Nguyen, The M. (2006) *Optimal cutpoint determination via design theory for regression models.*

PhD thesis

OPTIMAL CUTPOINT DETERMINATION VIA DESIGN THEORY FOR REGRESSION MODELS

The M. Nguyen

A Dissertation Submitted to the

University of Glasgow

Faculty of Information and Mathematical Sciences

for the degree of

Doctor of Philosophy

Department of Statistics

October 2006

To My Family

ABSTRACT

The main focus of this thesis is how to optimally choose a set or sets of cut-points (in a categorized survey question) which are offered to respondents. In the case of several sets, a further issue is how to allocate sampled subjects to these sets (design points). Applications include Contingent Valuation (CV) studies (surveys on a population's willingness to pay for a service or public good) and market research studies which might include for example a question on individual incomes.

Chapter 1 starts with an introduction to linear and non linear design theory including properties of the information matrix of the design. Then, a general optimizing problem (P1) is stated along with the concepts of directional derivatives and optimality conditions. The chapter closes with the introduction of several design criteria and their properties. The criteria include A-, D-, G-, E-, D_A-, linear and E_A-optimality.

Chapter 2 considers the formation of the problem in our particular context which begins with the main idea of the problem and applications in contingent valuation (CV) studies (the primary aim is to estimate a population's willingness to pay (WTP) for some non market products or public goods) and market research studies. For example a survey is conducted and X, $X \in \mathcal{X} = [C, D]$ is a variable of interest. Suppose that responses are placed into one of k categories determined by cutpoints $x_0, x_1, x_2, \ldots, x_{k-1}, x_k$, $x_0 = C$ and

$x_k = D$. A generalized linear model of the form $P(X \leq x) = F(\alpha + \beta x)$, $x \in \mathcal{X}$ is adopted where α and β are reparameterized from a location parameter μ and a scale parameter σ, $\alpha = -\mu/\sigma$, $\beta = 1/\sigma$. We transform X to Z: $Z = (X - \mu)/\sigma = \alpha + \beta X$. We have a standardized form of the set of cutpoints $z_0, z_1, \ldots, z_k$, $z_i = \alpha + \beta x_i$, $Z \in \mathcal{Z} = [A, B]$, $z_0 = A, z_k = B$. Then some design objectives are introduced which relate to good estimation of either the single parameter μ or σ or both of them. The chapter ends with a review of the case $k = 2$ categories which was presented in Ford, Torsney and Wu (1992).

Chapter 3 focuses on a one point design problem. In this case, we assume there are three or more categories. It is possible to estimate all parameters using the same cutpoints for all respondents, i.e. the same design point. The formula of the Fisher information matrix is constructed. We use search methods to find the solutions for some optimal designs. Five standard optimal criteria ($D-, A-, e_1-, e_2-$ and E- optimality) and four distributions (logistic, probit, double exponential and double reciprocal) are considered.

Chapter 4 focuses on the cell probabilities $\theta_1, \theta_2, \ldots, \theta_k$ defined by $\theta_1 = F(z_1), \theta_i = F(z_i) - F(z_{i-1}), \theta_k = 1 - \sum_{j=1}^{k-1} \theta_j$. It proposes a multiplicative algorithm for determining their optimal values and hence those of $z_1, z_2, \ldots, z_{k-1}$. These iterations neatly satisfy the constraints on $\theta_1, \theta_2, \ldots, \theta_k$ i.e. $\theta_j \geq 0$, $\sum \theta_j = 1$. Some properties of the algorithm are shown. Using the algorithm, we verify the results found by search methods and extend these to determining results for asymmetrical distributions (complementary log-log and skewed logistic) cases.

Chapter 5 extends results to the more general case of a multiple design point problem. The problem is stated and the expected information matrix is constructed. We consider two main cases: multiple point designs with constraints on cell probabilities and equal design weighting and multiple point designs with arbitrary weights and no constraints. The multiplicative algorithm is extended to determining the several sets of cell probabilities defining the different design points. Finally, the choice of the number of design points is considered.

Chapter 6 explores a new approach to CV studies, namely the bivariate approach. The motivation of this approach comes from the two stage process or the double bound approach in contingent valuation studies in which a first bid is offered to a respondent and then a lower or higher bid depending on the response to the first bid. Allowing for some change in a respondent's willingness to pay, we denote by (WTP_1) and (WTP_2) the willingness to pay of the respondent at the first and second bid respectively. This generates an extension of our problem in which we wish to find a set or sets of cutpoints in each of two dimensions. Many authors assume the bivariate normal for the joint distribution of the two $WTPs$. In our case, we extend our analysis to alternative bivariate models, namely Copula models. The first part of the chapter introduces the construction of the problem, the concept and use of copula (in particular the Plackett copula) and the formula for the Fisher information matrix. The Plackett copula is characterized by a coefficient of association denoted by ψ. The second part of the chapter focuses on investigating two main cases: two parameter models and four parameter models.

The two parameter model arises when the marginal distributions of (WTP_1), (WTP_2) are identical in their parameters. If the two marginal distributions differ in their parameters, we have a four parameter model. In each case, we derive the Fisher information matrix and use a search method to find optimal solutions. The special case when the coefficient of association $\psi = 1$, is considered in each case too.

Chapter 7 concludes with a brief review of the main findings of the thesis and a discussion of potential future work.

ACKNOWLEDGEMENTS

First of all, I would like to express my deepest gratitude to my supervisor Dr. Ben Torsney for his suggestion of the topic, his enthusiastic guidance and advice, his huge support, patience and encouragement during the course of my research. Without him, this thesis would not been successfully completed.

Secondly, I would like to thank all the staff members and postgraduate students in the Department of Statistics. My special thanks are due to Professor Adrian Bowman, Professor Marian Scott, Professor Mike Titterington, Mrs Kathleen Mosson.

Thirdly, I acknowledge the financial support from Vietnamese Government, University of Glasgow Scholarship and a U.K. Government Overseas Research Student Award (ORS).

Finally and importantly, from the bottom of my heart, my gratitude is due to my family, especially my wife, my son and my parents. They are my motivation, determination and belief from which I can complete my research.

Contents

List of Tables

List of Figures

Chapter 1

The Theory of Optimal Designs

The purpose of designing an experiment is to answer a variety of questions of interest. To do this, experimenters have to assume a model and choose necessary inputs. After running the experiment, they will observe the measurements on some variables of interests. A design is also related to determining how many observations should we take at each combination of inputs.

Normally, interest is in obtaining estimates of the parameters and using the estimated model for other purposes such as statistical inference or prediction. To obtain good estimation of the parameters, the experiment will be designed so that it optimizes a chosen criterion. The way of doing this we call optimal experiment design.

In this chapter, we will focus on introducing the general theory of optimal design for linear and non-linear models with some fundamental concepts and definitions. Then, we will mention briefly a general problem (that we call problem (P1)) and the conditions for optimality of this problem. At the end of the chapter, some optimal criteria and their properties will be introduced.

1.1 Design for a linear model

Suppose we have a linear model for N observations with k explanatory variables $x_1, x_2, \ldots, x_k$ as follows:

$$\begin{aligned} Y_i &= \theta_1 f_1(\underline{x}_i) + \theta_2 f_2(\underline{x}_i) + \ldots + \theta_m f_m(\underline{x}_i) + \epsilon_i && (1.1) \\ &= \underline{\theta}^T \underline{f}(\underline{x}_i) + \epsilon_i && (1.2) \\ &= \underline{\theta}^T \underline{v}_i + \epsilon_i, && (1.3) \end{aligned}$$

where $i = 1, 2, \ldots, N$, $\underline{v}_i = \underline{f}(\underline{x}_i)$, $\underline{f}(\underline{x}_i)$ being regression functions.

We can write the above equation in the matrix form:

$$\underline{Y} = F\underline{\theta} + \underline{\epsilon} \quad (1.4)$$

In which:

$$F = \begin{bmatrix} \underline{f}^T(\underline{x}_1) \\ \underline{f}^T(\underline{x}_2) \\ \ldots \\ \underline{f}^T(\underline{x}_N) \end{bmatrix}$$

$\underline{x} = (x_1, x_2, \ldots, x_k)^T$ is a vector of k explanatory or control variables. We assume that the values of $\underline{x}$ can be chosen by experimenters from a set $\mathcal{X}$, i.e. $\underline{x} \in \mathcal{X} \subseteq \mathbb{R}^k$, also $\underline{\theta} \in \Theta \subseteq \mathbb{R}^m$. The set $\mathcal{X}$ is called the design space and the set Θ is called the parameter space where the m-dimensional vectors of unknown parameters $\underline{\theta} = (\theta_1, \theta_2, \ldots, \theta_m)^T$ take their values.

In most applications, $\mathcal{X}$ is taken to be compact. $f_j : \mathcal{X} \to \mathbb{R}$, a continuous function from $\mathcal{X}$ into $\mathbb{R}$, $(j = 1, 2, \ldots, m)$. $\underline{\epsilon}$ is a vector of error terms, independent of $\underline{x}$. $\underline{Y}$ is a vector of response variables, $\underline{Y} \in \mathbb{R}^N$. For each $\underline{x}_i \in \mathcal{X}$, an experiment can be performed whose outcome is the observed value of a random variable Y_i, where $var(Y_i) = var(\epsilon_i) = \sigma^2$, (provided that in this particular case, we consider the error terms ϵ_i to be independent and uncorrelated with zero mean and constant variance σ^2).

The matrix F is called the design matrix. Suppose that under a design the N observations are taken at n distinct points $\underline{x}_1, \underline{x}_2, \ldots, \underline{x}_n$ in $\mathcal{X}$ $(n \geq m)$ and that n_i $(i = \overline{1,n})$ are the numbers of observations taken at the point $\underline{x}_i$, so $N = \sum_{i=1}^{n} n_i$ and $p_i = n_i/N$ are the proportions or weights of observations taken at $\underline{x}_i$. We call the set of points $Supp(p) = \{\underline{x}_i, (i = \overline{1,n})\}$ the support points of the design, denoted by p.

Definition: Design Measure.

We can present the set of support points $\underline{x}_i$ and the set of proportions p_i by the following form:

$$\xi = \left\{ \begin{array}{cccc} \underline{x}_1 & \underline{x}_2 & \cdots & \underline{x}_n \\ p_1 & p_2 & \cdots & p_n \end{array} \right\}, \tag{1.5}$$

where $\sum_{i=1}^{n} p_i = 1,\ 0 < p_i \leq 1$. Then, ξ is defined to be the design measure.

More generally, a design will be characterized by some probability measure $\xi(\underline{x})$, given on the design space $\mathcal{X}$ and satisfying the conditions:

$$\int_{\underline{x} \in \mathcal{X}} d\xi(\underline{x}) = 1,\ \ \xi(\underline{x}) > 0,\ \ \underline{x} \in \mathcal{X}. \tag{1.6}$$

From now on, we will not distinguish notationally between a design and a design measure. This does not make for confusion. If ξ is a design measure or distribution, it is by definition defined on the design space. We always will be clear about what the design space is and about which design point receives which weight.

By appropriate methods, we can estimate the values of the unknown parameters $\underline{\theta}$. For the least squares method, the parameters $\underline{\theta}$ are estimated by the following formula:

$$\underline{\hat{\theta}} = (F^T F)^{-1} F^T \underline{Y}. \tag{1.7}$$

And the variance-covariance matrix of $\underline{\hat{\theta}}$ is:

$$Cov(\underline{\hat{\theta}}) = (F^T F)^{-1} \sigma^2. \tag{1.8}$$

It is notable that this matrix and the terms following are independent of $\underline{\theta}$. The predicted value of the response variable at $\underline{x}$ is:

$$\hat{Y} = \underline{\hat{\theta}}^T \underline{f}(\underline{x}) \tag{1.9}$$

and the corresponding variance of $\hat{Y}$ is:

$$Var(\hat{Y}) = \sigma^2 (\underline{f}(\underline{x}))^T (F^T F)^{-1} \underline{f}(\underline{x}). \tag{1.10}$$

The standardized variance is:

$$Var(\hat{Y})/\sigma^2 = (\underline{f}(\underline{x}))^T (F^T F)^{-1} \underline{f}(\underline{x}). \tag{1.11}$$

The matrix $F^T F$ is called Fisher information matrix of the N observations and denoted as matrix I. It can be presented by the following form:

$$F^T F = I = \sum_{i=1}^{n} n_i \underline{f}(\underline{x}_i)(\underline{f}(\underline{x}_i))^T \tag{1.12}$$

$$= N \sum_{i=1}^{n} p_i \underline{f}(\underline{x}_i)(\underline{f}(\underline{x}_i))^T \tag{1.13}$$

$$= NM(p). \tag{1.14}$$

The matrix:

$$M(p) = \sum_{i=1}^{n} p_i \underline{f}(\underline{x}_i)(\underline{f}(\underline{x}_i))^T$$

is the expected information matrix per-observation under the design. That is:

$$M(p) = M(\underline{\theta}, \xi) = \sum_{i=1}^{n} p_i I(\underline{\theta}, \underline{x}_i), \tag{1.15}$$

where $I(\underline{\theta}, \underline{x})$ is the expected information matrix of a single observation at $\underline{x}$ and under our linear model, $I(\underline{\theta}, \underline{x}) = \underline{f}(\underline{x})(\underline{f}(\underline{x}))^T$.
Now we have:

$$Cov(\hat{\underline{\theta}}) = \frac{\sigma^2}{N} M^{-1}(p) \tag{1.16}$$

So the design problem is how to choose the support points $\underline{x}_i$ to optimize the estimation of the parameters $\underline{\theta}$. In practice, we will focus on choosing the proportion p_i of observations at $\underline{x}_i$ for good estimation of $\underline{\theta}$, that is to minimize a function of the inverse of the information matrix (to make the covariance of $\hat{\underline{\theta}}$ small), or to maximize a suitable criterion function $\phi(p) = \psi\{M(p)\}$ which is a function of the information matrix. Since $Cov(\hat{\underline{\theta}})$ is independent of $\underline{\theta}$ in linear models, the criteria are independent of the parameters $\underline{\theta}$, so we can determine optimal designs before collecting the data or carrying out the experiment. This is not the case for non-linear models as we now see. We will come back to the concept of criteria later on in this chapter.

