BAYESIAN DETECTION OF ACTIVE EFFECTS IN
DESIGNED EXPERIMENTS MODELED WITH GLM´S

Román de la Vara and Victor Aguirre Torres

Comunicación Técnica No I-07-06/27-03-2007
(PE/CIMAT)

Bayesian Detection of Active Effects in Designed Experiments Modeled with GLM's.

ROMÁN de la VARA
Centro de Investigación en Matemáticas, Guanajuato, GTO 3600

VÍCTOR AGUIRRE TORRES
Instituto Tecnológico Autónomo de México, México, DF 01000

March 21, 2007

Abstract

Unreplicated fractional factorial experiments with response modeled with Generalized linear models (GLM) are found more and more frequently in industrial applications. GLM analysis relies heavily on large sample results. This paper presents a Bayesian method for detecting the active effects in unreplicated factorial experiments analyzed by a GLM that does not require the large sample assumption. The proposed method is based on Bayesian model selection. In the examples shown, the Bayesian method produces more consistent results than inference based on Wald's test, and in a simulated example the usual approach brakes down while the Bayesian method identifies the significant effects correctly. The method is presented for the 2^k experiments, but it can easily be generalized to other designs.

Key Words: *Bayesian Information Criteria, Bayesian Selection of Models, Posterior Probabiliy, Quasi-Monte Carlo Simulation, Significant Effects, Small Sample Analysis, Unreplicated Factorial.*

1. Introduction

The problem of detecting the active effects in unreplicated two-level factorial experiments with normal responses is a challenge because, in the saturated case, there are not enough degrees of freedom to estimate the error variance. Consequently, standard F tests cannot be used to identify the active effects.

One of the earliest and better known methods for the analysis of unreplicated designs with normal response was proposed by Daniel (1959) and consists of drawing the estimated effects on a normal probability plot. On the graph, the effects that fall along a straight line through the origin are assumed to be inactive and normally distributed with mean zero and constant variance, while the active effects tend to fall off the line. The subjectivity of Daniel´s method, has attracted much attention since the 1980´s, see Hamada and Balakrishnan (1998). There are now more than 40 methods that attempt the objective detection of the active effects in unreplicated factorial experiments with normal response. It is important to say that most of the methods rely on the sparsity principle, according to which only a small proportion of the effects are expected to be active.

One of such methods is by a Bayesian approach originally proposed by Box and Meyer (1986), which computes the posterior probability of each possible model that can be constructed from the estimated effects, and from here, marginalizing, obtains the posterior probability of each effect or coefficient. The coefficients with posterior probabilities greater than 0.5 are considered as candidates to be active effects, see also Box and Meyer (1993).

In many industrial experiments the response variable is not normaly distributed. Notably, when the data are counts or proportion of defectives, or the response may have an skewed distribution, see Lewis, *et al.* (2001). The traditional approach for analyzing such data is to apply a variance-stabilizing transformation to the response variable, and then use ordinary least squares with the transformed data. Hopefully the transformation will also induce normality and constancy of variance on the response and simplify the empirical model. Another approach is to use a generalized linear model (GLM), in which the normality and constant variance is no longer required, Myers and Montgomery (1997); Hamada and Nelder (1997).

When the GLM approach is used in the analysis of nonnormal responses in unreplicated (or replicated) factorial experiments the inference about the effects is based on the asymptotic properties of the maximum likelihood estimator. Typically the unreplicated experiments have a small number of observations and it is not clear how the asymptotic inference performs in this situation. With the purpose of detecting the active effects, Myers, et. al. (2002) recommend the normal probability plot of the coefficients estimates divided by their standard errors. Although the interpretation of the graph is subjective, it is very useful for identifying active effects, except when the correlations between the estimates is large, because in that case the estimator of an active effect may pull the corresponding estimator of a non significant effect and make it look as if it were significant. The reverse situation may also happen. In GLM the estimators may not be independent even if the design matrix is orthogonal. Myers, *et al.* (2002) show some examples of this situation, see for instance Example 7.5, where the response is number of grille defects in an unreplicated 2^{n-p} experiment, the correlation matrix for the log link (given in Table 7.25) contains several "ones" which should be interpreted as very highly correlated estimators .

This paper proposes a Bayesian method for detecting active effects in an unreplicated factorial experiment analyzed by a GLM. The idea is to generalize the Box and Meyer (1993) method to the case GLM. The proposed method is based on Bayesian model selection, an important research topic in the Bayesian literature, see for example, Raftery (1995, 1999), also used for doing Bayesian model averaging, Hoeting, *et. al.* (1999), Clyde (1999).

Section 2 gives a brief account of the GLM. Section 3 presents the basics of Bayesian model selection, a procedure that requires the computation of the prior predictive density (ppd) for a series of models. Section 4 provides two methods for computing the ppd. The first one aims at the direct computation of the ppd and hence requires the elicitation of a prior distribution of the parameters of the model, while the second one uses a Laplace approximation for the ppd and avoids the explicit definition of a prior distribution. Section 5 shows the implementation of the approach to experiments where the response has a Poisson distribution. This section also shows the application of the procedure to the car grille example. Section 6 gives the instrumentation of the idea to the case when the response has a binomial distribution. This section gives two examples, in the first one the frequentist and Bayesian analysis agree, but in the second one the frequentist analysis brakes down while the Bayesian analysis correctly detects the significant effects . Section 7 deals with the gamma distribution, this model differs from the other two because requires the elicitation of an extra parameter. Section 8 provides the concluding remarks.

2. Generalized Linear Model

The generalized linear model (GLM) was introduced by Nelder and Wedderburn (1972) and discussed in detail in McCullagh and Nelder (1989). This approach allows for regression modeling when the responses are distributed as one of the members of the exponential family. The normal, Poisson, binomial, exponential, gamma and negative binomial distributions are all members of the exponential family. Except for the normal distribution, in a GLM the variance is a function of the mean.

We have $\mathbf{y}^t = [y_1, y_2, \ldots, y_n]$ a vector of independent observations, with vector of means $[\mu_1, \mu_2, \ldots, \mu_n]$. The observation y_i has a distribution that is a member of the exponential family, that is

$$f\left(y_i \,|\zeta_i, \phi\right) = \exp\left\{r(\phi)[y_i\zeta_i - b(\zeta_i)] + c(y_i, \phi)\right\},\ (i = 1, 2, ..., n),$$

where $r(), b()$, and $c()$ are specific functions depending on the family of distributions. The parameter ζ_i is called the natural location parameter, while ϕ is called the dispersion parameter. The systematic part of the model involves the factors of the experiment represented by the variables $x_1, ...x_k$. The model is built around the linear predictor $\eta = \beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k$. The model is found through the use of a link function $\eta_i = g(\mu_i)$, which relates the linear predictor with the mean (μ_i) of the specified distribution for the response variable. The link function is monotonic and differentiable. The canonical link is such that $\eta_i = \zeta_i$. The variance $Var(y_i)$ is a function of μ_i.

The significance of the coefficients in generalized linear models is usually verified via the Wald test statistic, which is based on the asymptotic normality of the maximum likelihood estimator. Each coefficient divided by the standard error is distributed asymptotically standard normal. The square of these ratios is distributed asymptotically chi-square with one degree of freedom (χ_1^2). The performance of the asymptotic quantile of these tests in small samples is not clear.

