Estimation and Inference for 2^{k-p} Experiments with Beta Response

Luis Fernando Grajales Hernández
Matemático, MSc Estadística
Code: 196619

Universidad Nacional de Colombia
Facultad de Ciencias
Departamento de Estadística
Bogotá, D.C.
February 2015

Estimation and Inference for 2^{k-p} Experiments with Beta Response

Luis Fernando Grajales Hernández
Matemático, MSc Estadística
Code: 196619

Dissertation Presented for the Degree of
PhD in Sciences-Statistics

Advisor
Luis Alberto López Pérez, Ph.D.
Ph.D. Experimental Statistics

Coadvisor
Óscar Orlando Melo Martínez,Ph.D.
Ph.D. Statistics

Universidad Nacional de Colombia
Facultad de Ciencias
Departamento de Estadística

Bogotá, D.C.
February 2015

Title in English

Estimation and Inference for 2^{k-p} Experiments with Beta response

Título en español

Estimación e inferencia para experimentos 2^{k-p} con respuesta beta

Abstract: Fractional factorial experiments are widely used in industry and engineering. The most common interest in these experiments is to identify a subset of the factors with the greatest effect on the response. With respect to data analysis for these experiments, the most used methods include linear regression, transformations, and the Generalized Linear Model (GLM). This thesis focuses on experiments whose response is measured continuously in the $(0, 1)$ interval (if $y \in (a, b)$, then $\frac{y-a}{b-a} \in (0, 1)$). Analyses for factorial experiments in $(0, 1)$ are rarely found in the literature. In this work, advantages and drawbacks of the three mentioned methods for analyzing data from experiments in (0,1) are described. Here, as the beta distribution assumes values in (0,1), the beta regression model (BRM) is proposed for analyzing these kinds of experiments. More specifically, the necessity of considering variable dispersion (VD) and using linear restrictions on parameters are justified in data from 2^k and 2^{k-p} experiments. Thus, the first result in this thesis is to propose, develop, and apply a restricted VDBRM. The restricted VDBRM is developed from frequentist perspective: a penalized likelihood (by means of Lagrange multipliers), restricted maximum likelihood estimators with their respective Fisher Information Matrix, hypothesis tests, and a diagnostic measure. Upon applying the restricted VDBRM, good results were obtained for simulated data, and it is shown that the hypothesis related to 2^k and 2^{k-p} experiments are a special case of the restricted model. The second result of this thesis is to explore an integrated Bayesian/likelihood proposal for analyzing data from factorial experiments using the (Bayesian and frequentist) simple BRM's. This was done upon employing flat *prior* distributions in the Bayesian BRM. Thus, comparisons between confidence intervals (frequentist case) and credibility intervals (Bayesian case) on the mean response are done with good and promisory results in real experiments. This work also explores a technique for choosing the best model among several candidates which combine the Half-normal plots (given by the BRM) and the inferential results. Starting from the active factors chosen from each plot, subsequently the respective regression models are fitted and, finally, by means of information criteria, the best model is chosen. This technique was explored with the following models: normal, transformation, generalized linear, and simple beta regression for real 2^k and 2^{k-p} experiments: into the greater part of the examples considered for the Bayesian and frequentist BRM's, results were very similar (using flat *prior* distributions). Moreover, four link functions for the mean response in the BRM are compared: results highlight the importance to study each problem at hand.

Resumen: Los experimentos factoriales fraccionados se usan ampliamente en la industria y en la Ingeniería. El interés más común en estos experimentos es identificar el subconjunto de factores que tiene mayor efecto sobre la respuesta. Con respecto al análisis de datos de dichos experimentos, los métodos más usados incluyen regresión lineal, transformaciones y Modelo Lineal Generalizado (MLG). Esta Tesis se enfoca en experimentos cuya respuesta está medida continuamente en el intervalo (0,1), (si $y \in (a, b)$, entonces $\frac{y-a}{b-a} \in (0, 1)$). En la literatura se encuentran pocos análisis de

experimentos con esta respuesta. En este trabajo, se describen ventajas y desventajas de las tres metodologías mencionadas en experimentos con esta respuesta. Acá, como la distribución beta asume valores en (0,1), se propone el modelo de regresión beta (MRB) para analizar estos datos. Más específicamente, se justifica la necesidad de modelar la dispersión variable y usar restricciones sobre los parámetros se justifican en datos de experimentos 2^k y 2^{k-p}. De este modo, el primer resultado de esta Tesis es proponer, desarrollar y aplicar un modelo de regresión beta con dispersión variable y restricciones en los parámetros (MRBDV restringido). El modelo es desarrollado desde la perspectiva clásica: una función de verosimilitud penalizada (con multiplicadores de Lagrange), estimadores de máxima verosimilitud restringidos con su respectiva matriz de Información de Fisher, tests de hipótesis y una medidad de bondad de ajuste. Al aplicar el MRBDV restringido, se obtuvieron buenso resultados para datos simulados y se mostró que las hipótesis asociadas con experimentos 2^k y 2^{k-p} son un caso especial del modelo restringido. El segundo resultado de esta Tesis es explorar una propuesta integrada bayesiana/verosimil para analizar datos de experimentos factoriales usando los dos MRB (bayesiano y clásico). Esto se hizo al emplear distribuciones *a priori* planas (poco informativas) en el modelo bayesiano. Así, las comparaciones entre intervalos de confianza y de credibilidad presentaron buenos resultados y promisorios en experimentos factoriales reales. Esta Tesis tambien explora una técnica para elegir el mejor modelo entre varios candidatos, el cual combina los Half-normal plots (dados por el BRM) y resultados inferenciales. Partiendo de los efectos activos sgún cada gráfico, posteriormente se ajustan los modelos de regresión respectivos y, finalmente, por medio de criterios de información, se escoge el mejor modelo. Esta técnica fue explorada con los siguientes modelos: normal, transformaciones, MLG y MRB simple para datos reales de experimentos 2^k y 2^{k-p}: en la mayor parte de los ejemplos considerados, los MRB bayesiano y clásico presentaron resultados muy similares (usando distribuciones *a priori* planas). Además, se compararon cuatro funciones de enlace para el submodelo de la media: los resultados resaltan la importancia de estudiar cada problema específico

Keywords: factorial designs, restricted variable dispersion beta regression model, confidence regions, credibility regions, transformations

Palabras clave: diseños factoriales, modelo de regresión beta de dispersión variable restringido, regiones de confianza, regiones de credibilidad, estimadores restringidos

Acceptation Note

Thesis Work
Aprobada

Jury
Raúl Edgardo Macchiavelli

Jury
Jesús López Fidalgo

Jury
Jorge Franco Durán

Advisor
Luis Alberto López Pérez

Coadvisor
Óscar Orlando Melo Martínez,Ph.D.

Bogotá, D.C., February 2015

Thanks to

- My Mum: unconditional love.
- My four sisters: infinite help.
- My five brothers: support.
- My wife Marta Ines: love and patience.
- Sara, my daughter, and Emanuel, my son: motivation and inspiration.
- My good friends: good energy.
- My Professors, my colleagues, and my classmates: specially the academic and human qualities of Luis Alberto Lopez P., Oscar Orlando Melo M., Raydonal Ospina (Universidade Federal de Pernambuco, Recife, Brasil), Pedro Nel Pacheco, Sergio Calderon Villanueva, Esneyder Guerrero, J. Alfredo Jimenez, and Carlos Alonso.

I consider my Mum, my sisters and brothers, my wife and my children, my good friends, my Professors, my colleagues, and my classmates as invaluable gifts from my Heavenly Father.

Contents

List of Tables

List of Figures

Introduction and motivation

Factorial experiments are widely used in industry, engineering, the sciences, and product and process improvement. The 2^k experiments contain k factors at two-levels each, and several factors are evaluated simultaneously. When all possible factors cannot be evaluated, it is usual to perform only a fraction of the complete experiment, which leads to a 2^{k-p} fractional factorial experiment. The most common interest in these experiments is to identify the subset of the factors that has the greatest effect on the response, y. With respect to the response, this thesis focuses on experiments whose response is measured continuously in the $(0, 1)$ interval (if $y \in (a, b)$, then $\frac{y-a}{b-a} \in (0, 1)$). A lot of empirical data with these conditions are encountered in the statistical literature. Usually, the first statistical analysis for these data is the normal model (using ANOVA). Although linear normal model is widely applied, it has the difficulty of achieving normality and constant variance for the errors in several empirical experiments.

The traditional method of solving is to employ data transformations to induce three properties: to obtain approximate normality in the response, to stabilize variance, and to simplify the empirical model. A few times the three properties are found in one transformation. There are some potential problems associated with using data transformation, for instance, there is no guarantee that a single transformation will induce all the desirable properties, and some difficulties have been found when the inverse transformation is made back to the original scale. Two examples of appropriate transformations for data in (0,1) are *arcsin* and *logit*, and they are applied in this thesis in order to compare approaches. In cases where the first two approaches do not fit well, the next alternative is the generalized linear model (GLM) which neither models errors nor transforms the response; it models

a function of the mean response.

Data analysis from factorial experiments by means of these three approaches, linear normal model, data transformations, and GLM can be encountered in the literature. Both, good and bad results for these approaches and comparisons between them can be studied in Myers & Montgomery (1997), Lewis et al. (2001b), and more recently in Patil & Kulkarni (2011), Myers et al. (2011), and Bonat et al. (2012).

Now, considering that y lies in (0,1) and the beta distribution assumes values there, the beta regression model (BRM) might be an option so as to improve analysis or to offer additional comparisons with regard to the three mentioned approaches.

The possibilities of the BRM in factorial and fractional factorial experiments are illustrated afterwards by two examples and one discussion: 1) the simple BRM; 2) the variable dispersion BRM (VDBRM); and 3) the restricted VDBRM.

Example 1. *Justifying the simple BRM in a* 2^4 *experiment. Half-normal plots.*

The *Drill* experiment consists of a 2^4 unreplicated factorial investigating the response variable *advance rate* which assumes values in (0, 100). This experiment is quite well known in the literature. The design matrix and response data are shown in Table 2 at the end of this chapter. 14 effects are analyzed (4 main effects, 6 two-interaction effects, and 4 three-interaction effects). Initially, Daniel (1976) analyzed this data to illustrate the usefulness of normal probability plotting in order to identify large and potentially important factor effects. Also, Torres (1993), Montgomery (2001), Lewis et al. (2001b), Box et al. (2005), Montgomery (2009), and Myers et al. (2011) reanalyzed this data using linear normal model, logarithm transformation, rank transformation and GLM with *gamma* link. Here, *Half-normal* plots for normal, *log(y)* transformation, and generalized linear models are shown in the three graphics of Figure 1. This Figure indicates that the B, C, and D main effects appear as *active* in the three models mentioned. Note that BC and CD effects only were active for the normal model. The active effects visually chosen coincide with those reported in the literature, and they are shown in the first part of Table 1. (Notation $A = x_1$,..., $BCD = x_2x_3x_4$).

Therefore, a problem (and opportunity) is latent: different methods can yield different active effects. Could the BRM be another option?. ☑. Now, in this thesis, taking $y/100$ as the response, data is reanalyzed by means of the simple beta regression model (simple BRM)(Ferrari & Cribari-Neto (2004)). Initially, upon fitting a simple BRM

TABLE 1. Analysis of the *Drill* experiment

Authors (year)	Response	Method	Active factors
Daniel (1976)	y	Linear reg.	B,C,D, BC, CD
Box et al. (1978), Montgomery (2001)	$log(y)$	Linear reg.	B, C, D
Torres (1993)	y rank transf.	Plot	B, C, D
Lewis et al. (2001b), Myers et al. (2011)	y	GLM.gamma	B, C, D

(fixing dispersion parameter), four link functions were employed for the mean response: *logit*, *probit*, *cloglog* and *cauchit*. The first three link functions present similar results to those considered in Table 1. Instead, the *cauchit* link yielded a very interesting result, the active effects were: B, C, D, BC, BD, CD, and BCD. Therefore, although this is an empirical result, at the same time, it is an encouraging opportunity to pose some questions. With respect to the data analysis from factorial experiments with response in (0,1), two questions arise: if the beta regression model turns out to be empirically appropriate, can the BRM be proposed as a good alternative?. Can the graphical results be supported by the statistical inference of the BRM?.

Analyses were done using R software (R Development Core Team (2014)). The computational routine is shown in the Appendix 7.1

The inferential analysis of this example will be presented in chapter 4.6 ☑

The two previous questions lead to an interesting challenge from theoretical and applied

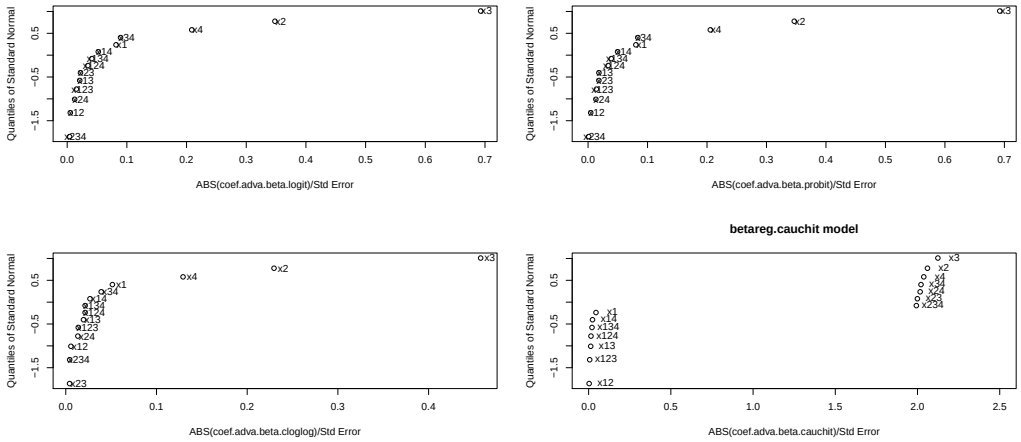

FIGURE 1. *Drill* data. Half-normal plots. Four simple BRM's

viewpoints. This challenge will be assumed in this thesis with two additional issues which are commented afterwards (variable dispersion, and restrictions on parameters).

An additional issue for the simple BRM: variable dispersion.

In the simple BRM fitted in Example 1, the dispersion parameter is taken as fixed. The generalization of this model allows for modelling the dispersion to try to explain the variability of the response through several covariates. It is known as the *variable (varying) dispersion beta regression model* (VDBRM). Several authors have discussed the advantages of this model with respect to the simple one. For instance, Bayer & Cribari-Neto (2014) present a simulation study to exemplify that efficiency loss takes place when the dispersion parameter is incorrectly taken as a constant. Therefore, in this thesis the variable dispersion BRM will be employed.

Example 2. *Justifying variable dispersion in the BRM for a* 2^4 *experiment.*

In example 1, the *Drill* data from a 2^4 experiment is described. Starting from Half-normal plots, the best model was the simple BRM with the *cauchit* link. Now, in order to illustrate the advantage of the Variable Dispersion BRM, two BRM's were fitted using the *cauchit* link (g_1) for the mean in both models, as follows:

- Simple BRM:

$$g_1(\mu) = \beta_0 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \beta_{23} x_2 x_3 + \beta_{24} x_2 x_4 \qquad log(\phi) = \alpha_0 \qquad (1)$$

- VDBRM:

$$g_1(\mu) = \beta_0 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \beta_{23} x_2 x_3 + \beta_{24} x_2 x_4 \qquad log(\phi) = \alpha_0 + \alpha_3 x_3 + \alpha_4 x_4 \quad (2)$$

After fitting the models and doing the residual plots, results are shown in figure 2. An inspection of figure 2 indicates that the VDBRM fits better than the simple one. ☑

Up to here, the variable dispersion beta regression model (VDBRM) has been motivated to be used in factorial experiments.

The last issue is commented on afterwards.

Restrictions on parameters for the VDBRM.

Under the normal linear model, a common methodology in 2^{k-p} experiments is to assume as *negligible* the higher-order interaction factors (3, 4, 5,... order: this assumption is related to the partial derivatives of Taylor series $\frac{\partial y}{\partial A \partial B \partial C}$, ..., Box et al. (2005)). When this

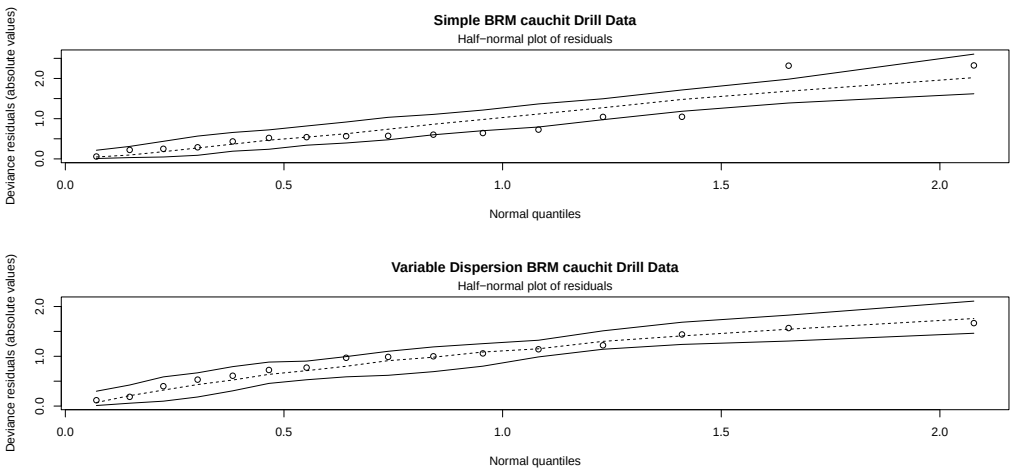

FIGURE 2. Half-normal residuals for *Drill* data. Simple and variable dispersion BRM

assumption is done, the estimations of main or low-order interactions effects can be expressed exactly. For instance, if we have the estimation $\widehat{A}+\widehat{BCD}$, and BCD is considered negligible, then we are assuming that BCD=0, and hence $\widehat{A}$ provides the estimation of the main effect A explicitly. In this thesis, using the VDBRM for analyzing data from 2^{k-p} experiments with response in (a,b), is proposed avoiding this assumption upon considering some restrictions on parameters associated to the higher-order interaction factors, that is to say, a *restricted VDBRM*. Thus, by means of hypothesis tests will be taken decisions as if certain interactions effect can be considered zero or not.

Finally, the objective of this thesis is to propose, develop, and apply a restricted variable dispersion BRM. After developing the general expressions for the inferential (restricted) results: estimation, hypothesis tests, and goodness of fit, the model will be applied in general problems, and in 2^{k-p} experiments with response in (a,b) also.

This thesis is organized as follows: Chapter 1 summarizes basic concepts, notation and literature review; also, models preceding the BRM for analyzing data from factorial experiments are presented. Chapter 2 introduces the proposed restricted variable dispersion BRM, with its specifications and estimates. Inferential aspects for the proposed model are detailed in Chapter 3 within a frequentist framework. Special cases (unrestricted models) for analyzing empirical and simulated data from fractional factorial experiments are discussed in Chapter 4.

Although discussions and conclusions given by for each considered method are presented throughout all the document, final conclusions are shown in Chapter 5. Chapter 6 indicates some future works based on the methods proposed, applied and discussed in this thesis;

also, some exploratory Bayesian solutions for the restricted VDBRM are presented. In the Appendix, some computational routines are presented; additional routines are available for the interested readers.

TABLE 2. *Drill* data: design matrix and response.

Run	x_1	x_2	x_3	x_4	y
1	-	-	-	-	1.68
2	+	-	-	-	1.98
3	-	+	-	-	3.28
4	+	+	-	-	3.44
5	-	-	+	-	4.98
6	+	-	+	-	5.70
7	-	+	+	-	9.97
8	+	+	+	-	9.07
9	-	-	-	+	2.07
10	+	-	-	+	2.44
11	-	+	-	+	4.09
12	+	+	-	+	4.53
13	-	-	+	+	7.77
14	+	-	+	+	9.43
15	-	+	+	+	11.75
16	+	+	+	+	16.30

CHAPTER 1

Framework

The primary purpose of this thesis is to propose and develop a restricted variable dispersion Beta regression model (Restricted VDBRM).
The secondary purpose is to apply this restricted VDBRM for analyzing data derived from 2^k and 2^{k-p} experiments whose response is measured continuously in (a, b) or $(0, 1)$. This chapter presents the basic concepts related to these kinds of experiments: the features of their design, and, the methods for analyzing data from these experiments. The restricted VDBRM will be introduced in the following chapter.

1.1 2^k and 2^{k-p} experiments: design and features.

In many scientific and industrial investigations, the interest lies in evaluating the effects of several factors, simultaneously. The two levels of a factor are denoted as +1 and -1. A 2^k design is an experiment with k factors of two-level each; this known as a *full* (complete) factorial design, and it has all combinations of the factor levels. A *run* is a combination of factor levels at which an experiment is to be carried out. Thus, with 8 factors, the full 2^8 factorial experiment requires 256 runs in a single replicate. In situations where a large number of factors is being investigated, and at the same time, it is necessary to take into account economy of space or material, it is not possible to perform a complete experiment. The solution is to evaluate a part of all possible factors. Performing a $(\frac{1}{2})^p$ fraction of

a 2^k experiment is extremely useful due to it leads to design a 2^{k-p} fractional factorial experiment. This fraction is favored because it needs only 2^{k-p} runs of the experiment. In this thesis, only *unreplicated* 2^k and 2^{k-p} experiments are studied. In works such as Fisher (1926), Box & Hunter (1961a), and Dey (1985) the advantages of the fractional factorial experiments are highlighted: i) some of the higher order interactions can be assumed to be zero from previous or historical knowledge, ii) in *screening* situations, it is expected that the effects of all but a few of the factors studied will be negligible, iii) groups of experiments are run in sequence and, ambiguities, which remain at a given stage of experimentation, can be resolved by a later group of experiments, iv) certain variables, which may interact, are studied simultaneously along with other variables whose influence, if any, can be described by main effects only, and v) using fewer runs, it is possible to detect which factors affect the variability of the response.

A summary of works related to experiments can be investigated in Steinberg & Hunter (1984) (optimal design, computer-aided design, robustness, response surfaces, mixtures, factorial design, blocking, nonlinear models, consulting, and teaching), and Dossou-Gbété & Tinsson (2005)(factorial experiments and generalized linear models).

A brief overview of the history of experiments designed under the normal theory is presented subsequently. The role of statistics in experimental design, and viceversa, was firmly established by Fisher's pioneering work in the United Kingdom (at Rothamsted Experimental Station) in the 1920's and 1930's. His monumental work was guided by the key insight that data analysis could be informative only if the data itself were informative, and that informative data could best be assured by applying statistical ideas to the way in which the data were collected. Fisher radically altered the role of the statistician: from one of after-the-fact technician to one of active collaborator at all stages of an investigation. He developed his insights concerning randomization, blocking, and replication; he invented new kinds of experimental designs; also, his contributions to genetical statistics, multivariate analysis, and Bioassays, among other areas, have been reported. He worked together with scientists who applied his ideas in their experiments; by mail, he advised experimenters in other places. According to Billard (2014), Fisher is considered as one of the two fathers of Statistics. Fisher was a Founder of the International Biometric Society. Several tributes have been carried out regarding Fisher's work, see Box (1978) and Billard (2014).

Factorial designs were first developed by Fisher and Yates at Rothamsted. Fisher (1926)

observed the advantages of factorial designs. A further advance was the introduction by Finney (1945) of fractional factorial designs. Plackett & Burman (1946) described a useful class of highly fractionated orthogonal designs, in which the main effects of $n-1$ two-level factors are estimated just using n runs. Box & Hunter (1961a) and Box & Hunter (1961b) described in detailed how to construct, apply, and analyze the 2^{k-p} fractional factorial designs. In the chemical, physical, and engineering sciences, the important contributions to experimentation made by factorial and fractional factorial designs were clearly evident. Several books describe and analyze two-level factorial and fractional factorial designs, for example, John (1971), Daniel (1976), Box et al. (1978), Box et al. (2005), and Montgomery (2009). With respect to some journals, *Biometrika*, *Biometrics*, *Technometrics* and *Journal of Quality Technology* have presented numerous papers regarding factorial experiments. Nowadays some papers suggest considering simulation studies for screening designs (see Anderson-Cook & Hamada (2014)).

In this short review, obviously, many contributors to the factorial experiments have not been intentionally omitted. The following step is to consider the data analysis from fractional factorial experiments but first the design matrix is described below.

The list of experimental runs in factorial experiments is called the *design matrix*, and is denoted by $\mathbf{X}$. There is a column for each of the k or $k-p$ variables, and each row gives the combinations of versions for each run. Here, $\boldsymbol{X}$ is a $n \times p_1$ matrix with $\mathbf{x}_i = (x_{i0}, x_{i1}, \ldots, x_{i,p_1-1})$ as its i-th row, and $x_0 = \mathbf{1}_n$ is an n-vector of ones. The rest of components of $\mathbf{X}$ belong to $\{-1, 1\}$ and their columns are orthogonal; $\mathbf{X}$ can be constructed from the columns of Hadamard matrices. An Hadamard matrix of order n, $\boldsymbol{M}$, is an orthogonal matrix with elements ± 1, that is, $\mathbf{M}^T\mathbf{M} = n\mathbf{I_n}$, where $\mathbf{I_n}$ is the nth-order identity matrix.

1.2 2^k and 2^{k-p} experiments: data analysis

The fractional and factorial experiments previously summarized are based on the normal theory, i.e., data from these experiments are analyzed mainly by linear regression, assuming normality for the errors, and this assumption is, consequently, associated with the normality of the observed responses.

Some examples of designs under the nonnormal perspective can be encountered in works such as Pearson (1931), Chaloner & Verdinelli (1995), and Wu et al. (2005). Under the

normal theory, efficiency and importance of 2^{k-p} trials have been quite justified and explained by Fisher (1926), Box & Hunter (1961a), and Dey (1985). Gaussian (normal) analysis of these trials appears in books such as Daniel (1976) (*Drill* experiment), Box et al. (2005) (Paint Trial Data), and Melo et al. (2007) (semiconductors data). The main feature of fractional factorial experiments is to yield models with more parameters than observations (supersaturated models). For instance, a full 2^4 experiment yields 16 observations and 16 effects can be estimated. However, these same 16 effects have to be estimated only with the 8 observations yielded by a 2^{4-1} experiment. Hence, a statistical problem appears, parameters are *not identifiable* (in the case of the 2^{4-1} experiment, it is clearly impossible to obtain 16 independent estimates with 8 observations). Theoretically, under normal designs, this problem is solved for *regular* fractional factorial experiments with a simple *aliasing* structure (two effects are either orthogonal or fully *aliased*). Two aliased effects are represented by the same parameter in the model. Knowing which effects are aliased depends on the *resolution* of each design and this has as a plausible consequence that the number of parameters has been reduced and, now, the data can be fitted by a non-saturated model with identifiable parameters.