1.2 Design for a non-linear model.

Consider a non-linear experimental design problem in which the scalar response variable y is distributed as a member of the exponential family $p(y, \eta)$. In particular, assume that model has the form:

$$E(y|\underline{x}, \underline{\theta}) \quad = \quad \eta(\underline{x}, \underline{\theta}), \tag{1.17}$$

$\eta(\underline{x}, \underline{\theta})$ is expected response. We also have the same explanation for independent variables $\underline{x}$, the parameters $\underline{\theta}$ and other notations as before.
For the exponential family of models as mentioned in equation 1.17, the Fisher information matrix for $\underline{\theta}$, given an observation at design point $\underline{x}$, is

defined to be:

$$I(\underline{\theta}, \underline{x}) \quad = \quad [a(\underline{\theta}, \underline{x})]^{-1} \underline{\eta}_\theta \underline{\eta}_\theta^T, \tag{1.18}$$

where $\underline{\eta}_\theta$ denotes the vector of partial derivatives:

$$\underline{\eta}_\theta^T \quad = \quad \left(\frac{\partial \eta}{\partial \theta_1}, \frac{\partial \eta}{\partial \theta_2}, \ldots, \frac{\partial \eta}{\partial \theta_m} \right) \tag{1.19}$$

and for the exponential family:

$$a(\underline{\theta}, \underline{x}) = var(y|\underline{x}). \tag{1.20}$$

The notation implies that these terms will in general depend on $\underline{\theta}$.

In the case of generalized linear models, the explanatory variable $\underline{x}$ and the parameter $\underline{\theta}$ appear together linearly , that is:

$$\eta = \eta(\underline{\theta}^T \underline{s}) = \eta(\mu), \quad \underline{s}^T = (1, \underline{x}^T), \quad \mu = \underline{\theta}^T \underline{s}. \tag{1.21}$$

Thus we have:

$$I(\underline{\theta}, \underline{x}) \quad = \quad w(\mu)[\underline{s}\, \underline{s}^T], \tag{1.22}$$

where:

$$w(\mu) = (\partial \eta / \partial \mu)^2 / var(y|\underline{x}). \tag{1.23}$$

The function $w(.)$ is playing the role of weight function . We assume it is measurable.

Because we may have more than one design point, it is necessary to introduce the concept of an expected information matrix. For the design with the support points $\underline{x}_i$ and the corresponding proportions p_i, the expected information matrix is defined to be:

$$M(\xi, \underline{\theta}) = \sum_{i=1}^{n} p_i I(\underline{\theta}, \underline{x}_i), \tag{1.24}$$

where ξ is the set of pairs $[\underline{x}_i, p_i]$.
We also may consider continuous design measures $\xi(.)$ satisfying $\int_{\mathcal{X}} \xi(d\underline{x}) = 1$ and $\xi(\underline{x}) \geq 0$ for which the expected per observation information matrix is

$$M(\xi, \underline{\theta}) = \int I(\underline{\theta}, \underline{x})\xi(d\underline{x}) \tag{1.25}$$

or

$$M(\xi, \underline{\theta}) = \int w(\underline{\theta}^T \underline{s})\underline{s}\,\underline{s}^T \xi(d\underline{x}). \tag{1.26}$$

Note: As we mentioned in the previous section, the main difference between linear and non-linear problems is that the choice of optimal design is straightforward in linear problem, because the information matrix is independent of $\underline{\theta}$. In non-linear design problem, on the other hand, the information matrix $M(\xi, \underline{\theta})$ or its function will depend on the parameters $\underline{\theta}$. In order to find out practical designs in this case, we need to have a prior estimate of the unknown parameters $\underline{\theta}$.
A design ξ which maximizes the function $\Psi(M(\xi, \underline{\theta}))$ for given $\underline{\theta}$ is called locally ϕ-optimal design. (See Ford, Torsney and Wu, (1992)).
Our main concern will deal with the non-linear design problem.

1.3 Approximate and exact designs

Suppose a design has N trials such that there are n_i replicates, $i = 1, \cdots, k$, at k distinct support points $\underline{x}_1, \underline{x}_2, \cdots, \underline{x}_k$. We call this an exact design. This can be represented by the notation:

$$\xi_E = \left\{ \begin{array}{cccc} \underline{x}_1 & \underline{x}_2 & \cdots & \underline{x}_k \\ n_1 & n_2 & \cdots & n_k \end{array} \right\}. \tag{1.27}$$

On the other hand, approximate designs are represented by a measure ξ_A on the design space $\mathcal{X}$. This can be expressed as follows:

$$\xi_A = \left\{ \begin{array}{cccc} \underline{x}_1 & \underline{x}_2 & \cdots & \underline{x}_k \\ p_1 & p_2 & \cdots & p_k \end{array} \right\}. \tag{1.28}$$

where:

$$p_i > 0 \quad \text{and} \quad \sum_{i=1}^{k} p_i = 1. \tag{1.29}$$

Exact and approximate designs differ in that every exact design can be expressed as an equivalent approximate design, but not every approximate design can be expressed as an exact one for a given number of runs. The equivalent approximate design ξ_R, for the above exact design can be written as follows:

$$\xi_R = \left\{ \begin{array}{cccc} \underline{x}_1 & \underline{x}_2 & \cdots & \underline{x}_k \\ \frac{n_1}{N} & \frac{n_2}{N} & \cdots & \frac{n_k}{N} \end{array} \right\}. \tag{1.30}$$

In practice, all designs have to be exact. In general, an approximate (or continuous) design ξ_C can coincide with an exact design if and only if p_i, $i = 1, 2, \cdots, k$, are rational numbers. Furthermore the exact design has $n_i = p_i N$ replicates at the corresponding support points $\underline{x}_i$, $i = 1, 2, \cdots, k$.

As is common practice, because the exact design problem is a non-trivial integer programming problem, we will focus on approximate or continuous designs.

1.4 Properties of the Fisher information matrix

We now list some basic properties of the Fisher information matrix by citing the theory given in Fedorov (1972).

<u>**Definition 1.1:**</u>

The set S is called convex if any point:

$$s = \alpha s_1 + (1-\alpha)s_2,$$

where

$$s_1 \in S, \quad s_2 \in S, \quad 0 \leq \alpha \leq 1$$

belongs to this set.

The set S^* of points:

$$s^* = \sum_{i=1}^{t} \alpha_i s_i$$

where

$$\sum_{i=1}^{t} \alpha_i = 1, \quad \alpha_i \geq 0, \quad s_i \in S \quad (i = 1, 2, \ldots, t)$$

is a convex set.

Such the set S^* is called the convex hull of the set S.

Theorem: Fedorov (1972):

Property 1: *For any design ξ, the information matrix $M(\xi, \underline{\theta})$ is a symmetric positive-semidefinite matrix.*

Property 2: *The matrix $M(\xi, \underline{\theta})$ is singular (that means $|M(\xi, \underline{\theta})| = 0$), if the support points of the design ξ contain less than m points (m is the number of unknown parameters).*

Property 3: *The family of matrices , corresponding to all possible normalized deigns is convex.*

Property 4: *For any design ξ, the matrix $M(\xi, \underline{\theta})$ can be represented in the form:*

$$M(\xi, \underline{\theta}) = \sum_{i=1}^{k} p_i w_i \underline{s}_i \underline{s}_i^T$$

where $k \leq m(m+1)/2+1, \quad 0 \leq p_i \leq 1, \quad \sum_{i=1}^{k} p_i = 1$, m is the dimension of the information matrix, k is number of support points, the set $\{\underline{s}_i : i = 1, 2, \ldots, k\}$ is the support of the design ξ, w_i is the value of the weight defined in 1.23.

We will consider choosing ξ to optimize some criteria or functions of $M(\xi)$.

<u>**Caratheodory's Theorem:**</u>

Each point s^* in the convex hull S^* of any subset S of the t-dimensional space can be represented as a convex combination of at most t+1 elements of S:

$$s^* = \sum_{i=1}^{t+1} \alpha_i s_i \tag{1.31}$$

where

$$\sum_{i=1}^{t+1} \alpha_i = 1, \ \alpha_i \geq 0, \ s_i \in S, \ i = 1, 2, \ldots, t+1$$

If s^* is a boundary point of set S^*, then α_{t+1} can be set equal to zero .

For the proofs of properties and theorems, see Fedorov (1972).

1.5 General problem (P1)

From now on we denote an approximate design ξ by p when it is defined by ξ_A in equation 1.28. Also we replace k by J. We now state a general problem (Problem (P1)) that we will deal with in the next chapters, (See Torsney and Mandal (2000)).

Problem(P1): Suppose that we choose proportions p_i to maximize some criterion $\phi(p)$ subject to the constraints $p_i \geq 0, \quad \sum_{i=1}^{J} p_i = 1, \quad i = 1, 2, \ldots, J$. The equality constraint $\sum_{i=1}^{J} p_i = 1$ renders the problem a constrained optimization one, the full constraint region being a closed bounded convex set. The above design problem is an example where $\phi(p) = \psi\{M(p)\}$. For this particular problem, $M(p)$ can be presented as follows:

$$M(p) = \sum_{j=1}^{J} p_j \underline{v}_j \underline{v}_j^T \tag{1.32}$$

where $\underline{v}_j$ are the induced support points.

In order to derive optimality conditions for this problem, we introduce here the concept of directional derivatives.

1.6 Directional derivatives

We can check for optimality (local or global) through a point to point directional derivative. These can be defined for a function $\phi(.)$ defined in a convex set.

1.6.1 Definition 1.6.1

Consider the function:

$$f(p, q, \epsilon) \quad = \quad \phi\{(1-\epsilon)p + \epsilon q\} \tag{1.33}$$

Define:

$$F_\phi\{p, q\} = \lim_{\epsilon \downarrow 0} \frac{f(p, q, \epsilon) - \phi(p)}{\epsilon} = \frac{df(p, q, \epsilon)}{d\epsilon}|\epsilon = 0^+ \tag{1.34}$$

and:

$$F_\phi^{(2)}\{p, q\} = \frac{d^2 f(p, q, \epsilon)}{d\epsilon^2}|\epsilon = 0^+ \tag{1.35}$$

Whittle (1973) call $F_\phi\{p, q\}$ the directional derivative of $\phi(p)$ at p in the direction of q. This derivative can exist even if $\phi(p)$ is not differentiable.

$$\begin{aligned} F_\phi\{p, q\} &= (q-p)^T \partial\phi/\partial p, \text{ if } \phi(p) \text{ is differentiable} \\ &= \sum_{i=1}^{J} (q_i - p_i) d_i \quad \text{where } d_i = \partial\phi/\partial p_i, \quad i = 1, 2, \ldots, J \end{aligned}$$

$$\begin{aligned} \text{Let } F_j &= F_\phi\{p, e_j\}, \quad \text{where } e_j \text{ is the } j^{th} \text{ unit vector in } R^J \\ &= d_j - \sum_{i=1}^{J} p_i d_i, \quad j = 1, 2, \ldots, J \end{aligned}$$

We call F_j a vertex directional derivative of $\phi(p)$ at p.

1.6.2 Definition 1.6.2

Consider the function:

$$g(p, m, \epsilon) \quad = \quad \phi\{p + \epsilon m\} \tag{1.36}$$

Define:

$$G_\phi\{p, m\} = \lim_{\epsilon \downarrow 0} \frac{g(p, m, \epsilon) - \phi(p)}{\epsilon} = \frac{dg(p, m, \epsilon)}{d\epsilon}|\epsilon = 0^+ \tag{1.37}$$

$G_\phi\{p, m\}$ is called *Gâteaux* derivative of $\phi(.)$ at p in the direction of m. If $m = q - p$, $F_\phi\{p, q\} = G_\phi\{p, m\}$ or $G_\phi\{p, m\} = F_\phi\{p, p + m\}$. We note that differentiability of $\phi(.)$ at p implies that G_ϕ is linear in its second argument (see Rockafellar (1970).