One recommendation given by Myers *et al.* (2002), p. 270, for the purpose of detecting the active effects in GLM, is to draw the standardized coefficients on a normal probability plot, but the usefulness of this graph depends on the size of the correlations between the coefficients, as will be discussed later in more detail with the examples. An alternative analysis would to use a link that leads to uncorrelated estimators, if such a link exists. This results in asymptotically uncorrelated estimators of the coefficients, which facilitates the interpretation of the normal probability plot.

3. Bayesian Model Selection

Consider a factorial experiment 2^k without replications, where $n = 2^k$ is the total number of observations. The problem is to detect the statistically significant effects in the set of $2^k - 1 = n - 1$ effects. If we let $m = 2^{n-1} - 1$, then with these $n - 1$ effects it is possible to construct 2^{n-1} models denoted by M_0, M_1, M_2,..., M_m where M_0 is the constant model without variables. The model i is associated with the vector of parameters $\boldsymbol{\theta}_i = (\beta_{i0}, \beta_{i1}, ..., \beta_{it_i})$. Let $\mathbf{y}$ be the experimental observations or data vector. If the distribution of $\mathbf{y}$ given the model is denoted by $f(\mathbf{y} \,| M_i, \boldsymbol{\theta}_i)$, the prior probability of M_i by $p(M_i)$ and the *prior* probability density of $\boldsymbol{\theta}_i$ is $f(\boldsymbol{\theta}_i \,| M_i)$, then the prior predictive density (ppd) of $\mathbf{y}$ or integrated likelihood, given the model M_i, is

$$f(\mathbf{y} \,| M_i) = \int\limits_{\Theta_i} f(\mathbf{y} \,| M_i, \boldsymbol{\theta}_i) f(\boldsymbol{\theta}_i \,| M_i) d\boldsymbol{\theta}_i \tag{1}$$

where Θ_i is the space set of $\boldsymbol{\theta}_i$. The posterior probability of the model M_i, given the data $\mathbf{y}$, is

$$p(M_i \,| \mathbf{y}) = \frac{p(M_i) f(\mathbf{y} \,| M_i)}{\sum\limits_{h=0}^{m} p(M_h) f(\mathbf{y} \,| M_h)}. \tag{2}$$

The prior probability $p(M_i)$ is computed as follows: let α be the prior probability that any one effect is active, then form empirical evidence on fractional factorial experiments, known as factor sparsity, $(0 < \alpha < 0.4)$, then the probability of observing a model with t_i significant effects will be taken as $\alpha^{t_i}(1-\alpha)^{n-t_i}$, Box and Meyer (1993), which is proportional to $(\alpha/(1-\alpha))^{t_i}$.

Then the posterior probability that the effect T_j is active is computed by adding the posterior probabilities of the models that contain this particular effect, that is

$$P_j = \sum_{M_i : T_j \; is \;\; active} p(M_i \,| \mathbf{y}) \,. \tag{3}$$

The application of this procedure requires the calculation of the ppd (1) for a large amount of models, which implies solving a multiple integral on the parameter space. Frequently this is a complicated problem soluble only by numerical methods or by using some analytic approximation Papageorgiou and Traub (1997). In this work we propose to approximate the ppd by using two methods: Quasi-Monte Carlo simulation and Bayesian information criteria (BIC). To get an idea of the amount of work consider that in the unreplicated design with 15 effects we need to evaluate $2^{15} = 32768$ integrals. Of course, it is possible to reduce the number of integrals by observing that, if the factor sparsity applies, the posterior probabilities of interest converge to their values considering models of at most 4 or 5 terms.

4. Computing the Prior Predictive Density

4.1. Quasi-Monte Carlo Approach

The value of the integrated likelihood is viewed as the expected value of the likelihood function with respect to the *prior* distribution of $\boldsymbol{\theta}_i$. This method approximates (1) by using low discrepancy sequences, Niederreiter (1992), instead of pseudo random numbers. Hence if N values of $\boldsymbol{\theta}_i$ are obtained from $f(\boldsymbol{\theta}_i \,|M_i)$ and the likelihood function is evaluated on each $\boldsymbol{\theta}_i$. The average of the N evaluations of the function approximates the integrated likelihood, that is,

$$f(\mathbf{y}\,|M_i) = \widehat{E}[f(\mathbf{y}\,|M_i, \boldsymbol{\theta}_i)] = \frac{1}{N}\sum_{j=1}^{N} f(\mathbf{y}\,|M_i, \boldsymbol{\theta}_{ij})\ . \tag{4}$$

Hence this approach requires the explicit specification of $f(\boldsymbol{\theta}_i \,|M_i)$. In order to do that we assume that the parameters are independently normally distributed. Consider first the parameter β_0 which is linked to original scale by the relationship

$$\beta_0 = g(\mu_0). \tag{5}$$

In order to obtain a prior distribution for β_0 notice that μ_0 could have the following two interpretations: first it could be considered the mean response when none of the effects are significant, or if the factors in the experiment are continuous, then it is the mean response when all the factors in the experiment are set to zero, that is it is the mean response in the central region of the experiment. Notice that μ_0 is related with an observable characteristic in the experiment. Whatever interpretation applies, this approach assumes that the experimenter has some broad idea about the value of this mean response and it is stated in terms of a probability interval of the form

$$P(L_\mu < \mu_0 < U_\mu) = 1 - \delta_\mu, \tag{6}$$

where L_μ and U_μ are a lower and upper bound for μ_0, and δ is a small fraction, say 5% or 1%. Assuming a strictly increasing link function from (5) and (6) it follows that

$$P(g(L_\mu) < \beta_0 < g(U_\mu)) = 1 - \delta_\mu \tag{7}$$

Then, one way to fulfill (7) with a normal distribution is to take the parameters for the prior density of β_0 as

$$\mu_{\beta_0} = \frac{g(L_\mu) + g(U_\mu)}{2}; \qquad \sigma_{\beta_0} = \frac{g(U_\mu) - \mu_{\beta_0}}{z_{1-(\delta_\mu/2)}} \tag{8}$$

where z_ξ is the $\xi - th$ percentile of the standard normal distribution. If the link function is strictly decreasing then μ_{β_0} remains the same but σ_{β_0} changes to

$$\sigma_{\beta_0} = \frac{g(L_\mu) - \mu_{\beta_0}}{z_{1-(\delta_\mu/2)}}.$$

For the rest of the parameters β_i, $1 \leq i \leq k$, we will assume a $N(0, \sigma^2_{\beta_0})$ distribution, where the zero mean is introduced since it is supposed that in advance there is no information about the sign of the effect.

In the examples that we discuss later, for the case of the experiment with 15 effects, we use the Quasi-Monte Carlo approach for models with at most 4 terms using $N = 1000$ quasi-random repetitions for each model. Thus, 1941 integrals were approximated. We wrote an R program, R Development Core Team (2006), for calculating (4) for each model: first we generate 1000 Halton numbers of dimension 5, that are used to obtain samples from the prior distributions for the parameters of models with 0-4 explanatory variables. The sample of parameters for a specific model are evaluated in the likelihood function accordingly to the model dimension and the average that represent the posterior probability of the model is calculated.

In the much smaller experiment with 8 runs and 7 effects the integral is approximated for each of the total 128 possible models.