In this thesis, designs of resolution III and IV are considered. More resolutions are detailed in Box & Hunter (1961a), Dey (1985), and Myers & Montgomery (1995), (i) A resolution-III plan is one in which no main-effect is aliased with any other main-effect, but main effects could be aliased with two-factor interactions. Also, two-factor interactions can be aliased amongst themselves; (ii) in a resolution-IV design the main-effects are not contaminated by the presence of main-effects and two-factor interactions, but could be aliased with three-factor interactions, (iii) resolution-V designs: these are designs in which no main effect or two-factor interaction is aliased with any other main effect or two-factor interaction, but two-factor interactions are aliased with three-factor interactions. In general, a design of resolution R is one in which no p factor effect is confounded with any other effect containing less than $R-p$ factors. According to Box & Hunter (1961a), particularly important designs are those for testing three variables in four runs, seven variables in eight runs and fifteen variables in sixteen runs. In this thesis, two real experiments, a 2^{8-4}_{IV} experiment (15 variables, 16 runs), and a 2^{4-1}_{III} experiment (7 variables, 8 runs) are analyzed. Moreover, some simulated data are based on these experiments. In the case of the *semiconductors* trial, a 2^{4-1}_{III} experiment, the importance of resolution-III can be summarized as follows: i) 16 original effects had to be estimated with only 8 observations

(runs); ii)the following 8 pairs of effects are *confounded*:

$$\mu \leftrightharpoons ABCD, \quad A \leftrightharpoons BCD, \quad B \leftrightharpoons ACD, \quad C \leftrightharpoons ABD, \quad D \leftrightharpoons ABC,$$
$$AB \leftrightharpoons CD, \quad AC \leftrightharpoons BD, \quad AD \leftrightharpoons BC. \tag{1.1}$$

The notation $A \leftrightharpoons BCD$ means that *the A effect is confounded with the BCD effect.*
For instance, $\mu \leftrightharpoons ABCD$ means that: i) μ is completely confounded with $ABCD$; ii) it is not possible to separate μ from $ABCD$; iii) when we estimate μ, we are really estimating μ + ABCD. (Myers & Montgomery (1995)). Denoting β_1 and β_{234} as the parameters associated to the effect of x_1 (or A) and $x_2x_3x_4$ (or BCD), respectively.
A special case: as $\mu \leftrightharpoons ABCD$, if the effect of ABCD, the four-factors interaction, is considered *negligible* then μ can be specified completely. Similarly, if BCD is considered negligible, the estimator of A should be completely specified.
As a notation: two or more effects that have the property shown in (1.1) are called *aliased.*
Thus, upon choosing one parameter from each pair of confounded effects in (1.1), and upon assuming that the effect of ABCD interaction factor is *negligible* (its effect is zero), only 7 effect have to be estimated with the same 8 observations.
With respect to analysis of 2^k and 2^{k-p} experiments in the literature, linear regression is the primary analysis for these data. Also, experiments with response in the (a, b) interval (Daniel (1976), Box et al. (2005), and Montgomery (2009)) are analyzed with the linear normal model. The linear regression is a method that involves two heavy conditions for the errors, normality and constant variance (*homoscedasticity*). The difficulty of achieving these assumptions was previously discussed in Chapter . Then, if a normal model turns out to be inappropriate for analyzing some data from experiments with response in (0,1), what if alternative models are proposed?.
Alternative methods such as transformations of y, GLM, and some approaches less employed or reported, can be encountered in Box & Cox (1964), McCullagh & Nelder (1989), and Bonat et al. (2012). One objective of this thesis is to propose the BRM as another option for analyzing factorial experiments with response in (0,1).
A summary of the alternative methods, transformations, GLM, and BRM are presented forthwith:
1) **Transformations**. In works such as Box & Cox (1964) and Aranda-Ordaz (1981), transformations mainly to y in order to get normality and homoscedasticity were pro-

posed. These transformations return responses with values on the real line, and subsequently they are modeled by means of linear normal model. For y in (0,1), there are two common transformations: *arcsin* and *logit*. These two transformations already have been applied in this thesis and are detailed in the *unrestricted models* chapter. Now, transformations have some drawbacks, for example: i) it is difficult to correct for both heteroscedasticity (non-constant variance) and normality, ii) the parameters of the model cannot be easily interpreted in terms of the original response, and iii) sometimes, inverse transformations yield nonsensical values.

2) **Generalized linear models (GLM)**. Nelder & Wedderburn (1972) and McCullagh & Nelder (1989) are benchmarks for advances in an wide class of regression models, unifying several model specifications under the flexible class of GLM for distributions belonging to the exponential family, including the normal distribution. Some authors have analyzed data from experiments in (a,b) by GLM and they have compared their results to other approaches. For instance, Lewis et al. (2001b) and Lewis et al. (2001a) analyzed real fractional and factorial experiments by means of the lengths of confidence intervals on the mean response. They present the advantages of using GLM instead of normal models. Also, Patil & Kulkarni (2011) present an analysis of 2^n experiments with response in the exponential family, comparing GLM against normal and transformations models; in Myers et al. (2011) normal, transformation and a GLM for some experiments with continuous response in (a,b) are compared, see the Table 1.

Due to the beta distribution (μ, ϕ) does not belong to the (one parameter) exponential family, it is difficult to find a satisfactory distribution within this family for modelling responses in (0,1).

3) The work of Bonat et al. (2012) deserves special attention. The authors present two lesser reported distributions, the Simplex and Kumaraswamy, which assume values in (0,1). They fit some regression models with these two mentioned distributions as the response and compare their fittings to three models for analyzing empirical data: the linear normal, the simple beta regression, and some transformations models.

4) When the response is in (0,1), it is natural to think of *beta distribution* because it assumes its values in the unity interval. The most common parametrization of its density involves μ, the mean parameter, and ϕ the dispersion parameter.

In the *simple BRM*, the mean response is related to a linear predictor that includes covariates and unknown regression parameters through a link function, and the dispersion

parameter is taken to be constant. The simple BRM was explicitly developed by many authors, such as Paolino (2001) and Ferrari & Cribari-Neto (2004). These latter provide a more detailed mathematical and computational description of the simple BRM developing the statistical inference from a frequentist perspective (parameter estimation, hypothesis tests, goodness of fit, and diagnostic measures, among others). Ferrari (2013) is a good review regarding the BRM, and according to her, many areas where the BRM model was applied such as percentage of time devoted to an activity, fraction of income spent on food, unemployment rate, fraction of surface covered by vegetation, can be found in the literature. The simple BRM has contributed to special topics such as optimal design (Wu et al. (2005)), diagnostics (Espinheira et al. (2008b), Espinheira et al. (2008a), and Chien (2011), and Figueroa Zuniga (2011)), bias correction (Ospina et al. (2006), Kosmidis & Firth (2010)), response involving ones and zeros (inflated BRM, Ospina & Ferrari (2010)), and size-corrected tests (Bayer & Cribari-Neto (2013)).

In this thesis, one of the proposals is to fit the simple BRM for data from 2^{k-p} experiments with response in (0,1). Results are presented in Chapter 4.

Thus far, with regard to the classical analysis using linear normal model, transformation, GLM or BRM, only *simple* models have been considered, i.e., only the mean of the response variable is explained by means of a structure regression. Thus, the dispersion (or precision) parameter ϕ is assumed constant across observations. It means that the variability of the response: i) it is not important to be considered, or ii) none of the covariates (factors) can explain the variability. This is not realistic in many cases.

With respect to the GLM, when the first edition of McCullagh & Nelder (1983) was reviewed by Pregibon (1984), he stated that to assume the dispersion parameter ϕ as a constant was overly restrictive and suggested further generalizations to consider variable dispersion. As a consequence, the authors followed the suggestion made by Pregibon, in the second edition, explored the consequences of constructing and fitting formal models for the dependence of both μ_i and ϕ on several covariates. Additionally, within the motivation for this chapter, McCullagh & Nelder (1989) stated that the impetus for studying these extended models derived from the surge of interest in industrial quality-improvement experiments, in which both mean and dispersion were of substantive interest. The aim very often was to select that combination of factor levels that keeps the mean at a predetermined *ideal* value, while at the same time keeping the variability in the product at a minimum. Then, it was necessary to study not just how the mean response is affected by

factors under study, but also how the variance is affected by these factors. That chapter was entitled *joint modelling of mean and dispersion*.

Now, an extension of the simple BRM is the Variable Dispersion BRM (VDBRM). In this model, the precision (dispersion) parameter is not constant for all observations but instead modelled in a similar fashion as the mean parameter. The variable dispersion BRM (VDBRM) has been justified by several authors; for example, Bayer & Cribari-Neto (2014) undertake an example using a Montecarlo simulation; they note that efficiency loss takes place when the dispersion parameter is incorrectly taken to be constant. Additionally, they comment on dispersion modelling may be of direct interest since it allows for the statistician to identify the sources of data variability.

In this thesis, the necessity of considering the VDBRM in factorial experiments was exemplified in the *Introduction and Motivation*.

The VDBRM has been used under frequentist and Bayesian perspectives, by Cepeda (2001), Smithson & Verkuilen (2006), Simas et al. (2010), Ferrari et al. (2011), and Bayer & Cribari-Neto (2014), among others.

In the studied literature, works applying beta regression models to data from experiments are rarely encountered. Smithson & Verkuilen (2006) analyzed a replicated 2^2 experiment applied in psychology ($n = 104$), where the original response was measured in (0,100) interval. They used a VDBRM and presented a solution from a classical perspective. Also, they compared the results against normal and transformation models.

Finally, the use of restrictions to handle the higher order interaction factors in 2^k and 2^{k-p} experiments was explained in the *Introduction and Motivation*. There, a restricted VDBRM is presented as an alternative strategy for analyzing data from 2^{k-p} experiments with response in (0,1). Although in the literature restricted models for different kinds of responses can be encountered, a the restricted VDBRM has not yet known in the studied literature.

Thus, it is necessary to propose and develop an explicit restricted variable dispersion beta regression model (*restricted VDBRM*). After solving the model, a specific goal will be to analyze some 2^{k-p} experiments with response in (0,1) using all parameters involved within a theoretical 2^k experiment.

The restricted VDBRM proposed here will be introduced in next Chapter (specifically, expressions (2.6)-(2.7)). But before, a brief review of restricted models in another scenario will be described. As a starting point, when restrictions on parameters are

needed, the common mathematical solution involves some additional parameters known as *Lagrange multipliers*, (see Stein (1984)). These *new* parameters constitute a key point in the solutions because when they can be separated from the parameters of interest. These multipliers do not have to be estimated; however, on the other hand, it is necessary to consider the multipliers in the solutions. In the statistical literature, these multipliers are a special case of *nuisance* parameters; so this concept will be dealt with in Chapter 2.

A brief overview of the restricted models for response different from beta is presented further on. The common restricted linear regression model, with equality constraints is given by

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} \qquad \text{subject to} \quad \boldsymbol{R}\boldsymbol{\beta} = \boldsymbol{\delta} \tag{1.2}$$

where $\boldsymbol{R}$ matrix and $\boldsymbol{\delta}$ are a matrix and a vector are given, respectively.

i) Under normal response and constant varibility, the common solution is to use the least square criterium to minimize $(\boldsymbol{y} - \mathbf{X}\boldsymbol{\beta})^T(\boldsymbol{y} - \mathbf{X}\boldsymbol{\beta}) + 2\boldsymbol{\lambda}^T(R\boldsymbol{\beta} - \boldsymbol{\delta})$ with respect to $\boldsymbol{\beta}$, $\boldsymbol{\lambda}$ being the vector associated with the Lagrange multipliers. Using appropriate algebra, the restricted solution $\widehat{\boldsymbol{\beta}}_r$ includes R and $\boldsymbol{\delta}$, but it does not depend on Lagrange multipliers.

ii) Silvey (1959), Gourieroux et al. (1982), Wolak (1987), and Neuenschwander & Flury (1997), among others, consider maximum likelihood estimates under different types of constraints in the parameter space; some of these restrictions are considered under the normal model, and their solutions include estimates of Lagrange multipliers.

iii) Edwards et al. (2001) present a normal mixed model with restrictions for fixed and random effects. If the model contains p parameters, they use the singular-value decomposition on the two constraints matrices achieving the expression of c parameters in terms of the remaining $p - c$ parameters only. With this, and, given the observations, they estimate these $p - c$ unrestricted (linearly independent) parameters by the usual least square method. Their methodology has as a key point the difference between two normally distributed vectors, which is normal. This result is not possible in vectors of beta distributed variable. iv) Tian (2010) used the restricted normal model. He applied linear algebra to avoid the Lagrange multipliers and compared some estimable functions between the unrestricted and the restricted models.

v) In the case of exponentially distributed response, Nyquist (1991) proposed a penalty function to solve a restricted GLM. Using matrix algebra and the orthogonality between mean and dispersion in the GLM, he found the restricted estimates for the parameters of

interest without estimating the Lagrange multipliers; his approach cannot be used here to solve the equations (2.6)-(2.7) because, in beta regression, mean and dispersion are non-orthogonal. In another work, Cysneiros & Paula (2005) used the Nyquist solution to their restricted model for symmetrical distributions where mean and dispersion are orthogonal.

Remark: An important subject, types of constraints (restrictions), appears discussed in Neuenschwander & Flury (1997). They distinguish between three types of restrictions, *model constraints*, *identifiability constraints*, and *basic constraints*. Their definitions are summarized as follows:

i) *Model constraints*, reduce an identifiable parameter space $\Theta \subset \mathbb{R}^s$ to a subspace of smaller dimension.

ii) *Identifiability constraints*, guarantee that for any two different parameters the two corresponding distributions are different.

iii) *Basic constraints* guarantee that for any $\theta \subset \Theta$, there exists a F_θ distribution.

In this thesis, using the previous definitions, the main problem is associated to the identifiability restrictions, because the set of parameters derived from a 2^{k-p} experiment is unidentifiable and only becomes identifiable when a restriction is imposed on parameters in order to get a reduced (identifiable) model.

In the following chapter, concepts and expressions required for the proposed model are detailed. But first, the Bayesian simple BRM will be commented on.

In general, fractional factorial experiments involve small sample sizes. Paradoxically, upon analyzing experimental by means of GLM or BRM, inferential results are based on the condition of large sample sizes. Therefore, it is necessary to consider some methods which do not depend on this condition. Bayesian statistics is a methodology which has this condition. A summary of this methodology is presented fortwith.

(*Bayesian simple BRM*). Within the Bayesian framework, parameters are modeled as random variables, and the concepts can be studied in books such as Bernardo & Smith (2000) and Kruschke (2011). As a summary, in the Bayesian simple BRM, parameters of the simple BRM are modelled as random variables by means of a *prior* distribution. Afterwards, using this distribution and the data likelihood, a *posterior* distribution is computed. Then, Bayesian inference is done employing the posterior distribution. The Bayesian simple BRM has been applied in works as Branscum et al. (2007) and Figueroa Zuniga (2011). With respect to the computational aspect, some packages in R

software perform the Bayesian simple BRM. For example, the *Bayesianbetareg* package (Marin et al. (2013)) can be used fixing the dispersion parameter in the algorithm. There are other softwares available for BRM, *gamlss* (Stasinopoulos & Rigby (2007)) and *betareg* package (Cribari-Neto & Zeileis (2010), Grün et al. (2012)) in R software (R Development Core Team (2014)). Also, according to Ferrari (2013), the Bayesian and frequentist BRM can be fitted in other softwares: SPLUS, SPSS, and SAS (the Procedure NLMIXED has the *Beta regression* macro, Swearingen et al. (2011) and Swearingen et al. (2012)). Now, frequentist or Bayesian analyses of factorial experiments by means of simple BRM are rarely found in the literature.

As a preliminary result of this thesis, for some 2^{k-p} experiments, the following models have been fitted: normal, two transformations, the simple BRM (with four link functions), and the Bayesian simple BRM (4 link functions). These models were compared by means of credibility and confidence intervals on mean response. Results and comparisons appear in Chapter 4. Upon using the restrictions yielded for the respective resolution of the experimental design, 11 *implicitly restricted* models have been fitted.

CHAPTER 2

Restricted Beta Regression Model.

2.1 Beta distribution

Beta distribution is a good alternative when response is in (0,1) because it assumes values in this interval. The common expression for its density contains the p and q parameters and is given by

$$f(y;p,q) = \frac{\Gamma(p+q)}{\Gamma(p)\Gamma(q)}\ y^{p-1}(1-y)^{q-1}, \quad y \in (0,1), \tag{2.1}$$

where $p > 0$, $q > 0$ and $\Gamma(\cdot)$ is the mathematical gamma function. The mean and variance of y are, respectively, $E(y) = \mu = \frac{p}{p+q}$ and $Var(y) = \frac{pq}{(p+q)^2}(p+q+1)$.
Beta distributions are very versatile and have different empirical uses and applications. Many practitioners have used the BRM when the response is in (0,1) and is related to covariates, (see Johnson et al. (1995), Bury (1999), and Ferrari (2013)). Some examples of beta densities are exemplified in 2.1.

In order to impose a regression structure for the beta response, several parameterizations for p and q have been proposed: Jorgesen (1997) (four parameters), Cepeda (2001) (Bayesian purposes), Paolino (2001), Kieschnick & McCullough (2003), and Ferrari & Cribari-Neto (2004), among others. With the parametrization $\mu = \frac{p}{p+q}$ and $\phi = p+q$, the probability density of a variable y in terms of its mean μ and its precision ϕ is given by

$$f(y;\mu,\phi) = \frac{\Gamma(\phi)}{\Gamma(\mu\phi)\Gamma((1-\mu)\phi)}\ y^{\mu\phi-1}(1-y)^{(1-\mu)\phi-1}, \quad y \in (0,1), \tag{2.2}$$

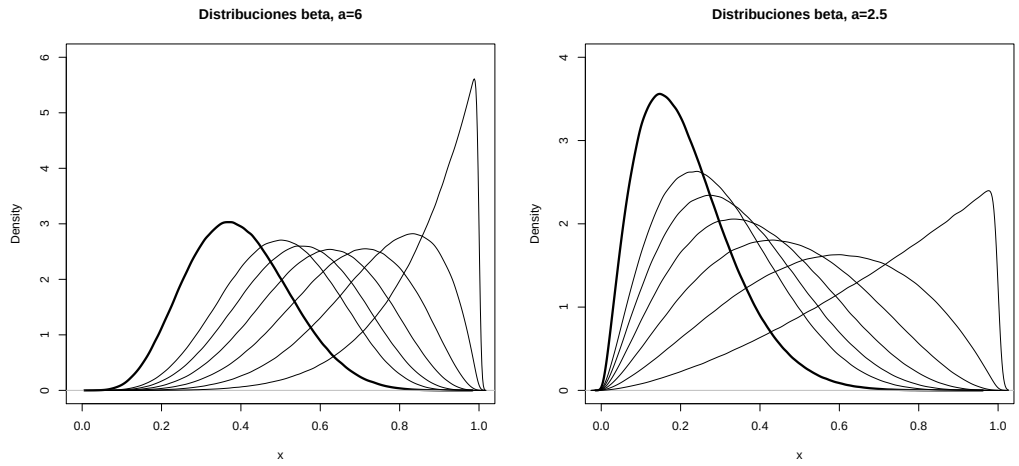

FIGURE 2.1. Some beta densities.

In that case, $E(y) = \mu$ $(0 < \mu < 1)$, $Var(y) = \frac{\mu(1-\mu)}{1+\phi}$, where $\phi > 0$ that can be interpreted as a precision parameter. If y has the density expressed in (2.2), and $y_1, \ldots, y_n$ denotes a random sample of size n, two situations are possible,

i) $y_i \sim beta(\mu_i\phi, (1-\mu_i)\phi)$, (variable mean and fixed dispersion, simple BRM).

ii) when each μ_i and ϕ_i can be different, it is written $y_i \sim beta(\mu_i\phi_i, (1-\mu_i)\phi_i)$, (Varying dispersion BRM).

2.2 Variable dispersion BRM.

In the variable dispersion beta regression model the mean μ_i as well as the precision ϕ_i are modelled through covariates. The VDBRM was employed by Smithson & Verkuilen (2006), formally introduced (along with further extensions) by Simas et al. (2010), and again justified by means of Monte Carlo simulation in Bayer & Cribari-Neto (2014), and in this thesis in example 2. The model is specified as follows.

Let $y_1, \ldots, y_n$ be a random sample such that $y_i \sim beta(\mu_i\phi_i, (1-\mu_i)\phi_i)$, $i = 1, ..., n$. The VDBRM is defined as

$$g_1(\mu_i) = \eta_{1i} = \sum_{l=0}^{p_1-1} x_{il}\beta_l = \boldsymbol{x}_i\boldsymbol{\beta} \qquad g_2(\phi_i) = \eta_{2i} = \sum_{l=0}^{p_2-1} z_{il}\alpha_l = \mathbf{z}_i\boldsymbol{\alpha} \tag{2.3}$$

where $\boldsymbol{\beta} = (\beta_0, \ldots, \beta_{p_1-1})^T$ and $\boldsymbol{\alpha} = (\alpha_0, \ldots, \alpha_{p_2-1})^T$ are the sets of regression coefficients in the two equations, with $(p_1 + p_2 < n)$. $\mathbf{x}_i = (x_{i0}, x_{i1}, \ldots, x_{ip_1-1})$, $\mathbf{z}_i = (z_{i0}, z_{i1}, \ldots, z_{ip_2-1})$ are regressor vectors associated with the means and the precisions, respectively. $\mathbf{z}_i$ usually is a set of some columns of $\mathbf{x}_i$ (then, $\mathbf{Z}$ is a $n \times p_2$ matrix of covariates). $g_1(\cdot)$ and $g_2(\cdot)$ are monotonic link functions, preferably with the property of mapping the range of μ_i and of ϕ_i to the real line, ($\mu_i \in (0, 1)$, and $\phi_i \in (0, \infty)$, i=1,...n.). A suitable candidate for $g_1(\cdot)$ is any inverse of a cumulative distribution function, and for $g_2(\cdot)$, the *log* function is the common election. In this thesis, the following four canonical link functions will be used for $g_1(\cdot)$:

logit: $g_1(\mu_i) = \log(\mu_i/(1-\mu_i))$.

probit: $g_1(\mu_i) = \Phi^{-1}(\mu_i)$, where $\Phi(\cdot)$ is the cumulative distribution function of a standard normal random variable.

cloglog: $g_1(\mu_i) = \log(-\log(1-\mu_i))$.

cauchit: $g_1(\mu_i) = \tan(\pi[\mu_i - 0.5])$, where $\pi = 3.14159$.

A graphical summary of these link functions is shown in figure 2.2, however, other canonical and non-canonical link functions are possible, see Cribari-Neto & Zeileis (2010).

For ϕ, the common link functions are $log(\phi_i)$, and $\sqrt{\phi_i}$. In this work, the $log(\phi_i)$ link is

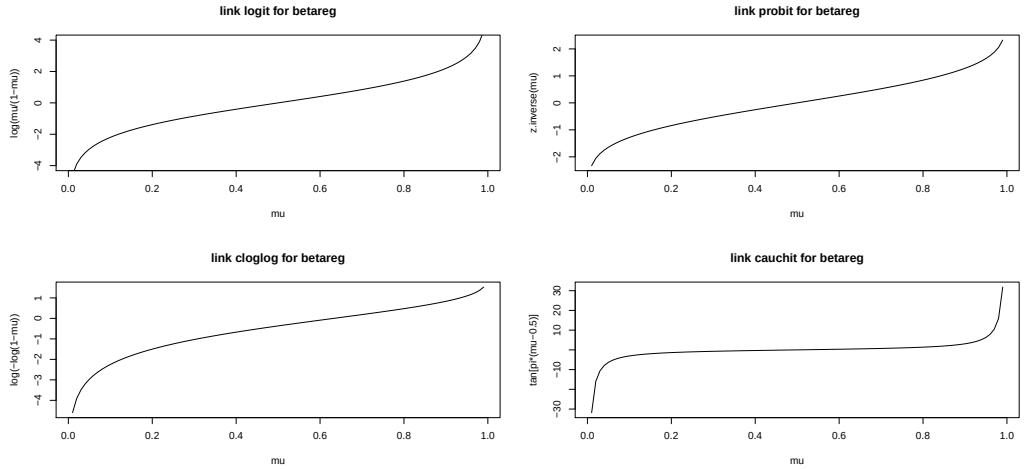

FIGURE 2.2. Four link functions for the mean.

employed.

The log-likelihood function for the VDBRM defined in figure (2.3) is given by

$$l(\boldsymbol{\beta}, \boldsymbol{\alpha}) = \sum_{i=1}^{n} l_i(\mu_i, \phi_i) \tag{2.4}$$

where

$$l_i(\mu_i, \phi_i) = log\Gamma(\phi_i) - log\Gamma[(\mu_i\phi_i)] - log\Gamma[(1-\mu_i)\phi_i] + (\mu_i\phi_i - 1)logy_i + [((1-\mu_i)\phi_i - 1)]log(1-y_i) \tag{2.5}$$

with $\mu_i = g_1^{-1}(\eta_{1i})$, and $\phi_i = g_2^{-1}(\eta_{2i})$ defined in (2.3). Within the frequentist perspective, theoretical developments for the estimations of parameters in the BRM can be found in Ferrari & Cribari-Neto (2004) (for fixed ϕ), Espinheira (2007), Cribari-Neto & Souza (2011), and Bayer & Cribari-Neto (2014). The frequentist estimation in the BRM is similar to the GLM, both based on the maximum likelihood estimation.

In the GLM, the parameters ϕ and β are orthogonal, which does not occur in the BRM. Inference for the BRM is based on asymptotic results under regularity conditions, the solutions do not have closed-form.

2.3 A Restricted Variable Dispersion Beta Regression Model: a proposal.

Suppose that, in addition to the VDBRM (2.3), there are q linearly independent restrictions on the parameter vector $\boldsymbol{\beta}$, yielding the Restricted VDBRM):

$$g_1(\mu_i) = \eta_{1i} = \mathbf{x}_i\boldsymbol{\beta} \qquad g_2(\phi_i) = \eta_{2i} = \mathbf{z}_i\boldsymbol{\alpha} \tag{2.6}$$

subject to

$$\mathbf{r}_j^T\boldsymbol{\beta} = \delta_j, \quad j = 1, \ldots, q \tag{2.7}$$

In matrix form,

$$g_1(\boldsymbol{\mu}) = \boldsymbol{X}\boldsymbol{\beta}, \quad g_2(\boldsymbol{\phi}) = \boldsymbol{Z}\boldsymbol{\alpha} \qquad \text{s.t.} \quad \boldsymbol{R}\boldsymbol{\beta} = \boldsymbol{\delta}$$

Dimension: $\boldsymbol{\beta}_{p_1\times 1}$, $\boldsymbol{\alpha}_{p_2\times 1}$, $\delta_{q\times 1}$ (known fixed numbers), $\boldsymbol{R}_{q\times p_1}$, with q rows linearly independent (known fixed numbers). Condition: $(p_1 + p_2 - q < n)$.

Remark: the columns of $\boldsymbol{Z}$ are usually some columns of $\boldsymbol{X}$.

In that follows: i) a penalized likelihood is proposed in order to estimate the (restricted) parameters, ii) inferential aspects of the model are presented: hypothesis tests, and goodness of fit.

The problem now with the restricted VDBRM is to maximize the log-likelihood (2.4) over $\boldsymbol{\beta}$ and $\boldsymbol{\alpha}$ under restrictions (2.7). Therefore, it is necessary to propose some solutions for the restricted VDBRM (2.6) - (2.7). Initially, the classical solution is introduced, which is based on Nyquist (1991). After, a Bayesian solution is explored as a future work.

2.4 Frequentist solution for the restricted VDBRM.

One approach to solve restricted optimization problems is the *penalty function* method, (Lange 2010), and it is used for the frequentist solution of (2.6)-(2.7). The first step is to consider the quadratic penalty function

$$P(\boldsymbol{\lambda}, \boldsymbol{\beta}, \boldsymbol{\alpha}) = \sum_{i=1}^{n} l_i(\mu_i, \phi_i) - \frac{1}{2} \sum_{k=1}^{q} \lambda_k (\delta_k - \mathbf{r}_k^T \boldsymbol{\beta})^2 \tag{2.8}$$

and the second step is to find a solution to the unrestricted problem $max_{\boldsymbol{\beta},\boldsymbol{\alpha}} P(\boldsymbol{\lambda}, \boldsymbol{\beta}, \boldsymbol{\alpha})$ for fixed and positive values of λ_k, $k = 1, ..., q$ (Lagrange multipliers). The notation is summarized as follows, and the concepts can be studied in Atkinson (1985):

1. $\mathbf{b}^T = (\boldsymbol{\lambda}^T, \boldsymbol{\beta}^T, \boldsymbol{\alpha}^T)^T$ is the parameter vector of $(p_1 + p_2 + q) \times 1$ order. In some discussions, $\mathbf{b}$ also will be written as $(\boldsymbol{\beta}, \boldsymbol{\alpha}, \boldsymbol{\lambda})$.

2. $\mathbf{Q}(\mathbf{b}) = \frac{\partial P(\boldsymbol{\lambda}, \boldsymbol{\beta}, \boldsymbol{\alpha})}{\partial \mathbf{b}}$ is the score vector of $(q + p_1 + p_2) \times 1$ order. The score vector can be partitioned as $\mathbf{Q}^T = (Q(\boldsymbol{\lambda})^T, Q(\boldsymbol{\beta})^T, Q(\boldsymbol{\alpha})^T)^T$.

3. Maximum likelihood estimate of $\mathbf{b}$ is defined by the $q + p_1 + p_2$ equations $\mathbf{Q}(\widehat{\mathbf{b}}) = \mathbf{0}$.

4. The *expected information* or Fisher information matrix is
   $K(\mathbf{b}) = K(\boldsymbol{\lambda}, \boldsymbol{\beta}, \boldsymbol{\alpha}) = -\mathbb{E}_{\mathbf{b}} \left[\frac{\partial^2 P(\boldsymbol{\lambda}, \boldsymbol{\beta}, \boldsymbol{\alpha})}{\partial \mathbf{b} \partial \mathbf{b}^T} \right]$.
   In some cases, it may be easier to use the *observed information*
   $I(\mathbf{b}) = - \left[\frac{\partial^2 P(\boldsymbol{\lambda}, \boldsymbol{\beta}, \boldsymbol{\alpha})}{\partial \mathbf{b} \partial \mathbf{b}^T} \right]$.
   Both information matrices are square of $q + p_1 + p_2$ dimension.