Whittle (1971) uses this alternative but equivalent definition of 1.6.1. Kiefer (1974) uses the concept of *Gâteaux* derivatives in his design theory though he did not call it a directional derivative. The definition 1.6.1, which allows the direction of interest to be determined by a point q, as above, is more useful and leads to a generalization of some standard calculus. The derivative $F_\phi\{p, q\}$ will serve our purpose better than $G_\phi\{p, m\}$

1.6.3 Properties of directional derivatives $F_\phi\{p, q\}$

1. Suppose that $\mathcal{S}$ is a convex set, if $p, q \in \mathcal{S}$, then so does $\{(1 - \epsilon)p + \epsilon q\}$, which is an advantage if one wishes to calculate $F_\phi\{p, q\}$ only for

$p, q \in S$. In contrast, $G_\phi\{p, m\}$ dose not particularly benefit from such convexity.

2. If $\phi(.)$ is concave, $F_\phi\{p, q\} \geq \phi(q) - \phi(p)$.

   Proof:

$$\begin{aligned} F_\phi\{p, q\} &= \lim_{\epsilon \downarrow 0} \frac{\phi\{(1-\epsilon)p + \epsilon q\} - \phi(p)}{\epsilon} \\ &\geq \lim_{\epsilon \downarrow 0} \frac{(1-\epsilon)\phi(p) + \epsilon\phi(q) - \phi(p)}{\epsilon} \\ &= \phi(q) - \phi(p) \end{aligned}$$

3. $F_\phi\{p, p\} = 0$, because no change is effected in $\phi(.)$ if one does not move from p. In contrast, $G_\phi\{p, p\} = F_\phi\{p, 2p\} \neq 0$.

4. $F_\phi\{p, q\}$ in some sense measures the rate of change in $\phi(.)$ at p in the direction of q. $F_\phi\{p, q\}$ depends on the distance between p and q and the rate of change as well.

Note that if we move from p in the direction of q, i.e. we move from p in the direction of the vector $m = q - p$. Thus if we have $c > 0$, the above movement is equivalent to the move from p in the direction of vector cm. If we pass along the full length of the vector cm from p, we will stop at $\{p + c(q - p)\}$. So $F_\phi\{p, p + c(q - p)\}$ measures the rate of change in $\phi(.)$ at p in all directions which remain the same for all $c > 0$.

We can prove above statement as follows:

$$\begin{aligned} F_\phi\{p, p+c(q-p)\} &= \lim_{\epsilon\downarrow 0} \frac{\phi\{(1-\epsilon)p+\epsilon[p+c((q-p)]\}-\phi(p)}{\epsilon} \\ &= \lim_{\epsilon\downarrow 0} \frac{\phi\{p+c\epsilon(q-p)\}-\phi(p)}{\epsilon} \\ &= \lim_{\epsilon\downarrow 0} \frac{c[\phi\{p+c\epsilon(q-p)\}-\phi(p)]}{c\epsilon} \\ &= \lim_{\epsilon\downarrow 0} \frac{c[\phi\{(1-c\epsilon)p+c\epsilon q\}-\phi(p)]}{c\epsilon} \end{aligned}$$

Hence:

$$F_\phi\{p, p+c(q-p)\} = cF_\phi\{p, q\} \tag{1.38}$$

We can denote $F_\phi\{p, q\}$ by $f'_+(0)$ where $f(\epsilon) = \phi\{(1-\epsilon)p+\epsilon q\}$. Since $f'_+(0)$ is the rate of change induced in the linear approximation to $f(.)$ at 0 by a unit increase in ϵ, it follows that $F_\phi\{p, q\}$ defines the rate of change induced in a corresponding linear approximation to $\phi(.)$ at p by a step toward q, the magnitude of which is the distance between p and q.

Thus it suggests that we should calculate $F_\phi\{p, q\}$ only for a q which is a unit distance from p. However, we face the problem that we will be presented with a q of interest which will not typically be a unit distance from p. Such distances must be standardized. We easily see that we should choose c so that $c(q-p)$ has unit length, say $c = \frac{1}{\|z\|}$, where $z = q - p$.

This gives rise the following concept.

1.6.4 Normalized directional derivatives

The normalized directional derivative is defined as follows:

$$F^I(\{p, q\}) = \frac{F_\phi\{p, q\}}{\sqrt{z^T z}} \tag{1.39}$$

This uses only one particular norm. A more general normalized directional derivative would be

$$F^A(\{p,q\}) = \frac{F_\phi\{p,q\}}{\sqrt{z^T A z}} \tag{1.40}$$

where A is a symmetric non-negative definite matrix.

1.6.5 Further properties of directional derivatives

As mentioned above, the directional derivatives can exist even if $\phi(p)$ is not differentiable. Now assume that $\phi(p)$ is differentiable. We state some other properties for directional derivatives; (see Kiefer (1959))..

1. **(FP1)**

$$G_\phi(p,m) = m^T \frac{\partial \phi}{\partial p} = m^T d \text{ where } d = \frac{\partial \phi}{\partial p} \tag{1.41}$$

and

$$\begin{aligned} F_\phi(p,q) &= G_\phi\{p,(q-p)\} \\ &= G_\phi(p,q) - G_\phi(p,p) \\ &= (q-p)^T d \end{aligned} \tag{1.42}$$

2. **(FP2)**

In the above property, if we replace m or q by a unit vector e_j, then

$$G_\phi(p,m) = e_j^T d \tag{1.43}$$

and

$$F_\phi(p,e_j) = (e_j - p)^T d = \frac{\partial \phi}{\partial p_j} - p^T d. \tag{1.44}$$

As mentioned in subsection 1.6.1, $F_\phi(p,e_j)$ is called the vertex directional derivative of $\phi(.)$.

In our calculations later on, we need to take the criterion ϕ to be functions of information matrices. We now construct the formula for directional derivatives in these cases. In general, we take the criteria ϕ in the form:

$$\phi(p) = \psi\{A(p)\}$$

where $A(p)$ is a symmetric non-negative definite matrix. In the case of design problems, $A(p)$ is the expected information matrix and can take the form:

$$A(p) = M(p) = \sum_{j=1}^{n} p_j v_j v_j^T$$

We now derive the formula for the derivatives $G_\phi(p, q)$ and $F_\phi(p, q)$. Based on the extensions of these derivatives to $\psi(A)$, we have:

$$G_\psi(A, B) = tr\left(B\frac{\partial\psi}{\partial A}\right) \tag{1.45}$$

and

$$F_\psi(A, B) = tr\left[\left(B - A\right)\frac{\partial\psi}{\partial A}\right] \tag{1.46}$$

Then:

$$G_\phi(p, q) = G_\psi\left(A(p), \sum_{i=1}^{k} q_i \frac{\partial A(p)}{\partial p_i}\right) \tag{1.47}$$

and

$$F_\phi(p, q) = F_\psi\left(A(p), A(p) + \sum_{i=1}^{k}(q_i - p_i)\frac{\partial A(p)}{\partial p_i}\right) \tag{1.48}$$

Proof of 1.47:

Allowing for a nonlinear dependence of A on p and using a first order Taylor expansion of the matrix $A(p + \epsilon q)$:

$$A(p + \epsilon q) = A(p) + \epsilon\left[\sum_{i=1}^{k} q_i \frac{\partial A(p)}{\partial p_i}\right]$$

So:

$$\begin{aligned} G_\phi\{p,q\} &= \lim_{\epsilon\downarrow 0} \frac{\phi(p+\epsilon q)-\phi(p)}{\epsilon} \\ &= \lim_{\epsilon\downarrow 0} \frac{\psi\Big(A(p+\epsilon q)\Big)-\psi\Big(A(p)\Big)}{\epsilon} \\ &= \lim_{\epsilon\downarrow 0} \frac{\psi\Big(A(p)+\epsilon\sum_{i=1}^{k} q_i \frac{\partial A(p)}{\partial p_i}\Big)-\psi\Big(A(p)\Big)}{\epsilon} \\ &= G_\psi\left(A(p), \sum_{i=1}^{k} q_i \frac{\partial A(p)}{\partial p_i}\right) \end{aligned} \tag{1.49}$$

Proof of 1.48:

$$\begin{aligned} F_\phi\{p,q\} &= G_\phi(p, q-p) \\ &= G_\psi\left(A(p), \sum_{i=1}^{k} (q_i-p_i) \frac{\partial A(p)}{\partial p_i}\right) \\ &= F_\psi\left(A(p), A(p)+\sum_{i=1}^{k} (q_i-p_i) \frac{\partial A(p)}{\partial p_i}\right). \end{aligned} \tag{1.50}$$

1.7 Conditions for local optimality

In problem (P1), for optimizing a concave criterion function like $\phi(p) = \psi\{M(p)\}$, we need some optimality conditions for checking and constructing optimal designs.

We will now state two theorems which will allow us to use an algorithm to construct and check conjectured optimal designs. Then, the general equivalence theorem will be introduced as a special case of these two theorems.

1.7.1 Theorem 1.7.1

In problem (P1), for a concave criterion function $\phi(p)$, p^* is optimal if and only if:

$$F_\phi\{p^*, q\} \leq 0 \qquad \forall q \in \mathcal{P} \tag{1.51}$$

$$i.e. \quad \max_{q \in \mathcal{P}} F_\phi\{p^*, q\} = 0, \tag{1.52}$$

where $\mathcal{P}$ is the probability simplex in J dimensions; (see Whittle (1973).

1.7.2 Theorem 1.7.2 Vertex direction optimality theorem

If $\phi(p)$ is differentiable at p^*, 1^{st} order conditions for a local maximum at p^* in the feasible region of problem (P1) are:

$$F_j^* = F_\phi\{p^*, e_j\} \begin{cases} = 0 & : \text{ if } \; p_j^* > 0 \\ \leq 0 & : \text{ if } \; p_j^* = 0, \; j = 1, 2, \ldots, J \end{cases} \tag{1.53}$$

So, p^* will minimize $\max_j F_j(p)$

1.7.3 General Equivalence Theorem

In theorem 1.7.2 above, if $\phi(p)$ is concave on its feasible region then the first order stationarity condition 1.53 is both necessary and sufficient for a solution to problem (P1). This is the General Equivalence Theorem in Optimal Design; see Whittle (1973). The theorem can be stated as follows: Suppose the criteria function is the function of an information matrix, say $\phi(p) = \psi\{M(p)\}$. The derivative of $\phi(p)$ at $M(p_1)$ in the direction of $M(p_2)$ is:

$$F_\phi\{M(p_1), M(p_2)\} = \lim_{\epsilon \to 0^+} \left[\frac{\psi\{(1-\epsilon)M(p_1) + \epsilon M(p_2)\} - \psi\{M(p_1)\}}{\epsilon} \right]$$

The General Equivalence Theorem states the equivalence of the following three conditions on p^* (and the respective design problem):
If $\psi\{M(p)\}$ is strictly concave on the set of symmetric positive definite matrices, then:

1. p^* maximizes $\psi\{M(p)\}$
2. p^* minimizes the maximum over $\underline{x} \in \mathcal{X}$ of $F_\phi\{M(p^*), I(\underline{x}, \underline{\theta})\}$, i.e. the minimum of $F_\phi\{M(p^*), I(\underline{x}, \underline{\theta})\} \leq 0$
3. The derivative $F_\phi\{M(p^*), I(\underline{x}, \underline{\theta})\}$ achieves its maximum of zero at the support points of the design with respect to p^*, say $p^*(\underline{x})$, i.e. $F_\phi\{M(p^*), I(\underline{x}, \underline{\theta})\} = 0$ if $p^*(\underline{x}) > 0$.

1.8 Criteria of optimality

We now consider examples of problem (P1) arising in optimal design. We wish to choose the proportions p_i to make the matrix $M(p)$ as large as possible. In general, we will consider various ways in which to make the matrix $M(p)$ large, namely by maximizing some real valued function $\phi(p) = \psi\{M(p)\}$.
Note that the function ϕ is called the criterion function, and in turn, the criterion defined by the function ϕ is usually called ϕ-optimality. A design maximizing $\phi(p)$ is called a ϕ-optimal design.
We now consider some of the design criteria and their properties. In general, we can divide the set of criteria into two cases. Case 1 corresponds to the case in which interest is in inference about all of the parameters $\underline{\theta}$. The information matrix $M(p)$ must be positive definite. Possible criteria in this case

include D-optimality, A-optimality, E-optimality and G-optimality. Case 2 is about the criteria when the interest of experimenter is centered on some linear combination(s) of the unknown parameter $\underline{\theta}$. Such criteria include D_A-optimality, linear optimality, c-optimality and E_A-optimality.