4.2. The BIC Approach

The integrated likelihood given in (1) can be approximated by using the Bayesian information criteria. For details in the following development you can see Raftery (1995). For simplicity the integral that we want to approximate can be rewritten not mentioning the model, so (1) becomes

$$f(\mathbf{y}) = \int f(\mathbf{y}\,|\boldsymbol{\theta})\; f(\boldsymbol{\theta})d\boldsymbol{\theta} \tag{9}$$

Consider a Taylor series expansion of $g(\boldsymbol{\theta}) = \log\{f(\mathbf{y}\,|\boldsymbol{\theta})\ f(\boldsymbol{\theta})\}$ about $\widetilde{\boldsymbol{\theta}}$ the value of $\boldsymbol{\theta}$ that maximizes $g(\boldsymbol{\theta})$, i.e. the posterior mode, the integral of interest can be approximated by

$$\begin{aligned} f(\mathbf{y}) &= \int \exp\left[g(\boldsymbol{\theta})\right] d\boldsymbol{\theta} \\ &\approx \exp\left[g\left(\widetilde{\boldsymbol{\theta}}\right)\right] \int \exp\left[\frac{1}{2}\left(\boldsymbol{\theta}-\widetilde{\boldsymbol{\theta}}\right)^t \frac{\partial^2 g(\widetilde{\boldsymbol{\theta}})}{\partial\theta\partial\theta^T}\left(\boldsymbol{\theta}-\widetilde{\boldsymbol{\theta}}\right)\right] d\boldsymbol{\theta}. \end{aligned} \tag{10}$$

The integrand in equation (10) is proportional to a multivariate normal density, then

$$f(\mathbf{y}) \approx \exp\left[g\left(\widetilde{\boldsymbol{\theta}}\right)\right] (2\pi)^{d/2} |A|^{-1/2} \tag{11}$$

where d is the number of parameters in the model and $A = -\frac{\partial^2 g(\widetilde{\boldsymbol{\theta}})}{\partial\theta\partial\theta^T}$. This equation is the basis for the Laplace approximation approach, Bernardo and Smith (1994), page 341. The error in the approximation is $O(n^{-1})$, Tierney and Kadane (1986), and so

$$\log f(\mathbf{y}) = \log f(\mathbf{y}\,|\widetilde{\boldsymbol{\theta}}) + \log f(\widetilde{\boldsymbol{\theta}}) + (d/2)\log(2\pi) - (1/2)\log|A| + O(n^{-1}). \tag{12}$$

In large samples $\widetilde{\boldsymbol{\theta}} \approx \widehat{\boldsymbol{\theta}}$ where $\widehat{\boldsymbol{\theta}}$ is the MLE, $A \approx nI_F$ and $|A| \approx n^d |I_F|$, where I_F is the expected Fisher information matrix for one observation. This is a $d \times d$ matrix whose (i, j) element is $-E\left[\frac{\partial^2 \log f(y_1|\theta)}{\partial\theta_i\partial\theta_j}\Big|_{\theta=\widehat{\theta}}\right]$. These two approximations introduce an error of order $O(n^{-1/2})$ into equation (12), which becomes

$$\log f(\mathbf{y}) = \log f(\mathbf{y}\,|\widehat{\boldsymbol{\theta}}) + \log f(\widehat{\boldsymbol{\theta}}) + \frac{d}{2}\log(2\pi) - \frac{d}{2}\log n - \frac{1}{2}\log|I_F| + O(n^{-1/2}). \tag{13}$$

Assuming a multivariate normal prior $f(\boldsymbol{\theta})$ with mean $\widehat{\boldsymbol{\theta}}$ and variance matrix I^{-1}, called the unit information prior, Raftery (1999), we have

$$\log f(\widehat{\boldsymbol{\theta}}) = -(d/2)\log(2\pi) + (1/2)\log|I_F|, \tag{14}$$

and substituting this equation into (13) gives

$$\log f(\mathbf{y}) = \log f(\mathbf{y}\,|\widehat{\boldsymbol{\theta}}) - (d/2)\log n + O(n^{-1/2}). \tag{15}$$

This last expression for $\log f(\mathbf{y})$ can be used to approximate the Bayes factor $B_{21} = f(\mathbf{y} \mid \mathbf{M}_2)/f(\mathbf{y} \mid \mathbf{M}_1)$, which on the scale twice the logarithm is

$$2 \log B_{21} = 2 \left(\log f(\mathbf{y} \mid \widehat{\boldsymbol{\theta}}_2, M_2) - \log f(\mathbf{y} \mid \widehat{\boldsymbol{\theta}}_1, M_1) \right) - (d_2 - d_1) \log n + O(n^{-1/2}). \tag{16}$$

If M_1 is nested within M_2 this equation becomes

$$2 \log B_{21} \approx \chi^2_{21} - df_{21} \log n \tag{17}$$

where χ^2_{21} is the standard likelihood ratio test (LRT) statistic for testing M_1 against M_2 and $df_{21} = d_2 - d_1$ is the number of degrees of freedom associated with the test.

When several models are being compared, it is useful to compare each of them in turn with a baseline model, usually either the null model (M_0) with no independent variables or the saturated model (M_S) with all independent variables, into which each data point is fit exactly. When the saturated model M_S is the baseline model the LRT statistic in equation (17) is called the deviance.

The value of the Bayes information criterion (BIC) for model M_k, denoted by BIC_k, is the approximation to $2 \log B_{Sk}$ given by equation (17), where B_{Sk} is the Bayes factor for model M_S against model M_k. That is

$$BIC_k = D^2_k - df_k \log n \tag{18}$$

where $D^2_k = \chi^2_{Sk}$ is the deviance for model M_k and df_k is the corresponding number of degrees of freedom. Smaller BIC_k means that the fit of M_k is better. The built in function "deviance" from R was used to compute (18).

In our case, we have the family of models $\{M_0, M_1, \ldots, M_m\}$, and we want to approximate the posterior model probability $p(M_k \mid \mathbf{y})$ given in (2) where $f(\mathbf{y} \mid M_k)$ is the integrated likelihood. From (16) and (18), approximately, $f(\mathbf{y} \mid M_k) \propto \exp(-(1/2)BIC_k)$. Thus, the posterior probability $p(M_k \mid \mathbf{y})$ in terms of BIC_k´s is

$$p(M_k \mid \mathbf{y}) \approx \frac{p(M_k) \exp(-\frac{1}{2} BIC_k)}{\sum_{h=0}^{m} p(M_h) \exp(-\frac{1}{2} BIC_h)}. \tag{19}$$

Hence this procedure is expected to work better for larger samples and does not require the specification of a prior density. Also the computation of (19) is very fast.

