambda_i = -\log\left[-\left(-\phi_i^{-1}\right)\right] = -\log(-\xi_i)$$

and

$$t(y_i) = -\frac{1}{2}\log(2\pi y_i^3).$$

If we consider (4.5) as the likelihood of the dispersion submodel, we have the deviance components

$$d_i = 2\left(\frac{y_i}{2\mu_i^2} - \frac{1}{\mu_i} + \frac{1}{2y_i}\right) = \frac{(y_i - \mu_i)^2}{\mu_i^2 y_i}$$

which are distributed as $\phi_i \chi_1^2$, according to a property of the inverse Gaussian distributions. (Smyth, 1989).

Suppose $Y_i \sim IG(\mathbf{x}_i'\beta, \mathbf{z}_i'\gamma)$, so that $\mu_i = \mathbf{x}_i'\beta$ and $\lambda_i = \mathbf{z}_i'\gamma$. Obviously, the link function for the mean submodel is the identity link. It can also be easily shown that

$$\delta_i = \dot{s}(\xi_i) = -\frac{1}{\xi_i} = \frac{1}{\phi_i^{-1}} = \frac{1}{\lambda_i},$$

and so the link function for the dispersion submodel is the reciprocal link. As we show here, the choice of the link function depends on the model we use. These two link functions are used because they suit our simulation study in the next section. McCullagh and Nelder (1989) listed some canonical link functions and other characteristics of some common univariate distributions in the exponential family.

Rewriting (4.5), we have

$$l = \sum_{i=1}^{n}\left[-\lambda_i \frac{(y_i - \mu_i)^2}{2\mu_i^2 y_i} + \frac{1}{2}\log\lambda_i - \frac{1}{2}\log(2\pi y_i^3)\right]$$

$$= \sum_{i=1}^{n}\frac{1}{2}\left[-\gamma'\mathbf{z}_i \frac{(y_i - \beta'\mathbf{x}_i)^2}{(\beta'\mathbf{x}_i)^2 y_i} + \log\gamma'\mathbf{z}_i - \log(2\pi y_i^3)\right]$$

The first and the second derivatives with respect to β and γ are given below:

$$\frac{\partial l}{\partial \beta} = \sum_{i=1}^{n} \mathbf{x}_i \frac{\partial}{\partial \mu_i}\left[\lambda_i\left(-\frac{y_i}{2\mu_i^2}+\frac{1}{\mu_i}-\frac{1}{2y_i}\right)\right]$$

$$= \sum_{i=1}^{n}\lambda_i\left(\frac{y_i}{\mu_i^3}-\frac{1}{\mu_i^2}\right)\mathbf{x}_i = \sum_{i=1}^{n}\lambda_i\left(\frac{y_i-\mu_i}{\mu_i^3}\right)\mathbf{x}_i$$

$$= \sum_{i=1}^{n}\lambda_i\left(\frac{y_i-\beta'\mathbf{x}_i}{(\beta'\mathbf{x}_i)^3}\right)\mathbf{x}_i$$

$$\frac{\partial l}{\partial \gamma} = \sum_{i=1}^{n}\mathbf{z}_i\left[\frac{-(y_i-\mu_i)^2}{2\mu_i^2 y_i}+\frac{1}{2\lambda_i}\right]$$

$$= \sum_{i=1}^{n}\left(\frac{-(y_i-\mu_i)^2}{2\mu_i^2 y_i}+\frac{1}{2\gamma'\mathbf{z}_i}\right)\mathbf{z}_i$$

$$= \sum_{i=1}^{n}\frac{1}{2}\left(-d_i+\frac{1}{\gamma'\mathbf{z}_i}\right)\mathbf{z}_i$$

$$\frac{\partial^2 l}{\partial\beta\partial\beta'} = \sum_{i=1}^{n}\mathbf{x}_i\mathbf{x}_i'\left(\frac{\partial}{\partial\mu_i}\left[\lambda_i\frac{y_i-\mu_i}{\mu_i^3}\right]\right)$$