To solve the penalty function (2.8) from a classical perspective, the *Fisher scoring* method is employed. In this case, for the $(m+1)$ step

$$K(\mathbf{b}^{(m+1)},\boldsymbol{\lambda})\mathbf{b}(\boldsymbol{\lambda})^{(m+1)} = K(\mathbf{b}^{(m)},\boldsymbol{\lambda})\mathbf{b}(\boldsymbol{\lambda})^{(m)} + Q(\mathbf{b}^{(m)},\boldsymbol{\lambda}) \tag{2.9}$$

or equivalently

$$\mathbf{b}(\boldsymbol{\lambda})^{(m+1)} = \mathbf{b}(\boldsymbol{\lambda})^{(m)} + K^{-1}(\mathbf{b}^{(m)},\boldsymbol{\lambda})Q(\mathbf{b}^{(m)},\boldsymbol{\lambda}) \tag{2.10}$$

where $K^{-1}(\mathbf{b}^{(m)},\boldsymbol{\lambda})$ is the inverse of the K matrix. Calculus for the score vector and the K matrix is shown in the following sections.

2.4.1 Calculus for (Q) score vector in the restricted VDBRM.

Initially, in order to simplify the notation, $\mathbf{y}^*$ and $\boldsymbol{\mu}^*$ are defined as vectors of n dimension with components

$$y_i^* = \log\left(\frac{y_i}{1-y_i}\right) \qquad\qquad \mu_i^* = \psi(\mu_i\phi_i) - \psi[(1-\mu_i)\phi_i] \tag{2.11}$$

where $\psi(\cdot)$ indicates the digamma function, $\psi(z) = \frac{dlog\Gamma(z)}{dz}$ for $z > 0$. (See Abramowitz & Stegun (1965)).

Subsequently, it will be demonstrated that the score vector has the expression

$$Q(\boldsymbol{\lambda},\boldsymbol{\beta},\boldsymbol{\alpha}) = \begin{pmatrix} \mathbf{0}_{q\times 1} \\ \mathbf{X}^T T_1(\mathbf{y}^* - \boldsymbol{\mu}^*) + R^T\Lambda(\boldsymbol{\delta} - R\boldsymbol{\beta}) \\ Z^T T_2 \boldsymbol{\nu} \end{pmatrix} \tag{2.12}$$

where

$$T_1 = diag\left(\frac{\phi_i}{g_1'(\mu_i)}\right) \qquad T_2 = diag\left(\frac{1}{g_2'(\phi_i)}\right) \qquad \Lambda_{q\times q} = diag(\lambda_k) \tag{2.13}$$

$\boldsymbol{\nu}$ is a $n\times 1$ vector with components, $\nu_i = \psi(\phi_i) + \mu_i(y_i^* - \mu_i^*) - \psi((1-\mu_i)\phi_i) + \log(1-y_i)$.

First-order derivatives: score vector.

Rewriting the penalty function (2.8) as $P(\boldsymbol{\beta}, \boldsymbol{\lambda}, \boldsymbol{\alpha}) = P_1(\boldsymbol{\beta}, \boldsymbol{\alpha}) + P_2(\boldsymbol{\lambda}, \boldsymbol{\beta})$, where

$$\begin{aligned} P_1(\boldsymbol{\beta}, \boldsymbol{\alpha}) &= \sum_{i=1}^{n} l_i(\mu_i, \phi_i) \\ P_2(\boldsymbol{\lambda}, \boldsymbol{\beta}) &= -\frac{1}{2}\sum_{k=1}^{q} \lambda_k(\delta_k - \mathbf{r}_k^T\boldsymbol{\beta})^2 \\ &= -\frac{1}{2}\sum_{k=1}^{q} \lambda_k \left(\delta_k - \sum_{s=1}^{p_1} r_{sk}\beta_s\right)^2 \\ &= -\frac{1}{2}\sum_{k=1}^{q} \lambda_k \left[\delta_k^2 - 2\delta_k \sum_{s=1}^{p} r_{sk}\beta_s + \left(\sum_{s=1}^{p} r_{sk}\beta_s\right)^2\right] \end{aligned} \tag{2.14}$$

each component of the (partitioned) score vector (2.12) can be computed as

Q_λ**:**

$$\frac{\partial P(\lambda, \beta, \alpha)}{\partial \lambda_k} = \frac{\partial P_2(\lambda, \beta)}{\partial \lambda_k} = -\frac{1}{2}\left(\delta_k - \sum_{s=1}^{p_1} r_{sk}\beta_s\right)^2, \ k = 1, \ldots, q. \tag{2.15}$$

Q_β**:**

$$\frac{\partial P(\boldsymbol{\lambda}, \boldsymbol{\beta}, \boldsymbol{\alpha})}{\partial \boldsymbol{\beta}_j} = \frac{\partial P_1(\boldsymbol{\beta}, \boldsymbol{\alpha})}{\partial \boldsymbol{\beta}_j} + \frac{\partial P_2(\boldsymbol{\lambda}, \boldsymbol{\beta})}{\partial \boldsymbol{\beta}_j}, \ j = 1, \ldots, p_1. \tag{2.16}$$

- For $P_1(\boldsymbol{\beta}, \boldsymbol{\alpha})$: having the log-likelihood (2.4), some algebraic results and notations from Ferrari & Cribari-Neto (2004):

$$\frac{\partial P_1(\boldsymbol{\beta}, \boldsymbol{\alpha})}{\partial \boldsymbol{\beta}_j} = \sum_{i=1}^{n} \frac{\partial l_i(\mu_i, \phi_i)}{\partial \mu_i}\frac{\partial \mu_i}{\partial \eta_{1i}}\frac{\partial \eta_{1i}}{\partial \beta_j}$$

Denoting $g_1(\mu_i) = \eta_{1i}$:

$$\begin{aligned} \frac{\partial l_i(\mu_i, \phi_i)}{\partial \mu_i} &= \phi_i\left[\log\left(\frac{y_i}{1-y_i}\right) - \{\psi(\mu_i\phi_i) - \psi[(1-\mu_i)\phi_i]\}\right] \\ \frac{d\mu_i}{d\eta_{1i}} &= \frac{dg_1^{-1}(\eta_{1i})}{d\eta_{1i}} = \frac{1}{g_1'(\mu_i)} \\ \frac{\partial \eta_{1i}}{\partial \beta_j} &= x_{ij} \end{aligned}$$

Thus,

$$Q_j(\beta, \alpha) = \frac{\partial P_1(\boldsymbol{\beta}, \boldsymbol{\alpha})}{\partial \boldsymbol{\beta}_j} = \sum_{i=1}^{n} \phi_i[y_i^* - \mu_i^*][1/g_1'(\mu_i)][x_{ij}], \ j = 1, \ldots, p. \tag{2.17}$$

- Now, for $P_2(\boldsymbol{\lambda}, \boldsymbol{\beta})$ in (2.14):

$$\frac{\partial P_2(\boldsymbol{\lambda}, \boldsymbol{\beta})}{\beta_j} = \sum_{k=1}^{q} \lambda_k \delta_k r_{j'k} - \sum_{k=1}^{q} \lambda_k r_{j'k} \left(\sum_{s=1}^{p_1} r_{sk} \beta_s \right) \tag{2.18}$$

$$= \sum_{k=1}^{q} \lambda_k \delta_k r_{j'k} - \left(\sum_{k=1}^{q} \lambda_k r_{sk}^2 \right) \beta_s - \sum_{k=1}^{q} \sum_{s \neq j'=1}^{p_1} \lambda_k r_{sk} r_{j'k} \beta_{j'} \tag{2.19}$$

$$= \sum_{k=1}^{q} \lambda_k r_{j'k} \left(\delta_k - \sum_{s=1}^{p_1} r_{sk} \beta_s \right) \tag{2.20}$$

Thus, the complete expression for Q_β is:

$$Q_{\boldsymbol{\beta}_{j'}} = \sum_{i=1}^{n} \phi_i (y_i^* - \mu_i^*)[1/g_1'(\mu_i)][x_{ij'}] + \sum_{k=1}^{q} \lambda_k \delta_k r_{sk} - \left(\sum_{k=1}^{q} \lambda_k r_{sk}^2 \right) \beta_s - \sum_{k=1}^{q} \sum_{s \neq j'=1}^{p_1} \lambda_k r_{sk} r_{j'k} \beta_{j''} \tag{2.21}$$

The first part of $Q_{\boldsymbol{\beta}_{j'}}$ can be written in matrix form as:
$\sum_{i=1}^{n} \phi_i (y_i^* - \mu_i^*)[1/g_1'(\mu_i)][x_{ij'}] = \boldsymbol{x}_{j'}^T T_1 (\mathbf{y}^* - \boldsymbol{\mu}^*)$, and the second part of $Q_{\boldsymbol{\beta}_{j'}}$ as $\boldsymbol{r}_{j'}^T \Lambda (\boldsymbol{\delta} - R\boldsymbol{\beta})$. Then
$Q_{\boldsymbol{\beta}} = \mathbf{X}^T T_1 (\mathbf{y}^* - \boldsymbol{\mu}^*) + R^T \Lambda (\boldsymbol{\delta} - R\boldsymbol{\beta})$.

(Getting $Q_{\boldsymbol{\alpha}}$): Deriving (2.5) partly with respect to each α_s, $s = 1, ..., p_2$, we obtain

$$Q_s(\beta, \alpha) = \frac{\partial l(\beta, \alpha)}{\partial \alpha_s} = \sum_{i=1}^{n} \frac{\partial l_i(\mu_i, \phi_i)}{\partial \phi_i} \frac{\partial \phi_i}{\partial \eta_{2i}} \frac{\partial \eta_{2i}}{\partial \alpha_s} \tag{2.22}$$

where $\partial \eta_{2i} / \partial \alpha_s = z_{is}$, and from (2.4), $\partial \phi_i / \partial \eta_{2i} = 1/g_2'(\phi_i) = dg_2^{-1}(\eta_{2i})/d\eta_{2i}$. In addition, deriving partly (2.5) with respect to each ϕ_i, for $i = 1, \ldots, n$, and substituting y_i^* and μ_i^* given in (2.11), we have that

$$\begin{aligned} \frac{\partial l_i(\mu_i, \phi_i)}{\partial \phi_i} =& \frac{\partial \log \Gamma(\phi_i)}{\partial \phi_i} - \mu_i \frac{\partial \log \Gamma(\mu_i \phi_i)}{\partial \phi_i} - (1 - \mu_i) \frac{\partial \log \Gamma((1 - \mu_i)\phi_i)}{\partial \phi_i} \\ &+ \mu_i \log(y_i) + (1 - \mu_i) \log(1 - y_i) \\ =& \psi(\phi_i) - \mu_i (\psi(\mu_i \phi_i) - \psi((1 - \mu_i)\phi_i)) + \mu_i \psi((1 - \mu_i)\phi_i) \\ &+ \mu_i \log \left(\frac{y_i}{1 - y_i} \right) + \log(1 - y_i) \\ =& \psi(\phi_i) + \mu_i (y_i^* - \mu_i^*) - \psi((1 - \mu_i)\phi_i) + \log(1 - y_i) \end{aligned} \tag{2.23}$$

Therefore, the final expression for $Q_{\boldsymbol{\alpha}}$ is

$$Q_{\boldsymbol{\alpha}_s} = \sum_{i=1}^{n} \left[\psi(\phi_i) + \mu_i(y_i^* - \mu_i^*) - \psi((1-\mu_i)\phi_i) + \log(1-y_i)\right] \frac{1}{g_2'(\phi_i)} z_{is'} = \boldsymbol{z}_{s'}^T T_2 \boldsymbol{\nu} \tag{2.24}$$

Then, $Q_{\boldsymbol{\alpha}} = Z^T T_2 \boldsymbol{\nu}$, with $\boldsymbol{Z}$, T_2, and $\boldsymbol{\nu}$ previously defined in (2.12). Therefore, as already has been expressed, the maximum likelihood estimators of $\mathbf{b} = (\boldsymbol{\lambda}, \boldsymbol{\beta}, \boldsymbol{\alpha})$ are obtained from the equations $Q(\mathbf{b}) = \mathbf{0}$. In the current case, although solutions exist, they do not have closed-form. Hence, they need to be obtained by numerical methods, using some optimization procedure, such as Newton algorithm, (Lange (2010)).
The solutions are denoted as $\widehat{\mathbf{b}} = (\widehat{\boldsymbol{\lambda}}, \widehat{\boldsymbol{\beta}}, \widehat{\boldsymbol{\alpha}})$. In order to have some inferential results, it is necessary to compute the covariance matrix of $\mathbf{b}$, which is the same (symmetric) Fisher information matrix, and it can be expressed in partitioned form as,

$$K(\boldsymbol{\lambda}, \boldsymbol{\beta}, \boldsymbol{\alpha}) = \begin{pmatrix} K_{\lambda\lambda} & K_{\lambda\beta} & K_{\lambda\alpha} \\ K_{\beta\lambda} & K_{\beta\beta} & K_{\beta\alpha} \\ K_{\alpha\lambda} & K_{\alpha\beta} & K_{\alpha\alpha} \end{pmatrix} \tag{2.25}$$

In the following section, it is demonstrated that the matrices of K are,

$$K_{\lambda\lambda} = \mathbf{0}_{q\times q} \qquad K_{\lambda\beta} = K_{\beta\lambda}^T = \mathbb{A}R = \mathbf{0}_{q\times p_1} \tag{2.26}$$

$$K_{\lambda\alpha} = \mathbf{0}_{q\times p_2} = K_{\alpha\lambda}^T \qquad (K_{\beta\beta})_{p_1\times p_1} = \mathbf{X}^T W \mathbf{X} + R^T \Lambda R \tag{2.27}$$

$$(K_{\beta\alpha})_{p_1\times p_2} = \mathbf{X}^T V Z = K_{\alpha\beta}^T \qquad (K_{\alpha\alpha})_{p_2\times p_2} = Z^T D Z \tag{2.28}$$

where

$$\mathbb{A}_{q\times q} = diag(\delta_k - \mathbf{r}_k^T \boldsymbol{\beta}) \qquad W_{n\times n} = diag(w_i) \qquad V_{n\times n} = diag(c_i) \tag{2.29}$$

$$D_{n\times n} = diag\left(\frac{b_i}{(g_2'(\phi_i))^2}\right) \tag{2.30}$$

for $i = 1, ..., n$, and for $k = 1, ..., q$, and the components of the matrices are

$$a_i = \left(\psi'(\mu_i\phi_i) + \psi'((1-\mu_i)\phi_i)\right) \frac{1}{(g_1'(\mu_i))^2} \tag{2.31}$$

$$b_i = \mu_i^2\psi'(\mu_i\phi_i) + (1-\mu_i)^2\psi'[(1-\mu_i)\phi_i] - \psi'(\phi_i) \tag{2.32}$$

$$c_i = \phi_i \left(\mu_i a_i - \psi'[(1-\mu_i)\phi_i]\right) \left(\frac{1}{(g_1'(\mu_i))(g_2'(\phi_i))}\right) \tag{2.33}$$

$$w_i = \phi_i^2 a_i \tag{2.34}$$

where $\psi'(\cdot)$ indicates the trigamma function.

2.4.2 Calculus for the Fisher Information (K) in the restricted VDBRM.

This section presents the calculations done to obtain the components of K matrix. Submatrices of K are obtained starting from the $Q(\boldsymbol{\beta}, \boldsymbol{\alpha}, \boldsymbol{\lambda})$ vector.

1. $K_{\lambda\lambda} = \mathbf{0}_{q\times q}$, because $Q_{\boldsymbol{\lambda}} = \mathbf{0}$ (see (2.12)).

2. For $K_{\beta\beta}$, consider the sums for Q_β in (2.21):
   On the first sum of (2.21): using $E\left[\frac{\partial l_i(\mu_i,\phi_i)}{\partial \mu_i}\right] = 0$, the chain rule, and definition for η_{1i}:

$$\begin{aligned}\frac{\partial^2 P_1(\boldsymbol{\beta},\boldsymbol{\alpha})}{\partial\beta_{j\prime}\beta_j} &= \sum_{i=1}^{n} \frac{\partial}{\partial\mu_i}\left[\frac{\partial l_i(\mu_i,\phi_i)}{\partial\mu_i}\frac{d\mu_i}{d\eta_{1i}}\right]\frac{d\mu_i}{d\eta_{1i}}\frac{\partial\eta_{1i}}{\partial\beta_j}x_{ij} \\ &= \sum_{i=1}^{n}\left[\frac{\partial^2 l_i(\mu_i,\phi_i)}{\partial\mu_i^2}\frac{d\mu_i}{d\eta_{1i}} + \frac{\partial l_i(\mu_i,\phi_i)}{\partial\mu_i}\frac{\partial}{\partial\mu_i}\frac{d\mu_i}{d\eta_{1i}}\right]\frac{d\mu_i}{d\eta_{1i}}x_{ij\prime}x_{ij}\end{aligned}$$

   Upon deriving twice with respect to μ_i in (2.4),

$$\frac{\partial^2 l_i(\mu_i,\phi_i)}{\partial\mu_i^2} = -\phi_i^2\left(\psi'(\mu_i\phi_i) + \psi'[(1-\mu_i)\phi_i]\right) \tag{2.35}$$

   and taking expectations,

$$E\left(\frac{\partial^2 P_1(\lambda,\beta,\phi)}{\partial\beta_{j\prime}\beta_j}\right) = \sum_{i=1}^{n} E\left(\frac{\partial^2 l_i(\mu_i,\phi_i)}{\partial\mu_i^2}\right)\left(\frac{d\mu_i}{d\eta_{1i}}\right)^2 x_{ij\prime}x_{ij} = -\sum_{i=1}^{n} w_i x_{ij\prime}x_{ij} \tag{2.36}$$

Having defined w_i in (2.29).

For the remaining sums in (2.21)): note that $s \neq j'$ for the last sum, and it follows that:

$$\frac{\partial^2 P_2(\boldsymbol{\beta}, \boldsymbol{\lambda})}{\partial \beta_{j'} \beta_j} = -\sum_{k=1}^{q} \lambda_k r_{ik} r_{jk}$$

Finally, using expressions for $\boldsymbol{W}$ and Λ in (2.29), one can express in matrix form,

$$K_{\beta\beta} = -E\left[\frac{\partial^2 P(\lambda, \beta, \phi)}{\partial \beta \partial \beta^T}\right] = \mathbf{X}^T W \mathbf{X} + R^T \Lambda R \tag{2.37}$$

3. Now $K_{\alpha\beta} = -E\left[\frac{\partial^2 P(\lambda,\beta,\phi)}{\partial \alpha \partial \beta^T}\right]$. Starting from (2.21),

$$\begin{aligned}\frac{\partial^2 l(\beta, \alpha)}{\partial \alpha_s \partial \beta_j} &= \sum_{i=1}^{n} \frac{\partial}{\partial \phi_i}\left(\frac{\partial l_i(\mu_i, \phi_i)}{\partial \mu_i}\frac{d\mu_i}{d\eta_{1i}}\right)\frac{d\phi_i}{d\eta_{2i}}\frac{\partial \eta_{2i}}{\partial \alpha_s} x_{ij'} \\ &= \sum_{i=1}^{n}\left(\frac{\partial^2 l_i(\mu_i, \phi_i)}{\partial \phi_i \partial \mu_i}\frac{d\mu_i}{d\eta_{1i}} + \frac{\partial l_i(\mu_i, \phi_i)}{\partial \mu_i}\frac{\partial}{\partial \phi_i}\frac{d\mu_i}{d\eta_{1i}}\right)\frac{d\phi_i}{d\eta_{2i}} z_{is} x_{ij'}\end{aligned}$$

Taking expected values on both sides of the above expression, we have

$$E\left(\frac{\partial^2 l(\beta, \alpha)}{\partial \alpha_s \partial \beta_j}\right) = \sum_{i=1}^{n} E\left(\frac{\partial^2 l_i(\mu_i, \phi_i)}{\partial \phi_i \partial \mu_i}\right)\left(\frac{d\mu_i}{d\eta_{1i}}\right)\left(\frac{d\phi_i}{d\eta_{2i}}\right) z_{is} x_{ij'}$$

On the other hand, deriving (2.4) twice, partly with respect to ϕ_i, we obtain

$$\begin{aligned}\frac{\partial^2 l_i(\mu_i, \phi_i)}{\partial \phi_i \partial \mu_i} &= y_i^* - \mu_i^* - \phi_i\{\mu_i[\psi'(\mu_i\phi_i) + \psi'((1-\mu_i)\phi_i)] - \psi'((1-\mu_i)\phi_i)\} \\ &= y_i^* - \mu_i^* - \phi_i[\mu_i w_i - \psi'((1-\mu_i)\phi_i)]\end{aligned}$$

and then

$$E\left(\frac{\partial^2 l(\beta, \alpha)}{\partial \alpha_s \partial \beta_j}\right) = -\sum_{i=1}^{n} \phi_i[\mu_i a_i - \psi'[(1-\mu_i)\phi_i]]\left(\frac{d\mu_i}{d\eta_{1i}}\right)\left(\frac{d\phi_i}{d\eta_{2i}}\right) z_{is} x_{ij'} \tag{2.38}$$

where $a_i = \frac{w_i}{\phi_i^2}$, with w_i defined in (2.29).

4. For $K_{\alpha\alpha}$.

   Upon calculating the partial derivative of the expression (2.24) with respect to ϕ_i, we obtain

$$\begin{aligned}\frac{\partial^2 l(\beta,\alpha)}{\partial\alpha_s\partial\alpha_{s'}} =& \sum_{i=1}^{n}\frac{\partial}{\partial\phi_i}\left(\frac{\partial l_i(\mu_i,\phi_i)}{\partial\phi_i}\frac{d\phi_i}{d\eta_{2i}}\right)\frac{d\phi_i}{d\eta_{2i}}\frac{\partial\eta_{2i}}{\partial\alpha_s}z_{is'}\\ =& \sum_{i=1}^{n}\left(\frac{\partial^2 l_i(\mu_i,\phi_i)}{\partial\phi_i^2}\frac{d\phi_i}{d\eta_{2i}}+\frac{\partial l_i(\mu_i,\phi_i)}{\partial\phi_i}\frac{\partial}{\partial\phi_i}\frac{d\phi_i}{d\eta_{2i}}\right)\frac{d\phi_i}{d\eta_{2i}}z_{is}z_{is'}\end{aligned}$$

   Since $E(\partial l_i(\mu_i,\phi_i)/\partial\phi_i)=0$, and taking expected values on both sides of the above expression, we have

$$E\left(\frac{\partial^2 l(\beta,\alpha)}{\partial\alpha_s\partial\alpha_{s'}}\right)=\sum_{i=1}^{n}E\left(\frac{\partial^2 l_i(\mu_i,\phi_i)}{\partial\phi_i^2}\right)\left(\frac{d\phi_i}{d\eta_{2i}}\right)^2 z_{is}z_{is'} \tag{2.39}$$

   Deriving (2.4) twice, partly with respect to ϕ_i, we have

$$\begin{aligned}\frac{\partial^2 l_i(\mu_i,\phi_i)}{\partial\phi_i^2} =& \psi'(\phi_i)-\mu_i[\mu_i\psi'(\mu_i\phi_i)-(1-\mu_i)\psi'((1-\mu_i)\phi_i)]-(1-\mu_i)\psi'((1-\mu_i)\phi_i)\\ =& \psi'(\phi_i)-\mu_i^2\psi'(\mu_i\phi_i)-(1-\mu_i)^2\psi'((1-\mu_i)\phi_i)\end{aligned}$$

   and then

$$\begin{aligned}E\left(\frac{\partial^2 l(\beta,\alpha)}{\partial\alpha_s\partial\alpha_{s'}}\right) =& -\sum_{i=1}^{n}\left[(1-\mu_i)^2\psi'((1-\mu_i)\phi_i)+\mu_i^2\psi'(\mu_i\phi_i)-\psi'(\phi_i)\right]\left(\frac{d\phi_i}{d\eta_{2i}}\right)^2 z_{is}z_{is'}\\ =& -\sum_{i=1}^{n}b_i\left(\frac{d\phi_i}{d\eta_{2i}}\right)^2 z_{is}z_{is'}\end{aligned}$$

   where b_i defined in (2.31).

5. For $K_{\lambda\beta}$, starting from (2.21), considering terms depending on λ :

$$\frac{\partial}{\partial\lambda_k}\left(\frac{\partial P(\boldsymbol{\lambda},\boldsymbol{\beta},\boldsymbol{\phi})}{\partial\beta_j}\right)=\delta_k r_{jk}-r_{jk}^2\beta_j-\sum_{j\neq s=1}^{p}r_{sk}r_{jk}\beta_s=r_{jk}\left(\delta_k-\sum_{s=1}^{p}r_{sk}\beta_s\right)$$

which yields the following matrix:

$$(K_{\lambda\beta})_{q\times p} = \mathbb{A}R$$

where $R_{q\times p}$ is a matrix for restrictions and $\mathbb{A}$ is expressed in (2.29). It is noteworthy that $\mathbb{A}$ is a $q \times q$ zero matrix, taking into account the conditions in the restricted VDBRM (2.7).

Finally, the following results are easily obtained using definitions of second derivatives, and (2.13):

6. $K_{\alpha\lambda} = K_{\lambda\beta}^T = \mathbf{0}_{p_2\times q}$
7. $K_{\beta\alpha} = K_{\alpha\beta}^T$.

The expression for the K matrix in (2.26) has been demonstrated.

Note that the matrices of the first row and the first column of partitioned matrix (2.25) now are null. That is

$$K(\boldsymbol{\lambda}, \boldsymbol{\beta}, \boldsymbol{\alpha}) = \begin{pmatrix} \mathbf{0}_{q\times q} & \mathbf{0}_{q\times p_1} & \mathbf{0}_{q\times p_2} \\ \mathbf{0}_{p_1\times q} & K_{\beta\beta} & K_{\beta\alpha} \\ \mathbf{0}_{p_2\times q} & K_{\alpha\beta} & K_{\alpha\alpha} \end{pmatrix} \tag{2.40}$$

Observing the final K in (2.40) two aspects can be appreciated:

i) Using the Fisher-scoring method in (2.9) and extracting the first equations: $\mathbf{0}_{q\times q}\boldsymbol{\lambda}^{(m+1)} = \mathbf{0}_{q\times q}\boldsymbol{\lambda}^m + \mathbf{0}_{q\times 1}$ indicating that $\boldsymbol{\lambda}$ can be any vector in $\mathbb{R}^q$.

ii) K is singular, unfortunately.

After doing the appropriate algebra, the new Fisher-scoring system is generated,

$$\boldsymbol{\theta}^{(m+1)} = \boldsymbol{\theta}^{(m)} + \mathcal{K}^{-1}Q(\boldsymbol{\theta}^{(m)}) \tag{2.41}$$

where

- $\mathcal{K}$ is a $(p_1 + p_2) \times (p_1 + p_2)$ matrix. Explicitly,

$$\mathcal{K}(\boldsymbol{\lambda}, \boldsymbol{\beta}, \boldsymbol{\alpha}) = \begin{pmatrix} \mathbf{X}^T W\mathbf{X} + R^T\Lambda R & \mathbf{X}^T VZ \\ Z^T V^T\mathbf{X} & Z^T DZ \end{pmatrix} \tag{2.42}$$

with the same matrices established in (2.26) and (2.29) .

- $\boldsymbol{\theta}^T = (\boldsymbol{\beta}^T, \boldsymbol{\alpha}^T)^T$ is the parameter vector of $(p_1 + p_2) \times 1$ order.
- $\mathcal{Q}(\boldsymbol{\lambda}, \boldsymbol{\beta}, \boldsymbol{\alpha})$ is a $(p_1 + p_2) \times 1$ vector. Explicitly,

$$\mathcal{Q}(\boldsymbol{\lambda}, \boldsymbol{\beta}, \boldsymbol{\alpha}) = \begin{pmatrix} \mathbf{X}^T T_1(\mathbf{y}^* - \boldsymbol{\mu}^*) + R^T \Lambda(\boldsymbol{\delta} - R\boldsymbol{\beta}) \\ Z^T T_2 \boldsymbol{\nu} \end{pmatrix} \tag{2.43}$$

With the same matrices and vectors defined in (2.12).