5. Poisson Response

Let $\mathbf{y}^t = [y_1,\, y_2, \ldots, y_n]$ be the vector of observations, where y_j is the number of defects or events occurred in the treatment j. If $E(y_j) = \mu_j$ then the density of y_j is

$$f(y_j) = \frac{1}{y_j!} \exp(-\mu_j) \left(\mu_j\right)^{y_j}.$$

For this family, the functions are: $\zeta_j = \log(\mu_j), b(\zeta_j) = \mu_j, r(\phi) = 1, \phi = 1$, and $c(y, \phi) = -\log(y!)$. Hence the canonical link is $\eta_j = \log(\mu_j)$. Consider the model M_i with vector of parameters $\boldsymbol{\theta}_i = \left[\beta_{i0},\ \beta_{i1}, \ldots,\ \beta_{it_i}\right]$ and log link function given by

$$g(\mu_j) = \log \mu_j = \mathbf{x}_j^t \boldsymbol{\theta}_i. \tag{20}$$

Then under model M_i, and loglink function, the distribution of the observation y_j is

$$f(y_j \,|M_i, \boldsymbol{\theta}_i) = \frac{1}{y_j!} \exp\left(-e^{\mathbf{x}_j^t \boldsymbol{\theta}_i}\right) \left(\exp(\mathbf{x}_j^t \boldsymbol{\theta}_i)\right)^{y_j}. \tag{21}$$

The likelihood function considering the n observations is

$$\begin{aligned} f(\mathbf{y} \,|M_i, \boldsymbol{\theta}_i\,) \;&=\; \prod_{j=1}^{n} \frac{1}{y_j!} \exp\left(-e^{\mathbf{x}_j^t \boldsymbol{\theta}_i}\right) \left(\exp(\mathbf{x}_j^t \boldsymbol{\theta}_i)\right)^{y_j}. \qquad (22) \\ &\propto\; \exp\left(\sum_{j=1}^{n} \left(-e^{\mathbf{x}_j^t \boldsymbol{\theta}_i} + y_j \mathbf{x}_j^t \boldsymbol{\theta}_i\right)\right) \end{aligned}$$

As we have already mentioned, the use of the Monte Carlo approach assumes that the experimenter should have some broad idea about μ_0. We assume that this knowledge is stated in the form of (6). Then using formulas (8) provides values for the hyperparameters for the prior densities. In the case of the log link the hyperparameters result

$$\mu_{\beta_0} = \frac{\log(L_\mu) + \log(U_\mu)}{2}; \qquad \sigma_{\beta_0} = \frac{\log(U_\mu) - \mu_{\beta_0}}{z_{1-(\delta_\mu/2)}} \tag{23}$$

Thus, the if prior distributions are

$$\begin{aligned} \beta_0 &\sim N(\mu_{\beta_0}, \sigma^2_{\beta_0}) \\ \beta_i &\sim N(0, \sigma^2_{\beta_0}). \end{aligned} \tag{24}$$

Then the prior predictive distribution, $f(\mathbf{y}\,|M_i)$, for the GLM Poisson with log link is the expectation of

$$\frac{1}{\prod y_j!} \exp\left(\sum_{j=1}^{n}\left(-e^{\mathbf{x}_j^t\boldsymbol{\theta}_i} + y_j\mathbf{x}_j^t\boldsymbol{\theta}_i\right)\right),$$

when $\boldsymbol{\theta}_i$ has a density $N[(\mu_{\beta_0}, 0, ..., 0)^T, \sigma^2_{\beta_0} I]$, where I is the identity matrix of dimension $t_i + 1$. It is not possible to get a closed form expression of this expectation. Hence, in order to compute the posterior probabilities of the models, the value of the integral is approximated by using Quasi-Monte Carlo simulation. It is customarily for this distribution to use the square root link, that is $g(\mu) = \sqrt{\mu}$, then for this case the formulas for the hyperparameters become:

$$\mu_{\beta_0} = \frac{\sqrt{L_\mu} + \sqrt{U_\mu}}{2}; \quad \sigma_{\beta_0} = \frac{\sqrt{U_\mu} - \mu_{\beta_0}}{z_{1-(\delta_\mu/2)}} \tag{25}$$

and the likelihood function changes accordingly.

If the Bayesian information criteria approximation is used then no prior consideration is required.

5.1. Example: Car Grille Experiment

This example corresponds to the example 7.5 of Myers *et al.* (2002) page 272 and also is the example 1 from Myers and Montgomery (1997). The experiment is a fractional factorial 2^{9-5} with resolution III, which investigated the impact of nine factors on the number of observed defects in the finishing of sheet-molded grille opening panels. The experimental design, along with the observed defects and some information about the alias structure, is shown in Table 1. The c column contains the observed defects and the last column is the Freeman and Tukey $\left(FT = \left(\sqrt{c} + \sqrt{c+1}\right)/\mathbf{2}\right)$ modification to the square root transformation that was used in Bisgaard and Fuller (1994-1995) to analyze this data.

#	A	B	C	D	E	F	G	H	J	c	FT
1	−1	−1	−1	−1	1	−1	1	−1	1	56	7.52
2	1	−1	−1	−1	1	−1	−1	1	−1	17	4.18
3	−1	1	−1	−1	−1	1	1	−1	−1	2	1.57
4	1	1	−1	−1	−1	1	−1	1	1	4	2.12
5	−1	−1	1	−1	1	1	−1	1	1	3	1.87
6	1	−1	1	−1	1	1	1	−1	−1	4	2.12
7	−1	1	1	−1	−1	−1	−1	1	−1	50	7.12
8	1	1	1	−1	−1	−1	1	−1	1	2	1.57
9	−1	−1	−1	1	−1	1	1	1	1	1	1.21
10	1	−1	−1	1	−1	1	−1	−1	−1	0	0.50
11	−1	1	−1	1	1	−1	1	1	−1	3	1.87
12	1	1	−1	1	1	−1	−1	−1	1	12	3.54
13	−1	−1	1	1	−1	−1	−1	−1	1	3	1.87
14	1	−1	1	1	−1	−1	1	1	−1	4	2.12
15	−1	1	1	1	1	1	−1	−1	−1	0	0.50
16	1	1	1	1	1	1	1	1	1	0	0.50

$$l_1 = A + BJ + CG$$
$$l_2 = B + AJ + DE$$
$$l_3 = C + AG + EF$$
$$l_4 = D + BE + GH$$
$$l_5 = E + BD + CF$$
$$l_6 = F + CE + HJ$$
$$l_7 = G + AC + DH$$
$$l_8 = H + DG + FJ$$
$$l_9 = J + AB + FH$$
$$l_{10} = AD + CH + EG$$
$$l_{11} = AE + FG + JD$$
$$l_{12} = AF + BH + EG$$
$$l_{13} = AH + BF + CD$$
$$l_{14} = BC + DF + GJ$$
$$l_{15} = BG + CJ + EH$$

Table 1. Data Car Grille Experiment.

5.2. Frequentist Analysis

We present the results of the frequentist analysis for reference. The analysis of FT by Bisgaard and Fuller (1994-1995) declares active the effects D, F and the $BG + CJ + EH$ alias chain. Myers and Montgomery (1997) analyze the data with a Poisson model with log link function and they found these same significant effects, using Wald inference. Myers *et al.* (2002) redo the analysis using log and square root links functions. Figure 1 shows the corresponding half-normal probability plot of the standardized coefficients for the log link.