$$= \sum_{i=1}^{n}\lambda_i\left(\frac{-3y_i+2\mu_i}{\mu_i^4}\right)\mathbf{x}_i\mathbf{x}_i'$$

$$= \sum_{i=1}^{n}-\lambda_i\left(\frac{3y_i-2\beta'\mathbf{x}_i}{(\beta'\mathbf{x}_i)^4}\right)\mathbf{x}_i\mathbf{x}_i'$$

$$\frac{\partial^2 l}{\partial\gamma\partial\gamma'} = \sum_{i=1}^{n} \mathbf{z}_i\mathbf{z}_i' \frac{\partial}{\partial\lambda_i}\left(\frac{1}{2\lambda_i}\right)$$

$$= \sum_{i=1}^{n} \frac{-1}{2\lambda_i^2}\mathbf{z}_i\mathbf{z}_i'$$

$$= \sum_{i=1}^{n} -\frac{\mathbf{z}_i\mathbf{z}_i'}{2\left(\gamma'\mathbf{z}_i\right)^2}$$

$$\frac{\partial^2 l}{\partial\gamma\partial\beta'} = \sum_{i=1}^{n}\left(\frac{y_i - \beta'\mathbf{x}_i}{\left(\beta'\mathbf{x}_i\right)^3}\right)\mathbf{z}_i\mathbf{x}_i'.$$

Then we have the estimated expected information matrix

$$\hat{I} = -E\begin{bmatrix} \frac{\partial^2 l}{\partial\beta\partial\beta'} & \frac{\partial^2 l}{\partial\beta\partial\gamma'} \\ \frac{\partial^2 l}{\partial\gamma\partial\beta'} & \frac{\partial^2 l}{\partial\gamma\partial\gamma'} \end{bmatrix}$$

$$= \begin{bmatrix} X'WX & 0 \\ 0 & Z'VZ \end{bmatrix}$$

where

$$W = \operatorname{diag}\left(\lambda_i/\mu_i^3\right) = \operatorname{diag}\left(Var\left(Y_i\right)\right)^{-1}$$

and

$$V = \operatorname{diag}\left(1/2\lambda_i^2\right).$$

This is equivalent to equation (4.3) and (4.4) since

$$\dot{g}\left(\mu_i\right) = 1,$$

$$\dot{f}\left(\delta_i\right) = -\delta_i^{-2} = -\lambda_i^2,$$

and

$$Var\left(d_i\right) = 2\ddot{s}\left(\xi_i\right) = 2\xi_i^{-2} = 2\lambda_i^{-2}.$$

To obtain the maximum likelihood estimates for the mean and dispersion, we should first fit an unweighted regression to give an initial estimate β^0, and use β^0 (or equivalently the predicted values $\hat{\mu}_i$) to obtain the estimate γ^0, and we then use the scoring iterations F_1 and F_2 alternatively until the sequence converges.

4.3 Devising the GLIM 4 macros

Based on the mean and dispersion submodels described in the last section, we modify Aitkin's (1987) macro to fit our model of a generalized linear model with varying dispersions under the inverse Gaussian distribution in GLIM 4.

First, we must change the macro **mean**. The error should be the inverse Gaussian (denoted by *i* in GLIM 4), the link is still the identity (denoted by *i* in GLIM 4), and the initial value of the linear predictors is 1. We also have to calculate the deviances and the values of $t(.)$ which are used in the next macro **var**.

The macro **var** uses the deviances d_i as the responses (specified by *yvariate* command in GLIM 4). The link is now reciprocal (denoted by *r*) instead of log, the maximum number of iteration cycles has been increased to 25, and the overall log-likelihood value is calculated by equation (4.1).

The rest of the changes control the stopping criteria of the iterations, so that iterations stop when successive log-likelihood values differ by less than 0.00001 or when 20 iterations have been performed.

A listing of the macros can be found in Appendix A.

4.4 Model fitting

Before everything begins, we must first input the macros and the data through a series of GLIM commands. It seems rather awkward, especially when specifying the design matrix, to type the whole series of commands every time, so we suggest to put these commands into a separate file, and import it into GLIM using the $input command. Our input file is given in Appendix B.