Therefore, in the new system,
i) As the $\mathbf{X}$ matrix has full column rank, and $\boldsymbol{W}$ is a nonsingular diagonal matrix, then $\mathbf{X}^T W \mathbf{X} + R^T \Lambda R$ is also a nonsingular matrix. This result can be demonstrated doing an analysis of the Λ diagonal matrix. On one hand, if Λ is a null matrix, the restricted VDBRM is reduced to the VDBRM, which has a solution, see Espinheira (2007) or Bayer & Cribari-Neto (2014). On the another hand, if only one λ_k is nonzero, then $\mathbf{X}^T W \mathbf{X} + R^T \Lambda R$ is a scalar nonzero. Also, if more than one λ_k is nonzero, the matrix will be nonsingular.
ii)As a consequence of i), $\mathcal{K}^{-1}$ exists (applying the result 2.1.3 of Ravishanker & Dey (2003)). Then, calculations for $\mathcal{K}^{-1}$ are considered.

2.4.3 Obtaining $\mathcal{K}^{-1}$

Writing the $\mathcal{K}$ matrix (2.42) as

$$\mathcal{K} = \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix} \tag{2.44}$$

we can write the inverse of $\mathcal{K}$, $\mathcal{K}^{-1}$ as

$$\mathcal{K}^{-1} = \begin{pmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{pmatrix} \tag{2.45}$$

$$B_{11} = A_{11}^{-1} + A_{11}^{-1} A_{12} B_{22} A_{21} A_{11}^{-1} \tag{2.46}$$

$$B_{12} = -A_{11}^{-1} A_{12} B_{22} \tag{2.47}$$

$$B_{21} = B_{12}^T \tag{2.48}$$

$$B_{22} = (A_{22} - A_{21} A_{11}^{-1} A_{12})^{-1} \tag{2.49}$$

And specifically

$$\mathcal{K}^{-1} = \begin{pmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{pmatrix} \tag{2.50}$$

$$B_{11} = \boldsymbol{\Omega}^{-1} + \boldsymbol{\Omega}^{-1}\mathbf{X}^T V Z B_{22} Z^T V^T \mathbf{X}\boldsymbol{\Omega}^{-1} \tag{2.51}$$

$$B_{12} = -\boldsymbol{\Omega}^{-1}\mathbf{X}^T V Z B_{22} \tag{2.52}$$

$$B_{21} = B_{12}^T \tag{2.53}$$

$$B_{22} = (Z^T D Z - Z^T V^T \mathbf{X}\boldsymbol{\Omega}^{-1}\mathbf{X}^T V Z)^{-1} \tag{2.54}$$

where $\boldsymbol{\Omega} = \mathbf{X}^T W \mathbf{X} + R^T \Lambda R$. Several expressions can be found for $(\mathbf{X}^T W \mathbf{X} + R^T \Lambda R)^{-1}$; for instance, as $\mathbf{X}^T W \mathbf{X}$ is nonsingular, applying the equation (17) from Henderson & Searle (1981), the following inverse matrix is obtained:

$$(\mathbf{X}^T W \mathbf{X} + R^T \Lambda R)^{-1} = (\mathbf{X}^T W \mathbf{X})^{-1} - (\mathbf{X}^T W \mathbf{X})^{-1} R^T [\Lambda^{-1} + R(\mathbf{X}^T W \mathbf{X})^{-1} R^T]^{-1} R (\mathbf{X}^T W \mathbf{X})^{-1}$$

It is noteworthy that the estimated parameters by means of frequentist method involve the $\boldsymbol{\lambda}$ parameter vector. This parameter vector is known in the statistical inference as a nuisance parameter.

Remark:**Nuisance parameters**. When the penalty function (2.8) was proposed to solve the restricted VDBRM (2.6) - (2.7), it contained the $(\boldsymbol{\lambda}, \boldsymbol{\beta}, \boldsymbol{\alpha})$ parameters. According to Basu (1977), due to the inference problem at hand relates only to $\boldsymbol{\beta}$ and $\boldsymbol{\alpha}$ parameters, and because information gained on $\boldsymbol{\lambda}$ is of no direct relevance to the problem, then $\boldsymbol{\lambda}$ can be classified as the *nuisance* parameter, ($\boldsymbol{\beta}$ and $\boldsymbol{\alpha}$ are named the *parameters of interest*). The same author expresses that the methods for eliminating these parameters began with Fisher. With respect to nuisance parameters there are several works from frequentist and Bayesian perspectives; for instance, Cox (2006) deals the subject in the starting point of his book and, moreover he introduces some analyses for these parameters. Aitkin (2010) presents a wide analysis integrating the frequentist and Bayesian perspectives regarding this subject. (Bernardo & Smith (2000), p.479) indicate that, within a Bayesian framework, the presence of nuisance parameters does not have any theoretical problems. However, they consider that the problem posed by the presence of nuisance parameters can only be satisfactorily solved within a pure frequentist framework in those few cases

where the optimality criterion used leads to a procedure which depends on a statistics whose sampling distribution does not depend on the nuisance parameter.

Next chapter deals with the frequentist inferential aspects of the restricted VDBRM.

CHAPTER 3

Inferential aspects for the Restricted VDBRM.

This chapter deals with the inferential aspects related to the solutions for the restricted VDBRM proposed in the expressions (2.6) - (2.7). Solutions are considered from a frequentist perspective, based on asymptotic results, which is common in fitted models for nonnormal response. In the first part, a simulated restricted BRM is used to exemplify the inferential concepts. In the second part, a real 2^{k-p} experiment is introduced to show a special case of hypothesis for restricted models. Some Bayesian ideas to analyze the restricted VDBRM are presented in Chapter 6.

A comment on notation. In order to solve the restricted VDBRM (2.6) - (2.7), the penalty function (2.8) was proposed, which contains the $\mathbf{b}$ vector, which was defined in (2.4). This parameter vector can be broken into two parts, $(\boldsymbol{\beta}^T, \boldsymbol{\alpha}^T, \boldsymbol{\lambda}^T)^T = (\boldsymbol{\theta}^T, \boldsymbol{\lambda}^T)^T$, where $\boldsymbol{\theta} = (\boldsymbol{\beta}^T, \boldsymbol{\alpha}^T)^T$ is the vector with $p_1 + p_2$ *parameters of interest* and $\boldsymbol{\lambda}$ is a vector that contains q *nuisance parameters*. (This notation is based on Aitkin (2010)). The proposal of the inference for the restricted VDBRM is developed afterwards. But before, a summary of the regularity conditions is discussed.

Remark: The regularity conditions assumed in this work can be encountered in texts as Gourieroux et al. (1982) (Vol. 1, p. 181 and 184):

1. The variables Y_i, i=1,...,n are independent and identically distributed with density $f(y; \boldsymbol{\theta})$, $\boldsymbol{\theta} \in \boldsymbol{\Theta} \subset \Re^p$.

2. The parameter space $\boldsymbol{\Theta}$ is compact.

3. The true but unknown parameter value parameter $\boldsymbol{\theta}_0$ is identified.

4. The log-likelihood function

$$L_n(y;\boldsymbol{\theta}) = \sum_{i=1}^{n} \log f(y_i, \boldsymbol{\theta})$$

is continuous in $\boldsymbol{\theta}$.

5. $E_{\boldsymbol{\theta}_0} \log f(Y_i, \boldsymbol{\theta})$ exists.

6. The log-likelihood function is such that $(1/n)L_n(y;\boldsymbol{\theta})$ converges almost surely to $E_{\boldsymbol{\theta}_0} \log f(Y_i, \boldsymbol{\theta})$ uniformly in $\boldsymbol{\theta} \in \boldsymbol{\Theta}$.
Property 1: under assumptions 1-6, there exists a sequence maximum likelihood estimators converging almost surely to the true parameter value $\boldsymbol{\theta}_0$.

7. The log-likelihood function $L_n(\boldsymbol{\theta})$ is continuously differentiable in an open neighborhood of $\boldsymbol{\theta}_0$

8. The matrix

$$I(\boldsymbol{\theta}_0) = E_{\boldsymbol{\theta}_0}\left(-\frac{\partial^2 \log f(Y_i, \boldsymbol{\theta}_0}{\partial(\theta)\boldsymbol{\theta}^T}\right)$$

exists and is nonsingular.

Property 2: Under the assumptions of *Property 1*, suppose that assumptions 7-8 hold. Then a consistent sequence $\widehat{\boldsymbol{\theta}}_n$ of local maxima is such that $\sqrt{n}(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0)$ converges in distribution to a normal distribution with mean zero and covariance matrix $I(\boldsymbol{\theta}_0)^{-1}$.

3.1 Large sample inference to the restricted VDBRM.

Denoting the solution of the restricted VDBRM (2.6) - (2.7) as $\widehat{\boldsymbol{\theta}}_r = (\widehat{\boldsymbol{\beta}}_r^T, \widehat{\boldsymbol{\alpha}}_r^T)^T$, and assuming that the regularity conditions 3 are also satisfied under the restricted model, then $\widehat{\boldsymbol{\theta}}_r$ is asymptotically normal with mean $\boldsymbol{\theta}_r$ and covariance matrix given by (2.50) (here it depends upon $\boldsymbol{\lambda}$). When the large sample inference is applied to analyze data without nuisance parameters, three tests are widely employed: *likelihood ratio*(LR), *Wald*(W), and *score statistics*(S). The asymptotic equivalence of the three tests is shown in books such as Cox & Hinkley (1974), and Shao (2003). Some geometric relationships

between these three tests are commented on in Atkinson (1985), and additionally graphed in Cox (2006). Additional details can be found in Atkinson (1985) and Cox (2006), and these can be summarized as follows:

i) The asymptotic properties of the score test are not changed if the observed information is substituted by the expected information.

ii) Unlike the other two tests, the score test requires only quantities calculated under the null hypothesis.

iii) Some guidance on the choice among the three tests for finite samples is given by Cox & Hinkley (1974) in Section 9.3, they do not recommend the Wald test because it depends on the parameterizations of the problem.

Now, in the presence of nuisance parameters, the respective three tests are analogous to those given in absence of the nuisance parameters (Atkinson (1985), p. 95)

3.2 Hypothesis about restrictions.

3.2.1 A special case: Hypothesis $\mathbf{R}\boldsymbol{\beta} = \mathbf{0}$.

(Useful for testing hypothesis about effects in 2^k and 2^{k-p} experiments). Starting from the VDBRM (2.6) - (2.7), the interest focuses in the hypothesis

$$\mathbf{R}\boldsymbol{\beta} = \mathbf{0} \tag{3.1}$$

with $\mathbf{R}$ a $q \times p_1$ matrix, with $(q \leq p_1)$, and $\quad rank\,(\mathbf{R}) = q$

Remark. The rows of matrix $\mathbf{R}$ are linearly independent in order to avoid redundant hypothesis. Subsequently, it is possible to find a matrix $\mathbf{R_0}$ of $(p_1 - q) \times p_1$ order, and to construct the partitioned $\mathbf{F}_{p_1 \times p_1}$ matrix, with $F = \begin{bmatrix} \mathbf{R} \\ --- \\ \mathbf{R_0} \end{bmatrix}$ so that the first part of (2.6) can be reparameterized as follows

$$g_1(\boldsymbol{\mu}) = \mathbf{X}\boldsymbol{\beta} = \mathbf{X}F^{-1}F\boldsymbol{\beta} \tag{3.2}$$

Moreover, upon writing $F^{-1} = \begin{bmatrix} S^T & \vdots & S_0^T \end{bmatrix}$, (3.2) can be written as

$$g_1(\boldsymbol{\mu}) = \mathbf{X}\left[\begin{array}{ccc} S^T & \vdots & S_0^T \end{array}\right]\left[\begin{array}{c} \mathbf{R} \\ --- \\ \mathbf{R}_0 \end{array}\right]\boldsymbol{\beta}$$

and,

$$g_1(\boldsymbol{\mu}) = \mathbf{X}S^T\mathbf{R}\boldsymbol{\beta} + \mathbf{X}S_0^T\mathbf{R}_0\boldsymbol{\beta} \tag{3.3}$$

upon doing $\mathbf{X}_{(1)} = \mathbf{X}S^T$, $\mathbf{X}_{(2)} = \mathbf{X}S_0^T$, $\mathbf{R}\boldsymbol{\beta} = \boldsymbol{\beta}_{(1)}$, and $\mathbf{R}_0\boldsymbol{\beta} = \boldsymbol{\beta}_{(2)}$, it is possible to reparameterize (3.3) as

$$g_1(\boldsymbol{\mu}) = \mathbf{X}_{(1)}\boldsymbol{\beta}_{(1)} + \mathbf{X}_{(2)}\boldsymbol{\beta}_{(2)} \tag{3.4}$$

Two details with respect to equation (3.4) can be highlighted:
i) The test of $H_0 : \mathbf{R}\boldsymbol{\beta} = \mathbf{0}$ is then equivalent to the test $H_0 : \boldsymbol{\beta}_{(1)} = \mathbf{0}$, i.e. whether the constructed variable $\mathbf{X}_{(1)}$ should be included in the model or not.
ii) Under H_0, $g_1(\boldsymbol{\mu}) = \mathbf{X}_{(2)}\boldsymbol{\beta}_{(2)}$. Now, denote the matrix B as

$$B = X\left[\begin{array}{ccc} S^T & \vdots & S_0^T \end{array}\right] = XF^{-1}$$

$$B^TB = \left(F^{-1}\right)^T X^TXF^{-1} = \left[\begin{array}{cc} S\left(X^TX\right)S^T & S\left(X^TX\right)S_0^T \\ S_0\left(X^TX\right)S^T & S_0\left(X^TX\right)S_0^T \end{array}\right] \tag{3.5}$$

$$\left(B^TB\right)^{-1} = F\left(X^TX\right)^{-1}F^T = \left[\begin{array}{cc} R\left(X^TX\right)^{-1}R^T & R\left(X^TX\right)^{-1}R_0^T \\ R_0\left(X^TX\right)^{-1}R^T & R_0\left(X^TX\right)^{-1}R_0^T \end{array}\right] \tag{3.6}$$

and finally, $B\left(B^TB\right)^{-1}B^T = XF^{-1}F\left(X^TX\right)^{-1}F^T\left(F^T\right)^{-1}X^T$

$$= X\left(X^TX\right)^{-1}X^T \tag{3.7}$$

the common *Hat* matrix in linear normal model. (Ravishanker & Dey (2003)). Therefore, it is verified that $rank\,(B) = rank\,(X) = p_1$

If, additionally $\mathbf{R}_0^T$ is constructed so that $R\left(X^TX\right)^{-1}R_0^T = 0$, then in (3.6)

$$\left(B^TB\right)^{-1} = \begin{bmatrix} R\left(X^TX\right)^{-1}R^T & 0 \\ 0 & R_0\left(X^TX\right)^{-1}R_0^T \end{bmatrix}$$

and then, using expressions in (3.4)

$$B^TB = \begin{bmatrix} S\left(X^TX\right)S^T & 0 \\ 0 & S_0\left(X^TX\right)S_0^T \end{bmatrix} \tag{3.8}$$

$$= \begin{bmatrix} X_{(1)}^TX_{(1)} & 0 \\ 0 & X_{(2)}^TX_{(2)} \end{bmatrix} \tag{3.9}$$

Hence, the model (3.4) can be expressed in an orthogonal form.

Note: If the expression for F matrix is constructed as $F^\star = \begin{bmatrix} \mathbf{R_0} \\ --- \\ \mathbf{R} \end{bmatrix}$ an orthogonal expression similar to (3.8) is obtained.

A very important remark. The equality in (3.8) indicates the independence between the subsets of covariates $X_{(1)}$ and $X_{(2)}$ with respect to the estimation spaces. This result is very important to show that in 2^{k-p} experiments, when we need to know if effects are active or not (parameters significantly different to zero or not), we have a hypothesis of the kind $R\boldsymbol{\beta} = \mathbf{0}$, and statistical analysis for the fractional experiments is a special case of the restricted VDBRM. The next examples can exemplify this.

Example 3. *Hypothesis $R\boldsymbol{\beta} = \mathbf{0}$ for Semiconductors data.*

The *Semiconductors* is an experiment described in Melo et al. (2007). It is a 2_{III}^{4-1} design with a resolution-III. It contains the main effects A, B, and C. (D is confounded with ABC, AB is confounded with CD, ...). The design matrix and response are presented in Table 3.1. In the inferential part, in order to know if effects are active or not, consider the restrictions on parameters associated to the two-factor interactions, AB, AC, and AD. Then, the hypothesis can be presented as $H_0 : R\boldsymbol{\beta} = \mathbf{0}_{3\times1}$ versus $H_1 : R\boldsymbol{\beta} \neq \mathbf{0}_{3\times1}$ where $\boldsymbol{\beta} = (\beta_0, \beta_1, \beta_2, \beta_3, \beta_4, \beta_{12}, \beta_{13}, \beta_{14})^T$. Thus, the relationship among R, R_0, and F matrices is

TABLE 3.1. Semiconductors: design matrix and response.

Run	x_1	x_2	x_3	x_4	y
1	-	-	-	-	0.07
2	+	-	-	+	0.10
3	-	+	-	+	0.32
4	+	+	-	-	0.55
5	-	-	+	+	0.18
6	+	-	+	-	0.20
7	-	+	+	-	0.40
8	+	+	+	+	0.61

i) $R_{3\times 8} = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}$,

R is a *nilpotent* matrix. (Ravishanker & Dey (2003)). ii)$R_0 = \left[\begin{array}{c|c} I_5 & \mathbf{0}_{5\times 3} \end{array} \right]$; thus,

iii) $F_{8\times 8} = I_8$. Then $F^T F = I_8$, and $F^{-1} = F^T$, i.e.,

$$F^{-1} = \left[\begin{array}{c|c} R^T & R_0^T \end{array} \right] \tag{3.10}$$

Hence, iv) $S = R$, and $S_0 = R_0$, which simplifies the calculations for matrices $X_{(1)}$, and $X_{(2)}$.

v) The matrix $X_{(1)}$ has 5×8 order, and the matrix $X_{(2)}$ has 8×3 order. ☑.

Then, a general test to contrast $H_0 : R\boldsymbol{\beta} = \boldsymbol{\delta}$ versus $H_1 : R\boldsymbol{\beta} \neq \boldsymbol{\delta}$, for $\boldsymbol{\delta} \neq \mathbf{0}$ is considered afterwards.

3.2.2 Hypothesis: Significance of the restricted BRM.

Consider the hypothesis $H_0 : R\boldsymbol{\beta} = \boldsymbol{\delta}$ versus $H_1 : R\boldsymbol{\beta} \neq \boldsymbol{\delta}$. A common procedure for testing hypotheses about parameters in GLM is the Wald Test Statistics:

$$\xi^S = (Q(\boldsymbol{\theta}_0, \widehat{\boldsymbol{\lambda}}_0))^T \mathfrak{K}^{-1}(\boldsymbol{\theta}_0, \widehat{\boldsymbol{\lambda}}_0) Q(\boldsymbol{\theta}_0, \widehat{\boldsymbol{\lambda}}_0)$$

Under H_0 true, ξ^S in (3.2.2) follows an asymptotic χ^2 distribution with q degrees of freedom.

3.3 Hypothesis: Significance of the restricted BRM.

Consider the null hypothesis

$$H_0 : \mathbf{b} = \mathbf{b}_0^T = (\boldsymbol{\theta}_0)^T = (\beta_0, \alpha_0)^T \tag{3.11}$$

(where H_0 indicates that the mean and dispersion of the response are not explained for any covariate.)

Under the null hypothesis, the constrained maximum likelihood estimator of $\boldsymbol{\lambda}$ will be denoted as $\widehat{\boldsymbol{\lambda}}_0$. The vector of score statistics for $\mathbf{b}$ evaluated at the null hypothesis is $Q(\boldsymbol{\theta}_0, \widehat{\boldsymbol{\lambda}}_0)$. In general, the information matrix of parameter vector $(\boldsymbol{\theta}, \boldsymbol{\lambda})$, partitioned according to the partition of the parameter vector, is the following square matrix of $p_1 + p_2 + q$ order

$$\mathfrak{K}(\boldsymbol{\theta}) = \begin{pmatrix} \mathfrak{K}_{11}(\boldsymbol{\theta}) & \mathfrak{K}_{12}(\boldsymbol{\theta}) \\ \mathfrak{K}_{12}^T(\boldsymbol{\theta}) & \mathfrak{K}_{22}(\boldsymbol{\theta}) \end{pmatrix} \tag{3.12}$$

Then, specifically, the proposed statistics to evaluate the hypothesis (3.11), Likelihood Ratio, Wald, and Score tests are given upon calculating

$$\xi^{LR} = 2\{L(\widehat{\boldsymbol{\theta}}, \widehat{\boldsymbol{\lambda}}) - L(\boldsymbol{\theta}_0, \widehat{\boldsymbol{\lambda}}_0)\} \tag{3.13}$$

$$\xi^{W} = (\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta_0})^T \mathfrak{K}^{-1}(\boldsymbol{\theta}_0)(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta_0}) \tag{3.14}$$

$$\xi^{S} = (Q(\boldsymbol{\theta}_0, \widehat{\boldsymbol{\lambda}}_0))^T \mathfrak{K}^{-1}(\boldsymbol{\theta}_0, \widehat{\boldsymbol{\lambda}}_0) Q(\boldsymbol{\theta}_0, \widehat{\boldsymbol{\lambda}}_0) \tag{3.15}$$

where $\mathfrak{K}^{-1}$ denotes the inverse matrix given in (3.12), and it can be computed by means of the expression for the inverse of partitioned matrix given in (2.45). In all cases, the null hypothesis is rejected at the 100α level if the computed statistic is greater than $\chi^2_(p_1 + p_2, \alpha)$. According to (Shao (2003), p. 435), in a particular application, one may choose one of these tests that is easy to compute. Therefore, for the hypothesis (3.11), it is more appropriate to choose the score test, which does not require to estimate $\boldsymbol{\theta}$ nor $\boldsymbol{\lambda}$. In the current work, the score test was used for the examples.

For a restricted GLM, the asymptotic equivalence of the three tests in 3.13 is shown in Gourieroux et al. (1982), and Cysneiros (1997).

3.4 A goodness-of-fit measure

After fitting the model, in order to check the goodness-of-fit of the restricted estimated model, it is necessary to perform a diagnostic analyss. A global measure to explain the percentage of total variation due to the relationship between the mean of y and the covariates given by the linear predictor (2.3) can be obtained by computing $R^2_{rest,p}$ (a restricted pseudo R^2), which here is defined as

$$R^2_{rest,p} = r^2(\widehat{\boldsymbol{\eta}}_{1,rest}, g_1(\mathbf{y})), \quad 0 \leq R^2_{rest,p} \leq 1 \tag{3.16}$$

That is, the square of the sample correlation coefficient (of Spearman) between $\widehat{\boldsymbol{\eta}}_{1,rest} = \mathbf{X}\widehat{\boldsymbol{\beta}}_r$ and $g_1(\mathbf{y})$. When there is a perfect agreement between $\widehat{\boldsymbol{\eta}}_{1,rest}$ and $g_1(\mathbf{y})$, $R^2_{rest,p} \approx 1$ (then a perfect agreement between $\widehat{\boldsymbol{\mu}}$ and $\mathbf{y}$).

Example 4. *A simulated restricted simple BRM.*

Suppose that $y_1, \ldots, y_n$ are random variables where $y_i \sim beta(\mu_i\phi, (1-\mu_i)\phi)$, and consider the following restricted simple BRM:

$$g_1(\mu_i) = log\left(\frac{\mu_i}{1-\mu_i}\right) = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{3i} + \beta_4 x_{4i} \tag{3.17}$$

subject to

$$\beta_1 + 5\beta_2 = 0, \text{and} \quad \beta_3 + \beta_4 = -0.2 \tag{3.18}$$

$i = 1, \ldots, n$.

Hence, according to the notation used in (2.7),

- $p_1 = 5, q = 2$, $\beta = (\beta_0, \beta_1, \beta_2, \beta_3, \beta_4)^T$, $\lambda = (\lambda_1, \lambda_2)^T$,
- $R = \begin{pmatrix} 0 & 1 & 5 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 \end{pmatrix}$, and $\delta = (0, -0.2)^T$

- Specifically, the simulations for the covariates were done as follows:

$$\mathbf{x_1} \sim uniform(0,1)$$
$$\mathbf{x_2} \sim exponential(2)$$
$$\mathbf{x_3} \sim normal(7,1)$$
$$\mathbf{x_4} \sim \chi^2_6$$

- Initial values for parameters were: $(\beta_0, \beta_1, \beta_2, \beta_3, \beta_4) = (1, -1/2, 1/10, 0.3, -0.5)$.

The routines for simulations are shown in (7.2). An example of the $n = 8$ simulated observations are shown in Table (3.2). ☑

In order to evaluated the simulated observations, two examples are presented. In the first,

TABLE 3.2. Simulated restricted VDBRM: response and covariates.

Observation	y	x_1	x_2	x_3	x_4
1	0.46	0.47	0.19	6.99	8.33
2	0.24	0.21	2.39	8.58	9.18
3	0.76	0.80	0.29	6.64	2.66
4	0.29	0.65	1.24	7.22	10.64
5	0.74	0.32	0.87	7.27	2.25
6	0.20	0.72	0.30	5.47	4.52
7	0.38	0.29	1.58	7.07	7.49
8	0.67	0.93	0.22	6.37	3.19

relative bias of the estimators for different sample sizes and values of Lagrange multipliers are computed. In the second example, the evaluation for observations is done by means of a Bayesian BRM with a flat prior distribution to be compared against a frequentist BRM.

Example 5. *Invariance for estimations in the simulated simple BRM.*

For simulated data, in order to evaluate the property of invariance for estimates, with regards to $\boldsymbol{\lambda}'s$ values, firstly, a VDBRM model was fitted, and estimates were obtained; secondly, the relative biases of estimates were computed. The strategy was as follows:

1. The sample sizes considered were: 8, 20, 60, 100, and 250.

2. The penalty function proposed was:

$$\sum_{i=1}^{n} l_i(\mu_i, \alpha_0) - \frac{1}{2}\sum_{k=1}^{2} \lambda_k(\delta_k - \mathbf{r}_k^T\beta)^2 \tag{3.19}$$

3. The *logit* link for μ was employed. (The simulated data used this same link.)

4. The pairs of values for $\boldsymbol{\lambda} = (\lambda_1, \lambda_2)$ were $(1,1)$, $(2,1)$, $(2,3)$, $(10,3)$, and $(1,20)$.

5. For each estimate, its relative bias was computed as: $rel.bias(\widehat{\beta}_i) = \frac{\widehat{\beta}_i - \beta_{i,true}}{\beta_{i,true}}$, and $rel.bias(\widehat{\alpha}_0) = \frac{\widehat{\alpha}_0 - \alpha_{0,true}}{\alpha_{0,true}}$

The computed relative biases are shown in Table (3.3).

Remark. Although each pair of values for (λ_1, λ_2) yields 25 estimates, and hence 25 relative biases, they can be summarized with only 5 relative biases because the estimates were almost the same for the different pairs of (λ_1, λ_2).

TABLE 3.3. Simulated Restricted BRM: relative bias (r.b.). *logit* link.

n	$r.b.(\widehat{\beta}_0)$	$r.b.(\widehat{\beta}_1)$	$r.b.(\widehat{\beta}_2)$	$r.b.(\widehat{\beta}_3)$	$r.b.(\widehat{\beta}_4)$	$r.b.(\widehat{\alpha}_0)$
8	2.93	1.27	-0.33	-1.20	0.18	0.58
20	0.36	2.14	2.47	0.24	0.29	0.32
60	0.21	0.33	0.06	-0.25	-0.17	0.03
100	0.06	-0.08	0.08	-0.12	-0.06	0
250	-0.01	0.07	0.49	0.04	0.05	0.03

An inspection of the relative biases of estimates in Table (3.3) indicate that, for all the pairs of Lagrange multipliers, only when a large sample size is used, it is possible to achieve the true parameters (excepting for $\widehat{\beta}_2$ in 250 observations, although the variation coefficient for the five estimates $\widehat{\beta}_{2,8}$,..., $\widehat{\beta}_{2,250}$ was lower than 0.0001, the relative bias for the estimate presents an unsatisfactory value. This result was found for several seeds used in the computational routines.). However, it is possible to see the importance of having presented the inferential aspects for the VDMRB based on asymptotic results. ☑.