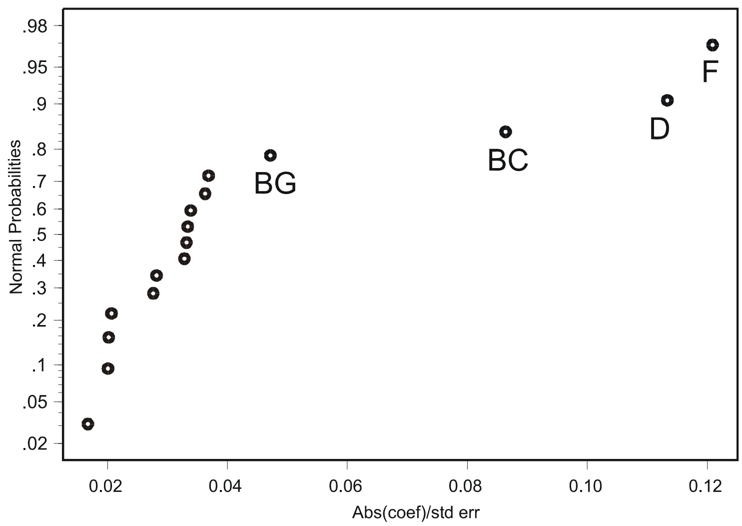

Figure 1: Car Grille Experiment. Half-normal probability plot of standardized effects. Poisson log link.

Figure 1 suggests that the effects F, D, and BC are clearly significant, while the effect BG is barely significant. They conclude, with the log link, that the active effects are D, F and $BC + DF + GJ$. Figure 2 shows the corresponding half-normal probability plot of the standardized coefficients for the square root link.

Figure 2 shows as active effects F and D as before, but $BG + CJ + EH$ that was not very significant before now it is clearly significant. The effect BC that was clearly significant before now it is not significant, and a new effect AD appears barely significant. Hence the results are different for the two link function. Myers *et al.* (2002) conclude, based on the residual analysis of each model, that the model with square root link fits the data better.

Myers *et al.* (2002) also include the covariance and correlation matrices for both links, but they do not seem to use them. According to the covariance matrix for the log link, Figure 1, $corr(D, BC) = 0.9999788$, hence it is very likely that BC is not significant but it was pulled out by D which is significant. Also, in the same Figure 1, BG which is significant could have been pulled in by A which is non significant since also $corr(A, BG) = 0.9999788$. For Figure 2 Myers *et al.* (2002) report that the correlation matrix is the identity and hence no pulling effect is expected.

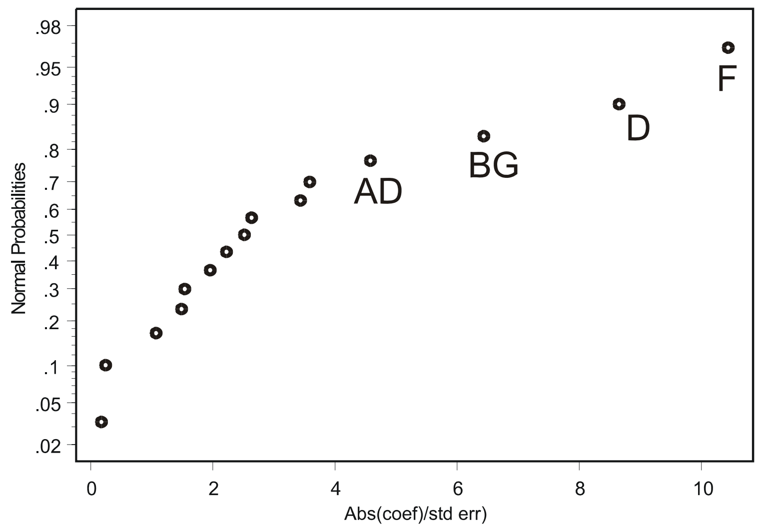

Figure 2: Car Grille Experiment. Half-normal probability plot of standardized effects. Poisson square root link.

5.3. Bayesian Analysis.

For the Quasi Monte Carlo approach let us assume a broad interval for μ_0 say $0.5 < \mu_0 < 50$ with 99% probability, hence $\delta_\mu = 0.01$. Then considering the Poisson GLM with log link and formulas (23) give

$$\mu_{\beta_0} = 1.61; \sigma_{\beta_0} = 0.89,$$

similarly, with the square root link (25) give the corresponding hyper-parameters for the prior distribution of β_0 as

$$\mu_{\beta_0} = 3.89; \sigma_{\beta_0} = 1.24.$$

The integrated likelihood in each model was approximated with 1000 quasi-random number (Halton sequences) evaluations to simulate values for the parameters β_0 and β_i. Assuming the factor sparsity principle, only models up to 4 independent terms are considered (1941 models). Also a prior probability $\alpha = 0.2$ of active effect was assumed. The posterior probabilities of the effects assuming the Poisson model with log link are shown in the first row of Table 2. Clearly the active effects are F, D and BG. The second row of Table 2 shows the results obtained for the Poisson model with square root link. Hence with the Monte Carlo method, the same effects F, D and BG (and its alias), are detected as active and none of the other interactions AD or BC appeared to be significant.

Method	*0*	*A*	*B*	*C*	*D*	*E*	*F*	*G*
MClog	0.0	0.0	0.0	0.0	1.0	0.0	1.0	0.0
MCroot	0.0	0.0	0.0	0.0	0.99	0.0	0.98	0.0
BIClog	0.0	0.07	0.03	0.02	1.0	0.2	1.0	0.03
BICroot	0.0	0.0	0.0	0.0	1.0	0.0	1.0	0.0

Method	*H*	*J*	*AD*	*BC*	*CD*	*BG*	*AE*	*AF*
MClog	0.0	0.0	0.03	0.03	0.0	0.96	0.0	0.0
MCroot	0.0	0.0	0.0	0.01	0.0	0.98	0.0	0.1
BIClog	0.01	0.01	0.05	0.02	0.02	0.99	0.03	0.02
BICroot	0.0	0.0	0.97	0.01	0.0	0.99	0.0	0.01

Table 2. Car Grille Experiment. Poisson Model. Posterior Probabilities of Being Active.

The rows of Table 2 that represent the results of the BIC approximation for the log and square root links, were made by using also a prior probability $\alpha = 0.2$ of an effect being active. Although the BIC approximation is significantly less computer intensive, the posterior probabilities are based, as in the Quasi-Monte Carlo approach, in models from 0 to 4 independent terms (1,941 models). For the log link the detected active effects are D, F and BG. With the square root link function the same active effects are detected, but the effect AD appears also as active. This agrees with Figure 2 in the frequentist analysis, but disagrees with the quasi-Monte Carlo approach with the same link function. Hence the Bayesian methodology implemented with the Monte Carlo was the only method that showed consistency for both link functions.