Moreover, whenever we start fitting, we also have to specify the response variables and the factors in the mean and dispersion model. This can be done by the following commands:

```
$Macro MODM A+B+C+D+E+F+G+H $Endmacro
$Macro MODV A+B+C+D+E+F+G+H $Endmacro
$USE VMOD Y$
```

Knowing how to use the macros, we are then interested in the fitting procedure. It can be summarized into the following three steps:

1. fit a saturated model for the mean;
2. fit an adequate model for the dispersion using backward elimination;
3. with the dispersion model found in step 2, fit an adequate model for the mean.

We can use either the Wald's test or the likelihood ratio test as the selection criteria for the backward elimination.

For the adequate fitted model, all the dispersion factors will be called the control factors and those significant mean factors that are not control factors will be called the signal factors.

To minimize the dispersion while holding the mean at a preassigned value, we may set each control factor at the level that gives the minimum dispersion and adjust the levels of the highly significant signals so that the mean characteristic is on target. All other factors are set at levels determined by cost and/or convenience.

5. Simulation Study

Unfortunately, there are not enough practical examples to verify our macros. The only possible way out is to carry out a simulation study to examine the effectiveness of the macros. First, we see how random variates from the inverse Gaussian distribution are generated.

5.1 Generating random variates from the inverse Gaussian distribution

Michael, Schucany & Haas (1976) gave a method of generating random variates using a transformation with multiple roots. Their approach is to find a transformation of the random variable of interest that follows a nice distribution, and then to use the probabilities associated with the multiple roots of the transformation to choose one root for the random observation. Their method was successfully applied to the inverse Gaussian distribution.

For $X \sim IG(\mu,\lambda)$, the transformed variable

$$Y = \frac{\lambda(X-\mu)^2}{\mu^2 X} \qquad (5.1)$$

is distributed as the chi-square with one degree of freedom (Shuster, 1968). There are two roots for the equation (5.1), which are

$$X_1 = \frac{\mu}{2\lambda}\left(2\lambda + \mu Y - \sqrt{4\lambda\mu Y + \mu^2 Y^2}\right) \quad \text{and} \quad X_2 = \frac{\mu^2}{X_1}.$$

The conditional probability with which the smaller root X_1 should be selected is $\frac{\mu}{\mu + X_1}$, and therefore X_2 should be selected with probability $\frac{X_1}{\mu + X_1}$. The overall procedure for generating inverse Gaussian variates is as follows:

1. Generate random numbers from the chi-square distribution with one degree of freedom.
2. For each random value in step 1, compute the smaller root X_1 given above.
3. Generate uniform(0,1) random variates U.
4. If $U \le \mu/(\mu + X_1)$, the root X_1 is chosen for the random observation from the inverse Gaussian distribution; otherwise the larger root X_2 is chosen.

With this algorithm, we may generate the observations for our simulation model that is described in details in the next section.

5.2 Simulation model

In our simulation, a 2_{IV}^{8-4} fractional factorial design is employed. There are eight factors *A, B, C, D, E, F, G, H*. Each of the factors has two levels. The design matrix is shown in Table 5.1.

Table 5.1 The model matrix with the proposed means and dispersions

A	B	C	D	E	F	G	H	μ	λ
0	0	0	0	0	0	0	0	35	400
1	0	0	0	1	1	1	0	55	400
0	1	0	0	1	1	0	1	65	400
1	1	0	0	0	0	1	1	85	400
0	0	1	0	1	0	1	1	55	400
1	0	1	0	0	1	0	1	35	400
0	1	1	0	0	1	1	0	85	400
1	1	1	0	1	0	0	0	65	400
0	0	0	1	0	1	1	1	35	200
1	0	0	1	1	0	0	1	15	200
0	1	0	1	1	0	1	0	65	200
1	1	0	1	0	1	0	0	45	200
0	0	1	1	1	1	0	0	15	200
1	0	1	1	0	0	1	0	35	200
0	1	1	1	0	0	0	1	45	200
1	1	1	1	1	1	1	1	65	200

For each combination, observations are generated from $IG(\mu_i, \lambda_i)$. We arbitrarily choose

$$\mu = 35 + 30B - 20D + 20G \quad \text{and} \quad \lambda = 400 - 200D,$$

where B, D and G are equal to 0 at low level and 1 at high level. The values of the factor levels can take any value other than 0 and 1, but in GLIM, these have to be greater than 0. This is why the values of the factors are set to 1 and 2 in the input file given in Appendix B. The means μ_i and dispersion parameters λ_i for the sixteen combinations can be easily calculated and are given in Table 5.1.