Example 6. *Estimates (a) and hypothesis (b) about the restricted VDBRM.*

Using 100 observations of the response variable from the simulated restricted VDBRM, in order to evaluate the behaviour of LR test upon comparing both models, the restricted and the unrestricted, the following Restricted VDBRM was fitted:

$$g_1(\mu_i) = logit(\mu_i) = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{3i} + \beta_4 x_{4i} \qquad g_2(\phi_i) = \alpha_0 + \alpha_1 z_{1i} + \alpha_2 z_{2i} \quad (3.20)$$

subject to

$$\beta_1 + 5\beta_2 = 0, \beta_3 + \beta_4 = -0.2 \tag{3.21}$$

where $z_1 = x_1$, and $z_2 = x_2$.

a) Using the penalty function (2.8), with different values for (λ_1, λ_2), the parameter estimates where approximately the same. In Table (3.4) the estimates for both, the restricted VDBRM and the VDBRM are shown for $\lambda_1 = 1, \lambda_2 = 20$), and the *logit* link. Moreover, the computed $p - values$ for the restricted model are presented. Upon interpreting the p-values of this Table, it is possible to conclude that the simulated data are good because the parameters which were considered nonzero in the simulation, are significantly different to zero according to the p-values.

TABLE 3.4. Parameter Estimates: restricted and unrestricted. VDBRM.

Parameter	unrestricted estimate	restricted estimate ($p - value$)
β_0	0.855	1.061 (0.014)
β_1	-0.423	-0.461 (< 0.001)
β_2	0.137	0.108 (< 0.001)
β_3	0.287	0.263 (< 0.001)
β_4	-0.477	-0.470 (< 0.001)
α_0	2.063	2.320 (< 0.001)
α_1	-0.116	-0.005 (0.25)
α_2	0.150	-0.004 (0.17)

b) Now, consider the hypothesis about restrictions: $H_0 : R\boldsymbol{\beta} = \boldsymbol{\delta}$ versus $H_1 : R\boldsymbol{\beta} \neq \boldsymbol{\delta}$. After applying the penalty function (2.8), with different values for (λ_1, λ_2), the LR test introduced in Section (3.2) was applied, and the respective p-values were computed. A summary of some values are shown in Table (3.5) is obtained. Upon analyzing the Table, p-values indicate that i)H_0 is rejected for all combinations of λ_1, and λ_2 considered. This results indicate that does exist a restricted model. ☑.

NA: Not Apply.

Example 7. *A diagnostic measure for the simulated Restricted VDBRM.*

The estimated $R^2_{rest,p}$ for the fitted model was 0.81 indicating a *moderate* agreement between $\widehat{\boldsymbol{\mu}}$ and $\mathbf{y}$. $R^2_{rest,p}$ was computed by means of the square of the correlation coefficient between $\mathbf{X}\widehat{\boldsymbol{\beta}}_r$ and $logit(\mathbf{y})$ vectors. ☑.

Remark: Exploring the Bayesian Analysis for the Restricted VDBRM.

TABLE 3.5. p-values for Restrictions Hypothesis, VDBRM.

λ_1	λ_2	D	$p-value$
0	0	NA	NA
5	10	6.87	0.031
10	5	6.90	0.032
2	3	6.86	0.032
20	1	6.97	0.031
2	20	6.86	0.032
30	20	7.039	0.030
1	**0**	2.62	**0.105**
0	**1**	2.61	**0.106**

Due to the fractional factorial experiments have a small number of runs, the sample-size of experiment is small. Hence, a criticism to the development of the (2.4) for data from 2^{k-p} experiments is to base the inferential results on large-samples theory, which is not realistic in experimental designs. Hence, the next task is to search for new alternatives analyses for the restricted VDBRM which are not have any problems with the small sample-sizes. One option is encountered in the Bayesian Statistics, where the parameters are modeled as random variables. It uses conjointly prior information and the data likelihood in order to get the estimates. Although some Statistic users prefer not to combine their results with the frequentist ones, many authors do works where it is possible to integrated both perspectives, Casella & Berger (1987), Samaniego & Reneau (1994), Bayarri & Berger (2004), Aitkin (2010), and Xie & Singh (2013). In this work, some Bayesian ideas for the Restricted VDBRM are proposed in Chapter (6). On another hand, the unrestricted simple BRM, a Bayesian analysis will be considered for in Section (4.6), where the basic concepts and notation of the Bayesian statistics will be presented.

CHAPTER 4

Unrestricted models.

In order to present several options for analyzing and comparing data from several real and simulated experiments, the unrestricted models considered here are: i) linear regression for original response; ii) linear regression for the *arcsin* and *logit* transformations; iii) Frequentist simple beta regression and iv) Bayesian simple beta regression. For frequentist methods, comparing methods the estimated mean responses ($\mu_i = E(y_i|X_i = x_i)$) have been used by many authors, Myers & Montgomery (1997), Lewis et al. (2001b), and Lewis et al. (2001a), among others.

4.1 Linear regression model.

Denoting y_i as the response for the ith experimental run (unit), $i = 1, \ldots, n$, the normal regression model is expressed as:

$$y_i = \sum_{l=0}^{p_1-1} x_{il}\beta_l + \varepsilon_i \tag{4.1}$$

or in matrix form

$$\mathbf{y} = X\boldsymbol{\beta} + \varepsilon \tag{4.2}$$

where ε denotes the vector of random errors, and $\boldsymbol{\beta} = (\beta_0, \beta_2, \ldots, \beta_{p_1-1})^T$ is the vector of unknown regression parameters. For estimating the $\boldsymbol{\beta}$, the strategy of ordi-

nary least squares is used. It requires the minimization of the quadratic form $\varepsilon^T\varepsilon = (\mathbf{y}-\mathbf{X}\boldsymbol{\beta})^T(\mathbf{y}-\mathbf{X}\boldsymbol{\beta})$. Upon differentiating the quadratic form with respect to $\boldsymbol{\beta}$ and equating the derivative to zero, the following *normal equations* are obtained: $X^TX\boldsymbol{\beta} = X^T\mathbf{y}$. If $rank(X^TX) = p_1$, a unique solution of normal equations is

$$\widehat{\boldsymbol{\beta}} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{y} \tag{4.3}$$

When the ε_i errors are assumed as independent and normally distributed with mean zero and constant variance σ^2, $\mathbf{y}$ and $\widehat{\boldsymbol{\beta}}$ are random vectors normally distributed (in multivariate form). The covariance matrix for $\widehat{\boldsymbol{\beta}}$ is given by $Cov(\widehat{\boldsymbol{\beta}}) = \sigma^2(\mathbf{X}^T\mathbf{X})^{-1}$, and an unbiased estimator of σ^2 is $\widehat{\sigma^2} = \frac{\mathbf{y}^T\mathbf{y}-\widehat{\boldsymbol{\beta}}^T\mathbf{X}^T\mathbf{y}}{n-p}$.

In a normal model, a confidence interval of $(1-\alpha)\times 100\%$ for μ_i, the mean response, can be obtained as follows: i) using the normality of $\widehat{\boldsymbol{\beta}}$; and ii) having $\widehat{y}(x_i) = X_i\widehat{\boldsymbol{\beta}}$ as an unbiased estimator of μ_i, the confidence interval for the mean response is obtained as:

$$\mathbf{x}_i\widehat{\boldsymbol{\beta}} \mp t_{(\alpha/2,n-p)}\sqrt{\widehat{\sigma^2}(\mathbf{x}_i(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{x}_i^T)}, \tag{4.4}$$

where $t_(\alpha/2, n-p)$ is the percentile of the t distribution, with $n-p$ degrees of freedom, corresponding to a cumulative probability of $(1-\alpha/2)$, α is the significance level. In linear regression models, for i=1,...,n:

$$h_{ii} = \mathbf{x}_i(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{x}_i^T \tag{4.5}$$

are the diagonal entries of *Hat* matrix. Due to the characteristics of X in 2^{k-p} experiments analyzed as classical linear models, the length of confidence intervals (4.4) is constant (it is due to observations are within the same combination of the levels of the factors).

The confidence interval in (4.4) was constructed based on hard assumptions for the errors in (4.1); however, when the response y is in (0,1), usually the variance is not constant, and it is difficult of achieving normality. A proposal to solve this problem is the transformation method, which is detailed as follows.

4.2 Transformations.

Transformations have been handled in different works, such as Box & Cox (1964) and Aranda-Ordaz (1981) where transformations families mainly to y are proposed, in order to get normality and homoscedasticity; generally it is desirable to find a function $f(y)$ (transformation of y) which will stabilize the variance. In this thesis, for y in (0,1), two appropriate transformations are used, *arcsin* and *logit*, which return values on the real line and then are modelled by the Gaussian model; the strategy is: i) to transform y appropriately, ii) to apply the normal model (4.1) for the transformed y, and iii) to present and to interpret in the original scale the results, doing the appropriate inverse transformation. Although, transformations have some drawbacks in some situations, Myers & Montgomery (1997), Lewis et al. (2001b), and Lewis et al. (2001a), it is also possible to find some positive results as in Patil & Kulkarni (2011), and Bonat et al. (2012). In this thesis, in order to provide more discussions regarding this subject for 2^{k-p} experiments, two transformations for y in (0,1) are analyzed: *logit*, $\widetilde{y} = \log(y/(1-y))$ and *arcsin*, $y^{\diamond} = \arcsin(\sqrt{y})$. With these transformed variables as response, the normal models were applied, and the confidence intervals for mean response, μ_i, were computed. The confidence interval of $(1-\alpha)\times 100\%$ for μ_i using the normal model and *logit* transformation is given by:

$$\mathbf{x}_i\widehat{\widetilde{\beta}} \mp t_{(\alpha/2,n-p)}\sqrt{\widehat{\widetilde{\sigma}^2}h_{ii}} \tag{4.6}$$

Similarly, the confidence interval of $(1-\alpha)\times 100\%$ for μ_i using the normal model and *arcsin* transformation is given by:

$$\mathbf{x}_i\widehat{\beta^{\diamond}} \mp t_{(\alpha/2,n-p)}\sqrt{\widehat{\sigma^2_{\diamond}}h_{ii}} \tag{4.7}$$

Finally, the solution is presented in the original scale, using the appropriate inverse transformation in each case: $y = \frac{1}{1+exp(-\widetilde{y})}$ for the *logit*, and $y = sin^2(y^{\diamond})$ for the *arcsin* transformation. Also, h_{ii} is presented in 4.5.

Remark: with respect to transformations, as suggested by Bonat et al. (2012), it is necessary to analyze each specific problem.

Nelder & Wedderburn (1972) and McCullagh & Nelder (1989) are benchmarks for advances

in generalized linear models, unifying several model specifications under the flexible class of distributions belonging to the exponential family, including the normal distribution. Some authors have analyzed the data with response on (a,b) by GLM (nonnormal link) and compare their results with the normal approaches. For instance, Lewis et al. (2001b) and Lewis et al. (2001a) analyzed real fractional and factorial experiments throughout the length of the confidence intervals on the mean response, (they compared normal vs GLM). Also, Patil & Kulkarni (2011) presented an analysis of 2^n experiments with response in the exponential family, comparing GLM against normal and transformations models; in Myers et al. (2011) several models are compared for some experiments with continuous response in (a,b). In this thesis, some of those results have been summarized and augmented in Table (1), Chapter ().

Now, considering different options for modeling response in (0,1), a logical approach is the beta distribution, which has its values in the (0,1) interval and is versatile for modeling.

4.3 Frequentist simple BRM.

In the simple BRM, only μ is modelled, and the parameter ϕ is taken as fixed ($\phi = exp(\alpha_0)$ in the VDBRM. Thus,

$$g_1(\mu_i) = \eta_{1i} = \sum_{l=0}^{p_1-1} x_{il}\beta_l = \mathbf{x}_i\boldsymbol{\beta} \tag{4.8}$$

The log-likelihood function for the simple BRM (4.8) is given by

$$l(\boldsymbol{\beta}, \phi) = \sum_{i=1}^{n} l_i(\mu_i, \phi) \tag{4.9}$$

where

$$l_i(\mu_i, \phi) = log\Gamma(\phi) - log\Gamma[(\mu_i\phi)] - log\Gamma[(1-\mu_i)\phi] + (\mu_i\phi - 1)logy_i + [((1-\mu_i)\phi - 1)]log(1-y_i) \tag{4.10}$$

with $\mu_i = g_1^{-1}(\eta_{1i})$ defined in (4.8). This model is completely detailed in Ferrari & Cribari-Neto (2004). As a special case in (2.3), inference in the simple BRM (4.8) is based on asymptotic results under regularity conditions (Because properties of estimators of BRM are similar to those of a GLM). The solutions do not have closed-form. Thus, when the

sample size is large, asymptotically,

$$(\widehat{\boldsymbol{\beta}}, \widehat{\phi})^T \sim N_{p+1}\left((\boldsymbol{\beta}, \phi)^T, K^{-1}\right) \tag{4.11}$$

where K^{-1} is the inverse of Fisher information matrix, so, if $\theta^T = ((\boldsymbol{\beta}, \phi)^T)$, and $\mu_i = g_1^{-1}(\mathbf{x}_i\beta)$, and using the likelihood (4.10), for $s, j = 1, \ldots, p+1$, then,

$$K(\boldsymbol{\beta}, \phi) = -\mathbb{E}_\theta\left[\frac{\partial^2\left[l(\boldsymbol{\beta}, \phi)\right]}{\partial\theta_{\mathbf{s}}\partial\theta_{\mathbf{j}}}\right] = \begin{pmatrix} K_{\beta\beta} & K_{\beta\phi} \\ K_{\phi\beta} & K_{\phi\phi} \end{pmatrix} \tag{4.12}$$

That is, the submatrix nonzero of matrix (2.40).

Now, for computing the confidence interval of μ_i in the simple BRM, the strategy is as follows: i) using the asymptotic results in (4.11), note that

$$\mathbf{x}_i\widehat{\boldsymbol{\beta}} \sim N(\mathbf{x}_i\boldsymbol{\beta}, Cov(\boldsymbol{x}_i s.e.(\widehat{\boldsymbol{\beta}})\boldsymbol{x}_i^T)) \tag{4.13}$$

where $s.e.(\widehat{\boldsymbol{\beta}})$ denotes the standard error of the estimated parameter.

ii) with appropriate algebra, the confidence interval of $(1-\alpha)\times 100\%$ for μ_i is obtained (it was proposed in Ferrari & Cribari-Neto (2004)). Explicitly, for each link function considered, the expressions for confidence intervals of μ_i are:

1. logit link:

$$\left(\frac{1}{1+\exp\{-\widehat{\eta}_i + Z_{\alpha/2}\,\text{s.e.}(\widehat{\eta}_i)\}},\ \frac{1}{1+\exp\{-\widehat{\eta} - Z_{\alpha/2}\,\text{s.e.}(\widehat{\eta})\}}\right)$$

2. probit link:

$$\left[\Phi(\widehat{\eta}_i - Z_{\alpha/2}\,\text{s.e.}(\widehat{\eta}_i)),\ \Phi(\widehat{\eta}_i + Z_{\alpha/2}\,\text{s.e.}(\widehat{\eta}_i))\right]$$

3. cloglog link:

$$\left[1-\exp\{-\exp(\widehat{\eta}_i - Z_{\alpha/2}\,\text{s.e.}(\widehat{\eta}_i))\},\ 1-\exp\{-\exp(\widehat{\eta}_i + Z_{\alpha/2}\,\text{s.e.}(\widehat{\eta}_i))\}\right]$$

4. cauchit link:

$$\left[\frac{1}{2} + \frac{\tan^{-1}(\widehat{\eta}_i - Z_{\alpha/2}\,\text{s.e.}(\widehat{\eta}_i))}{\pi}, \frac{1}{2} + \frac{\tan^{-1}(\widehat{\eta}_i + Z_{\alpha/2}\,\text{s.e.}(\widehat{\eta}_i))}{\pi}\right]$$

where: $Z_{\alpha/2} = \Phi^{-1}(1-\alpha/2)$ indicates the upper $\alpha/2$ percentage point of the standard normal distribution; $\widehat{\eta}_i = \mathbf{x}_i\widehat{\beta}$ is the fitted linear predictor; s.e.$(\widehat{\eta}_i) = \sqrt{\mathbf{x}_i \widehat{cov}(\widehat{\beta})\mathbf{x}_i^T}$; where, $\widehat{cov}(\widehat{\beta})$ is obtained from the inverse of the Fisher information matrix evaluated at the maximum likelihood estimates by excluding the row and column of this matrix corresponding to the precision parameter.

4.4 Frequentist variable dispersion BRM.

In the previous chapter the inference for the restricted VarD BRM (2.6) - (2.7) was developed. In the current section only the equations (2.6) are considered. Therefore, the inference for the unrestricted VarD BRM only refers to the $\boldsymbol{\theta} = (\boldsymbol{\beta}^T, \boldsymbol{\alpha}^T)^T$ vector, which contains the interest parameters, with $p_1 + p_2$ dimension.

One of first frequentist theoretical developments for the variable dispersion BRM is found in Espinheira (2007), although Smithson & Verkuilen (2006) already had applied the model without theoretical details. Later on works such as Cribari-Neto & Souza (2011), Cribari-Neto & Queiroz (2012), and Bayer & Cribari-Neto (2014) is dealt widely with this subject. The basic inferential aspects can be summarized as follows:

i) The log-likelihood function associated to the model is given by $\ell(\boldsymbol{\theta})$ as is expressed in (2.4).

ii) Similar to the calculus done in (2.13), the $p_1 + p_2$ partial derivatives of the P_1 penalty (the first part of (2.8)) form the vector of scores

$$Q(\boldsymbol{\beta}, \boldsymbol{\alpha}) = Q(\boldsymbol{\theta}) = \frac{\partial P_1(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \tag{4.14}$$

iii) Under regularity conditions the maximum likelihood estimates of $(\boldsymbol{\beta}, \boldsymbol{\alpha})$ are defined by the $p_1 + p_2$ equations

$$Q(\widehat{\boldsymbol{\theta}}) = \mathbf{0} \tag{4.15}$$

iv) The distribution of $\widehat{\boldsymbol{\theta}}$ is then asymptotically normal with mean $\boldsymbol{\theta}$ and the covariance matrix is the inverse of the $K(\boldsymbol{\theta})$ matrix. This $(K(\boldsymbol{\theta}))_{(p_1+p_2)\times(p_1+p_2)}$ matrix can be obtained from (2.40) doing $\boldsymbol{\lambda} = \mathbf{0}$. $K(\boldsymbol{\theta})$ is known as the *expected information* for the variable

dispersion BRM. In some cases, it may be easier to use the *observed information*

$$I(\boldsymbol{\theta}) = -\frac{\partial^2 P(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}_j \partial \boldsymbol{\theta}_s} \tag{4.16}$$

In practice, the value of $\boldsymbol{\theta}$ will not be known and so the estimate $\widehat{\boldsymbol{\theta}}$ will be substituted by the values $K(\widehat{\boldsymbol{\theta}})$ and $I(\widehat{\boldsymbol{\theta}})$.

v) To test the null hypothesis $\boldsymbol{\theta} = \boldsymbol{\theta}_0$, where $\boldsymbol{\theta}_0$ is a known vector, there are three tests which are asymptotically equivalent, *likelihood ratio* test, *Wald* test, and *Score* test. The expressions for the respective statistics are

$$2\{L(\widehat{\boldsymbol{\theta}}) - L(\boldsymbol{\theta_0})\} \tag{4.17}$$

$$(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta_0})^T K(\theta_0)(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta_0}) \tag{4.18}$$

$$(Q(\boldsymbol{\theta_0}))^T K^{-1}(\theta_0) Q(\boldsymbol{\theta_0}) \tag{4.19}$$

In all cases, the null hypothesis is rejected at the 100α % level if the computed statistic is greater than $\chi^2_{(p_1+p_2;\alpha)}$.

Now, considering the restricted VarD BRM, an additional notation will be employed. Let $\widehat{\boldsymbol{\beta}}_r$ and $\widehat{\boldsymbol{\alpha}}_r$ be the maximum likelihood estimates of the restricted VarD BRM (2.6) - (2.7), and then $\widehat{\boldsymbol{\theta}}_r = (\widehat{\boldsymbol{\beta}}_r, \widehat{\boldsymbol{\alpha}}_r)$. The maximum likelihood estimate of $\boldsymbol{\lambda}$ is denoted as $\widehat{\boldsymbol{\lambda}}$.

Due to 2^{k-p} experiments arising from situations with not many runs, using asymptotic results of the classical BRM can be questionable. Hence, in order to expand the options to analyze the mean response, here the Bayesian BRM is also considered, which takes the parameters as random variables and it does not require large sample sizes.

4.5 Bayesian simple BRM.

The main idea of this work is not to generate discussion regarding frequentist and Bayesian paradigms, but analyzing the data from two different perspectives, frequentist and Bayesian. Philosophical and methodological discussions between these two branches have been presented in many papers, see Casella & Berger (1987), Samaniego & Reneau (1994), Bayarri & Berger (2004), Aitkin (2010), and Xie & Singh (2013).

The basic concepts of Bayesian statistics can be studied in books such as Bernardo & Smith (2000), and Kruschke (2011). Here, is presented a summary of some concepts and

functions inherent to the Bayesian methodology. Other aspects can be investigated in the mentioned books.

- Parameter vector is denoted as $\boldsymbol{\theta}$.
- Before performing the experiment, users of the Bayesian Statistics employ some previous knowledge about $\boldsymbol{\theta}$, and it is specified by means of one *prior* distribution, denoted $p(\boldsymbol{\theta})$, and it is based on the history of a specific problem or application, and it is chosen before analyzing the real data or performing the experiment.
- $L(\mathbf{y}|\boldsymbol{\theta})$ is the likelihood function. It contains the information about the study or experiment.
- After applying (sometimes partly) the Bayes' theorem, $p(\boldsymbol{\theta}|\mathbf{y}) \propto L(\mathbf{y}|\boldsymbol{\theta})p(\boldsymbol{\beta})$, and then, the function $p(\boldsymbol{\theta}|\mathbf{y})$ is called the *posterior* distribution, and it indicates the distribution of $\boldsymbol{\theta}$, after using the data and the previous information in the model.
- The design matrix employed in the frequentist analysis, is used in the Bayesian analysis too.
- Generally, when a flat prior distribution is employed, Bayesian estimates are similar to the frequentist ones.

Initially, similar to the frequentist approach, the Bayesian BRM studies how, and how some covariates affect the response; in that sense, the linear predictor ($\mathbf{X}\boldsymbol{\beta}$) is common for both approaches. ($\boldsymbol{\beta}$ is considered as fixed parameter in the frequentist approach). *Remark*: In this thesis, the Bayesian examples are done for the unrestricted BRM and VDBRM. The complete development of the Bayesian analysis for the restricted BRM is left as *future work*. The illustrative Bayesian examples are done from the integrated Bayesian/likelihood framework.

4.6 Unrestricted Models: Applications.

Example 8. *Example 1 continued (BRM and Half-normal plots for a* 2^4 *experiment).*

In Example 1 was presented the *Drill* data experiment. Active effects yielded for several non-beta methods were summarized in Table 1. In this thesis, four link functions

for the BRM were employed to reanalyze the data and three aspects can be highlighted:
i) Among the four link functions, *cauchit* was the best in terms of the Akaike and Bayesian information criteria (AIC and BIC).
ii) Among the seven applied models for constructing the Half-normal plots, the *cauchit* simple BRM had more active effects than any other model (see Figure 1, bottom-right).
iii) Among the seven compared models, with respect to the inferential aspects, the *cauchit* simple BRM was the unique model that yielded the *p-values* lower than 0.05 for the *active* effects from the Half-normal plot. For the other six models, $p-values$ were much greater than 0.10 for all effects.
Then, the new summary for *Drill* data is shown in 4.1:

TABLE 4.1. New analysis of the *Drill* experiment

Authors (year)	Response	Method	Active	BIC (smaller ✓)
Daniel (1976)	y	Linear reg.	B,C,D, BC, CD	47.72
Montgomery (2001)	$log(y)$	Linear reg.	B, C, D	-36.54
Lewis et al. (2001a))	y	GLM.gamma	B, C, D	14.58
Grajales (2015)				
BRM	$y/100$	logit.link	B, C, D	-133.15
BRM	$y/100$	probit	B, C, D	-126.4
BRM	$y/100$	cloglog	B, C, D	-134.86
BRM	$y/100$	**cauchit**	7 effects	**-212.68**

The 7 active effects for BRM.cauchit: B, C, D, BC, CD, BD, BCD.
A conclusion taken of the 4.1 is that the BRM's presented the best fits for this specific experiment, which allows for proposing the simple BRM as a alternative for analyzing factorial experiments using the simple BRM from a inferential viewpoint also. Analyses were done using R software (R Development Core Team (2014)). The computational routine is shown in the Appendix 7.1. ☑

Some examples for the unrestricted models are presented. Two real fractional factorial experiments: *semiconductors*, and *paint trial* are analyzed. Seven specific models are described and applied for real and simulated data. The *Semiconductors* is a 2^{4-1}_{III} experiment, and it was introduced in chapter 3.
The *paint trial* experiment.
The *paint trial* is a 2^{8-4}_{IV} experiment and it can be encountered in Box et al. (2005). In order to develop a paint for certain vehicles a customer requires that the paint has high glossiness and acceptable abrasion resistance. Glossiness was measured on a scale of 1 to

100, then we consider $\frac{y-1}{99}$ which lies on (0,1). The experimenters were helped by paint technologist to decide that was necessary to include eight factors each with two levels. To carry out the complete factorial experiment 2^8, they required 256 runs or experimental units; however, they only had sixteen experimental units. The solution was to do a fractional factorial, a 2^{8-4}_{IV} experiment, which requires exactly sixteen experimental units, and all (8) factors can be analyzed. In Box et al. (2005), this data were analyzed by a linear regression model where with β_0, β_1, and β_2 as parameters significantly different to zero. The design matrix taken from Box et al. (2005), and glossiness (appropriately modified) for the *paint trial* experiment are shown in Table 4.2. Seven fitted models.

TABLE 4.2. Design matrix and response for *paint trial* data.

Run	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	$(y-1)/99$
1	-	-	-	-	-	-	-	-	0.525
2	+	-	-	-	+	+	+	-	0.596
3	-	+	-	-	+	+	-	+	0.677
4	+	+	-	-	-	-	+	+	0.778
5	-	-	+	-	+	-	+	+	0.475
6	+	-	+	-	-	+	-	+	0.667
7	-	+	+	-	-	+	+	-	0.545
8	+	+	+	-	+	-	-	-	0.778
9	-	-	-	+	-	+	+	+	0.485
10	+	-	-	+	+	-	-	+	0.677
11	-	+	-	+	+	-	+	-	0.606
12	+	+	-	+	-	+	-	-	0.808
13	-	-	+	+	+	+	-	-	0.515
14	+	-	+	+	-	-	+	-	0.697
15	-	+	+	+	-	-	-	+	0.646
16	+	+	+	+	+	+	+	+	0.818

Seven models were fitted, and compared: three normal (original response, *arcsin* and *logit* transformations), and four simple BRM's (one for each link). Specifically:

1. $y_i = \eta_{1i} + \varepsilon_i$ (normal).
2. $\text{logit}(y_i) = \eta_{1i} + \varepsilon_i$ (logit tranformation).
3. $\arcsin(\sqrt{y_i}) = \eta_{1i} + \varepsilon_i$ (arcsin transformation).
4. $\log(\frac{\mu_i}{1-\mu_i}) = \eta_{1i}$ (beta logit).
5. $\Phi^{-1}(\mu_i) = \eta_{1i}$ (beta probit).
6. $\log(-\log(1-\mu_i)) = \eta_{1i}$ (beta cloglog).

7. $\tan(\pi[\mu_i - \frac{1}{2}]) = \eta_{1i}$ (beta cauchit).

The statistical methodologies applied in this Section can be summarized as follows. The first method is the combining of the Half-normal plots and statistical inference, as it was exemplified in (1); this method is applied for two fractional factorial experiments (*Semiconductors* and *paint trial*). The second method (*estimated mean response*) is based on comparing confidence intervals on mean response for observations; this methodology is justified in works such as Myers & Montgomery (1997), Lewis et al. (2001b), and Lewis et al. (2001a). Within a frequentist framework, these authors show that the coverage of asymptotic intervals on mean response for small samples remains very close to theoretical coverage claims. In the current work, for the simple BRM, Bayesian credibility intervals on mean response are proposed in order to compare against the frequentist confidence intervals. As the Bayesian credibility intervals were built using a flat prior in the simple BRM, its results should be very similar to the frequentist results, (Aitkin (2010)). The third method dealts with intervals coverage for parameters from a simulated model under different replications, it will be applied in the *paint trial* experiment. The fourth methodology explores to recover the mean response for a simulated experiment whose response is in (0,1).