6. Binomial Response

Consider an experiment with binary response, let y_j be the number successes observed out of n_j trials ($j = 1, 2, ..., n$), processed at treatment j. Let p_j be the probability of success with the treatment $\mathbf{x}_j$. The density of y_j is

$$f(y_j) = \begin{pmatrix} n_j \\ y_j \end{pmatrix} (p_j)^{y_j} \, (1\text{-}p_j)^{n_j - y_j}.$$

In this case the GLM applies to y_j/n_j and for this family, the functions are: $\zeta_j = \log(\frac{p_j}{1-p_j})$, $b(\zeta_j) = \log(1+\zeta_j)$, $r(\phi) = 1, \phi_j = 1/n_j$, and $c(y_j, \phi) = \log \begin{pmatrix} n_j \\ y_j \end{pmatrix}$. Hence the canonical link is $\eta_j = \log(\frac{p_j}{1-p_j})$. Therefore the logistic regression model can be written in the form

$$p(\mathbf{x}_j) = \frac{1}{1+e^{-\mathbf{x}_j^t \boldsymbol{\theta}}} = \frac{e^{\mathbf{x}_j^t \boldsymbol{\theta}}}{e^{\mathbf{x}_j^t \boldsymbol{\theta}} + 1}, \tag{26}$$

where $\boldsymbol{\theta} = (\boldsymbol{\beta}_0, \boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_k)$ is the parameter vector in the linear predictor. Alternatively, the model can be written in the linearized form given by

$$\ln \left[\frac{p(\mathbf{x}_j)}{1 - p(\mathbf{x}_j)} \right] = \mathbf{x}_j^t \boldsymbol{\theta}. \tag{27}$$

The distribution of y_j given the model M_i with parameter $\boldsymbol{\theta}_i$ is defined as

$$\begin{aligned} f\left(y_j \left| M_i \right., \boldsymbol{\theta}_i\right) &= \begin{pmatrix} n_j \\ y_j \end{pmatrix} (p_i)^{y_j} (1-p_i)^{n_j - y_j} = \begin{pmatrix} n_j \\ y_j \end{pmatrix} \left(\frac{p_i}{1-p_i} \right)^{y_j} (1-p_i)^{n_j} \qquad (28) \\ &= \begin{pmatrix} n_j \\ y_j \end{pmatrix} \left(e^{y_j \mathbf{x}_j^t \boldsymbol{\theta}_i} \right) \left(\frac{1}{1+e^{\mathbf{x}_j^t \boldsymbol{\theta}_i}} \right)^{n_j}. \end{aligned}$$

Ignoring the combinatorial factor, the likelihood function for the n observations is given by

$$f\left(\mathbf{y} \left| M_i \right., \boldsymbol{\theta}_i\right) \propto \prod_{j=1}^{n} \left(e^{y_j \mathbf{x}_j^t \boldsymbol{\theta}_i} \right) \left(\frac{1}{1+e^{\mathbf{x}_j^t \boldsymbol{\theta}_i}} \right)^{n_j} \tag{29}$$

Once again, we make the mild assumption that the experimenter has some vague idea on the proportion of defectives p_0. That is we expect an interval of the form $L_p < p_0 < U_p$ and some probability associated to this interval, say $1 - \delta_p$. In this case the link function is given in terms of

$$\eta_i = \log \left(\frac{p_i}{1-p_i} \right),$$

hence we have the relation

$$\beta_0 = \log\left(\frac{p_0}{1-p_0}\right) = g(p_0),$$

and the formulas (8) can be applied directly to obtain the hyperparameters as follows, since the link function is increasing:

$$\mu_{\beta_0} = \frac{\log(\frac{L_p}{1\text{-}L_p}) + \log(\frac{U_p}{1-U_p})}{2}; \qquad \sigma_{\beta_0} = \frac{\log(\frac{U_p}{1-U_p}) - \mu_{\beta_0}}{z_{1-(\delta_p/2)}}, \tag{30}$$

Notice that if the interval $L_p < p_0 < U_p$ is symmetric around 0.5 then $\mu_{\beta_0} = 0$. Considering again the prior normal distribution with mean zero and variance $\sigma^2_{\beta_0}$ for the parameters in the model M_i that multiply the independent variables, the prior predictive density, $f(\mathbf{y}\,|M_i)$, is the expectation of

$$\prod_{j=1}^{n}\left(e^{\mathbf{x}_j^t\boldsymbol{\theta}_i}\right)^{y_j}\left(\frac{1}{1+e^{\mathbf{x}_j^t\boldsymbol{\theta}_i}}\right)^{n_j}$$

when $\boldsymbol{\theta}_i$ has a density $N[(\mu_{\beta_0}, 0, ..., 0)^T, \sigma^2_{\beta_0} I]$.

6.1. Example: Survival of Sperm Experiment.

This experiment, Myers, *et al.* (2002) page 116, is about the survival of spermatozoa in a sperm bank, stored in sodium citrate and glycerol. The amounts of these substances were varied along with equilibrium time in a factorial array. The response variable is the number of samples that survive, meaning that the sample has the ability to impregnate. Fifty samples of material were used in each experimental point. The purpose of the study is to assess the effects of the factors on proportion of survival. The data are given in Table 2, where the column labeled Y gives the number of samples that survived.

A (Sodium Citrate)	B (Glycerol)	C (Equilibrium Time)	Y
-1	-1	-1	34
1	-1	-1	20
-1	1	-1	8
1	1	-1	21
-1	-1	1	30
1	-1	1	20
-1	1	1	10
1	1	1	25

Table 3. Data from Survival of Sperm Experiment.

In the frequentist analysis Myers *et al.* (2002) find, using Wald test, that the active effects are B (glycerol) and AB (interaction of sodium citrate and glycerol). Assuming a broad interval for p_0, say $0.1 < p_0 < 0.9$ with probability of 99% then the following hyperparameters are obtained using (30)

$$\mu_{\beta_0} = 0; \quad \sigma_{\beta_0} = 0.85.$$

Table 4 is obtained with a prior probability of active effect $\alpha = 0.2$. Clearly the effects B and AB are active, with posterior probabilities almost equal one, while the other effects have smallish posterior probabilities. This result agrees with the frequentist analysis. The BIC approximation detects the same active effects B and AB.

$Effect$	0	A	B	C	AB	AC	BC
MC	0.0	0.02	0.99	0.03	0.99	0.03	0.06
BIC	0.0	0.02	0.99	0.01	0.99	0.01	0.02

Table 4. Sperm Survival Experiment. Logistic Model. Posterior Probabilities of Being Active

6.2. Example: Simulated Binomial Experiment.

This example is presented to test the procedures in the case of a small sample within treatments but a large sample of treatments. The example is motivated by the discussion of the assumptions on the deviance function to work as a goodness of fit test given in section 4.4.3 of McCullagh and Nelder (1989). In that section they mention that a crucial assumption is that n the dimension of the vector of observations remains fixed and that for each j, $n_j \to \infty$ and $n_j p_j(1-p_j) \to \infty$.

In our example we use an unreplicated 2^{5-1} experiment, the fraction considered is $ABCDE = +1$, hence $n = 16$ and $n_j = 10$ for every j. The design matrix along with the simulated responses are given in Table 5. The linear predictor is given by the following equation

$$\eta = 2A - 3B + 3C + 2BC$$

run	**A**	**B**	**C**	**D**	**E**	**Y**
1	1	1	1	1	1	1
2	-1	-1	1	1	1	3
3	-1	1	-1	1	1	10
4	1	-1	-1	1	1	0
5	-1	1	1	-1	1	3
6	1	-1	1	-1	1	0
7	1	1	-1	-1	1	10
8	-1	-1	-1	-1	1	5
9	-1	1	1	1	-1	4
10	1	-1	1	1	-1	0
11	1	1	-1	1	-1	10
12	-1	-1	-1	1	-1	3
13	1	1	1	-1	-1	1
14	-1	-1	1	-1	-1	0
15	-1	1	-1	-1	-1	10
16	1	-1	-1	-1	-1	0

Table 5. Simulated Binomial Experiment.

Although there is a clear effect of the explanatory variables on the response, the frequentist approach via GLM does not work well, since the built in function in R that fits the GLM converged to unseasonable results, particularly for the models that contained the significant effects.