Using the procedure mentioned in section 5.1, we generate sixty-four replicates for each combination, and repeat using different random seeds to generate one hundred different sets of data. For each set of data, we fit adequate models with the first four replicates as the responses for our macros. Then we use the first eight replicates, first sixteen replicates, first thirty-two replicates, and finally use all the sixty-four replicates as our responses. The purpose of using different numbers of replicates is to investigate the effectiveness of the macros under different conditions.

As our aim is to see how well our macros work, we will look at the number of times that the designated models are found as the adequate models by using the macros.

5.3 Results

In the simulation study, we observed that all the fitted mean models contain factors *B*, *D* and *G* no matter how many replicates we use, while some of the dispersion models do not involve the factor *D* at all. The latter cases occur more often when fewer replicates are used. Moreover, most of these cases contain only the constant term, showing that the dispersion model is not necessary and only a single dispersion parameter is to be estimated.

The simulation results are summarized in Table 5.2. We see that if more observations are used in each combination, more designated models can be found. However, when the number of observations becomes large, doubling the sample size does not effectively increase the proportion that the designated models are found.

Table 5.2

Number of replicates in each combination	Relative frequency of finding the designated models	Relative frequency of having dispersion models which do not involve factor *D*
4	0.21	0.45
8	0.49	0.24
16	0.56	0.02
32	0.64	0.00
64	0.68	0.00

In fact, using only four replicates in each treatment combination may cause errors sometimes (21 cases reported). As reported in the GLIM output, these errors are probably due to inappropriate models. If we increase the number of replicates, no error is reported then. This indicates that the macros do not work very well if too few observations are given.

When sixty-four replicates were used, we have sixty-eight cases in which designated models are obtained. Using these sixty-eight cases, we calculate the sample mean square errors (MSEs) for the estimates $\hat{\mu}$ and $\hat{\phi}$. The estimates for the eight different combinations are calculated from the following formulae:

$$\hat{\mu} = \hat{\beta}_0 + \hat{\beta}_1 B + \hat{\beta}_2 D + \hat{\beta}_3 G,$$

$$\hat{\lambda} = \hat{\gamma}_0 + \hat{\gamma}_1 D,$$

$$\hat{\phi} = \hat{\lambda}/\hat{\mu},$$

and the sample MSEs are then calculated using $MSE(\hat{\mu}) = \frac{1}{68}\sum_{1}^{68}(\hat{\mu} - \mu)^2$ and $MSE(\hat{\phi}) = \frac{1}{68}\sum_{1}^{68}(\hat{\phi} - \phi)^2$. The results are shown in table 5.3.

Table 5.3 The estimated MSEs of the estimates

B	D	G	μ	λ	ϕ	MSE($\hat{\mu}$)	MSE($\hat{\phi}$)
0	0	0	35	400	11.43	0.617022	0.547316
0	0	1	55	400	7.27	1.010249	0.210950
1	0	0	65	400	6.15	1.162774	0.157883
1	0	1	85	400	4.71	1.368432	0.088838
0	1	0	15	200	13.33	0.113159	0.713469
0	1	1	35	200	5.71	0.930615	0.167385
1	1	0	45	200	4.44	1.150622	0.076750
1	1	1	65	200	3.08	1.780510	0.040124

The variances of the maximum likelihood estimators of μ and ϕ are given by equation (3.9) and (3.10). Their values are given in Table 5.4.