1.8.1 D-optimality

The most important design criterion in applications is D-optimality. A design is called D-optimal if it maximizes the value of the following functions:

$$\phi(p) = \psi\{M(p)\} = det\{M(p)\} \tag{1.54}$$

or

$$\phi(p) = \psi\{M(p)\} = log\ det\{M(p)\} \tag{1.55}$$

or

$$\phi(p) = \psi\{M(p)\} = -log\ det\{M^{-1}(p)\} \tag{1.56}$$

that means the generalized variance of the parameter estimates is minimized. We can explain the meaning of D-optimality in term of a confidence region for the vector of unknown parameters. Suppose that the model (1.1) is linear with the error terms normally distributed. Then the general form of the joint confidence region for the vector of unknown parameters $\underline{\theta} \in \Theta$ is:

$$(\underline{\theta} - \hat{\underline{\theta}})^T\ M^{-1}(p)\ (\underline{\theta} - \hat{\underline{\theta}}) \leq c, \text{ with some constant } c \tag{1.57}$$

where $\hat{\underline{\theta}}$ is the least square estimate or the maximum likelihood estimate of $\underline{\theta}$ and c is proportional to a percentage point of a distribution e.g. χ^2 distribution. This confidence region is an ellipsoid. The volume of this ellipsoid is proportional to $[det\{M(p)\}^{-1}]^{1/2}$. So a D-optimal design is a design which minimizes the volume of this ellipsoid. An advantage of D-optimality is that the optimal design for quantitative factors do not depend on the scale of the variable. That means the criterion is invariant with respect to a linear transformation of the parameters. Some of the authors who studied this criterion are: Kiefer (1959), Fedorov (1972), Silvey (1980), Pazman (1986), Farrell *et al.*(1967), Kiefer and Wolfowitz (1961), Kiefer (1961) and Atkinson and Donev (1992) (including D_s-optimality)

Properties of D-optimality:

Assuming that $\phi(p) = \psi\{M(p)\} = log\, det\{M(p)\}$, so $\phi(p)$ has following properties:

1. $\psi\{M(p)\}$ is an increasing function over the set of positive definite symmetric matrices. That is for $M_1, M_2 \in \mathcal{M}$,

$$\psi\{M_1 + M_2\} \geq \psi\{M_1\}$$

   where $\mathcal{M}$ is the set of all positive definite symmetric matrices.

2. $\psi\{M(p)\}$ is a strictly concave function on the set $\mathcal{M}$; see Fedorov(1972).

3. $\psi\{M(p)\}$ is differentiable whenever it is finite, and the first derivative has the form:

$$\frac{\partial \phi}{\partial p_j} = \underline{v}_j^T M^{-1}(p) \underline{v}_j$$

   Where $\underline{v}_j = \underline{f}(\underline{x}_j), \quad j = \overline{1,k}$.

4. D-optimal designs are invariant with respect to any non-singular linear transformation of the parameters and of design space; see Fedorov(1972).

1.8.2 A-optimality

A design is called A-optimal if it maximizes the value of the following function:

$$\phi(p) = \psi\{M(p)\} = -Trace\{M^{-1}(p)\} \tag{1.58}$$

From the above function, we see that an A-optimal design wants to minimize the sum of the variances of the estimated parameters or their average variance. However, an A-optimal design does not take correlations between these estimates into account. A-optimal designs were studied by Elfving (1952) and Chernoff (1953).

Properties of A-optimality:

1. $\psi\{M(p)\}$ is an increasing function over the set of positive definite symmetric matrices.
2. $\psi\{M(p)\}$ is concave function on the set of positive definite symmetric matrices.
3. $\psi\{M(p)\}$ is differentiable whenever it is finite, and the first derivative has the form:

$$\frac{\partial \phi}{\partial p_j} = \underline{v}_j^T M^{-2}(p) \underline{v}_j$$

1.8.3 G-optimality

A design is called G-optimal if it maximizes the value of the following function:

$$\phi(p) = \psi\{M(p)\} = \{-\max_{\underline{v}} \underline{v}^T M^{-1} \underline{v}\} \tag{1.59}$$

This criterion seeks to minimize the maximum value of $\underline{v}^T M^{-1} \underline{v}$ which is proportional to the variance of $\underline{v}^T \underline{\theta}$. Kiefer and Wolfowitz (1960) prove the equivalence of G-optimality and D-optimality.

Properties of G-optimality:

1. $\psi\{M(p)\}$ is an increasing function over the set of positive definite symmetric matrices.

2. $\psi\{M(p)\}$ is concave function on the set of positive definite symmetric matrices.

3. Suppose that there is unique

$$\underline{v}_i^T M^{-1} \underline{v}_i = -\max_t \underline{v}_t^T M^{-1} \underline{v}_t$$

then $\phi(p)$ has unique partial derivatives corresponding to positive weights:

$$\frac{\partial \phi}{\partial p_j} = [\underline{v}_j^T M^{-1}(p) \underline{v}_j]^2$$

otherwise $\phi(p)$ is not differentiable.

4. G-optimal designs are invariant with respect to any non-singular linear transformation of the parameters and of the design space.

1.8.4 E-optimality

In E-optimality, the variance of the least well-estimated contrast $\underline{a}^T \underline{\theta}$ is minimized subject to the constraint $\underline{a}^T \underline{a}=1$. So, a design is called E-optimal if it maximizes the value of the following function:

$$\phi(p) = \psi\{M(p)\} = -\lambda_{max}[M^{-1}(p)]. \tag{1.60}$$

where $\lambda_{max}[M^{-1}(p)]$ denotes the largest eigenvalue of $M^{-1}(p)$; see Kiefer (1974).

Properties of E-optimality:

1. $\psi\{M(p)\}$ is an increasing function over the set of positive definite symmetric matrices.

2. $\psi\{M(p)\}$ is concave function on the set of positive definite symmetric matrices.

3. Let $\lambda_1, \lambda_2, \ldots, \lambda_k$ denote the eigenvalues of $M(p)$. If λ_{max} is unique then $\phi(p)$ has unique partial derivatives corresponding to positive weights. Otherwise, $\phi(p)$ is not differentiable. We can present the three criteria D-,A- and E- by eigenvalues of the information matrix. Let $\lambda_1, \lambda_2, \ldots, \lambda_k$ denote the eigenvalues of the information matrix $M(p)$, we have:

   - D-optimality
   $$\max_p \prod_{i=1}^{k} \lambda_i.$$
   - A-optimality
   $$\min_p \sum_{i=1}^{k} \frac{1}{\lambda_i}.$$
   - E-optimality
   $$\min_p \max_i \left(\frac{1}{\lambda_i}\right) \qquad i = 1, \cdots, k.$$

In the case when the interest is not in all parameters of the model, we will use the criteria that take in to account a subset or a linear combination of the parameters of the model. Suppose that we are interested in s linear combinations of the parameters $\theta_1, \theta_2, \ldots, \theta_k$ which are elements of the vector $\underline{\alpha} = A\underline{\theta}$. A is an $s \times k$ matrix of rank $s \leq k$. In particular, $A = [I_s : O]$ where I_s is the $s \times s$ identity matrix and O is the $s \times (k-s)$ zero matrix. In this case interest is only in estimating the first parameters $\theta_1, \theta_2, \ldots, \theta_s$ of $\underline{\theta} \in \Theta$.

If matrix $M(p)$ is non-singular, then the variance matrix of the least squares estimator of $A\underline{\theta}$ is proportional to the matrix $AM^{-1}(p)A^T$. However, the information matrix $M(p)$ can now be singular since the basic requirement for estimating the vector $\underline{\alpha} = A\underline{\theta}$ is that the row space of A is in the range space (column space) of $M(p)$ which results in the invariance of the matrix $AM^{-}(p)A^T$ to the choice of generalized inverse matrix $M^{-}(p)$ of $M(p)$ (see Graybill (1969)).

Note that a generalized inverse of a matrix M is defined as any matrix M^{-} satisfying the condition $MM^{-}M = M$. Such a generalized inverse exists for each matrix M, but it is not unique except when M is a square non-singular matrix; in this case $M^{-} = M^{-1}$ uniquely. A particular example is when $M^{-} = M^{+}$, where M^{+} is the Moore-Penrose generalized inverse which not only satisfies $MM^{+}M = M$, but also $M^{+}MM^{+} = M^{+}$ and symmetry of $M^{+}M$ and MM^{+}. M^{+} is unique.

So, a good design will be one which makes the matrix $AM^{-}(p)A^T$ as small as possible. There are some criteria which have been proposed as follows.

1.8.5 D_A-optimality

A design is called D_A-optimal if it maximizes the value of the following function:

$$\phi(p) = \psi\{M(p)\} = -log\ det\{AM^-(p)A^T\} \tag{1.61}$$

Sibson (1974) called this criterion D_A-optimality to emphasize the dependence of the design on the matrix of coefficients A.

Properties of D_A-optimality:

1. $\psi\{M(p)\}$ is an increasing function over the set of positive definite symmetric matrices.

2. $\psi\{M(p)\}$ is a concave function on the set of positive definite symmetric matrices.

3. $\psi\{M(p)\}$ has unique partial derivatives corresponding to positive weights:

$$\frac{\partial \phi}{\partial p_j} = \underline{v}_j^T M^-(p)A^T[AM^-(p)A^T]^{-1}AM^-(p)\underline{v}_j$$

These derivatives are invariant for any generalized inverse $M^-(p)$ of $M(p)$ if $\underline{v}_j$'s and A are in the column space of $M(p)$ (see Graybill (1969)).

We now consider an important special case of D_A-optimality.

If $A = [I_s\ :\ O]$ and we can partition the matrix $M(p)$ as follows:

$$M(p) = \begin{bmatrix} M_{11}^{s\times s} & M_{12}^{s\times(k-s)} \\ M_{12}^T & M_{22}^{(k-s)\times(k-s)} \end{bmatrix}$$

then the matrix $(AM^-(p)A^T)^{-1}$ can be expressed as $(M_{11} - M_{12}M_{22}^- M_{12}^T)$ (see Rhode (1965) and Torsney (1981)) and our design criterion becomes that of choosing p to maximize the determinant of this matrix.

So maximize $\phi(p)$ in this case is equivalent to maximizing:

$$\phi(p) = log\, det\{M_{11} - M_{12}M_{22}^{-}M_{12}^{T}\}$$

which is known as the D_s-optimality. See Karlin and Studden (1996), Adwood (1969), Silvey and Titterington (1973) and Silvey (1980).

1.8.6 Linear optimality

Let L be a $k \times k$ matrix of coefficients. A design is linear optimal if it maximizes the value of the following function:

$$\phi(p) = \psi\{M(p)\} = -tr\{M^{-}(p)L\} \qquad (1.62)$$

It has its name since it is linear in the elements of the covariance matrix $M^{-}(p)$.

If L is of rank $s \leq k$, it can be expressed in the form $L = A^T A$ where A is a $s \times k$ matrix of rank s. Then the criterion function turns out to be:

$$\phi(p) = -tr\{M^{-}(p)L\} = -tr\{M^{-}(p)A^T A\} = -tr\{AM^{-}(p)A^T\} \qquad (1.63)$$

This form stresses the relationship with D_A-optimality where the determinant, rather than the trace, of $-\{AM^{-}(p)A\}$ was maximized.

Another special case arises when $A = \underline{c}^T$, where $\underline{c}$ is a $k \times 1$ vector. The criterion function can be expressed as follows:

$$\phi(p) = \psi\{M(p)\} = -\underline{c}^T M^{-}(p)\underline{c} \qquad (1.64)$$

This is the case of a criterion called c-optimality. If we let $\underline{c} = e_1 = (1,0)$ and $\underline{c} = e_2 = (0,1)$, we will have the special cases of c-optimality which we call e_1- and e_2-optimality respectively. We will use these criteria to find optimal designs later on.

Properties of linear optimality:

1. $\psi\{M(p)\}$ is an increasing function over the set of positive definite symmetric matrices.

2. $\psi\{M(p)\}$ is a concave function on the set of positive definite symmetric matrices.

3. $\psi\{M(p)\}$ has unique partial derivatives corresponding to positive weights:

$$\frac{\partial \phi}{\partial p_j} = \underline{v}_j^T M^-(p) A^T A M^-(p) \underline{v}_j, \quad p_j \geq 0$$

4. As D-optimality, the criterion function in c-optimality is invariant under non-singular linear transformation of the design variable $\underline{x}$.

1.8.7 E_A-optimality

A design is called E_A-optimal if it maximizes the value of the following function:

$$\phi(p) = \psi\{M(p)\} = -\lambda_{max}[AM^-(p)A^T] \tag{1.65}$$

λ_{max} denotes the largest eigenvalue of the matrix $AM^-(p)A^T$; see Pazman (1986).

Properties of E_A-optimality:

1. $\phi(p)$ is an increasing function over the set of positive definite symmetric matrices.

2. $\phi(p)$ is a concave function on the set of positive definite symmetric matrices.

3. The differentiability properties of this criterion are similar to those of E-optimality.

Chapter 2

Optimal Cutpoints Defined

In this chapter, we present the main focus of the problem which we will deal with in the remaining part of the thesis. Some practical contexts resulting from applying the problem will be introduced. We then construct the formal problem as a generalized linear model. Finally, a special case, namely the two category case will be reviewed.