4.6.1 Applied unrestricted models: *Paint trial*, a 2^{8-4}_{IV} experiment

Example 9. *The* paint trial *experiment. Half-normal plot and Inference.*

In *paint trial* data seven models were fitted: three normal (original response, *arcsin* and *logit* transformations), and four simple BRM's (one for each link). Initially, the Half-normal plots were done and they are shown in Figure 4.1, for normal and transformations models, and Figure 4.2 presents the plots for the BRM's.

Afterwards, active effects extracted from each plot to fit the three normal models were used. For the simple BRM was dropped the two-factor interaction x1x4 because its estimate was near to zero in the Half-Normal plots for the normal models. A summary of results appears in Table (4.3). Two aspects to highlight from this Table:
1) For the simple BRM, the $x7$ effect was not used.
2) Using the AIC and BIC criteria, the best model was the BRM with *cloglog* link.

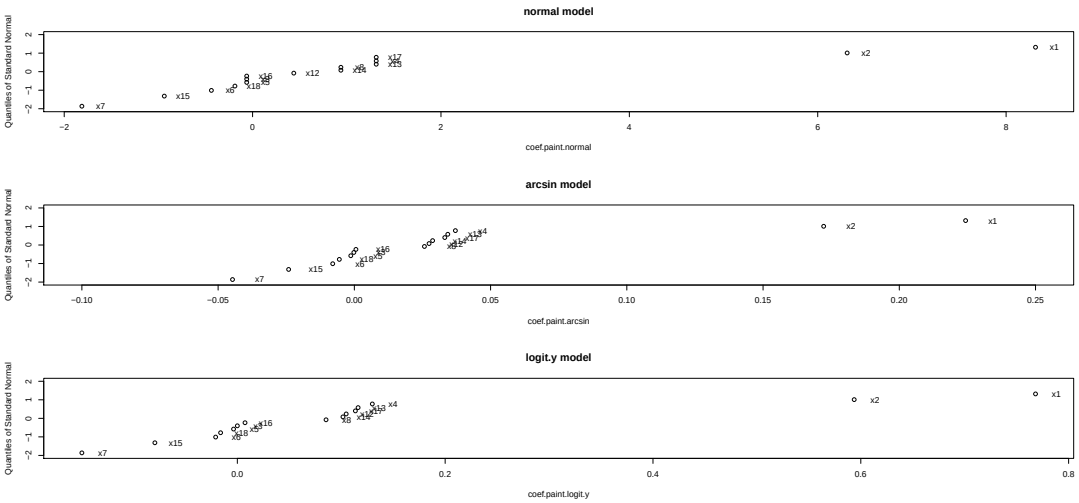

FIGURE 4.1. *paint trial* data. Half-normal plots. Normal and transformations

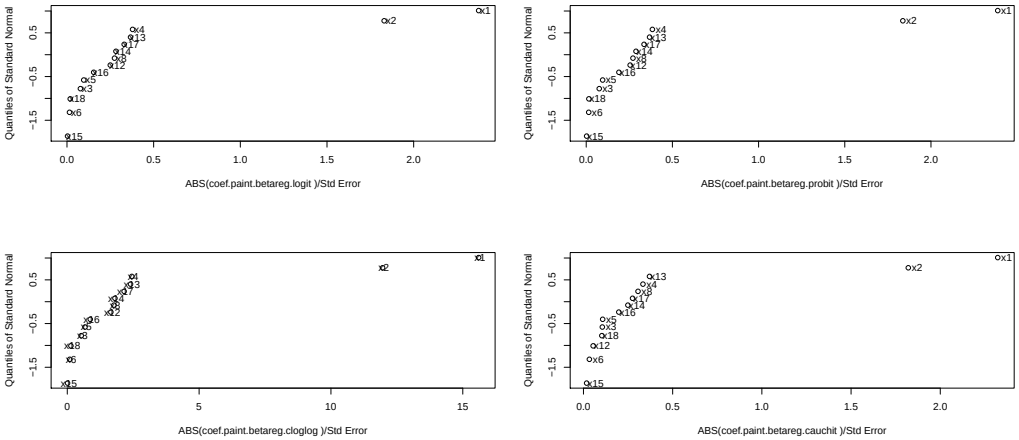

FIGURE 4.2. *paint trial* data. Half-normal plots: four BRM's

TABLE 4.3. *paint trial*. Half-normal plots and inference

Model	*p-value* < 0.05	AIC	BIC
Normal	All	92.51	95.60
arcsin	All	-44.93	-41.84
logit	All	-5.4	-2.3
beta.logit	All	-54.14	-51.05
beta.probit	All	-54.54	-51.45
beta.cloglog	**All**	**-55.53**	**-52.44**
beta.cauchit	All	-50.80	-47.71

☑.

An additional analysis, comparing Bayesian results against Half-normal plots for *paint trial* data is done. The objective of next example is to explore the simple Bayesian BRM as an alternative to identify active effects in 2^{k-p} experiments. Theoretically, this example

is thought as follows: i) a simple Bayesian BRM is fitted for data from a 2^{k-p} experiment, using a flat prior distribution; ii) the credible intervals for the parameters are constructed; iii) We will call *Bayesian active effects* to those effects associated to parameters whose credibility intervals do contain to zero. Subsequently, a comparison with *active effects* from Half-normal plot can be done.
Here, the method is exemplified using data from the *paint trial* experiment.

Example 10. *The* paint trial *experiment. Bayesian and frequentist analysis, and Half-normal plot.*

First, $\boldsymbol{\beta}$ the parameter vector with 14 dimension (the parameter for x_7 was not included for the simple BRM) was considered. Second, a simple Bayesian BRM for data from *paint trial* was applied with the *logit* link for the mean parameter, and using a flat prior distribution for $\boldsymbol{\beta}$. Third, the 14 credibility intervals are computed. Summary of results can be seen in Table (4.4). An interpreting of the results indicates that:

- Both methods, Bayesian and Half-normal yield the same active effects, A and B.
- Only two effects present discordant results: the two-factor interactions, AE and AH. In the first case, the credibility interval does not contains to zero but its parameter was not significantly different to zero (*p-value*= 0,80) within the frequentist inference. In the case of the AH factor, although it is considered as *active effect* for two methods (the credibility interval for its associated parameter does not include to zero, and the respective *p-value* is 0,07), this effect is not considered as *active* by means of the Half-normal plot.

Finally, from an integrated Bayesian/likelihood approach, the Bayesian BRM is a potential method for choosing *best models* based on the active effects yielded by its credibility intervals. ☑.

Example 11. *The* paint trial *experiment. Mean response by confidence intervals.*

For the *paint trial* experiment Box et al. (2005) fitted a normal model with original y as response and the systematic part $\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2}$. Based on this model, in this thesis

TABLE 4.4. Table *paint trial*. Bayesian and frequentist BRM active effects. *logit* link.

FACTOR	$\widehat{\beta}_{Bayes}$	CredIntLow	CredIntUpp	$\widehat{\beta}_{freq}$	*p-value*	graphically *active*?
A	0.36	**0.27**	**0.44**	**0.34**	**0.01**	**YES**
B	0.26	**0.19**	**0.33**	**0.27**	**0.07**	**YES**
C	0.003	-0.10	0.08	-0.002	0.98	NO
D	0.06	-0.03	0.152	0.06	0.70	NO
E	0.007	-0.09	0.13	-0.002	0.98	NO
F	-0.005	-0.11	0.07	-0.01	0.92	NO
H	0.06	-0.04	0.17	0.04	0.78	NO
AB	0.02	-0.09	0.13	0.02	0.88	NO
AC	0.05	-0.04	0.16	0.05	0.72	NO
AD	0.05	-0.04	0.16	0.04	0.78	NO
AE	**-0.04**	**-0.10**	**0.04**	**-0.03**	**0.80**	**NO**
AF	0.002	-0.08	0.07	0	0.99	NO
AG	0.06	-0.008	0.17	0.05	0.17	NO
AH	0.006	**-0.08**	**0.07**	-0.01	**0.07**	**NO**

were computed the confidence intervals on mean response for each observation. Also, in order to compare the mean response between models, the seven model presented in (4.6) were fitted using the same systematic part. The lengths of confidence intervals (CI) on mean response were computed for these seven models. Lengths of CI are shown in Table 4.5. ⊤.

TABLE 4.5. Lengths of the 95% confidence Intervals on the Mean Response in *paint trial* Data. Normal, transformed and beta models.

RUN	Normal	Transformed logit	Transformed arcsin	Beta link=logit	Beta probit	Beta cloglog	Beta cauchit
1	17.73	20.27	18.81	6.43	6.29	5.82	7.49
2	17.73	17.98	17.72	5.94	5.86	5.73	6.67
3	17.73	18.92	18.19	6.15	6.04	5.84	7.04
4	17.73	13.77	15.35	4.95	4.96	4.99	4.94
5	17.73	20.27	18.81	6.43	6.29	5.82	7.49
6	17.73	17.98	17.72	5.94	5.86	5.73	6.67
7	17.73	18.92	18.19	6.15	6.04	5.84	7.04
8	17.73	13.77	15.35	4.95	4.96	4.99	4.94
9	17.73	20.27	18.81	6.43	6.29	5.82	7.49
10	17.73	17.98	17.72	5.94	5.86	5.73	6.67
11	17.73	18.92	18.19	6.15	6.04	5.84	7.04
12	17.73	13.77	15.35	4.95	4.96	4.99	4.94
13	17.73	20.27	18.81	6.43	6.29	5.82	7.49
14	17.73	17.98	17.72	5.94	5.86	5.73	6.67
15	17.73	18.92	18.19	6.15	6.04	5.84	7.04
16	17.73	13.77	15.35	4.95	4.96	4.99	4.94

Example 12. *The* paint trial *experiment. The mean response by Bayesian and frequentist confidence intervals.*

In this example both approaches, Bayesian and frequentist are integrated upon comparing the confidence intervals with credibility intervals on mean response for the 16 observations in the *paint trial* experiment.

The Bayesian BRM fitted here can be summarized as follows:

i) The linear predictor $\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2}$, i=1,..., 16, was considered.

ii) the normal was chosen as the prior distribution for parameter vector, thus

$$(\boldsymbol{\beta})^T \sim N_p\left(\mathbf{0}_{px1}, 10^6 I_p\right) \tag{4.20}$$

where I_p is the Identity matrix of p order. This (chosen) covariance matrix indicates that a flat prior distribution for the parameter vector

iii) with independent data, starting from (4.9), the likelihood function is

$$L(\boldsymbol{\beta}, \phi) = \prod_{i=1}^{n} \frac{\Gamma(\phi)}{\Gamma(F(x_i^T\boldsymbol{\beta})\phi)\Gamma(\phi(1-F(x_i^T\boldsymbol{\beta}))} y_i^{(\phi F(x_i^T\boldsymbol{\beta}))^{-1}} (1-y_i)^{\phi(1-F(x_i^T\boldsymbol{\beta}))^{-1}} \tag{4.21}$$

where $F(\cdot)$ represent the inverse link function $g_1^{-1}(\cdot)$.

iv) The joint *posterior* distribution is given by

$$p(\boldsymbol{\beta}, \phi|\mathbf{y}) = \frac{L(\boldsymbol{\beta}, \phi)p(\boldsymbol{\beta}, \phi)}{\int L(\boldsymbol{\beta}, \phi)p(\boldsymbol{\beta}, \phi)d\beta d\phi} \tag{4.22}$$

and eliminating terms not depending on $\mathbf{y}$:

$$p(\boldsymbol{\beta}, \phi|\mathbf{y}) \propto p(\boldsymbol{\beta}, \phi)[\Gamma(\phi)]^n \prod_{i=1}^{n} [\Gamma(F(x_i^T\boldsymbol{\beta})\phi)\Gamma(\phi(1-F(x_i^T\boldsymbol{\beta})))]^{-1} y_i^{(\phi F(x_i^T\boldsymbol{\beta}))^{-1}} (1-y_i)^{\phi(1-F(x_i^T\boldsymbol{\beta}))^{-1}} \tag{4.23}$$

which is analytically untractable, i.e., it does not present closed form. Thus, it must be approximated numerically with methods based on, for instance, Markov chains Monte Carlo (MCMC), such as Metropolis-Hasting or the Gibbs sampler, see Robert & Casella (2004). According to Branscum et al. (2007), and Figueroa Zuniga (2011), Gibbs sampling can be used to generate a Monte Carlo sample from (4.23). In this context, the Gibbs sampler involves iteratively sampling the *full conditional distributions*:

$$p(\boldsymbol{\beta}|\phi, \mathbf{y}) \propto L(\boldsymbol{\beta}, \phi)p(\boldsymbol{\beta}) \tag{4.24}$$

$$p(\phi|\boldsymbol{\beta}, \mathbf{y}) \propto L(\boldsymbol{\beta}, \phi)p(\phi) \tag{4.25}$$

In this Example, the Random Walk for Metropolis Hasting (RWMH)'s method (Robert & Casella (2004)) was used, and it was implemented in software R Development Core Team (2012). The general WRMH's method appears detailed in (7.3.1).

The Bayesian BRM was used to compute the (Bayesian)credibility intervals of $(1-\alpha)\times 100\%$ for μ_i in *paint trial* experiment. The process to compute the length of credibility intervals on mean response for the four link functions considered in BRM can be detailed, thus: i) 50000 Monte Carlo iterations were considered, and the first 5000 were taken account of burning-time, therefore the analyses were done with the last 45000 simulated *posterior* distributions, and having 45000 (Bayesian) estimated $\boldsymbol{\beta}$'s, say $\beta_{\mathbf{B}}$, that is, values of the estimated *posterior* distribution, were computed 45000 mean responses which are percentiles of an unknown distribution. Finally, those 45000 mean response are sorted, and the 2.5 and 97.5 percentiles are taken; these percentiles are the limits of the credibility interval.

The necessary diagnostics tests were performed (convergence, autocorrelation, history): desirable behaviors were observed; Figure 4.3 shows the diagnostics for the β_1 parameter with *cauchit* link: the upper graph presents the convergence of the chain; on another hand, because we observed high autocorrelation (20th order) in chains, we used each 20 observations and got low autocorrelation, see Figure 4.3 (middle and lower graphs). A similar behaviour was encountered in the other parameters.

The four links described in (2.2) were used for both Bayesian and frequentis models. Results and comparisons are shown in Table 4.6. The analysis on estimated mean response by Bayesian BRM is different to the frequentist one: at taking each β_i as random variable, an estimation of β_i is not a real value but a percentile of a (*posterior*) distribution. In that case, on Bayesian statistics is not possible to speak about *the true value of the parameter*, a term used in the interpretation of a confidence interval. Hence, an appropriate way to estimate the mean response by means of the Bayesian model is through a *credibility interval* of $(1-\alpha)\times 100\%$ for μ_i, (denoted as $\mu_{i,B}$) with limits $L1$ and $L2$, whose interpretation is: *the probability that the mean response lies between* $L1$ *and* $L2$ *is* $1-\alpha$. Note that is not necessary to say *after repeating many times the experiment*, which must be expressed in the interpretation regarding the (frequentist) confidence interval. Inspecting Table 4.6 we can observe similar values for Bayesian and frequentist intervals for the *logit* link. The Bayesian BRM with *probit* link had the worst lengths because they are greater than the normal lengths. This bad result is surprisingly because the Akaike's Information Criterium

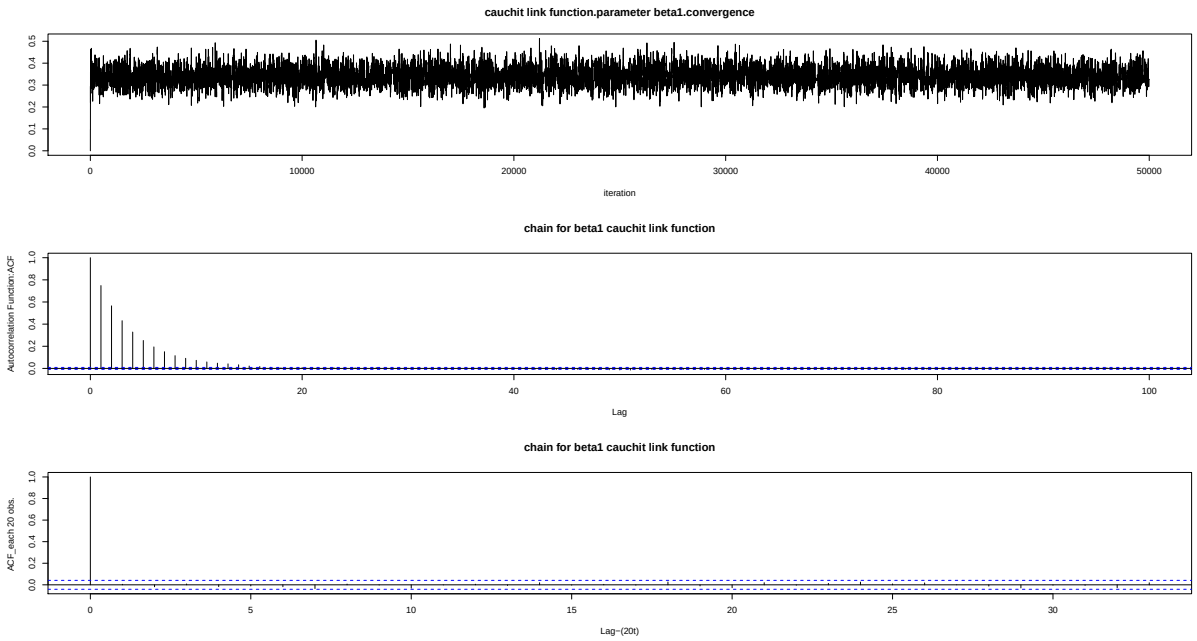

FIGURE 4.3. Some diagnostic tests for Bayesian analysis.

(AIC) for the frequentist fit, in the BRM, presented similar values for both link functions: *logit* (AIC = -54.14) and *probit* (AIC = -54.55). These contradictory results leads to suggest a more detailed study using the *probit* link for the BRM.

The Bayesian models with *cloglog* and *cauchit* links had intermediate results: wider lengths and also narrower intervals than frequentist's lengths for some experimental units. It is contradictory because the *cloglog* link achieved the best AIC, (-55.53) and the *cauchit* link got the worst AIC, (-50.80) which could lead to judging these link functions as unstable in Bayesian analysis on mean response. ☑. Due to in this work only unreplicated 2^k and 2^{k-p} experiments are considered, the next example deals with simulated data in order to analyze the effect yielded in the estimated parameters when two, three, and four replications of the experiment are considered. The simulated data are based on the paint trial experiment. The seven models presented in (4.6) are fitted in the simulation scenario.

Example 13. *The* paint trial *experiment. Coverage for parameters.*

For the *paint trial* experiment a 16×1 response vector distributed beta was generated, based on an *target* BRM with covariate matrix $X_{pt} = (\mathbf{1}, x_1, x_2)$, x_1 and x_2 being the first and second columns of the design matrix in *paint trial* data, i.e., the systematic part considered was: $\eta_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2}$, and for this case was used the *logit* link function, $\mu_i = 1/(1 + \exp(-\eta_i))$. Then, the model contains a known mean μ_i for each experimental

TABLE 4.6. Lengths of 95% frequentist and Bayesian confidence intervals on Mean. *paint trial* Data.

		logit	logit	probit	probit	cloglog	cloglog	cauchit	cauchit
RUN	Normal	freq	Bayes	freq	Bayes	freq	Bayes	freq	Bayes
1	17.73	**6.43**	**6.83**	6.29	22.47	5.82	10.02	**7.49**	**7.54**
2	17.73	5.94	6.42	5.86	20.14	5.73	7.67	6.67	6.76
3	17.73	6.15	6.69	6.04	21.35	5.84	8.87	7.04	6.87
4	17.73	4.95	5.33	4.96	16.92	**4.99**	**3.02**	**4.94**	**4.47**
5	17.73	6.43	6.83	6.29	22.47	5.82	10.02	7.49	7.54
6	17.73	5.94	6.42	5.86	20.14	5.73	7.67	6.67	6.76
7	17.73	6.15	6.69	6.04	21.35	5.84	8.87	7.04	6.87
8	17.73	4.95	5.33	4.96	16.92	4.99	3.02	4.94	4.47
9	17.73	6.43	6.83	6.29	22.47	5.82	10.02	7.49	7.54
10	17.73	5.94	6.42	5.86	20.14	5.73	7.67	6.67	6.76
11	17.73	6.15	6.69	6.04	21.35	5.84	8.87	7.04	6.87
12	17.73	4.95	5.33	4.96	16.92	4.99	3.02	4.94	4.47
13	17.73	6.43	6.83	6.29	22.47	5.82	10.02	7.49	7.54
14	17.73	5.94	6.42	5.86	20.14	5.73	7.67	6.67	6.76
15	17.73	6.15	6.69	6.04	21.35	5.84	8.87	7.04	6.87
16	17.73	4.95	5.33	4.96	16.92	4.99	3.02	4.94	4.47

run (sixteen runs in a replication of the experiment). By Monte Carlo simulation, the actual simulated response y_i is obtained by adding an error drawn at random from a specified distribution to the linear predictor, $y_i = 1/(1 + \exp(-\eta_i)) + \xi_i$ is generated with some initial values for $(\beta_0, \beta_1, \beta_2)$. When the beta regression model is applied throughout the sixteen rows of X_{pt}, the 16×1 response vector, y is finally generated. This 16×1 simulated vector is called *a replication* of the experiment. Each of the replications of the experiment is simulated 5000 times, and after, the seven models presented in (4.6) were fitted. The precision parameter used was $\phi = 50$ for all experimental scenario. Summarizing, the *ideal* beta regression model has *logit* link and the *inappropriate* use of other models is measured by coverage of confidence intervals on the parameters. For each of the mentioned seven models, and for each of the three β parameters, 95% confidence intervals are examined; therefore, for one replication of the experiment (16 runs) a total of 80000 points observational are used in the coverage calculations for each parameter, and for each model. Moreover, in order to study how much the results are affected for the sample size; we considered simulations for 2, 3, and 4 replications of the experiment, concatenating the design matrix one, two and three times horizontally leading to have 32, 48, and 64 observations, respectively. Also, each experimental situation was simulated 5000 times and the final result is, again, coverage of confidence intervals. For instance, in the experimental design situation for 64 observations, 5000×7 models are fitted. The

coverage of confidence intervals on β parameters are shown in Table (4.7) (16, 32, 48, and 64 runs). (For the following four models: normal, *arcsin* transformation, and BRM

TABLE 4.7. Coverage (%) for the three parameters in fitted models for simulated data. 1, 2, 3, and 4 replications of the experiment.

		Transformed	Beta	Beta
Runs	β_i	**logit**	*logit*	*cauchit*
16	β_0	92.22	89.62	78.60
16	β_1	92.98	90.20	87.02
16	β_2	92.84	90.28	88.10
32	β_0	93.54	92.78	70.82
32	β_1	93.36	92.90	86.24
32	β_2	93.50	92.74	87.62
48	β_0	92.42	93.02	60.78
48	β_1	93.12	93.04	84.26
48	β_2	93.92	93.38	86.56
64	β_0	92.52	93.70	52.44
64	β_1	93.18	93.56	81.44
64	β_2	94.20	93.98	84.92

with probit and cloglog links had bad results, its coverage were lower than 37%, and their values are not shown). The Table (4.7) indicates that, in spite of *logit* transformation model had yielded a bad performance when the lengths of confidence intervals for mean response were analyzed (see the Table (4.5)), this model yielded the best coverage, near to the target percentage, in the three parameters for each sample size. The beta model with *logit* link, although was the *target* model, had only a satisfactory performance because it needed more sample sizes to achieve better coverage. At this point, as some asymptotic results are assumed for the distribution of $\widehat{\beta}$ in the BRM, this could justify the latter result. On another hand, the logit transformation could have achieved the normality and homogeneous variance for response which would lead to $\widehat{\beta}$ normally distributed with variance constant, and in that case, a good coverage should be achieved with a little sample size, too. The BRM with *cauchit* link had an acceptable coverage for the β_1 and β_2 parameters for all sample sizes; however, for the intercept parameter, his coverage was illogically decreasing as sample size was increasing. Three models had bad results: normal, beta with *probit* and *cloglog* links. Their intervals never achieved coverage greater than 37% and some of them had coverage equal to zero%. A special analysis deserves the behavior of the *arcsin* transformation model because in all simulations for all sample sizes, and, for the three parameters, the model had a coverage of zero percent indicating that all confidence intervals never included the true parameter. This result is surprisingly, taking

into account that the *logit* transformation model presented the best performance and both transformations have the same theoretical objectives, getting normality and homogeneous variance; this result does not surprise upon analyzing that this model yielded bad lengths of the confidence intervals on mean response. (see the Table (4.5). ☑
The next example studies the four link functions of the BRM presented in (2.2) with regards to its ability of recovering the mean response throughout subintervals of the (0,1).

Example 14. *The* paint trial *experiment. Recovering the mean response.*

This simulation has as objective to study the behaviour of the link functions for the simple BRM that was fitted to the *paint trial* data. The links' behaviour is analyzed throughout the subintervals of unity interval. The procedure can be summarized as follows:

1. The systematic part was $\eta_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2}$, with x_1 and x_2 covariates taken from the *paint trial* data. Notation in matrix form: $\eta_{pt} = X_{pt}\boldsymbol{\beta}$, pt:*paint trial*.
2. Initial values were given for parameters: $(\beta_0, \beta_1, \beta_2)^T = (1, 0.2, 2)^T$.
3. Therefore, the η vector can be computed as $\eta_{true} = g(\mu) = X_{pt}\boldsymbol{\beta}$
4. Consequently, the *true* mean response for each link is computed as $\mu_{true} = g^{-1}(X_{pt}\boldsymbol{\beta})$. for example, for *logit* link: $\mu_{true} = 1/(1 + exp(-\eta_{true}))$.
5. Parameter ϕ was fixed as 10.
6. Interval (0,1) is partitioned in four subintervals as: (0,0.25), [0.25, 0.5), [0.5,0.75), and [0.75,1)
7. 2000 random numbers from a beta(a,b) distribution are simulated ($a = \mu\phi$, $b = \phi(1-\mu)$), and they are assigned to the respective subinterval.
8. In each subinterval a simple BR Model $g(\mu_i) = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2}$ is fitted, with its values of link, simulated y_i, and concatenating X_{pt} to equalize the sample size of simulated $\mathbf{y}$.
9. Each fit yields a parameter estimate $(\widehat{\beta}_0, \widehat{\beta}_1, \widehat{\beta}_2)^T$. Hence, there are four estimated vector for each link, (one for each subinterval).

10. For each link the mean response is estimated as $\widehat{\mu} = g^{-1}(X_{pt}\widehat{\boldsymbol{\beta}})$, in each subinterval. Summarizing, for a specific link, there are four estimates on mean response, one for each subinterval.

11. For each link, within each subinterval, is computed the Euclidean distance between $\widehat{\mu}$ and its respective μ_{true}.

The results are shown in Table 4.8. Upon analyzing the results of this simulation, the mean response can be recovered easier when values of y are in [0.75,1), modeling the mean with *logit*, *probit*, and *cloglog* links. When the mean was modeled by the *cauchit* link, it was impossible to recovery the mean response in the four subintervals. Also, at subinterval (0,0.25) it was not possible to recover the mean response for the four link functions. ☑.

TABLE 4.8. Euclidean distance between estimated vs *true* mean response for links and subintervals. Simulated Data from a simple BRM.

Link	(0,0.25)	[0.25,0.5)	[0.5,0.75)	[0.75,1)
logit	NC	0	106.83	3.16
probit	NC	0	106.83	3.16
cloglog	NC	47.63	106.86	3.16
cauchit	NC	NC	NC	NC
NC: Not converged.				

4.6.2 Applied unrestricted models: *semiconductors*, a 2^{4-1}_{III} experiment

Example 15. *The* Semiconductors *experiment. Half-normal plot and inference.*

In *semiconductors* data the seven models introduced in (4.6) were fitted. Initially, the Half-normal plots were performed and they are shown in Figure 4.4 for normal and transformations models, and the Figure 4.5 presents the plots for the BRM. Subsequently, the active effects extracted from each plot were used to fit the three normal models. For the simple BRM, the x1x4 effect was dropped because it presented an estimate near to zero in the Half-Normal plots for the normal models. A summary of results appears in Table (4.9). An inspection of this Table indicates that:

1) From Half-normal plots, the same four active effects for all models. A,B,C, and AB.