Table 5.4 The variances of estimates of μ and ϕ

B	D	G	μ	λ	ϕ	$Var(\hat{\mu})$	$Var(\hat{\phi})$
0	0	0	35	400	11.43	1.674805	4.260204
0	0	1	55	400	7.27	6.499023	1.766529
1	0	0	65	400	6.15	10.72754	1.279586
1	0	1	85	400	4.71	23.98926	0.765571
0	1	0	15	200	13.33	0.263672	5.763889
0	1	1	35	200	5.71	3.349609	1.109694
1	1	0	45	200	4.44	7.119141	0.686728
1	1	1	65	200	3.08	21.45508	0.343935

There are large differences between the sample MSEs and the variances of the same treatment combination. Probably, this is because we ignore many cases in which other models than those designated are found. In these cases, the difference between the estimates and the true parameters must be very large, and so, if we include them in the calculations, the sample MSEs may then be close to the variances shown in Table 5.4.

Table 5.5 shows the GLIM outputs of one case in which the designated models are obtained. We notice that the standard errors of the estimates decrease as the number of observations

increase. This is not surprising at all because the accuracy of the estimation always depends on the sample size.

Table 5.5 GLIM output of final fit for one case

4 replicates

deviance = 64.000 at cycle 5
residual df = 60

Mean model

	estimate	s.e.	parameter
1	36.35	3.367	1
2	27.24	5.468	B(2)
3	-21.00	3.608	D(2)
4	20.06	4.634	G(2)

scale parameter 1.067

-- model changed
scaled deviance = 72.238 at cycle 6
residual df = 62

Variance model

	estimate	s.e.	parameter
1	436.4	109.1	1
2	-280.0	115.9	D(2)

scale parameter 2.000

Iteration number 4. , deviance is 536.610

8 replicates

deviance = 128.00 at cycle 5
residual df = 124

Mean model

	estimate	s.e.	parameter
1	34.12	2.271	1
2	28.90	3.634	B(2)
3	-19.54	2.387	D(2)
4	19.86	2.915	G(2)

scale parameter 1.032

-- model changed
scaled deviance = 156.00 at cycle 6
residual df = 126

Variance model

	estimate	s.e.	parameter
1	384.3	67.94	1
2	-184.4	76.57	D(2)

scale parameter 2.000

Iteration number 4. , deviance is 1058.39

16 replicates

deviance = 256.00 at cycle 4
residual df = 252

Mean model

	estimate	s.e.	parameter
1	33.90	1.660	1
2	32.43	2.753	B(2)
3	-18.33	1.743	D(2)
4	19.11	2.011	G(2)

scale parameter 1.016

-- model changed
scaled deviance = 299.19 at cycle 6
residual df = 254

Variance model

	estimate	s.e.	parameter
1	349.5	43.69	1
2	-121.9	52.13	D(2)

scale parameter 2.000

Iteration number 4. , deviance is 2144.79

32 replicates

deviance = 512.00 at cycle 4
residual df = 508

Mean model

	estimate	s.e.	parameter
1	35.88	1.282	1
2	29.05	1.859	B(2)
3	-20.10	1.332	D(2)
4	17.98	1.419	G(2)

scale parameter 1.008

-- model changed
scaled deviance = 652.81 at cycle 6
residual df = 510

Variance model

	estimate	s.e.	parameter
1	323.8	28.58	1
2	-104.9	34.51	D(2)

scale parameter 2.000

Iteration number 4. , deviance is 4299.31

64 replicates

deviance = 1024.0 at cycle 4
residual df = 1020

Mean model

	estimate	s.e.	parameter
1	35.08	0.8127	1
2	30.22	1.347	B(2)
3	-19.34	0.8622	D(2)
4	19.09	1.043	G(2)

scale parameter 1.004

-- model changed
scaled deviance = 1395.7 at cycle 6
residual df = 1022

Variance model

	estimate	s.e.	parameter
1	394.4	24.60	1
2	-197.2	27.50	D(2)

scale parameter 2.000

Iteration number 3. , deviance is 8591.38

5.4 Discussion

It is impractical to have many observations in a Taguchi experiment. Generally, we may have not more than 10 replicates for each treatment combination. Therefore, based on our simulation results, our macros do not seem to perform very well. However, in the simulation above, only one design and a specific model $IG(\mu_i,\lambda_i)$ are being considered. This is definitely not enough. We should do more simulations and consider more general conditions; for example, use other fractional factorial designs, choose other specifications for parameters μ and λ, or consider other link functions. In the last situation, we should change our macros correspondingly.