2.1 The main idea

In social sciences, we are interested in many aspects of social life that strictly relate to human being and the environment. In order to get the information about these aspects, we normally have to carry out a survey or investigation. Suppose that we are concerned about a characteristic of a population and a survey is conducted. We denote X, on a continuous scale, as the variable of interest. In practice, however, we can not measure this variable precisely on the sample members. An alternative is that we record only to which of a finite number of categories they belong, possibly determining this by a process of elimination. Our main task is how to determine these categories optimally.

2.2 Some applications

2.2.1 Market research studies

There are many kinds of market research studies such as population income, new product or service introduction. In this kind of research, the primary concern is about the customer needs and characteristics. For instance, we want to ask potential customers how much they often spend on a particular product or what is their average income. In the view of statisticians, it will be very costly and time consuming if the way of getting information is not designed efficiently. The categorical information as described above will be recorded in a market research investigation if respondents are likely to be reluctant to be very specific or to have poor memory recall. In this case, the best way to get information from respondents is to offer them consecutive ranges of values of the response variable with these ranges chosen in advance. So, the problem arises of how to choose such the ranges optimally. This kind of design is also applied in surveying general practitioners in respect of what percentage of patients they assign to a specific drug, or to a new market expansion in which a company wants to investigate the population expenditure potential for a new product.

2.2.2 Contingent valuation studies

Contingent Valuation (CV) study is the main application of our study. The primary aim of CV study is to assess a population's willingness to pay for some ecosystem, environmental services, non-market goods or towards an increase in charge for some public services. Some examples of these are willingness to pay for a fishing permit, or for access to a country park, or for new medical facilities. In other words, a CV study is used to estimate economic values for these kinds of goods and services.

The first such study focussed on pollution in the Delaware River Basin, USA in 1947. A more recent example is seen in Hanley (1989) which reported a study into the Willingness to Pay of Visitors to a part of the Queen Elizabeth Forest Park in Central Scotland. There was interest in four aspects: wildlife, landscape, recreation and all combined. Four WTP questions were asked. For the last category this was: " Suppose the government was considering selling the Queen Elizabeth Forest Park to a private forestry company. This would mean people would no longer be able to visit it. If the only way to prevent this happening was for the Forestry Commission to raise revenue by selling day tickets to visitors, how much would you be willing to pay, per person per visit?" This kind of question is known as an open ended question. An open ended CV study involves directly asking people, in a survey, how much they would be willing to pay for specific goods or services. The CV method is referred to as a "stated preference" method, since it asks people to directly state their values, rather than inferring values from actual choices. CV study is based on what people say they would do, as opposed to what people are observed to do.

Since respondents may never have considered such questions, it is unrealistic to expect them to state a specific "willingness to pay value". There are several variations of the WTP question.

- Closed ended format (or payment card):

  The respondent is offered a list (normally on a card) of possible payments and ask to identify the one closest to his/her maximum willingness to pay. This variation was also used in the Hanley (1989). An apparent alternative is that, often used in market research studies, where we offer the respondents a consecutive range of values of a variable. We call each range of these values a category and the limits defining these ranges the cutpoints.

- Dichotomous choice format:

  The respondent is offered a single payment or "bid" question , e.g. "are you willing to pay £20?". Then the respondent simply responds YES or No depending on his/her willingness to pay this "bid". This form of asking question is also known as a Discrete choice or single bounded question.

- Double bounded format:

  After the first "bid" as in a single bounded format, the respondent would then be offered a second "bid", lower, e.g. £10, if their response to the first "bid" is NO and higher, e.g. £30, otherwise. We would then know into which of four ranges, below £10, between £10 and £20, between £20 and £30 and above £30, a respondent's willingness to pay falls. This is known as a double bounded question.

- Iterative biding:

  In this case, the respondent is offered a sequence of dichotomous choice questions, increasing or decreasing in "bid" value offered according as the response to the first question is YES or NO respectively. The process stops when the response changes or the list of the bids is exhausted. We note that the payment card method could be viewed as a variation of this. A respondent's true WTP value should lie between the circled one and the next higher value.

In all the variations of the WTP question above, we need to choose the "bids" offered. In the other words, we have to choose in advance the categories or the cutpoints. Our task is how to choose them optimally. This is the focus of this thesis.

2.3 Establishing the formal problem-A Generalized Linear Model

Based on the idea above, we now set up the formal problem using a generalized linear model and then transform it to a standardized form.

2.3.1 A Generalized Linear Model

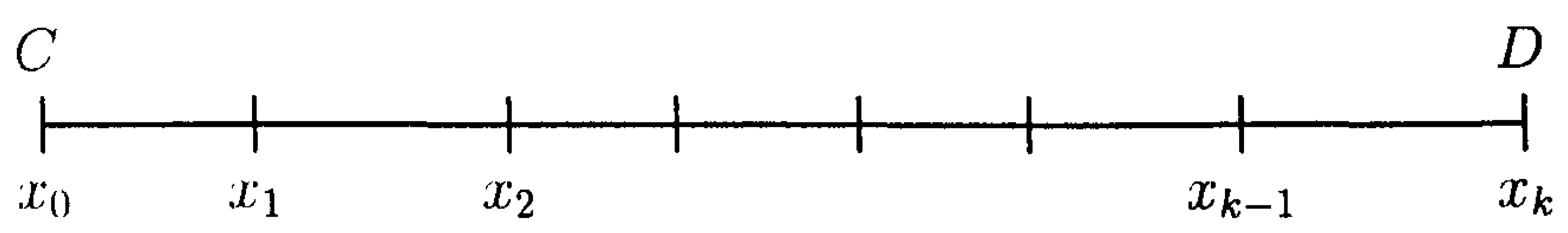
As stated above, we are interested in the characteristic of a population that we denote by variable X. Suppose we know that the variable $X \in \mathcal{X} = [C, D]$. We call the range $[C, D]$ a sample space. In order to get information about X, we carry out a survey. In the survey, we invite respondents to answer a categorical choice question. We also suppose that we wish to place responses in to one of k categories determined by cutpoints:

$$x_0, x_1, x_2, \dots, x_{k-1}, x_k$$

If we set $x_0 = C$ and $x_k = D$ of the sample space, we only need to determine the set of cutpoints $x_1, x_2, \dots, x_{k-1}$. These cutpoints have to be chosen in advance and satisfy the condition:

$$C = x_0 < x_1 < x_2 < \dots < x_{k-1} < x_k = D$$

The situation can be described by the following diagram.



The main idea of the problem is what sets of values should be chosen for these cutpoints?

This defines a non-linear regression design problem, in which the design variable is the vector:

$$\underline{x} = (x_1, x_2, \ldots, x_{k-1}).$$

The solution should depend on the underlying distribution of X in the population of interest.

We now assume that X has distribution function:

$$F(x) = P(X \leq x). \tag{2.1}$$

If we denote by μ a location parameter and by σ a scale parameter of X, both assumed unknown, we can transform $F(x)$ as follows:

$$P(X \leq x) = P\left(\frac{X-\mu}{\sigma} \leq \frac{x-\mu}{\sigma}\right) = F\left(\frac{x-\mu}{\sigma}\right), \quad x \in \mathcal{X} \tag{2.2}$$

Let:

$$Z = \frac{X-\mu}{\sigma} \text{ and } z = \frac{x-\mu}{\sigma}.$$

Hence:

$$F(z) = F\left(\frac{x-\mu}{\sigma}\right) \tag{2.3}$$

is a standardized distribution function.

If we let $\alpha = -\mu/\sigma$ and $\beta = 1/\sigma$ then equation 2.2 will turn out to be:

$$P(X \leq x) = F(\alpha + \beta x), \quad x \in \mathcal{X}. \tag{2.4}$$

The form 2.4 is a generalized linear model in the parameters α and β. For convenience, we let:

$$\underline{\gamma} = (\alpha, \beta)^T.$$

2.3.2 Standardization/Characterization

We now carry out a parameter dependent transformation which transforms the above problem to a standardized problem as follows:

Let:

$$Z = \frac{X - \mu}{\sigma} = \alpha + \beta X.$$

We discuss the fact that α, β are unknown below.

Then:

$$z = \frac{x - \mu}{\sigma} = \alpha + \beta x$$

and:

$$A = \frac{C - \mu}{\sigma} = \alpha + \beta C, \quad B = \frac{D - \mu}{\sigma} = \alpha + \beta D.$$

Thus:

$$F(x) = P(X \leq x) = P(Z \leq z) = F(z), \quad z \in \mathcal{Z} = [A, B], \tag{2.5}$$

where $[A, B]$ is the new sample space. We have Z a transformed standardized version of X.

Here is the statement of the standardized problem:

Determine cutpoints $\mathbf{z_1, z_2, \ldots, z_{k-1}}$

satisfying $\mathbf{A = z_0 < z_1 < z_2 < \ldots < z_{k-1} < z_k = B}$

We now have a design problem with design vector:

$$\underline{z} = (z_1, z_2, \ldots, z_{k-1}).$$

We know that $z_j = \frac{(x_j - \mu)}{\sigma} = \alpha + \beta x_j, \quad j = 0, 1, 2, \ldots k$. So whenever we determine optimal cutpoints z_j^*, we must transform back to the original variables x_j^*.

For the moment, we note that for non-linear models like this, optimal designs typically depend on the unknown parameters of such models. They are called locally optimal designs. Provisional estimates of parameters are needed for these to be of practical value. We will focus on construction of such designs. Ford, Torsney and Wu (1992) used this approach for the two-category case which we will review later on.

2.4 The canonical problem

As we mentioned in section 1.2, in a non-linear design problem, the information matrix depends on the parameters $\underline{\theta}$. To find the optimal design in this case, we will use the so-called canonical version of the design problem what in effect solves the design problem for all parameters $\underline{\theta}$.

Assume that we can choose the design variables from its design space $\mathcal{X}$. Let ξ be design measure. We have:

$$M_x(\underline{\theta}, \xi) = \int_{x \in \mathcal{X}} w(\underline{\theta}^T \underline{s}) \underline{s}\, \underline{s}^T \xi(d\underline{x}),$$

where

$$\underline{s}^T = (1, x).$$

The design problem usually involves seeking a design which maximizes some concave functions, say ϕ of the expected information matrix $M = M(\xi, \underline{\theta})$. The optimal design will depend on $\underline{\theta}$ since M depends on $\underline{\theta}$.

Suppose that we have a design criterion invariant under the linear transformation of the form:

$$\underline{s} \rightarrow \underline{t} = B\underline{s},$$

where B is a non-singular 2×2 matrix chosen such that its first row is $(1, 0)$ and its second row is $\underline{\theta}^T$. Thus $t_1 = 1$ and hence $\underline{t} = (1, z)$ for some z, while $t_2 = \underline{\theta}^T \underline{s}$, i.e. $\underline{\theta}^T \underline{s}$ is transformed to the last component of $\underline{t}$, and hence to z. So x is mapped to z and hence $\mathcal{X}$ is mapped to an induced design space $\mathcal{Z}$ for z.

We can see that the linear transformation from $\underline{s}$ to $\underline{t}$ and the choice of the matrix B will lead to a canonical version of the design problem, which can be solved independently of $\underline{\theta}$. We have some comments:

1. The design variable of this standardized problem is the image of x under the mapping, namely z, where $\underline{t} = (1, z)^T$.

2. The design space $\mathcal{Z}$ is the image of design space $\mathcal{X}$ under the mapping.

3. The expected information matrix of the standardized problem is:

$$M_z(\underline{\theta}, \xi) = \int_{z \in \mathcal{Z}} w(z) \underline{t}\, \underline{t}^T \xi(dz), \qquad (2.6)$$

   where $\underline{t} = (1, z)$.

The very important property of the transformation from $\mathcal{X}$ to $\mathcal{Z}$ is that the dependence of the optimal design on the true value of $\underline{\theta}$ for given design space $\mathcal{X}$ is replaced, in the transformed problem, by a design space which varies with $\underline{\theta}$. Thus, if we can solve the transformed problem for arbitrary Z, we can also solve the optimal design problem for arbitrary $\mathcal{X}$ and $\underline{\theta}$. See Ford, Torsney and Wu (1992).

We will come back to this problem later on.

2.5 Some design objectives

In chapter 1, we have mentioned two groups of criteria of optimality, the group of criteria when we want a good estimation for all parameters in the model and the remaining group is used when we need a good estimation for a subset or (linear) transformations of parameters.

Models 2.2 or 2.4 are two-parameter models (μ and σ) and our objective is good estimation of some aspects of these parameters by choosing a design which will ensure this objective. We could be interested in efficient estimation of either the parameter μ alone or the parameter σ alone or both μ and σ. Based on these objectives, we derive some criteria that we will use for constructing optimal designs.