2) The three fitted normal models achieved normality (Shapiro test was nonsignificant in

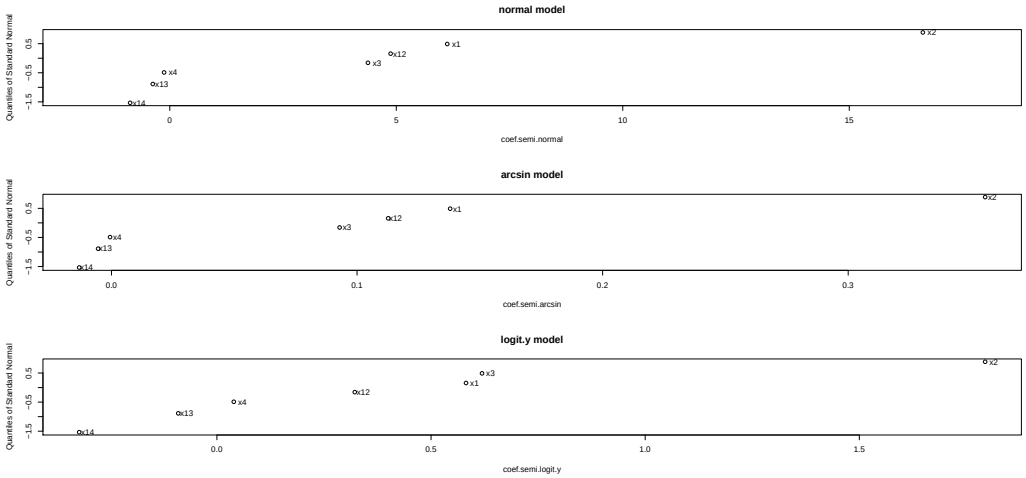

FIGURE 4.4. Semiconductors data. Half-Normal plots. Normal and transformations

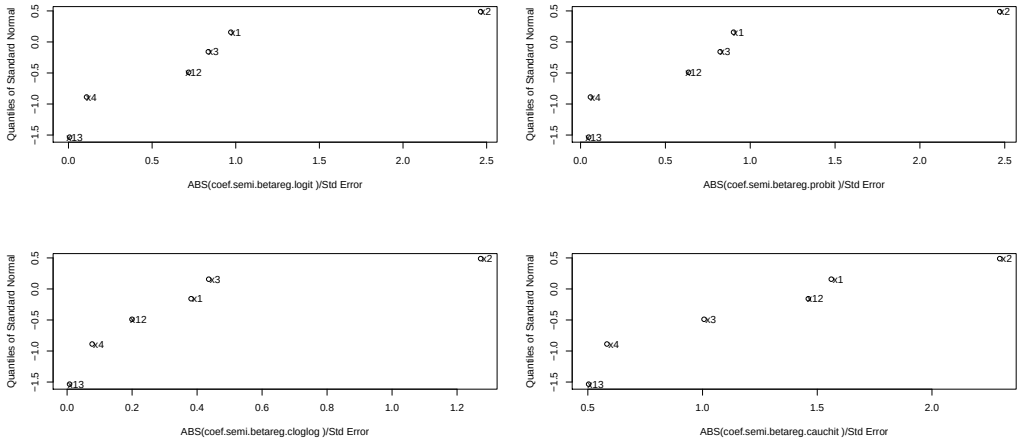

FIGURE 4.5. Semiconductors data. Half-Normal plots. Four Beta Regression Models

all models) and homoscedasticity (Breusch-Pagan test).

3) According to AIC and BIC criteria, the transformation *arcsin* was the best model. ☑.

TABLE 4.9. Semiconductors. Half-normal plots and inference

Model	$p\text{-}value < 0.06$	AIC	BIC
Normal	All	34.05	34.52
arcsin	**All**	**-44.39**	**-43.90**
logit	Not AB	6.18	6.65
beta.logit	All	-24.19	-23.71
beta.probit	All	-26.13	-25.65
beta.cloglog	All	-21.41	-20.93
beta.cauchit	Not AB	-17.77	-17.3

Example 16. *The* Semiconductors *experiment. Mean response by confidence intervals.*

For *semiconductors* data, Melo et al. (2007) applied a common linear regression model using the response on a 0 to 100 scale. Their fitted normal model had the covariates x_1, x_2, x_3, and the interaction x_1x_2. In the current work, using the same covariates, response was modeled as $(y/100)$, and using the same seven models studied in previous section, and the 95% confidence intervals on mean response were calculated. Hence, using the notation $\zeta_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3} + \beta_{12} x_{i1} x_{i2}$ for the linear predictor, and $\mathrm{E}(y_i) = \mu_i$, we explicitly fitted the following seven models for the *semiconductors* data:

1. $y_i = \zeta_i + \varepsilon_i$ (normal)
2. $\mathrm{logit}(y_i) = \zeta_i + \varepsilon_i$ (logit tranformation)
3. $\arcsin(\sqrt{y_i}) = \zeta_i + \varepsilon_i$ (arcsin transformation)
4. $\log(\frac{\mu_i}{1-\mu_i}) = \zeta_i$ (beta logit)
5. $\Phi^{-1}(\mu_i) = \zeta_i$ (beta probit)
6. $\log(-\log(1-\mu_i)) = \zeta_i$ (beta cloglog)
7. $\tan(\pi[\mu_i - \frac{1}{2}]) = \zeta_i$ (beta cauchit)

The lengths of confidence intervals on mean response for *semiconductors* data are shown in Table 4.10. Upon analyzing this Table, the best results were encountered in the beta models. The *logit* transformation was the worst, producing greater lengths than the normal model; however, it is not surprisingly because the parameter for interaction, β_{12}, was not significant (p-*value*=0.20). This characteristic was also encountered in the beta model with *cauchit* link, ($p-$value=0.70) for β_{12}; this fact explains why some lengths are greater than in the normal model. A little consideration should be had with regards to the BRM with the *cloglog* link: two experimental units had lengths greater than the normal model, ($p-$value=0.057) for β_{12}. ☑

TABLE 4.10. Lengths of the 95% confidence intervals on the mean response in *semiconductors* data. Systematic part for the seven models: $\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_{12} x_1 x_2$.

	Normal	Transform	Transform	BRM	BRM	BRM	BRM
RUN		logit	arcsin	**logit**	**probit**	**cloglog**	*cauchit*
1	0.13	0.19	0.16	0.03	0.03	0.04	0.05
2	0.13	0.23	0.18	0.07	0.06	0.07	0.08
3	0.13	0.44	0.26	0.12	0.10	0.15	0.30
4	0.13	0.51	0.29	0.12	0.10	0.16	0.36
5	0.13	0.29	0.21	0.07	0.06	0.08	0.11
6	0.13	0.34	0.22	0.08	0.07	0.09	0.12
7	0.13	0.50	0.28	0.10	0.09	0.12	0.16
8	0.13	0.47	0.28	0.10	0.08	0.12	0.15

CHAPTER 5

Conclusions.

In this work, a restricted variable dispersion beta regression model (restricted VDBRM) is justified, developed, and applied. Using a restricted likelihood by Lagrange multipliers, a solution for the restricted VDBRM from the frequentist perspective is explicitly done. Some solutions from the Bayesian framework are explored and left as a future work. Although the model is justified within the theory of the resolution for 2^{k-p} designs, the model can be applied in areas where the mean and the dispersion of a response variable in (0,1) need be explained by means of covariates in scenario when parameters are restricted: for example, relationship between covariates, higher order effects considered as negligible in factorial experiments.

A simulated restricted VDBRM was employed to exemplify the inferential aspects of the model. The main result indicates that when the model is applied in large sample sizes are obtained good estimates, in terms of the relative biases. Also is shown that the hypothesis about the 2^k and 2^{k-p} are special cases of the general hypothesis of the restricted VDBRM.

As the unrestricted VDBRM is used for analyzing data from unreplicated 2^{k-p} experiments, in this thesis is shown that the unrestricted VDBRM is a special case of the restricted VDBRM. It was done using a general expression for the hypothesis about parameters.

In this thesis, data from 2^{k-p} experiments with response in (0,1) under several kinds of

resolution were analyzed. Thus, unrestricted VDBRM and simple BRM where applied in order to analyze real and simulated fractional factorial experiments. Using these BRM's, three real data were analyzed: a complete factorial 2^4 experiment, and two fractional factorial experiments, a 2^{4-1}_{III}, and a 2^{8-4}_{IV}. Also, simulated data based on the fractional experiments were analyzed. With respect to the link functions, for the mean response, four functions were employed: *logit*, *probit*, *cloglog*, and *cauchit*. For the dispersion of the response, only the *log* link was used.

The results of these four BRM's were compared with other three models known in the literature: the linear normal, and the *logit* and *arcsin* transformations. In some specific cases, both, frequentist and Bayesian analyses were compared in a integrated way, similar to Aitkin (2010). For instance, the main integrating concept between Bayesian/likelihood approaches was obtained in the greater part of the examples: the frequentist analysis yielded similar results when a flat *prior* distribution was used in the Bayesian BRM's.

In the *paint trial* data (a 2^{8-4}_{IV} experiment), three analysis were done: frequentist and Bayesian beta regression for empirical data and frequentist analysis for simulated data in scenarios based on the empirical data. In the paint trial data.For Comparing the frequentist results, the lengths of confidence intervals from beta models were better than normal and transformed models for empirical Paint Trail data in all observations; for the semiconductors data, the *cloglog* and *cauchit* links for beta models had some greater lengths than the normal models for some observations, which is considered an unexpected result. After simulating with 1, 2, 3 and 4 replications of the beta *logit* model, using the systematic part of the *paint trial* data, were computed the intervals coverage for βs parameters applying the same seven models already mentioned; results for the simulated data can be summarized as follows: i) the *logit* transformation model presented the best result because coverage was greater than 90% for the three parameters of the fitted model in all sample sizes (replications) of the experiment; ii) the beta *cauchit* model was barely acceptable; iii) the other four models: normal, *arcsin* , beta *probit* , and beta *cloglog* presented very bad results. Finally, considering the bad results for frequentist BRM's with respect to the small sample sizes, also was fitted a Bayesian BRM for the *paint trial* data computing the lengths of credibility intervals on mean responses. The lengths were compared against the lengths of frequentist confidence intervals for the same observations produced for the four mentioned BRM's. The *logit* link presented similar results for Bayesian and frequentist fits; the *cloglog* and *cauchit* links had contradictory results and, unfortunately, the *probit*

link had credibility intervals greater than the normal confidence intervals for almost observations. Upon joining these results with the AIC values for the frequentist models, it is suggested to do further studies in order to elucidate some contradictory findings in the current simulated and real data. Studying the behaviour of link functions with respect to subintervals of (0,1) for y, in simulated data was found that the mean response is easily recovered within the subinterval [0.75, 1) for 3 out of 4 link functions; besides, the subinterval (0, 0.25) did not allow to recovery the mean response for any link function. As a final conclusion, one objective of this work was fulfilled: to present the restricted and unrestricted VDBRM as alternative methods for analyzing data from 2^{k-} experiments in (0,1). Also, in order to decide what method is the best, we suggest to analyze each particular application with the different methods. One method that is good in one specific data can not be good in other data. This conclusion also was obtained upon analyzing the *semiconductors* data (a 2^{4-1}_{III} experiment). This conclusion is similar to those obtained in works such as Bonat et al. (2012), and Patil & Kulkarni (2011), these results highlight the importance to study each problem at hand. Finally, I expect to see some promissory uses for the restricted variable dispersion beta regression model in several areas. Some additional subjects related to the restricted VDBRM will be considered in Chapter (6).

CHAPTER 6

Future work.

This chapter presents some Bayesian ideas to solve the restricted VDBRM and additional subjects to work based on the restricted VDBRM. Some results of these ideas are being developed in a next paper and the initial proposals were presented in Grajales et al. (2014).

6.1 Bayesian ideas for the restricted VDBRM

.

6.1.1 Complete Bayesian analysis for the restricted VDBRM

To develop completely the Bayesian analysis of the restricted VDBRM. To propose prior distributions, inferential results and diagnostic measures. The next example explores a comparison between a Bayesian analysis for a specific simulated restricted VDBRM against a frequentist analysis.

Example 17. *A simulated restricted simple BRM, Bayesian vs frequentist evaluations.*

This example uses the data from 3.17. Two models were fitted, one frequentist an unrestricted variable dispersion beta regression models one frequentist, and another Bayesian

were fitted. The models employed were,

$$g_1(\mu_i) = \beta_0 + \beta_3 x_{3i} + \beta_4 x_{4i} + \beta_{34} x_{3i} x_{4i} \qquad log(\phi_i) = \alpha_0 + \alpha_4 x_{4i}, \quad z0 = x0 \quad , \quad \text{and} \quad z4 = x4.$$

for $i = 1, ..., 8$. As the *logit* link was used for the mean response in the simulated model, here, in both frequentist and Bayesian models, the same link was employed. In the Bayesian case, a flat prior distributions for $(\beta_0, \beta_3, \beta_4, \beta_{34})^T$ and $(\alpha_0, \alpha_4)^T$ were employed $(N_6\left(\mathbf{0}, 10^6 I_6\right))$. The classical VDBRM was fitted by the *betareg* package (see Cribari-Neto & Zeileis (2010)), and the Bayesian fit by the *Bayesianbetareg* package, which uses the methodology presented in Cepeda & Gamerman (2005). Results are shown in Table (6.1) and the routine in R is presented in Appendix 4. On one hand, frequentist results in (6.1) indicate that the mean response is explained by means of the x_3, x_4, and x_{34} covariates. With respect to the dispersion parameter of the response, it can explained by the x_4 covariate, using the *log* link.

Also, the Bayesian results indicate that the credibility intervals for the six parameters do not include to zero. This coincides with $p - values$ lower than 0.05 in the frequentist analysis for the same parameters. Because a flat prior distribution were considered for the parameters, Bayesian inference based on the *posterior* distributions present results closer to the frequentist inference (based on the likelihood function)), see Aitkin (2010).

Note. Only the intercept in the dispersion submodel presented an unsatisfactory result (the credibility interval does not contain the zero, even it was not significatively different to zero in the frequentist analysis.)

☑

TABLE 6.1. Evaluating data from simulated restricted VDBRM

Parameters	Frequentist estimates	p.value	Bayesian estimates	CredIntLow	CredIntUpp
β_0	-12.68	< 0.05	-11.32	-14.64	-7.75
β_3	1.97	< 0.05	1.81	1.31	2.27
β_4	1.65	< 0.05	1.46	0.90	1.87
β_{34}	-0.26	< 0.05	-0.24	-0.30	-0.16
α_0	**0.002**	< 0.05	**1.64**	0.79	2.57
α_4	0.65	< 0.05	0.40	0.21	0.52

6.1.2 A Bayesian analysis for the nuisance parameters

Consider the restricted VDBRM (2.6) - (2.7), the penalty function (2.8), also $(\boldsymbol{\beta}, \alpha)^T$ as the parameters of interest, and λ the nuisance parameter. Within the Bayesian framework, in order to have inferential results, it is necessary to have a posteriori distribution of the parameters of interest. A strategy appears in books as Aitkin (2010) and Bernardo & Smith (2000). First to get a posterior distribution for the complete vector, $(\lambda, \boldsymbol{\beta}, \alpha)^T$, that is $p(\lambda, \boldsymbol{\beta}, \alpha|\mathbf{y})$. Second, the posterior (of interest) distribution can be computed as,

$$p(\boldsymbol{\beta}, \alpha|\mathbf{y}) = \int_0^\infty p(\lambda, \boldsymbol{\beta}, \alpha|\mathbf{y}) d\mathbf{y} \tag{6.1}$$

Now the questions arising are how to model the prior for λ and how to justify the mandatory orthogonality between λ and $\boldsymbol{\beta}$, taking account of the restrictions yielded for the confounded effects in fractional factorial experiments.

6.2 Additional topics

6.2.1 A library in R for the restricted VDBRM

To do a library for the R package using the computational routines presented in this work.

6.2.2 To extend the restricted VDBRM

- *Restrictions on parameters for the dispersion parameter.* When the parameters associated to the dispersion of the response variable are restricted, this information generates a more general model than the restricted VDBRM:

$$g_1(\boldsymbol{\mu}) = \boldsymbol{X}\boldsymbol{\beta}, \quad g_2(\boldsymbol{\phi}) = \boldsymbol{Z}\boldsymbol{\alpha} \qquad \text{s.t.} \quad \boldsymbol{R}_1\boldsymbol{\beta} = \boldsymbol{\delta}_1 \text{ and } \boldsymbol{R}_2\boldsymbol{\alpha} = \boldsymbol{\delta}_2$$

Information matrix (square of $(p_1 + p_2 + q_1 + q_2)$ order) to develop and compute:

$$K(\boldsymbol{\lambda}_1, \boldsymbol{\lambda}_2, \boldsymbol{\beta}, \boldsymbol{\alpha}) = \begin{pmatrix} K_{\lambda_1\lambda_1} & K_{\lambda_1\lambda_2} & K_{\lambda_1\beta} & K_{\lambda_1\alpha} \\ K_{\lambda_2\lambda_1} & K_{\lambda_2\lambda_2} & K_{\lambda_2\beta} & K_{\lambda_2\alpha} \\ K_{\beta\lambda_1} & K_{\beta\lambda_2} & K_{\beta\beta} & K_{\beta\alpha} \\ K_{\alpha\lambda_1} & K_{\alpha\lambda_2} & K_{\alpha\beta} & K_{\alpha\alpha} \end{pmatrix}$$

- To propose and develop a restricted VDBRM which includes some random effects.
- *Restricted Dirichlet regression.* To propose and develop a restricted VDBRM for a multivariate response which can be modelled by means of the Dirichlet distribution. Beta distribution is a special case of it. In that case, an unrestricted multivariate BRM can be encountered in the literature.
- A bivariate variable dispersion beta regression model with restrictions (frequentist and Bayesian approach).

6.2.3 The restricted VDBRM in 2^{k-p} experiments

In 2^{k-p} experiments with response in (0,1), some aspects can be explored using the restricted VDBRM: methods to select covariates, more diagnostics measures, outliers detecting, and response surface, among others.

6.2.4 Theoretical aspects

In this thesis the regularity conditions are assumed to do the inferential analysis. With respect to it, the regularity conditions in the restricted may be studied (proved) from the theoretical viewpoint.

CHAPTER 7

Appendix

Some computational routines of R are presented here. They are organized according to each Chapter. For interested readers, all routines are available .

7.1 Appendix 1. Introduction and motivation.

7.1.1 Program for Example 1.

```
#=========================================================================
# -------------------------------------------------------------------------
# ------------ BETA REGRESSION MODELS FOR 2^4 EXPERIMENT: DRILL DATA ----
# -------------------------------------------------------------------------
# Luis Fernando Grajales Hernandez
# Department of Statistics
# Universidad Nacional de Colombia, Bogota - Colombia
# contact: lfgrajalesh@unal.edu.co
##
#=========================================================================
library(MASS)
library(betareg)
library(lmtest)
library(Matrix)
```

```
library(Bayesianbetareg)
Advance<-read.csv2("d:/Advance_2a4.csv")
Advance <- data.frame(data_Advance)
attach(Advance)
Advance
summary(y)
y_origin <- y*100
y_origin
x0= as.matrix(rep(1,16)) # Intercepto
x1<-as.matrix(Advance[,2])
x2<-as.matrix(Advance[,3])
x3<-as.matrix(Advance[,4])
x4<-as.matrix(Advance[,5])
x12<-x1*x2
x13<-x1*x3
x14<-x1*x4
x23<-x2*x3
x24<-x2*x4
x34<-x3*x4
x123<-x1*x2*x3
x124<-x1*x2*x4
x134<-x1*x3*x4
x234<-x2*x3*x4
x1234<-x1*x2*x3*x4

###############
########################### Zi for Half normal plots 14 observations
index <-as.vector(1:14)
index
probab<-(index -0.5)/16
probab
Zi <- qnorm((index -0.5)/16)
Zi
######################
par(mfrow=c(2,2))
####################
######################### 1) NORMAL MODEL WITH ORIGINAL RESPONSE
```

```
############ SIMILAR RESULTS TO TABLE 3 LEWIS ET AL. 2001 a) (2*B4+x3ETA.HERE)
fit.adva.normal <-lm(y_origin ~
-1+x1+x2+x3+x4+x12+x13+x14+x23+x24++x34+x123+x124+x134+x234, data=Advance)
summary(fit.adva.normal)
summary(fit.adva.normal)$coefficients
### ESTIMATIONS ARE THE HALF OF THE TRUE VALUE. Montgomery (2004) Example 6.3
betas.adva.normal<-2*summary(fit.adva.normal)$coefficients[,1]
coef.adva.normal <-as.vector(sort(betas.adva.normal))
coef.adva.normal

## PLOT WITH ONLY ESTIMATIONS (LIKE MONTGOMERY (2004))

plot(coef.adva.normal, Zi,  main="normal model",
xlab="coef.adva.normal", ylab="Quantiles of Standard Normal")
text(coef.adva.normal +0.15,  Zi, names(sort(betas.adva.normal)))

################### 2) NORMAL MODEL TRANSFORMATION LOG(Y)
y_log<- log(y_origin)
y_log
fit.adva.log.y <-lm(y_log ~
 -1+x1+x2+x3+x4+x12+x13+x14+x23+x24++x34+x123+x124+x134+x234, data=Advance)
summary(fit.adva.log.y)
summary(fit.adva.log.y)$coefficients
### ESTIMATIONS ARE THE HALF OF THE TRUE VALUE. Montgomery (2004) Example 6.3
betas.adva.log.y<-2*summary(fit.adva.log.y)$coefficients[,1]
betas.adva.log.y
coef.adva.log.y <- as.vector(sort(abs(betas.adva.log.y)))
coef.adva.log.y

## PLOT ONLY WITH ESTIMATIONS
plot(coef.adva.log.y, Zi, main="log(y) model",
 xlab="coef.adva.log.y", ylab="Quantiles of Standard Normal")
text(coef.adva.log.y +0.02,  Zi, names(sort(betas.adva.log.y)))

## logLik(fit.adva.log.y)
## AIC(fit.adva.log.y)
## BIC(fit.adva.log.y)
```

```
## shapiro.test(residuals(fit.adva.log.y))

######################### 3) GLM WITH Gamma LINK
fit.adva.GLM.gamma <- glm(y_origin ~
 -1+x1+x2+x3+x4+x12+x13+x14+x23+x24+x34+x123+x124+x134+x234,
 family=Gamma(link="log"), data =Advance)
summary(fit.adva.GLM.gamma)
summary(fit.adva.GLM.gamma)$coefficients
coef.adva.GLM <- summary(fit.adva.GLM.gamma)$coefficients[,1]
s.e.coef.adva.GLM <- summary(fit.adva.GLM.gamma)$coefficients[,2]
ratio.coef.adva.GLM<-as.vector(sort(abs(coef.adva.GLM)/s.e.coef.adva.GLM))
ratio.coef.adva.GLM

plot(ratio.coef.adva.GLM, Zi, xlim=c(0,0.25), main="GLM.gamma model",
 xlab="ABS(coef.GLM.gamma)/Std Error", ylab="Quantiles of Standard Normal")
text(ratio.coef.adva.GLM +0.01,  Zi, names(sort(coef.adva.GLM)))

##logLik(fit.adva.GLM.gamma)
##AIC(fit.adva.GLM.gamma)
##BIC(fit.adva.GLM.gamma)

################## 4) SIMPLE BETA REGRESSION MODELS FOUR LINKS
fit.adva.beta.logit <-betareg(y ~
-1+x1+x2+x3+x4+x12+x13+x14+x23+x24+x34+x123+x124+x134+x234, data=Advance)
fit.adva.beta.logit
summary(fit.adva.beta.logit)$coefficients
coef.adva.beta.logit <-summary(fit.adva.beta.logit)$coefficients$mean[,1]
coef.adva.beta.logit
s.e.coef.adva.beta.logit <- summary(fit.adva.beta.logit)$coefficients$mean[,2]
s.e.coef.adva.beta.logit
ratio.coef.adva.beta.logit<-
as.vector(sort(abs(coef.adva.beta.logit)/s.e.coef.adva.beta.logit))
ratio.coef.adva.beta.logit

plot(ratio.coef.adva.beta.logit, Zi,
xlab="ABS(coef.adva.beta.logit)/Std Error", ylab="Quantiles of Standard Normal")
text(ratio.coef.adva.beta.logit +0.01,  Zi, names(sort(coef.adva.beta.logit)))
```

```
################### OTHERS LINKS
fit.adva.beta.probit <-betareg(y ~
 -1+x1+x2+x3+x4+x12+x13+x14+x23+x24+x34+x123+x124+x134+x234,
 link="probit", data=Advance)
fit.adva.beta.probit
summary(fit.adva.beta.probit)$coefficients
coef.adva.beta.probit <-summary(fit.adva.beta.probit)$coefficients$mean[,1]
coef.adva.beta.probit
s.e.coef.adva.beta.probit <- summary(fit.adva.beta.probit)$coefficients$mean[,2]
s.e.coef.adva.beta.probit
ratio.coef.adva.beta.probit
<-as.vector(sort(abs(coef.adva.beta.probit)/s.e.coef.adva.beta.probit))
ratio.coef.adva.beta.probit

plot(ratio.coef.adva.beta.probit, Zi,
xlab="ABS(coef.adva.beta.probit)/Std Error", ylab="Quantiles of Standard Normal")
text(ratio.coef.adva.beta.probit +0.01,  Zi, names(sort(coef.adva.beta.probit)))

########################### CLOGLOG LINK

fit.adva.beta.cloglog <-betareg(y ~
-1+x1+x2+x3+x4+x12+x13+x14+x23+x24+x34+x123+x124+x134+x234,
 link="cloglog", data=Advance)
fit.adva.beta.cloglog
summary(fit.adva.beta.cloglog)$coefficients
coef.adva.beta.cloglog <-summary(fit.adva.beta.cloglog)$coefficients$mean[,1]
coef.adva.beta.cloglog
s.e.coef.adva.beta.cloglog <- summary(fit.adva.beta.cloglog)$coefficients$mean[,2]
s.e.coef.adva.beta.cloglog
ratio.coef.adva.beta.cloglog<-
as.vector(sort(abs(coef.adva.beta.cloglog)/s.e.coef.adva.beta.cloglog))
ratio.coef.adva.beta.cloglog

plot(ratio.coef.adva.beta.cloglog, Zi,
xlab="ABS(coef.adva.beta.cloglog)/Std Error", ylab="Quantiles of Standard Normal")
```

```
text(ratio.coef.adva.beta.cloglog +0.01,  Zi, names(sort(coef.adva.beta.cloglog)))

########################## CAUCHIT LINK

fit.adva.beta.cauchit <-betareg(y ~
 -1+x1+x2+x3+x4+x12+x13+x14+x23+x24+x34+x123+x124+x134+x234,
 link="cauchit", data=Advance)
fit.adva.beta.cauchit
summary(fit.adva.beta.cauchit)$coefficients
coef.adva.beta.cauchit <-summary(fit.adva.beta.cauchit)$coefficients$mean[,1]
coef.adva.beta.cauchit
s.e.coef.adva.beta.cauchit <- summary(fit.adva.beta.cauchit)$coefficients$mean[,2]
s.e.coef.adva.beta.cauchit
ratio.coef.adva.beta.cauchit<-
as.vector(sort(abs(coef.adva.beta.cauchit)/s.e.coef.adva.beta.cauchit))
ratio.coef.adva.beta.cauchit

plot(ratio.coef.adva.beta.cauchit, Zi, xlim=c(0,2.5), main="betareg.cauchit model",
xlab="ABS(coef.adva.beta.cauchit)/Std Error", ylab="Quantiles of Standard Normal")
text(ratio.coef.adva.beta.cauchit +0.1,  Zi + 0.02, names(sort(coef.adva.beta.cauchit)))
```