Now consider the Bayesian approach using Monte Carlo to compute the posterior probabilities. We take the interval for p_0 as $0.1 < p_0 < 0.9$ with probability of 99%, then the following hyperparameters are obtained using (30)

$$\mu_{\beta_0} = 0; \quad \sigma_{\beta_0} = 0.85.$$

Table 6 shows that in this case the MC approach identified correctly all of the significant effects.

$Eff.$	0	A	B	C	D	E	AB	AC	AD	AE	BC	BD	BE	CD	CE	DE
MC	0.0	0.93	1.0	0.99	0.0	0.0	0.0	0.0	0.0	0.0	0.99	0.0	0.0	0.0	0.0	0.0
BIC	0.0	0.98	1.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.98	0.0	0.0	0.0	0.0	0.0

Table 6. Simulated Binomial Experiment. Logistic Model. Posterior Probabilities of Being Active.

Using the BIC approach caused some problems, because this procedure fits a GLM every time that computes a BIC measure, then in some cases the program gives senseless adjusted models, particularly when the model is one of the 12 models that contains three out of the four significative effects. Nevertheless, because the procedure is averaging over 1940 models, in the end the posterior probabilities shown in Table 5 identify the correct simulated effects.

7. Gamma Response

Let y_j be the response of the experiment under treatment j. It is assumed that the density of y_j is given by

$$f(y_j) = \frac{1}{\Gamma(r)\lambda_j} e^{-y_j/\lambda_j} y_j^{r-1} \tag{31}$$

then Myers et. al (2002) give $a(\phi) = r^{-1}, \zeta_j = -\frac{1}{r\lambda_j}, b(\zeta_j) = -\log(-\zeta_j)$ and $c(y_j, \phi) = r\log(r) - \Gamma(r) + (r-1)\log(y_j)$. This model assumes that the parameter r is the same for all observations. But $E(y_j) = \mu_j = r\lambda_j$ hence the canonical link is $g(\mu_j) = \frac{1}{\mu_j} = \eta_j$ but this link is not used since some estimates may give negative values for η_j. Then a log link will be used, that is $\log(\mu_j) = \eta_j$.

In order to obtain a prior distribution for β_0 the procedure is the same as the one given in section 4.1 using the log link. Hence formulas (23) apply here too. We considered again the prior normal distribution with mean zero and variance $\sigma^2_{\beta_0}$ for the parameters in the model M_i that multiply the independent variables.

But now we have the additional parameter r. It is well known that $Var(y_j) = \frac{\mu_j^2}{r}$ hence the coefficient of variation of y_j is given by: $CV(y_j) = \frac{\sqrt{Var(y_j)}}{\mu_j} = \frac{1}{\sqrt{r}}$. To get a prior distribution for r we assume that a probability interval for the coefficient of variation of the response is available, that is there are constants L_c, U_c , and δ_c such that

$$P(L_c < CV(y) < U_c) = 1 - \delta_c,$$

then

$$P(\frac{1}{U_c^2} < r < \frac{1}{L_c^2}) = 1 - \delta_c, \tag{32}$$

since r is a positive, we assume a two parameter $\Gamma(a, b)$ distribution. In order to fulfill (32) we need to solve in (a, b) the following system of nonlinear equations:

$$P(\Gamma(a,b) < \frac{1}{U_c^2}) = w\delta_c \quad \text{and} \quad P(\Gamma(a,b) > \frac{1}{L_c^2}) = (1-w)\delta_c \tag{33}$$

where $0 < w < 1$ is a weight that in the example we took as 0.5. We solved (33) by minimizing

$$Q_w(a,b) = \left[P(\Gamma(a,b) < \frac{1}{U_c^2}) - w\delta_c\right]^2 + \left[P(\Gamma(a,b) > \frac{1}{L_c^2}) - (1-w)\delta_c\right]^2,$$

with respect to (a, b). A possible strategy is to start with $w = 0.5$ and if a solution is not found then iterate on w.

In order to obtain the prior predictive density in terms of the GLM we reparameterize (31) as a function of μ_j :

$$f(y_j) = \frac{r^r}{\Gamma(r)\mu_j^r} e^{-ry_j/\mu_j} y_j^{r-1}$$

and since $\mu_j = e^{\mathbf{x}_j^t \boldsymbol{\theta}_i}$ then

$$f\left(y_j \mid M_i\ , \boldsymbol{\theta}_i, r\right) = \frac{r^r}{\Gamma(r)} e^{-r\mathbf{x}_j^t \boldsymbol{\theta}_i} e^{-r y_j e^{-\mathbf{x}_j^t \boldsymbol{\theta}_i}} y_j^{r-1}$$

finally the prior predictive density, $f(\mathbf{y} \mid M_i)$, is the expectation of

$$\left(\frac{r^r}{\Gamma(r)}\right)^n e^{-r\Sigma_{j=1}^n \mathbf{x}_j^t \boldsymbol{\theta}_i} e^{-r\Sigma_{j=1}^n y_j e^{-\mathbf{x}_j^t \boldsymbol{\theta}_i}} \prod\nolimits_{j=1}^n \mathbf{y}_j^{r-1} \tag{34}$$

when $\boldsymbol{\theta}_i$ has a density $N[(\mu_{\beta_0}, 0, ..., 0)^T, \sigma_{\beta_0}^2 I]$, and r is independent and has a $\Gamma(a,b)$ density. In the case of a 2^k fraction, (34) slightly simplifies to:

$$\left(\frac{r^r}{\Gamma(r)}\right)^n e^{-rn\beta_0} e^{-r\Sigma_{j=1}^n y_j e^{-\mathbf{x}_j^t \boldsymbol{\theta}_i}} \prod\nolimits_{j=1}^n \mathbf{y}_j^{r-1}$$

7.1. Example: Drill experiment.

This example appeared in Daniel (1976), it is an unreplicated 2^4 experiment, the factors define drilling conditions: A (Load), B (Flow), C (Speed), and D (Mud type), and the response is drill advance. The data is given in Table 7.

Table 7. Data from Drill Experiment.

Run	A	B	C	D	y
1	−1	−1	−1	−1	1.68
2	1	−1	−1	−1	1.98
3	−1	1	−1	−1	3.28
4	1	1	−1	−1	3.44
5	−1	−1	1	−1	4.98
6	1	−1	1	−1	5.70
7	−1	1	1	−1	9.97
8	1	1	1	−1	9.07
9	−1	−1	−1	1	2.07
10	1	−1	−1	1	2.44
11	−1	1	−1	1	4.09
12	1	1	−1	1	4.53
13	−1	−1	1	1	7.77
14	1	−1	1	1	9.43
15	−1	1	1	1	11.75
16	1	1	1	1	16.30

The analysis of the original response favors strongly the significance of C and B effects, and then there is a slight evidence of significance of D, BC and CD. A graphical analysis of the residuals of the model suggests that the variance depends on the level of the response. Box, Hunter and Hunter (1978) use this example to illustrate the technique of power transformation to induce constancy of variance, they arrive to the log transformation. The normal plot of the effects of log drill advance suggests that effects C, B and D are significant. Aguirre (1993) analyzed this data using the rank transformation of the original data and using the Daniel plot of the effects taking the ranks as response, and found the effects C, D and B to be significant. Then Lewis et. al. (2001) analyzed the data with a GLM and a log link, using the half normal probability plot of the standardized effects, they find evidence that effects B, D and C are clearly significant.