1. Efficient estimator of μ:

We minimize $Var(\hat{\mu})$, where $\hat{\mu}$ is the estimation of μ. We have:

$$\mu = -\alpha/\beta \Rightarrow \hat{\mu} = -\hat{\alpha}/\hat{\beta}$$

$$\text{and} \quad Var(\hat{\mu}) \cong Var(\underline{c}^T \hat{\underline{\gamma}}),$$

where:

$$\underline{c} = \frac{\partial \mu}{\partial \underline{\gamma}} \propto \frac{-(1, \mu)^T}{\beta}.$$

2. Efficient estimation of σ :

We minimize $Var(\hat{\sigma})$.

$$\sigma = 1/\beta \Rightarrow \hat{\sigma} = 1/\hat{\beta}$$

$$\text{and} \quad Var(\hat{\sigma}) \cong Var(\underline{c}^T \hat{\underline{\gamma}}),$$

where:

$$\underline{c} = \frac{\partial \sigma}{\partial \underline{\gamma}} \propto \frac{-(0, 1)^T}{\beta^2}.$$

The two cases above are examples of the c-optimal criterion with the vector $c_1 = \frac{-(1,\mu)^T}{\beta}$ and $c_2 = \frac{-(0,1)^T}{\beta^2}$ respectively. We will come back to these cases later on to expand to two other criteria that we call e_1-optimality and e_2-optimality respectively.

3. Efficient estimation of both μ and σ:

We make $V = cov(\hat{\underline{\gamma}})$ "small". So we can minimize:

$$det(V)(D - \text{optimality})$$

or minimize:

$$tr(V)(A - \text{optimality})$$

or minimize:

$$Maximum\ eigenvalue \text{ of } V\ (E - \text{optimality})$$

We will return to construction of these.

2.6 Case of $k = 2$ categories

We now review the the work of Ford, Torsney and Wu (1992) on construction of optimal designs in the case of two categories. This is the simplest case where the vector $\underline{x} = x_1$ is scalar, which means there is only one cut-point and consequently we have two categories. Let $x_1 = x \in \mathcal{X} = [C, D]$.

We focus on construction of design measures ξ, because if both parameters need to be estimated, at least two support points are needed. That is we seek a distribution ξ_x on $\mathcal{X}$ which will identify the optimal proportions of observations to take at each point in $\mathcal{X}$.

Note that we are assuming that we are free to take x to be any value in $\mathcal{X} = [C, D]$, even if $\mathcal{X} = \mathbb{R}$. This can be permissible. However, we could be restricted to a subset of $\mathcal{X}$, say $[c, d]$.

We denote $M(\xi_x)$ the expected information matrix per observation. We have:

$$Cov(\underline{\hat{\gamma}}) \propto M^{-1}(\xi_x), \tag{2.7}$$

where $\underline{\hat{\gamma}} = (\hat{\alpha}, \hat{\beta})^T$ as denoted above.

If the distribution ξ_x assigns weight ξ_i to a discrete set of values $x_1, x_2, \ldots$ and $\xi_i \geq 0$, $\sum \xi_i = 1$, then:

$$M(\xi_x) = E_{\xi_x}(I_x) = \sum \xi_i I_{x_i} \tag{2.8}$$

where I_x is the expected information matrix of a single observation at x or a one point design at x.

From the formula 1.14 and 1.15 in chapter 1, we have:

$$I(\underline{\theta}, \underline{x}) \quad = \quad w(z)[\underline{s}\,\underline{s}^T], \quad z = \underline{\theta}^T \underline{s}$$

In our case, let $\underline{s} = (1,\ x)^T$ and $\underline{\theta}^T = (\alpha, \beta)$, we have:

$$I_x = w(z) \begin{pmatrix} 1 \\ x \end{pmatrix} (1\ \ x)$$

where the function $w(.)$ is playing the role of weight function. We assume it is measurable. It has the form:

$$w(z) = \frac{\{f(z)\}^2}{\{F(z)[1 - F(z)]\}}, \quad f(z) = F'(z) \text{ and } z = \alpha + \beta x$$

We are now considering a standardized problem under the parameter dependent transformation:

$$\begin{pmatrix} 1 \\ z \end{pmatrix} = B \begin{pmatrix} 1 \\ x \end{pmatrix} \tag{2.9}$$

$$B = \begin{pmatrix} 1 & 0 \\ \alpha & \beta \end{pmatrix}$$

So, we have:

$$I_x = w(z)\,(B^{-1}) \begin{pmatrix} 1 \\ z \end{pmatrix} (1 \;\; z)(B^{-1})^T \tag{2.10}$$

Hence

$$I_x = B^{-1}\,I_z\,(B^{-1})^T \tag{2.11}$$

Where

$$I_z = w(z)\,(1, z)^T\,(1, z) \tag{2.12}$$

Extending these results to the expected information matrix per observation, we have:

$$M(\xi_X) = B^{-1}M(\xi_Z)(B^{-1})^T \tag{2.13}$$

where ξ_Z is the distribution induced on $Z = [A, B]$ by ξ_X on $X = [C, D]$. Hence we have:

$$M(\xi_Z) = E_{\xi_z}\{I_z\} = \sum \xi_i I_{z_i}$$

and

$$det\{M(\xi_X)\} \propto det\{M(\xi_Z)\}$$

$$\underline{c}^T M(\xi_x)\underline{c} = \underline{c}_B^T M(\xi_z)\underline{c}_B$$

$$\underline{c}_B = B\underline{c}, \quad B = \begin{pmatrix} 1 & 0 \\ \alpha & \beta \end{pmatrix}$$

Thus D-optimal and c-optimal criteria, as functions of ξ_X, transform respectively to the D-optimal and other c-optimal criteria as functions of ξ_Z. Thus, we focus on finding the design ξ_Z which either

$$maximizes\ det[M(\xi_Z)] \Rightarrow D - optimality$$

or

$$minimizes\ \underline{c}_B^T M^{-1}(\xi_Z)\underline{c}_B \Rightarrow c - optimality$$

We consider two cases related to the previous section:

- If $\underset{\sim}{c} = \frac{-(1,\mu)^T}{\beta} \Rightarrow \underset{\sim}{c}_B = B\underset{\sim}{c} = (\frac{-1}{\beta}, 0)^T$

So, we minimize $\underset{\sim}{c}_B^T M^{-1}(\xi_Z)\underset{\sim}{c}_B$ which is equivalent to minimizing $(1,0)M^{-1}(\xi_z)(1,0)^T$ i.e. $\underset{\sim}{c}_B \propto \underset{\sim}{e}_1 = (1,0)^T$.

- If $\underset{\sim}{c} = \frac{-(0,1)^T}{\beta^2} \Rightarrow \underset{\sim}{c}_B \propto \underset{\sim}{e}_2 = (0,1)^T$

We define two other criteria as follows:

1. **e_1-optimality:** A design is called e_1-optimal if it maximizes the value of the function:

$$-\underset{\sim}{e}_1^T M^{-1}(\xi_Z)\underset{\sim}{e}_1.$$

2. **e_2-optimality:** A design is called e_2-optimal if it maximizes the value of the function:

$$-\underset{\sim}{e}_2^T M^{-1}(\xi_Z)\underset{\sim}{e}_2,$$

where $\underset{\sim}{e}_1 = (1,0)^T$ and $\underset{\sim}{e}_2 = (0,1)^T$ respectively.

Note: e_1-optimality and e_2-optimality above actually are the transformed c-optimal criteria.

Ford, Torsney and Wu (1992) exploited the fact that these (non-linear) design problems are equivalent to corresponding weighted linear design problems with weight function w(z). Tools for constructing designs for linear models can be invoked. For example there are geometrical characterizations of D-optimality and c-optimality relating to the design locus:

$$G = \{(g_1, g_2) : g_1 = \sqrt{w(z)}, g_2 = z\sqrt{w(z)}, z \in Z = [A, B]\}. \tag{2.14}$$

They established, that, for several choices of $F(z)$, D-optimal designs need to take observations at only two distinct points (support points) in Z, in which case optimal weights are $(\frac{1}{2}, \frac{1}{2})$. Optimal designs are then of the form:

z	z_1	z_2
ξ_Z	$\frac{1}{2}$	$\frac{1}{2}$

For the *logistic* and *normal/probit* choices of $F(z)$, (for which $Z = \Re$), $z_1 = -z$ and $z_2 = z$, with $z = 1.543$ and $z = 1.138$ respectively. These two values are well established in the literature.

For the cases where the distribution of $F(z)$ are the ***double-exponential*** and *double-reciprocal* distribution functions (for which $Z = \mathbb{R}$), three support points are needed and optimal designs are of the form:

z	z_1	z_2	z_3
ξ_Z	ξ_{z_1}	ξ_{z_2}	ξ_{z_3}

In these cases, optimal weights are not uniform. Torsney and Murasti(1993) report the following optimal designs:

Double-exponential:

z	-1.594	0	1.594
ξ_Z	0.282	0.436	0.282

Double-reciprocal:

z	$-\sqrt{2}$	0	$\sqrt{2}$
ξ_Z	0.262	0.476	0.262

In the case of c-optimality, either one or two support points are needed. If only one is needed it is the value z such that $\underline{c}_B \propto \sqrt{w(z)}(1, z)^T$. If two are needed these are fixed for all $\underline{c}_B$ and there is an explicit solution for the optimal weights which vary with $\underline{c}_B$. Thus optimal designs are of the form:

z	z_1	z_2
ξ_Z	ξ_{z_1}	ξ_{z_2}

For the *logistic* and *normal/probit* choices of $F(z)$, (for which $Z = \Re$), $z_1 = -z$ and $z_2 = z$, with $z = 2.339$ and $z = 1.157$ respectively. These are the support points for e_2-optimality, each having equal weighting. For e_1-optimality, there is one support point; namely $z = 0$.

We now focus on investigating the case of optimal designs when there are at least 3 categories (at least 2 cut-points).

Chapter 3

One Point Design: k Categories

3.1 Introduction

In the two-category case (only one cutpoint) described above, to ensure the estimation of both parameters in the model, we need at least two support points. That is why we can not use the same cutpoints for all respondents. We have to use at least two cutpoints and the respondents will be divided in to the same number of groups as cutpoints according to optimal weights. Thus the problem is to determine these cutpoints and their optimal design weights. This is the case of multiple design points that we will consider in a later chapter. In the context of one design point, we assume that there are three or more categories. The main difference from the two-category case is that it is possible to estimate all parameters using the same cutpoints for all respondents, i.e. the same design point. This is a one point design.

In general, we assume that there are k categories, so there are $k-1$ cutpoints, (actually, there will be $k+1$ cutpoints but we assume that the first and the last cutpoint will be the lower limit and upper limit of the design space). Let the cutpoints be $x_1, x_2, \ldots, x_{k-1}$ and $x_0 = C, x_k = D$ with $[C, D]$ being

the design space. The condition for the cutpoints is $x_1 \leq x_2 \leq x_3 \ldots \leq x_{k-1}$
Let:

$$\theta_1 = P(X \leq x_1). \tag{3.1}$$

θ_1 is the probability that the variable of interest X falls in to the category $[x_0, x_1]$. We call θ_1 a cell probability.
Through the standardization, we have:

$$\theta_1 = P(X \leq x_1) = F(\alpha + \beta x_1) = F(z_1), \tag{3.2}$$

where α and β are the parameters of the generalized linear model as described in chapter two.
Similarly:

$$\theta_i = P(x_{i-1} \leq X \leq x_i) = F(z_i) - F(z_{i-1}) \quad i = 2, 3, \ldots, k-1. \tag{3.3}$$

Finally, let:

$$\theta_k = 1 - F(z_{k-1}). \tag{3.4}$$

$(\theta_1, \theta_2, \ldots, \theta_k)$ is an exhaustive set of cell probabilities in that $\sum_{j=1}^{k} \theta_j = 1$.
Now, our problem turns out to be determining the set of cutpoints $z_1, z_2, \ldots, z_{k-1}$ or the set of cell probabilities $(\theta_1, \theta_2, \ldots, \theta_k)$ optimally. Because they are invertible, we need only determine either one of the sets. We now construct the formula for information matrix of our problem.

3.2 The formula of the Fisher information matrix

In our problem, we need to place the response from the respondents into the categories or between two cutpoints. Thus, our model is a multinomial response one, for which we now construct the formula for the Fisher information matrix.

Let θ_i be the probability that a response falls between cutpoints x_{i-1} and x_i, i.e.

$$\theta_i = P(x_{i-1} \leq X \leq x_i), \quad i = 1, 2, \ldots, k.$$

Denote:

$$\underline{\theta} = (\theta_1, \theta_2, \ldots, \theta_k)^T, \quad \underline{\theta} = \underline{\theta}(\underline{\gamma}), \quad \sum_{i=1}^{k} \theta_i = 1.$$

In our case, $\underline{\gamma} = (\gamma_1, \gamma_2)^T = (\alpha, \beta)^T$.