Although the macros do not perform satisfactorily, they do provide a way to analyze Taguchi's experiments with inverse Gaussian distribution.

Appendix A GLIM 4 macros for fitting models

```
!
! This macro fits the mean regression model with weights lamba
!
$macro  mean
$yvar %1 $error i $link i $weight ww $offset ofsm
$ca %lp=1 $
$fit #modm $
$ca di=(%yv-%fv)**2/%fv**2/%yv : t=2*%pi*%yv**3 $
$endmac
!
! This macro fits the dispersion model
!
$macro var
$yvar di $error g $scale 2 $weight $link r $offset ofsv
$ca %lp=1$
$cycle 25 $
$fit #modv$
$ca %d=%cu(di/%fv+%log(%fv*t))$
$ca ww=1/%fv$
$endmac
!
! This is an initialising macro
!
$macro setup
$ca %k=20 : %n=%m=%k-1 : %e=1
$ca ofsm=0 : ofsv=0 : ww=1 : %c=0$
$endmac
!
! This macro drives the model fitting
!
$macro driver
$arg mean %1$
$use mean$
$use var$
$ca %e=(1-%le((%c-%d)/%d,.00001)*%ge((%c-%d)/%d,-.00001)):%n=%n-1:%c=%d $
$ca %e=%e*%gt(%n,0) $
$ca %u=%m-%n $output %poc $
$pr 'iteration number' *-1 %u ' , deviance is ' *6 %d $
$output $
$endmac
```

```
!
! This is the macro you actually use, with argument the response variable
!
$macro vmod
$arg driver %1$
$use setup$
$output
$ca %e=%gt(%m,0)$
$ca %u=%m-%n $
$while %e driver $
$output %poc $
$arg mean %1$
$use mean$
$pr 'Mean model' $d e$
$use var$
$ca %u=%u+1$
$pr 'Variance model' $d e$
$print 'Iteration number' *-1 %u ' , deviance is ' *6 %d$
$output %poc
$endmac
!
$return
```

Appendix B File for data input

```
!
!       This file is for data input. The data are stored in 'OBS.DAT'
!         The replicates for a combination appear in a cycle of 16.
!   If less data are to be used, ignore the commands after appropriate remark.
!
$unit 16
$data a b c d e f g h $
$read
 1 1 1 1 1 1 1 1
 2 1 1 1 2 2 2 1
 1 2 1 1 2 2 1 2
 2 2 1 1 1 1 2 2
 1 1 2 1 2 1 2 2
 2 1 2 1 1 2 1 2
 1 2 2 1 1 2 2 1
 2 2 2 1 2 1 1 1
 1 1 1 2 1 2 2 2
 2 1 1 2 2 1 1 2
 1 2 1 2 2 1 2 1
 2 2 1 2 1 2 1 1
 1 1 2 2 2 2 1 1
 2 1 2 2 1 1 2 1
 1 2 2 2 1 1 1 2
 2 2 2 2 2 2 2 2
$factor a 2 b 2 c 2 d 2 e 2 f 2 g 2 h 2 $
$unit 64
$data y
$dinput 'obs.dat'$
$ass a=a,a,a,a:b=b,b,b,b:c=c,c,c,c:d=d,d,d,d
  : e=e,e,e,e:f=f,f,f,f:g=g,g,g,g:h=h,h,h,h $
! use only 4 replicates up to here
!
$data y1
$din 'obs.dat'$
$ass a=a,a:b=b,b:c=c,c:d=d,d:e=e,e:f=f,f:g=g,g:h=h,h:y1=y1,y1$
$unit 128
! use only 8 replicates up to here
!
```

```
$data y2
$din 'obs.dat'$
$ass a=a,a:b=b,b:c=c,c:d=d,d:e=e,e:f=f,f:g=g,g:h=h,h:y1=y1,y2$
$unit 256
!  use only 16 replicates up to here
!
$data y3
$din 'obs.dat'$
$ass a=a,a:b=b,b:c=c,c:d=d,d:e=e,e:f=f,f:g=g,g:h=h,h:y1=y1,y3$
$unit 512
!  use only 32 replicates up to here
!
$data y4
$din 'obs.dat'$
$ass a=a,a:b=b,b:c=c,c:d=d,d:e=e,e:f=f,f:g=g,g:h=h,h:y1=y1,y4$
$unit 1024
!  use all 64 replicates up to here
!
$return
```