For the Bayesian analysis, using the Monte Carlo approach we assume an interval for μ_0 of $[0.5, 12]$ with 95% probability, this interval seems reasonable from the data, (23) produces the following parameters

$$\mu_{\beta_0} = 0.90; \quad \sigma_{\beta_0} = 0.81.$$

For the coefficient of variation we computed first a global value with $\frac{s_y}{\bar{y}} = 0.68$, hence we chose the interval of $[0.15, 3.6]$ with 95% probability. This interval allows a variation of 5.3 times above $(U_c/0.68)$ and 4.5 times below $(0.68/L_c)$ which gives a broad range of variation for the unknown CV. After solving (33) with $w = 0.5$, we get the hyperparameters

$$a = 0.72; \quad b = 14.47.$$

Using these values for the quasi Monte Carlo method produces the results in table 8.

$Eff.$	0	A	B	C	D	AB	AC	AD	BC	BD	CD
MC	0.0	0.0	0.99	0.99	0.01	0.0	0.0	0.0	0.0	0.0	0.0
BIC	0.3	0.06	0.1	0.44	0.07	0.06	0.06	0.06	0.06	0.06	0.06

Table 8. The Drill Experiment. Gamma Model. Posterior Probabilities of Being Active.

Consider the MC approach, the results are as expected with respect to the significance of C and B, however they are a bit surprising regarding the effect

D. Somehow when all possible models are considered the effect of D vanishes. Then how come D was clearly significant using the frequentist analysis? the asymptotic correlation matrix of the effects is diagonal for the gama distribution and the loglink, but sixteen observations is hardly a large sample, perhaps the small sample correlation matrix differs significantly from a diagonal matrix and we are observing a pull out effect. We conducted a sensitivity analysis of the posterior probabilities by changing the length of the intervals but the results were basically the same.

The results from the BIC approach in this case were completely different to the results of the previous analysis, very likely this outcome has to do with the fact that the BIC approach approximates the likelihood with a multivatiate normal density, something that seems to be grossly incorrect with the gamma model.

8. Concluding Remarks

In this paper we propose a method based on Bayesian model selection for detecting the active effects in unreplicated factorial experiments when the response variable is modeled by a GLM. For computing the posterior probabilities that the effects are active, the difficult step is to solve the multiple integral that defines the integrated likelihood. Once this integral is solved it is quite easy to obtain the posterior probabilities of all possible models and from here to compute the posterior probabilities for the effects.

Two approximations of the integrated likelihood were considered. The Quasi-Monte Carlo simulation approach and the BIC approach. Both approximations to the integrated likelihood gave good results in the discussed Poisson and binomial examples. The Quasi-Monte Carlo approach is more accurate because it is a direct integration of the function of interest, while the BIC approach is the Laplace approximation for the integrated likelihood. The Quasi-Monte Carlo approach requires the mild assumption that the user has some vague idea about the mean μ_0, the BIC approach does not. For the gamma model the MC approach additionally requires a probability interval for the coefficient of variation of the response.

The BIC approach has an advantage in the computing time compared with the Quasi-Monte Carlo approach. The former needs half a minute to compute the posterior probabilities for the 1941 models, while the second takes 23 minutes for the same calculation with 1000 simulation points for each integral. In the BIC approach it is not necessary to elicit any prior distribution. Implicitly the procedure is assuming a multivariate normal prior distribution for the parameters, centered on the maximum likelihood estimate and variance equal to the inverse Fisher information matrix. The BIC approach did not work well in the case of the gamma response, most likely due to the skewness of this distribution.

In general one may say that the Bayesian approach showed a better performance than the usual method of analysis, particularly in those situations where the large sample assumptions do not hold.

9. Acknowledgments

The second author was partially supported by Asociación Mexicana de la Cultura A.C. and he also conducted part of this research while on sabbatical leave at the Probability and Statistics Department of CIMAT.

References

[1] Aguirre-Torres V. (1993). "A Simple Analysis of Unreplicated Factorials with Possible Abnormalities". *Journal of Quality Technology*, 25, pp. 183-187.

[2] Bernardo, J. M. and Smith, A. F. M. (1994). *Bayesian Theory.* Wiley, New York.

[3] Box, G. E. P. and Meyer, R. (1986). "An Analysis for Unreplicated Fractional Factorials". *Technometrics,* 28, pp. 11-18.

[4] Box, G. E. P. and Meyer, R. (1993). "Finding the Active Factors in Fractionated Screening Experiments". *Journal of Quality Technology,* 25, pp. 94-104.

[5] Clyde, M. (1999). "Bayesian Model Averaging and Model Search Strategies" (with discussion). In *Bayesian Statistics* 6. J. M. Bernardo, A. P. Dawid, J. O. Berger and A. F. M. Smith Eds. Oxford University Press, pp. 157-185.

[6] Daniel, C. (1959). "Use of Half Normal Plots in Interpreting Factorial Two-level Experiments". *Technometrics*, 1, pp. 311-341.

[7] Hamada, M. and Nelder, J. A. (1997). "Generalized Linear Models for Quality-improvement Experiments". *Journal of Quality Technology,* 29, pp. 292-303.

[8] Hamada M. and N. Balakrishnan (1998). "Analyzing Unreplicated Factorial Experiments: a Review with Some New Proposals". *Statistica Sinica,* 8, pp. 1-41.

[9] Hoeting, J. A., Madigan, D., Raftery, A. E. and Volinsky, C. T. (1999). "Bayesian Model Averaging: a Tutorial". *Statistical Science,* 14, pp. 382-417.

[10] Lewis, S. L. and Montgomery, D. C. (2001). "Examples of Designed Experiments with Nonnormal Responses. *Journal of Quality Technology,* 33, pp. 265-278.

[11] McCullagh, P. and Nelder, J. A. (1989). *Generalized Linear Models*, 2nd ed., Chapman and Hall, New York.

[12] Myers, R. H. and Montgomery, D. C. (1997). "A Tutorial on Generalized Linear Models". *Journal of Quality Technology,* 29, pp. 274-291.

[13] Myers, R. H., Montgomery, D. C. and G. Vining (2002). *Generalized Linear Models.* Wiley, New York.

[14] Nelder, J. A. and Wedderburn, R. W. M. (1972). "Generalized Linear Models". *Journal of the Royal Statistical Society A,* 135, pp. 370-384.

[15] Niederreiter, H. (1992). *Random Numbers Generation and Quasi-Monte Carlo Methods. CBMS-NSF Regional Conference Series in Applied Mathematics*, Vol. 63, SIAM, Philadelphia.

[16] Papageorgiou, A. and Traub, J. F. (1997). "Faster Evaluation of Multidimensional Integrals". *Computers in Physics*, November, pp. 574-578.

[17] Raftery, A. E. (1995). "Bayesian Model Selection in Social Research" (with discussion) *Sociological Methodology,* 25, pp. 111-196.

[18] Raftery, A. E. (1999). "Bayes Factors and BIC: Comment on Weakliem". *Sociological Methods and Research* 27, pp. 411-427.

[19] R Development Core Team (2006). "R: A Language and Environment for Statistical Computing". R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org.

[20] Tierney, L. and Kadane, J. B. (1986). "Accurate Approximation for Posterior Moments and Marginal Densities". *Journal of the American Statistical Association,* 81, pp. 82-86.