Now let:

$$\underline{Y} = (Y_1, Y_2, \ldots, Y_k)^T,$$

where:

$$Y_i = \begin{cases} 1 & : \text{ if } \; x_{i-1} \leq X \leq x_i \\ 0 & : \text{ if } \; \text{otherwise.} \end{cases}$$

Then:

$$\underline{Y} \sim \mathcal{M}(1, \underline{\theta}), \tag{3.5}$$

$$E(Y_i) = \theta_i, \quad E(\underline{Y}) = \underline{\theta}, \quad Cov(\underline{Y}) = D_{\underline{\theta}} - \underline{\theta}\,\underline{\theta}^T,$$

where:

$$D_{\underline{\theta}} = diag(\theta_1, \theta_2, \ldots, \theta_k) = \begin{pmatrix} \theta_1 & 0 & \ldots & 0 \\ 0 & \theta_2 & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & \theta_k \end{pmatrix}.$$

The likelihood function is:

$$L = \prod_{i=1}^{k} \theta_i^{Y_i}.$$

The log-likelihood function is:

$$Ln\,L = \sum_{i=1}^{k} Y_i\, ln(\theta_i).$$

We know that one general form for the Fisher information matrix is:

$$I_X(\underline{\gamma}) = Cov\left(\frac{\partial lnL}{\partial \underline{\gamma}}\right). \tag{3.6}$$

First derivatives are:

$$\frac{\partial lnL}{\partial \gamma_j} = \sum_{i=1}^{k} Y_i \frac{\partial \theta_i / \partial \gamma_j}{\theta_i}, \quad j = 1, 2.$$

Let vector $\underline{a}_j$ be such that:

$$(\underline{a}_j)_i = \frac{\partial \theta_i / \partial \gamma_j}{\theta_i}.$$

We have:

$$\frac{\partial lnL}{\partial \gamma_j} = \underline{a}_j^T \underline{Y}.$$

So for $\underline{\gamma} = (\gamma_1, \gamma_2)^T = (\alpha, \beta)^T$:

$$\frac{\partial lnL}{\partial \underline{\gamma}} = (\underline{a}_1\ \underline{a}_2)^T \underline{Y} = A \underline{Y}$$

where $A = (\underline{a}_1\ \underline{a}_2)$, i.e.,

$$A_{ji} = \frac{\partial \theta_i}{\partial \gamma_j} \frac{1}{\theta_i}$$

We can express matrix A in the form:

$$A = \frac{\partial \underline{\theta}}{\partial \underline{\gamma}} D_{\underline{\theta}}^{-1} = E\, D_{\underline{\theta}}^{-1}$$

where

$$E_{ji} = \frac{\partial \theta_i}{\partial \gamma_j}$$

Note that

$$E \times \underline{1} = \frac{\partial \underline{\theta}}{\partial \underline{\gamma}} \times \underline{1} = \sum_{i=1}^{k} \frac{\partial \theta_i}{\partial \underline{\gamma}} = \underline{0}$$

since

$$\sum_{i=1}^{k} \frac{\partial \theta_i}{\partial \underline{\gamma}} = \frac{\partial \sum_{i=1}^{k} \theta_i}{\partial \underline{\gamma}} = \frac{\partial 1}{\partial \underline{\gamma}} = \underline{0}$$

and

$$D_{\underline{\theta}}^{-1}\,\underline{\theta} = \underline{1}$$

So we have

$$\begin{aligned} I_X(\underline{\gamma}) &= Cov\left(\frac{\partial lnL}{\partial \underline{\gamma}}\right) \\ &= Cov(A\underline{Y}) \\ &= ACov(\underline{Y})A^T \\ &= A(D_{\underline{\theta}} - \underline{\theta}\,\underline{\theta}^T)A^T \\ &= ED_{\underline{\theta}}^{-1}(D_{\underline{\theta}} - \underline{\theta}\,\underline{\theta}^T)D_{\underline{\theta}}^{-1}E^T \\ &= ED_{\underline{\theta}}^{-1}D_{\underline{\theta}}D_{\underline{\theta}}^{-1}E^T - ED_{\underline{\theta}}^{-1}\underline{\theta}\,\underline{\theta}^T D_{\underline{\theta}}^{-1}E^T \\ &= ED_{\underline{\theta}}^{-1}E^T - E\underline{1}\,\underline{1}^T E^T \\ &= ED_{\underline{\theta}}^{-1}E^T - \underline{0} \end{aligned}$$

So the formula for the Fisher information matrix is:

$$I_X(\underline{\gamma}) = \sum_{i=1}^{k} \frac{\partial \theta_i}{\partial \underline{\gamma}} \frac{1}{\theta_i} \left(\frac{\partial \theta_i}{\partial \underline{\gamma}}\right)^T \qquad (3.7)$$

We can develop this formula in more detail as follows:

$$\begin{aligned} \theta_i &= P(x_{i-1} \leq X \leq x_i), \quad i = 1, 2, \ldots, k \\ &= F(\alpha + \beta x_i) - F(\alpha + \beta x_{i-1}), \quad \underline{\gamma} = (\alpha, \beta) \end{aligned}$$

$$\begin{aligned} \frac{\partial \theta_i}{\partial \alpha} &= f(\alpha + \beta x_i) - f(\alpha + \beta x_{i-1}), \quad f(.) = F'(.) && (3.8) \\ &= -f(z_{i-1}) + f(z_i), \quad z_i = \alpha + \beta x_i && (3.9) \end{aligned}$$

$$\begin{aligned} \frac{\partial \theta_i}{\partial \beta} &= x_i f(\alpha + \beta x_i) - x_{i-1} f(\alpha + \beta x_{i-1}) && (3.10) \\ &= -x_{i-1} f(z_{i-1}) + x_i f(z_i) && (3.11) \end{aligned}$$

So:

$$\begin{aligned}
\frac{\partial \theta_i}{\partial \underline{\gamma}} &= \begin{pmatrix} 1 & 1 \\ x_{i-1} & x_i \end{pmatrix} \begin{pmatrix} -f(z_{i-1}) \\ f(z_i) \end{pmatrix} \\
&= \begin{pmatrix} 1 & 1 \\ x_{i-1} & x_i \end{pmatrix} \begin{pmatrix} f(z_{i-1}) & 0 \\ 0 & f(z_i) \end{pmatrix} \begin{pmatrix} -1 \\ 1 \end{pmatrix} \\
&= XD_f(0\ 0\ \ldots -1\ 1\ \ldots 0\ 0)^T \\
&= XD_f d_i
\end{aligned}$$

where:

$$D_f = diag\{f(z_1), f(z_2), \ldots, f(z_{i-1}), f(z_i), \ldots, f(z_{k-1})\}$$

with $f(z_i)$ being the pdf function of Z at $z = z_i$, $f(z_i) = F'(z_i)$.

or:

$$D_f = \begin{pmatrix} f(z_1) & \ldots & 0 & 0 & \ldots & 0 \\ \vdots & \ddots & \vdots & \vdots & & \vdots \\ 0 & \ldots & f(z_{i-1}) & 0 & & 0 \\ 0 & & 0 & f(z_i) & \ldots & 0 \\ \vdots & & \vdots & \vdots & \ddots & \vdots \\ 0 & \ldots & 0 & 0 & \ldots & f(z_{k-1}) \end{pmatrix}$$

$$X = \begin{pmatrix} 1 & 1 & \ldots & 1 & 1 & \ldots & 1 \\ x_1 & x_2 & \ldots & x_{i-1} & x_i & \ldots & x_{k-1} \end{pmatrix}$$

$$d_i = (0\ 0\ \ldots \underbrace{-1}_{(i-1)th}\ \underbrace{1}_{i-th}\ \ldots 0\ 0)^T$$

In general, we have:

$$\begin{aligned}
I_X(\underline{\gamma}) &= \sum_{i=1}^{k} \frac{\partial \theta_i}{\partial \underline{\gamma}} \frac{1}{\theta_i} \left(\frac{\partial \theta_i}{\partial \underline{\gamma}}\right)^T \\
&= XD_f \left(\sum_{i=1}^{k} d_i \frac{1}{\theta_i} d_i^T\right) D_f X^T \\
&= XD_f H D_\theta^{-1} H^T D_f X^T \\
&= XQX^T
\end{aligned}$$

where:

$$Q = D_f H D_\theta^{-1} H^T D_f \tag{3.12}$$

$$D_\theta = diag(\theta_1, \theta_2, \dots, \theta_k) = \begin{pmatrix} \theta_1 & 0 & \dots & 0 \\ 0 & \theta_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \theta_k \end{pmatrix}$$

$$H = (I_{k-1} | \underline{0}_{k-1}) - (\underline{0}_{k-1} | I_{k-1})$$

$\underline{0}_n = (0, 0, \dots, 0)^T \in \mathcal{R}^n$ and I_n is identity matrix of order n.

or:

$$H = \begin{pmatrix} 1 & -1 & 0 & 0 & 0 & \dots & 0 \\ 0 & 1 & -1 & 0 & 0 & \dots & 0 \\ 0 & 0 & 1 & -1 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \dots & \vdots \\ 0 & 0 & 0 & 0 & \dots & 1 & -1 \end{pmatrix}$$

Further, through the standardization $z_i = \alpha + \beta x_i$,

$$I_X = B^{-1} I_Z (B^{-1})^T \tag{3.13}$$

where:

$$B = \begin{pmatrix} 1 & 0 \\ \alpha & \beta \end{pmatrix}$$

The Fisher information matrix at $\underline{z} = (z_1, z_2, \dots, z_{k-1})^T$ is:

$$I_Z = Z Q Z^T \tag{3.14}$$

where:

$$Z = \begin{pmatrix} 1 & 1 & \cdots & 1 \\ z_1 & z_2 & \cdots & z_{k-1} \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ \alpha & \beta \end{pmatrix} X = BX$$

Note: Z is a $2 \times (k-1)$ matrix, D_f is a $(k-1) \times (k-1)$ matrix, D_θ is a $k \times k$ matrix and H is a $(k-1) \times k$ matrix.

We also note that I_Z is a standardized Fisher Information matrix (corresponding to $\alpha = 0, \beta = 1$). We will show later on that for some standardized criteria, optimizing a criterion of I_X is equivalent to optimizing a criterion of I_Z.

3.3 Search method and numerical results

We use simple search methods to find the solutions for some optimal designs. Five standard optimal criteria and four symmetric distributions will be considered.

3.3.1 Criteria and distributions considered

The five criteria considered are:

1. D-optimality : $maximize\{log\, det(I_z)\}$
2. A-optimality : $maximize\{-tr\,(I_z^{-1})\}$
3. $e1$-optimality : $maximize\{-e_1^T\, I_z^{-1}\, e_1\}$
4. $e2$-optimality : $maximize\{-e_2^T\, I_z^{-1}\, e_2\}$
5. E-optimality : $maximize\{\,-\lambda_{max}\}$
   where λ_{max} is maximum eigenvalue of I_z^{-1}

We can see that D-, A- and E-optimality will be used when we are concerned about optimally estimating both parameters α and β (or μ and σ originally). On the other hand, e_1- and e_2-optimality will be used when we want to optimally estimate α or β respectively.

The reason we first use symmetric distributions is that we can argue that there should be symmetry in the optimal cutpoints. Let $\underline{z}^*$ be the vector of optimal cutpoints. We should have $\underline{z}^*$ for up to 6 categories as follows:

$$\begin{aligned}
\underline{z}^* &= (-z^*, z^*) & k = 3 \\
\underline{z}^* &= (-z^*, 0, z^*) & k = 4 \\
\underline{z}^* &= (-z_2^*, -z_1^*, z_1^*, z_2^*) & k = 5 \\
\underline{z}^* &= (-z_2^*, -z_1^*, 0, z_1^*, z_2^*) & k = 6
\end{aligned}$$

Determining $\underline{z}^*$ then reduces to a one or two variable maximization i.e. of $\phi(z) = \psi(I_{\underline{z}})$ or $\phi(z_1, z_2) = \psi(I_{\underline{z}})$ where $\psi(.)$ is our design criterion.

The four symmetric distributions are listed in table 3.1

Table 3.1: Four symmetric distributions considered

Case	Distribution	$f_i(z)$	$F_i(z)$
1	Logit	$exp(-z)/[1 + exp(-z)]^2$	$[1 + exp(-z)]^{-1}$
2	Probit	$\frac{1}{\sqrt{(2\pi)}} exp(-z^2/2)$	$\Phi(z)$
3	Double exponential	$\frac{1}{2} exp(-\|z\|)$	$\frac{(1+s)}{2} - \frac{s}{2} exp(-\|z\|)$
4	Double reciprocal	$\frac{1}{2}(1 + \|z\|)^{-2}$	$\frac{(1+s)}{2} - \frac{s}{2}(1 + \|z\|)^{-1}$

$Z \in (-\infty, \infty), f_i(z) = F_i'(z), s = sign(z)$

We simply evaluate the criterion at a set of values of z or (z_1, z_2) (defined by a grid) and determine the maximum by inspection (helped by plotting).