References

Aitkin, M. (1987) Modelling variance heterogeneity in normal regression using GLIM. *Applied Statistics*, **36**, 332-339.

Barndorff-Nielsen, O. & Blasesild, P. (1983) Exponential models with affine dual foliations. *Annals of Statistics*, **11**, 753-769.

Blaesild, P. & Jensen, J.L. (1985) Saddlepoint formulas for reproductive exponential models, *Scandinavian Journal of Statistics*, **12**, 193-202.

Chan, N.N., Li, K.H. & Sharpe, K. (1992) Analyzing Taguchi's experiments using GLIM, *Total Quality Management*, **3**, 189-196.

Chhikara, R.S. & Folks, J.L. (1989) *The Inverse Gaussian Distribution: Theory, Methodology, and Applications*, (Dekker).

Dobson, A.J. (1990) *An Introduction to Generalized Linear Models*, (Chapman and Hall).

Healy, M.J.R. (1988) *GLIM: An Introduction*, (Oxford, Clarendon Press).

Johnson, N.L. & Kotz, S. (1970) *Distributions in Statistics: Continuous Univariate Distributions 1.* (Boston: Houghton Mifflin).

Jorgensen, B. (1982) Statistical Properties of the Generalized Inverse Gaussian Distribution, *Lecture Notes in Statistics*, **9** (Springer-Verlag, New York).

Logothetis, N. & Wynn, H.P. (1989) *Quality Through Design* (Oxford, Clarendon Press).

Madhav S. Phadke (1989) *Quality Engineering Using Robust Design* (Prentice Hall).

Michael, J.R., Schucany, W.R., & Haas, R.W. (1976) Generating random variables using transformation with multiple roots. *Amer. Statist.*, **30**, 88-90.

McCullagh, P. & Nelder J.A. (1989) *Generalized Linear Models*, 2nd ed., (London, Chapman and Hall).

Montgomery, D.C. (1991) *Design and Analysis of Experiments*, (John Wiley & Sons).

Nelder, J.A. & Wedderburn, R.W.M. (1972) Generalized linear models, *Journal of the Royal Statistical Society, Series A*, **135**, 370-384.

Patel, R.C. (1965) Estimates of parameters of truncated inverse Gaussian distribution, *Annals of the Institute of Statistical Mathematics, Tokyo*, **17**, 29-33.

Schrödinger, E. (1915) Zur Theorie der Fall- und Steigversuche an Teilchenn mit Brownscher Bewegung. *Physikalische Zeitschrift*, **16**, 289-295.

Shuster, J.J. (1968) On the inverse Gaussian distribution function, *Journal of the American Statistical Association*, **63**, 1514-1516.

Smyth, G.K. (1989) Generalized linear models with varying dispersion, *Journal of the Royal Statistical Society, Series B*, **51**, 47-60.

Tweedie, M.C.K. (1941) A mathematical investigation of some electrophoretic measurements on colloids, M.Sc. Thesis, University of Reading, England.

Tweedie, M.C.K. (1957a) Statistical properties of inverse Gaussian distributions I. *Annals of Mathematical Statistics*, **28**, 362-377.

Tweedie, M.C.K. (1957b) Statistical properties of inverse Gaussian distributions II. *Annals of Mathematical Statistics*, **28**, 696-705.