7.2 Appendix 3. Frequentist Inference for the Restricted BRM.

7.2.1 Simulated Restricted VDBRM. Routines: Evaluating data. Example 7

```
beta_rest<-function(datos_r, formula, formula_phi, R, lambda,
    link="logit", prec=0.001){
X<-model.matrix(formula, data=datos_r)
Z<-model.matrix(formula_phi, data=datos_r)
Y<-c()
for(i in 1:length(colnames(datos_r))){
  if(formula[[2]]==colnames(datos_r)[i]){
    Y<-datos_r[,i]}
}
```

```
p1<-ncol(X)
p2<-ncol(Z)
fit.r = betareg(formula, data=datos_r)

theta_old<-c(coef(fit.r)[1:p1], log(coef(fit.r)[p1+1]), rep(0, p2-1))
### 1)Creation of mu's with diferents link functions
muF<- function(beta, X, link=link){
 eta <- X %*% beta
 etas<-c()
for (i in 1:length(eta)){
  if(link=="logit"){
    etas[i] <- 1/(1+exp(-eta[i]))
  }else{
    if(link=="probit"){
      etas[i]<-pnorm(eta[i])
    }else{
      if(link=="cloglog"){
        etas[i]<-1-exp(-exp(eta[i]))
      }else{
        if(link=="cauchit"){
          etas[i]<-atan(eta[i])/pi+1/2
        }else{
          print("link function not supported!")
        }
      }
    }
  }
 }
 etas
 }

###################### 2) For g'(mu)

gpF<-function(mu, link=link){
  gp<-c()
  for(i in 1:length(mu)){
```

```
    if(link=="logit"){
    gp[i]<-1/(mu[i]*(1-mu[i]))
    }else{
      if(link=="probit"){
        gp[i]<-dnorm(mu[i])
      }else{
        if(link=="cloglog"){
          gp[i]<--mu[i]/((1-mu[i])*log(1-mu[i]))
        }else{
          if(link=="cauchit"){
            gp[i]<-pi*1/(cos(pi*(mu[i]-1/2))^2)
          }else{
            print("link function not supported!")
          }
        }
      }
    }
  }
gp
}


###### 3) Diagonal Matrix W 8x8

KINV<-function(theta, X, Y, lambda, Z){
  mu<-muF(beta = theta[1:p1], X = X, link = link)
  phi<-exp(Z%*%theta[(p1+1):(p1+p2)])
  d<-c()
  for(i in 1:length(mu)){
    d[i]<-((mu[i]^2)*trigamma(mu[i]*phi[i])+
  ((1-mu[i])^2)*trigamma((1-mu[i])*phi[i])-
      trigamma(phi[i]))*(phi[i]^2)
  }
  D<-diag(d)
  w<-c()
  for(i in 1:length(mu)){
    w[i]=(phi[i]^2/(gpF(mu[i], link = link)^2))*
```

```
  (trigamma(mu[i]*phi[i])+trigamma((1-mu[i])*phi[i]))
  }
  W<-diag(w)
  Lamb<-diag(lambda)
  a<-c()
  for(i in 1:nrow(X)){
    a[i]<-(1/(gpF(mu[i], link = link)^2))*(trigamma(mu[i]*phi[i])
    +trigamma((1-mu[i])*phi[i]))
  }
  c<-c()
  for(i in 1:nrow(X)){
    c[i]<-phi[i]*(mu[i]*a[i]-trigamma((1-mu[i])*phi[i]))*
      (phi[i]/gpF(mu[i], link = link))
  }
  V<-diag(c)
  B22<-solve(t(Z)%*%D%*%Z-t(Z)%*%t(V)%*%X%*%solve(t(X)%*%W%*%X+
      t(R)%*%Lamb%*%R)%*%t(X)%*%V%*%Z)
  B11<-solve(t(X)%*%W%*%X+t(R)%*%Lamb%*%R)+
    solve(t(X)%*%W%*%X+t(R)%*%Lamb%*%R)%*%t(X)%*%
V%*%Z%*%B22%*%t(Z)%*%t(V)%*%X%*%
  solve(t(X)%*%W%*%X+t(R)%*%Lamb%*%R)
  B12<--solve(t(X)%*%W%*%X+t(R)%*%Lamb%*%R)%*%t(X)%*%V%*%Z%*%B22
  cbind(rbind(B11, t(B12)),rbind(B12, B22))
}



Q<-function(theta, X, Y, Z, lambda){
  mu<-muF(beta = theta[1:p1], X = X, link = link)
  phi<-exp(Z%*%(theta[(p1+1):(p1+p2)]))
  T1<-diag(as.vector(phi/gpF(mu, link = link)))
  T2<-diag(as.vector(phi))
  y_ast<-log(Y/(1-Y))
  mu_ast<-digamma(mu*phi)-digamma((1-mu)*phi)
  nu<-c()
  for(i in 1:nrow(X))
    nu[i]<-digamma(phi[i])+mu[i]*(y_ast[i]-mu_ast[i])
      -digamma((1-mu[i])*phi[i])+log(1-Y[i])
```

```
  Q1<-t(X)%*%T1%*%(y_ast-mu_ast)
  Q2<-t(Z)%*%T2%*%nu
  c(Q1, Q2)
}

i<-1
theta<-theta_old+5*prec
while(max(abs(theta-theta_old))>=prec){
  theta<-theta_old+KINV(theta_old, X, Y, lambda, Z)%*%
      Q(theta_old, X, Y, Z, lambda)
  theta_old<-theta
  i<-i+1
  print(i)
}
names(theta)<-c(colnames(X), colnames(Z))

VAR<-abs(KINV(theta, X, Y, lambda, Z))
estand<-c()
for(i in 1:length(theta))
  estand[i]<-theta[i]/sqrt(VAR[i,i])
p_value<-(1 - pnorm(abs(estand), 0, 1))/2

list(parameters=data.frame(params=theta[1:p1],
    z=estand[1:p1] ,p.value=p_value[1:p1]),
     parametrs_phi=data.frame(params=theta[(p1+1):(p1+p2)],
    z=estand[(p1+1):(p1+p2)],p.value=p_value[(p1+1):(p1+p2)]), niter=i)
}
```

7.2.2 Simulated Restricted VDBRM. Routines: Bayesian and frequentist analysis. Example 8

```
# Frequentist BRM
fit.simula.vardbeta.logit<-betareg(y~x3+x4+x34 | x4, data=simulated1,
link = "logit", link.phi = "log")
summary(fit.simula.vardbeta.logit)
fit.simula.vardbeta.probit<-betareg(y~x3+x4+x34 | x4, data=simulated1,
```

```
link = "probit", link.phi = "log")
summary(fit.simula.vardbeta.probit)
fit.simula.vardbeta.clog<-betareg(y~x3+x4+x34 | x4, data=simulated1,
link = "cloglog", link.phi = "log")
summary(fit.simula.vardbeta.clog)
fit.simula.vardbeta.cau<-betareg(y~x3+x4+x34 | x4, data=simulated1,
link = "cauchit", link.phi = "log")
summary(fit.simula.vardbeta.cau)

# Bayesianbetareg
X_Bayes=cbind(x0,x3,x4,x34)
Z0 <-  x0
Z <- cbind(x0,x4)
n <- length(y)
burn <- 0.3
jump <- 4
nsim <- 600
bpri <- c(0,0,0,0)
Bpri <- diag(100,nrow=ncol(X_Bayes),ncol=ncol(X_Bayes))
gpri <- c(0,0)
Gpri <- diag(10,nrow=ncol(Z),ncol=ncol(Z))
Bayes.vard.simula2<-Bayesianbetareg(y,X_Bayes,Z,
 nsim,bpri,Bpri,gpri,Gpri,0.3,3,graph1=FALSE,graph2=FALSE)
summary(Bayes.vard.simula2)
```

7.3 Appendix 4. Unrestricted models.

7.3.1 Random Walk Metropolis-Hasting for Example 10: Description.

Although is difficult to order the concepts, here they are introduced according to Robert & Casella (2004). A Markov chain Monte Carlo method, (MCMC), for the simulation of a distribution f is any method producing an ergodic Markov chain $(W^{(t)})$ whose stationary distribution is f. (Current use: It is possible to obtain a sample $W_1, \ldots, W_n$ approximately distributed from f without directly simulating from f). An ergodic Markov chain is a particular stochastic process. The Metropolis-Hasting (MH) is an algorithm to sample MCMCs, and is preferred rather Gibbs sampling when the prior $p(\theta)$ and the Likelihood $L(\mathbf{y}|\theta)$ do not belong to the same distributional

family, i.e., they are not a *conjugate pair*. The Metropolis-Hasting algorithm (MH) starts with the (target) objective density f. A conditional density $q(y|w)$, defined with respect to the dominating measure for the model, is then chosen. The MH algorithm can be implemented when $q(\cdot|w)$ is easy to simulate from and is either explicitly available or symmetric ($q(y|w) = q(w|y)$). The target density f must be available to some extent; a general requirement is that the ratio $\frac{f(y)}{q(y|w)}$ is known up to a constant independent of w. The MH algorithm associated with the target density f and the conditional density q produces a Markov chain $(W^{(t)})$ through the following transition, named Algorithm.24 in Robert & Casella (2004). Given $w^{(t)}$,

1. Generate $Y_t \sim q(y|w^{(t)})$
2. Take $W^{(t+1)} = \begin{cases} Y_t & \text{with probability } \rho(w^{(t)}, Y_t) \\ w^{(t)} & \text{with probability } 1 - \rho(w^{(t)}, Y_t) \end{cases}$

where $\rho(w, y) = \min\{\frac{f(y)}{f(w)} \frac{q(w|y)}{q(y|w)}, 1\}$. The q distribution is called the *proposal* distribution and the probability $\rho(w, y)$ the Metropolis-Hasting *acceptance probability*. The Random walk Metropolis-Hasting (RWMH) is a modification of the (MH) algorithm taking into account the value previously simulated to generate the following value; that is, to consider a *local* exploration of the neighborhood of the current value of the Markov chain. The change consists in the first choice to simulate Y_t, according to

$$Y_t = X^{(t)} + \epsilon_t \tag{7.1}$$

where ϵ_t is a random perturbation, independent of $X^{(t)}$. Then the Markov chain in the MH algorithm associated with q is a *random walk* on the support of f density; in the RWMH method, is chosen a symmetric function g (that is, such that $g(t) = g(-t)$), which leads to the following simpler algorithm:
Given $w^{(t)}$,

1. Generate $Y_t \sim g(|y - w^{(t)}|)$
2. Take $W^{(t+1)} = \begin{cases} Y_t & \text{with probability } \min\{1, \frac{f(Y_t)}{f(w^{(t)})}\} \\ w^{(t)} & \text{otherwise.} \end{cases}$

Bibliography

Abramowitz, M. & Stegun, I. A. (1965). *Handbook of Mathematical Functions*, Dover, New York.

Aitchison, J. & Silvey, S. (1958). Maximum-likelihood estimation of parameters subject to restraints, *The Annals of Mathematical Statistics* **29**: 813–828.

Aitchison, J. & Silvey, S. (1960). Maximum-likelihood estimation procedures and associated tests of significance, *Journal Royal of the Statistical Society, Series B (Methodological)* **22**: 154–171.

Aitkin, M. (2010). *Statistical Inference, An Integrated Bayesian/Likelihood Approach*, CRC Press.

Akaike, H. (1974). A new look at the statistical model identification, *IEEE Transactions on Automatic Control* **19**: 716–723.

Anderson-Cook, C. & Hamada, M. (2014). Comment: Toward guidelines for practitioners on screening design and analysis, *Technometrics* **56**: 16–19.

Aranda-Ordaz, F. J. (1981). On two families of transformations to additivity for binary response data, *Biometrika* **68**: 357–363.

Atkinson, A. (1985). *Plots, tranformations and regression: An Introduction to Graphical Methods of Diagnostic Regression Analysis*, Clarendon Press, Oxford.

Atkinson, A., Donev, A. & Tobias, R. (2007). *Optimum Experimental Designs, with SAS*, Oxford University Press.

Basu, D. (1977). On the elimination of nuisance parameters, *Journal of The American Statistical Association* **72**: 355–366.

Bayarri, M. & Berger, J. (2004). The interplay of bayesian and frequentist analysis, *Statistical Sciences* **19**: 58–80.

Bayer, F. & Cribari-Neto, F. (2013). Bartlett corrections in beta regression models, *Journal of Statistical Planning and Inference* **143**: 531–547.

Bayer, F. & Cribari-Neto, F. (2014). Model selection criteria in beta regression with varying dispersion. arXiv:1405.371v1.

Bernardo, J. & Smith, A. (2000). *Bayesian Theory*, Wiley.

Billard, L. (2014). Sir Ronald A. Fisher and The International Biometric Society, *Biometrics* **70**: 259–265.

Bonat, W., Ribeiro, P. & Zeviani, W. (2012). Regression models with responses on the unity interval: specification, estimation and comparison, *Revista Brasileira de Biometria* **30**: 415–431.

Box, G. E. & Cox, D. R. (1964). An analysis of transformations, *Journal of the Royal Statistical Society Series B* **26**: 211–246.

Box, G. E. P. & Hunter, J. S. (1961a). The 2^{k-p} Fractional Factorial Designs I, *Technometrics* **3**: 311–351.

Box, G. E. P. & Hunter, J. S. (1961b). The 2^{k-p} Fractional Factorial Designs Part II, *Technometrics* **3**: 449–458.

Box, G., Hunter, J. S. & Hunter, W. G. (1978). *Statistics for Experimenters*, John Wiley & Sons.

Box, G., Hunter, J. S. & Hunter, W. G. (2005). *Statistics for Experimenters*, Wiley.

Box, J. F. (1978). *R. A. Fisher, the Life of a Scientist.*, John Wiley.

Branscum, A., Johnson, W. & Thurmond, M. (2007). Bayesian beta regression: Applications to household expenditure data and genetic distance between foot-and-mooth disease viruses, *Australian and New Zealand Journal of Statistics* **3**: 287–301.

Bury, K. (1999). *Statistical Distributions in Engineering*, Cambridge University Press, New York.

Carlin, B. & T.A., L. (2009). *Bayesian methods for data analysis*, CRC Press.

Casella, G. & Berger, R. (1987). Reconciling Bayesian and frequentist evidence in the one-sided testing problem, *Journal of the American Statistical Association* **82**: 106–11.

Cepeda, E. (2001). *Variability modeling in generalized linear models*, PhD thesis, Universidade Federal do Rio de Janeiro.

Cepeda, E. & Gamerman, E. (2005). Bayesian methodology for modeling parameters in the two parameter exponential family, *ESTADISTICA* **157**: 93–105.

Chaloner, K. & Verdinelli, I. (1995). Bayesian experimental design: A review, *Statistical Science* **10**: 273–304.

Chien, L.-C. (2011). Diagnostics plots in beta-regression models., *Journal of Applied Statistics* **38**: 1607–1622.

Cox, D. (2006). *Principles of Statistical Inference*, Cambridge University Press.

Cox, D. & Hinkley, D. (1974). *Theoretical Statistics*, Chapman and Hall.

Cribari-Neto, F. & Queiroz, M. (2012). On testing inference in beta regressions, *Journal of Statistical Computation and Simulation* **1**: 1–18.

Cribari-Neto, F. & Souza, T. (2011). Testing inference in variable dispersion beta regressions, *Journal of Statistical Computational Statistics and Simulation* **DOI:10.1080/00949655.2011.599033**: 1–17.

Cribari-Neto, F. & Vasconcellos, K. (2002). Nearly unbiased maximun likelihood estimation for the beta distribution, *Journal of Statistical Computation and Simulation* **72**: 107–118.

Cribari-Neto, F. & Zeileis, A. (2010). Beta regression in R, *Journal of Statistical Software* **34**(2): 1–24.

Cysneiros, F. (1997). *Estima(c)(a)o e testes em modelos lineares generalizados com restri(c)(o)es nos par(a)metros na forma de desigualdades lineares*, Master's thesis, Universidade de S(a)o Paulo.

Cysneiros, F. & Paula, G. (2005). Restricted methods in symmetrical linear regression models., *Computational Statistics and Data Analysis* **49**: 689–708.

Daniel, C. (1976). *Applications of Statistics to Industrial Experimentation*, Wiley.

Dey, A. (1985). *Orthogonal Fractional Factorial Designs*, Wiley.

Diaz, L. G. (2002). *Estadística Multivariada: Inferencia y Métodos*, Bogotá, D. C.

Dossou-Gbété, S. & Tinsson, W. (2005). Factorial experiments designs and generalized linear models., *SORT* **29**: 249–268.

Edwards, L., Stewart, P., Muller, K. & Helms, R. (2001). Linear equality constraints in the general linear mixed model, *Biometrics* **57**: 1185–1190.

Efron, B. (1979). Bootstrap methods: Another look at the jackknife, *Annals in Statistics* **7**: 1–26.

Espinheira, P. (2007). *Regressao Beta*, PhD thesis, Universidade de Sao Paulo.

Espinheira, P. L., Ferrari, S. L. P. & Cribari-Neto, F. (2008a). Influence diagnostics in beta regression, *Computational Statistics & Data Analysis* **52**: 4417–4431.

Espinheira, P. L., Ferrari, S. L. P. & Cribari-Neto, F. (2008b). On beta regression residuals, *Journal of Applied Statistics* **35**(4): 407–419.

Fahrmeir, L. & Tutz, G. (1994). *Multivariate statistical modeling based on generalized linear models*, Springer Verlag, New York.

Ferrari, S. (2013). Beta regression modeling: recent advances in theory and applications, 13a. Escola de Modelos de Regress(a)o. Maresias, S(a)o Paulo.

Ferrari, S. & Cribari-Neto, F. (2004). Beta regression for modelling rates and proportions, *Journal of Applied Statistics* **31**: 799–815.

Ferrari, S. L. P., Espinheira, P. L. & Cribari-Neto, F. (2011). Diagnostic tools in beta regression with varying dispersion, *Statistica Neerlandica* **65**.

Figueroa Zuniga, J. (2011). *Modelos de Regresion beta con efectos mixtos: enfoque Byesiano*, PhD thesis, Pontificia Universidad Catolica de Chile.

Finney, D. J. (1945). Fractional replication of factorial arrangements," annals of eugenics, 12, 291-301. fisher, r. a. (1922), "on the mathematical foundations of theoretical statistics," philosophical transactions of the royal society of london, ser. a, 222, 309-368. (1926), "the arrangement of field experiments," journal of the ministry of agriculture, 33, 503-513., *Annals of Eugenics* **12**: 291–301.

Fisher, R. (1926). The arrangement of field experiments., *Journal of the Ministry of Agriculture.* **33**: 503–513.

Gourieroux, C., Holly, A. & Monfort, A. (1982). Likelihood ratio test, wald test, and kuhn-tucker test in linear models with inequality constraints on the regression parameters, *Econometrica* **50**: 63–80.

Gourieroux, C. & Monfort, A. (1989a). *Statistics and Econometric Models.*, Vol. 1, Cambridge University Press.

Gourieroux, C. & Monfort, A. (1989b). *Statistics and Econometric Models.*, Vol. 2, Cambridge University Press.

Grajales, L. & Calderón, S. (2012). Bayesian estimation on mean response for $2^(k-p)$ experiments with beta response: empirical data., *XXII Simposio Internacional de Estadística, Bucaramanga, Colombia.*

Grajales, L., López, L. & Melo, O. (2014). Analysis of $2^(k-p)$ experiments with beta response using linear restrictions, *XXIV Simposio Internacional de Estadística*, Universidad Nacional de Colombia.

Grajales, L., Ospina, R. & López, L. A. (2012). On estimated means for $2^(k-p)$ experiments with beta response, *Journal of Statistics: Advances in Theory and Applications* **8**: 105–130.

Grün, B., Kosmidis, I. & Zeileis, A. (2012). Extended beta regression in r: Shaken, stirred, mixed, and partitioned, *Journal of Statistical Software* **48**: 1–25.

Hamada, M. & Nelder, J. (1997). Generalized linear models for quality improvement experiments, *Journal of Quality Technology* **29**: 292–304.

Henderson, H. & Searle, S. (1981). On deriving the inverse of a sum of matrices, *SIAM Review* **23**: 53–60.

John, P. (1971). *Statistical Design and Analysis of Experiments*, Mac Millan Co.

Johnson, N. L., Kotz, S. & Balakrishnan, N. (1995). *Continuous Univariate Distributions*, John Wiley & Sons, New York.

Jorgesen, B. (1997). Proper dispersion models (with discussion), *Brazilian Journal of Probability and Statistics* **11**: 89–140.

Kieschnick, R. & McCullough, B. (2003). Regression analysis of variates observed on (0,1): Percentages, proportions, and fractions, *Statistical Modelling* **3**: 193–213.

Kosmidis, I. & Firth, D. (2010). A generic algorithm for reducing bias in parametric estimation., *Electronical Journal of Statistics* **4**: 1097–1112.

Kruschke, J. (2011). *Doing Bayesian Data Analysis: A tutorial with R and BUGS*, Academic Press.

Lange, K. (2010). *Numerical analysis for statisticians*, Springer Verlag.

Lewis, S., Montgomery, D. & Myers, D. E. (2001a). Confidence interval coverage for designed experiments analyzed with GLM, *Journal of Quality Technology* **3**: 279–292.

Lewis, S., Montgomery, D. & Myers, R. (2001b). Examples of designed experiments with nonnormal responses., *Journal of Quality Technology* **3**: 279–292.

Liang, K. Y. & Zeger, S. L. (1986). Longitudinal data analysis using generalized linear models, *Biometrika* **73**: 13–22.

Lieberman, O. (2010). Asymptotic theory for empirical similarity models, *Econometric Theory* **26**(4): 1032–1059.

Lin, D. K. J. & Tu, W. (1995). Dual Response Surface Optimization, *Journal of Quality Technology* **27**: 34–39.

Marin, M., Rojas, J., Jaimes, D., Gutierrez, A. & Cepeda-Cuervo, E. (2013). *Bayesianbetareg: Bayesian Beta regression-joint mean and precision modeling*, Universidad Nacional de Colombia, Bogotá, Colombia.
*http://cran.r-project.org/web/packages/Bayesianbetareg/index.html

McCullagh, P. & Nelder, J. (1983). *Generalized Linear Models*, Chapman Hall, London.

McCullagh, P. & Nelder, J. (1989). *Generalized Linear Models*, Chapman Hall, London.

McCulloch, C. E., Searle, S. R. & Neuhaus, J. M. (2008). *Generalized, Linear, and Mixed Models*, second edn, John Wiley & Sons, New Jersey.

Melo, O. O., López, L. A. & Melo, S. (2007). *Diseño de Experimentos (métodos y aplicaciones)*, Universidad Nacional de Colombia.

Montgomery, D. (2001). *Design and Analysis of Experiments*, 5th Edition, México.

Montgomery, D. (2009). *Design and Analysis of Experiments*, Wiley.

Montgomery, D. C. (2003). *Diseño y Análisis de Experimentos*, Segunda Edición, Grupo Editorial Limusa, S.A, México.

Montgomery, D. C. (2004). *Diseño y Análisis de Experimentos*, Segunda Edición, Grupo Editorial Limusa, S.A, México.

Montgomery, D., Peck, E. & Vining, G. (2006). *Introduction to Linear Regression Analysis*, John Wiley and Sons Inc.

Myers, R. H., Montgomery, D. C. & Vinning, G. G. (2002). *Generalized Linear Models. With Applications in Engineering and the Sciences*, John Wiley & Sons, New York.

Myers, R. & Montgomery, D. (1995). *Response Surface Methodology: Process and Product Optimization Using Designed Experiments.*, Wiley.

Myers, R. & Montgomery, D. (1997). A tutorial on generalized linear models, *Journal of Quality Technology* **29**: 274–291.

Myers, R., Montgomery, D., Vining, G. & Robinson, T. (2011). *Generalized Linear Models: with Applications in Engineering and the Sciences*, Wiley.

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

Neter, J., Kutner, M. H., Nachtsheim, C. J. & Wasserman, W. (1996). A tutorial on generalized linear models, *Applied Linear Statistical Models* .

Neuenschwander, B. & Flury, B. (1997). A note on Silvey 's (1959) theorem, *Statistics and Probability Letters* **36**: 307–317.

Nocedal, J. & Wright, S. J. (1999). *Numerical Optimization*, Springer-Verlag, New York.

Nyquist, H. (1991). Restricted estimation of generalized linear models, *Journal Royal of the Statistical Society. Series C (Applied Statistics)* **40**: 133–141.

Ospina, R., Cribari-Neto, F. & Vasconcellos, K. L. P. (2006). Improved point and interval estimation for a beta regression model, *Computational Statistics & Data Analysis* **51**: 960–981. Erratum at Computational Statistics and Data Analysis, 55, 2445.

Ospina, R. & Ferrari, S. (2010). Inflated beta distributions, *Statistical Papers* **51**: 111–126.

Paolino, P. (2001). Maximum likelihood estimation of models with beta-distributed dependent variables, *Political Analysis* **9**: 325–346.

Patil, S. & Kulkarni, H. (2011). Analysis of 2^n factorial experiments with exponentially distributed response., *Applied Mathematical Sciences* **10**: 459–476.

Pearson, E. (1931). The analysis of variance in cases of non-normal variation., *Biometrika* **23**: 114–133.

Plackett, R. L. & Burman, J. P. (1946). The design of optimum multifactorial experiments., *Biometrika* **33**: 305–325.

Pregibon, D. (1984). Review of generalized linear models, *Annals of Statistics* **12**: 1589–1596.

R Development Core Team (2012). *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria, URL http://www.R-project.org/.

R Development Core Team (2014). *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria, `http://www.R-project.org/`.

Rao, C. R. (1973). *Linear Statistical Inference and Its Applications*, Second Edition, John Wiley & Sons, New York.

Rao, C. R. & Mitra, S. K. (1971). *Generalized Inverse of Matrices and Its Applications*, John Wiley & Sons, New York.

Ravishanker, N. & Dey, D. (2003). *A first course in linear model theory.*, Chapman Hall.

Robert, C. & Casella, G. (2004). *Monte Carlo Statistical Methods*, Springer Verlag.

Samaniego, F. & Reneau, D. (1994). Toward a reconciliation of the bayesian and frequentist approaches to point estimation, *Journal of the American Statistical Association* **89**: 947–957.

Searle, S. R. (1971). *Linear Models*, John Wiley & Sons, New York.

Shao, J. (2003). *Mathematical Statistics*, Springer Verlag.

Shapiro, S. S. & Wilk, M. B. (1965). An Analysis of Variance Test for Normality (complete samples), *Biometrika* **52**: 591–611.

Shilov, G. (1973). *Elementary Real and Complex Analysis*, Dover.

Silvey, S. (1959). The Lagrangian multiplier test., *The Annals of Mathematical Statistics* **30**: 389–407.

Simas, A. B., Barreto-Souza, W. & Rocha, A. V. (2010). Improved estimators for a general class of beta regression models, *Computational Statistics & Data Analysis* **54**(2): 348–366.

Siqueira, A. L. (1983). *Uso de Transformaciones en Análisis de Varianza y Análisis de Regresión*, Disertación USP/IME, .

Smithson, M., Merkle, E. C. & Verkuilen, J. (2011). Beta finite mixture regression models of polarization and priming, *Journal of Educational and Behavioral Statistics* .

Smithson, M. & Verkuilen, J. (2006). A better lemon squeezer? maximum-likelihood regression with beta-distributed depedent variables, *Psychological Methods* **11**(1): 54–71.

Stasinopoulos, D. & Rigby, R. (2007). Generalized additive models for location scale and shape (GAMLSS) in r., *Journal of Statistical Software* **23**: 7.

Stein, S. (1984). *Cálculo y Geometría Analítica*, McGraw-Hill.

Steinberg, D. & Hunter, W. (1984). Experimental design: Review and comment., *Technometrics* **26**: 71–97.

Swearingen, C., Castro, M. & Bursac, Z. (2011). Modeling percentages outcomes: the beta regression macro., *SAS Global Forum* . Paper 335.

Swearingen, C., Castro, M. & Bursac, Z. (2012). Inflated beta regression: zero, one, and everything in between., *SAS Global Forum* . Paper 325.

Tian, Y. (2010). On equalities of estimations of parametric functions under a general linear model and its restricted models, *Metrika* **72**: 313–330.

Torres, V. (1993). A simple analysis of unreplicated factorials with possible abnormalities, *Journal of Quality Technology* **25**: 183–187.

Vasconcellos, K. L. P. & Cribari-Neto, F. (2005). Improved maximum likelihood estimation in a new class of beta regression models, *Brazilian Journal of Probability and Statistics* **19**: 13–31.

Wedderburn, R. (1974). Quasi-likelihood functions, generalized linear models and the Gauss Newton method, *Biometrika* **61**: 439–447.

Wei, B. C., Hu, Y. Q. & Fung, W. K. (1998). Generalized leverage and its applications, *Scandinavian Journal of Statistics* **25**: 25–37.

Wolak, F. (1987). An excat test for multiple inequality and equality constraints in the linear regression model, *Journal of the American Statistical Association* **82**: 782–793.

Wu, Y., Fedorov, V. & Propert, K. (2005). Optimal design for dose response using beta distributed responses, *Journal of Biopharmaceutical Statistics* **15**: 753–771.

Xie, M. & Singh, K. (2013). Confidence distribution, the frequentist distribution estimator of a parameter: A review, *International Statistical Review* **81**: 3–39.

Zhao, W., Zhang, R., Yazhao, L. & Liu, J. (2014). Variable selection for varying dispersion beta regression model, *Journal of Applied Statistics* **41**: 95–108.