3.3.2 Numerical results

Table 3.2: Numerical results for logistic distribution, k=3

	k=3		
Criterion	z^*	$F(z^*)$	$\phi(z^*)$
D-optimality	1.4700	0.8131	-1.5567
A-optimality	1.1600	0.7613	-5.0182
e_1-optimality	0.6900	0.6660	-3.3750
e_2-optimality	2.1700	0.8975	-1.0226
E-optimality	0.6900	0.6660	-3.3750

Table 3.3: Numerical results for logistic distribution, k=4

	k=4		
Criterion	z^*	$F(z^*)$	$\phi(z^*)$
D-optimality	1.9800	0.8787	-1.2483
A-optimality	1.7100	0.8468	-4.3789
e_1-optimality	1.1000	0.7503	-3.2000
e_2-optimality	2.1700	0.8975	-1.0226
E-optimality	1.1000	0.7503	-3.2000

Table 3.4: Numerical results for logistic distribution, k=5

	k=5				
Criterion	z_1^*	z_2^*	$F(z_1^*)$	$F(z_2^*)$	$\phi(z_1^*, z_2^*)$
D-optimality	0.8500	2.5100	0.7006	0.9248	-1.0709
A-optimality	0.6100	2.1600	0.6479	0.8966	-4.1245
e_1-optimality	0.4100	1.3900	0.6011	0.8006	-3.1251
e_2-optimality	1.5900	3.1700	0.8306	0.9597	-0.8284
E-optimality	0.4100	1.3900	0.6011	0.8006	-3.1251

Table 3.5: Numerical results for logistic distribution, k=6

	k=6				
Criterion	z_1^*	z_2^*	$F(z_1^*)$	$F(z_2^*)$	$\phi(z_1^*, z_2^*)$
D-optimality	1.3300	2.9100	0.7908	0.9483	-0.9788
A-optimality	1.0500	2.5400	0.7408	0.9269	-3.9942
e_1-optimality	0.6900	1.6100	0.6660	0.8334	-3.0857
e_2-optimality	1.5900	3.1700	0.8306	0.9597	-0.8284
E-optimality	0.6900	1.6100	0.6660	0.8334	-3.0857

Table 3.6: Numerical results for normal/probit distribution, k=3

	k=3		
Criterion	z^*	$F(z^*)$	$\phi(z^*)$
D-optimality	1.1100	0.8665	-0.2070
A-optimality	1.0300	0.8485	-2.2784
e_1-optimality	0.6100	0.7291	-1.2348
e_2-optimality	1.4800	0.9306	-0.7666
E-optimality	0.6100	0.7291	-1.2348

Table 3.7: Numerical results for normal/probit distribution, k=4

	k=4		
Criterion	z^*	$F(z^*)$	$\phi(z^*)$
D-optimality	1.3900	0.9177	0.1001
A-optimality	1.3400	0.9099	-1.9426
e_1-optimality	0.9800	0.8365	-1.1331
e_2-optimality	1.4800	0.9306	-0.7666
E-optimality	0.9800	0.8365	-1.1331

Table 3.8: Numerical results for normal/probit distribution, k=5

	k=5				
Criterion	z_1^*	z_2^*	$F(z_1^*)$	$F(z_2^*)$	$\phi(z_1^*, z_2^*)$
D-optimality	0.6900	1.7000	0.7549	0.9554	0.3113
A-optimality	0.6000	1.6200	0.7257	0.9474	-1.7746
e_1-optimality	0.3800	1.2400	0.6480	0.8925	-1.0869
e_2-optimality	1.1400	2.0000	0.8729	0.9772	-0.6065
E-optimality	0.3800	1.2400	0.6480	0.8925	-1.0869

Table 3.9: Numerical results for normal/probit distribution, k=6

	k=6				
Criterion	z_1^*	z_2^*	$F(z_1^*)$	$F(z_2^*)$	$\phi(z_1^*, z_2^*)$
D-optimality	1.0000	1.8800	0.8413	0.9699	0.4130
A-optimality	0.9300	1.8200	0.8238	0.9656	-1.6923
e_1-optimality	0.6600	1.4500	0.7454	0.9265	-1.0615
e_2-optimality	1.1400	2.0000	0.8729	0.9772	-0.6065
E-optimality	0.6600	1.4500	0.7454	0.9265	-1.0615

Table 3.10: Numerical results for double exponential distribution, k=3

	k=3		
Criterion	z^*	$F(z^*)$	$\phi(z^*)$
D-optimality	0.6400	0.7374	-1.1277
A-optimality	0.6700	0.7452	-3.9972
e_1-optimality	0.0000 (*)	0.5000	-1.0000
e_2-optimality	1.600	0.8985	-1.5441
E-optimality	0.7100	0.7542	-2.0512

(*): *$\phi(z^*)$ reaches its maximum value when $z^* = 0$. Thus, the three category case reduces to two category case*

Table 3.11: Numerical results for double exponential distribution, k=4

	k=4		
Criterion	z^*	$F(z^*)$	$\phi(z^*)$
D-optimality	1.5900	0.8984	-0.4345
A-optimality	1.5900	0.8982	-2.5441
e_1-optimality	any z (*)		-1.0000
e_2-optimality	1.6000	0.8985	-1.5441
E-optimality	1.6000	0.8985	-1.5441

(*): *For any value of z, the criterion = -1*

Table 3.12: Numerical results for double exponential distribution, k=5

	k=5				
Criterion	z_1^*	z_2^*	$F(z_1^*)$	$F(z_2^*)$	$\phi(z_1^*, z_2^*)$
D-optimality	0.0000 (*)	1.5940	0.5000	0.8984	-0.4345
A-optimality	0.0000 (*)	1.5920	0.5000	0.8982	-2.5441
c_1-optimality	0.0000 (*)	any z	0.5000		-1.0000
c_2-optimality	1.0200	2.6100	0.8189	0.9632	-1.2191
E-optimality	0.3000	1.8900	0.6296	0.9245	-1.3568

(*): *$\phi(z_1^*, z_2^*)$ reaches its maximum value when $z_1^* = 0$. Thus, the five category case reduces to four category case*

Table 3.13: Numerical results for double exponential distribution, k=6

	k=6				
Criterion	z_1^*	z_2^*	$F(z_1^*)$	$F(z_2^*)$	$\phi(z_1^*, z_2^*)$
D-optimality	1.0200	2.6100	0.8197	0.9632	-0.1981
A-optimality	1.0200	2.6100	0.8190	0.9634	-1.2192
c_1-optimality	any z (*)	any z (*)			-1.0000
c_2-optimality	1.0200	2.6100	0.8190	0.9634	-1.2192
E-optimality	1.0200	2.6100	0.8190	0.9634	-1.2192

(*): *For any value of z_1^* and z_2^* the criteria = -1.*

Table 3.14: Numerical results for double reciprocal distribution, k=3

	k=3		
Criterion	z^*	$F(z^*)$	$\phi(z^*)$
D-optimality	0.2500	0.6000	-2.0722
A-optimality	0.3850	0.6390	-7.6492
e_1-optimality	0.0000 (*)	0.5000	-1.0000
e_2-optimality	1.0000	0.7500	-4.0000
E-optimality	0.6100	0.6894	-4.2493

(*): *$\phi(z*)$ reaches its maximum value when $z^* = 0$. Thus, the three category case reduces to two category case*

Table 3.15: Numerical results for double reciprocal distribution, k=4

	k=4		
Criterion	z^*	$F(z^*)$	$\phi(z^*)$
D-optimality	1.0000	0.7500	-1.1642
A-optimality	1.0000	0.7500	-4.8000
e_1-optimality	1.0000	0.7500	-0.8000
e_2-optimality	1.0000	0.7500	-4.0000
E-optimality	1.0000	0.7500	-4.0000

Table 3.16: Numerical results for double reciprocal distribution, k=5

	k=5				
Criterion	z_1^*	z_2^*	$F(z_1^*)$	$F(z_2^*)$	$\phi(z_1^*, z_2^*)$
D-optimality	0.0000 (*)	1.0000	0.5000	0.7500	-1.1642
A-optimality	0.0700	1.1400	0.5327	0.7664	-4.7506
c_1-optimality	0.0000 (*)	1.0000	0.5000	0.7500	-0.8000
c_2-optimality	0.5000	2.0000	0.6667	0.8333	-3.3750
E-optimality	0.5000	2.0000	0.6667	0.8333	-3.3750

(*): *$\phi(z_1^*, z_2^*)$ reaches its maximum value when $z_1^* = 0$. Thus, the five category case reduces to four category case*

Table 3.17: Numerical results for double reciprocal distribution, k=6

	k=6				
Criterion	z_1^*	z_2^*	$F(z_1^*)$	$F(z_2^*)$	$\phi(z_1^*, z_2^*)$
D-optimality	0.5000	2.0000	0.6667	0.8333	-0.9569
A-optimality	0.5000	2.0000	0.6667	0.8333	-4.1464
c_1-optimality	0.5000	2.0000	0.6667	0.8333	-0.7714
c_2-optimality	0.5000	2.0000	0.6667	0.8333	-3.3750
E-optimality	0.5000	2.0000	0.6667	0.8333	-3.3750

3.4 Comments and justification of the results

We have following comments on the numerical results using the search method above.

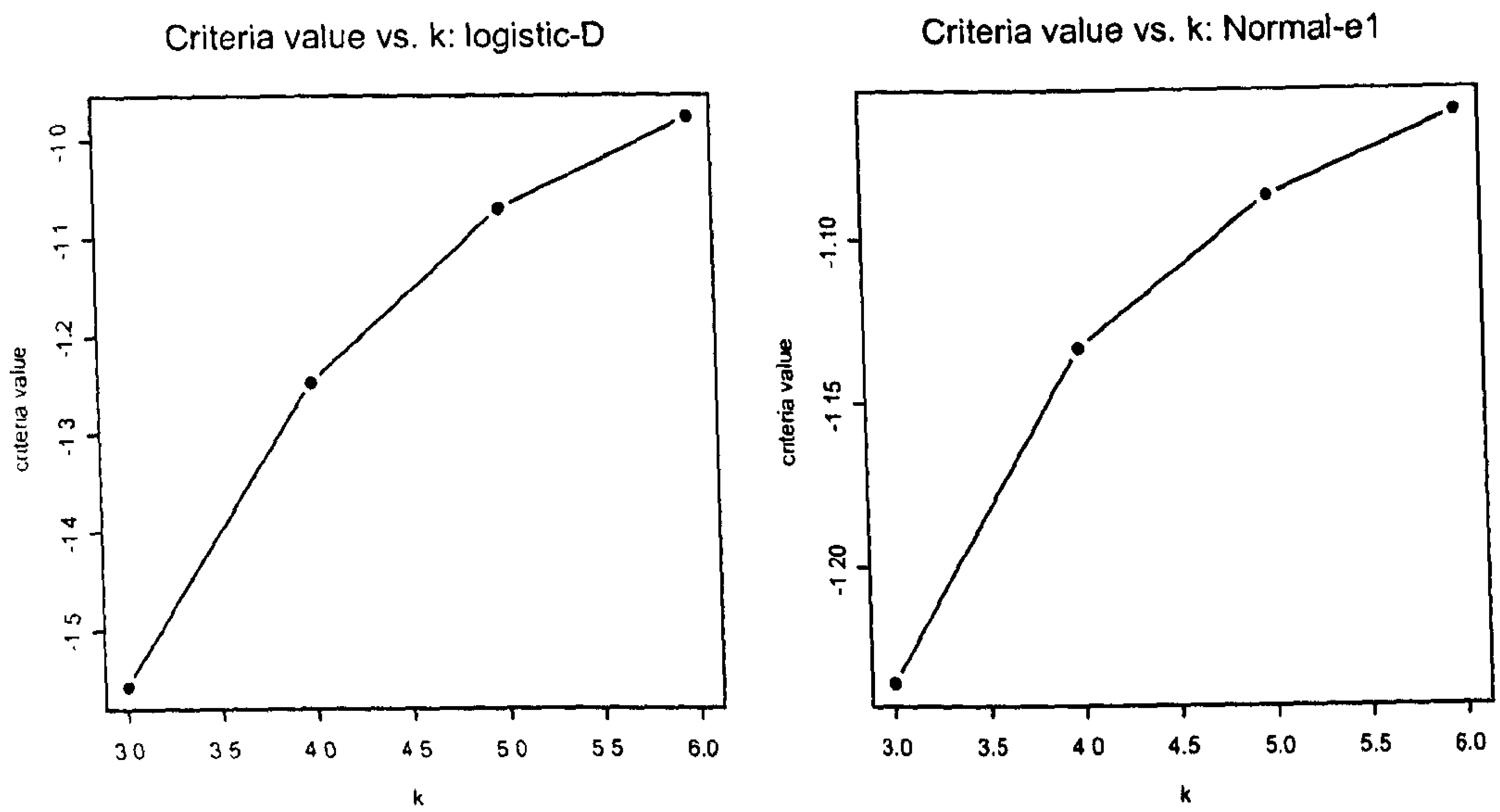
As mentioned above, these are results for symmetric designs or symmetric cutpoint sets for k=3, 4, 5 and 6 categories. We reiterate the forms for cutpoint sets as follows:

$$\underline{z}^* = (-z^*, z^*) \qquad k = 3$$
$$\underline{z}^* = (-z^*, 0, z^*) \qquad k = 4$$
$$\underline{z}^* = (-z_2^*, -z_1^*, z_1^*, z_2^*) \qquad k = 5$$
$$\underline{z}^* = (-z_2^*, -z_1^*, 0, z_1^*, z_2^*) \qquad k = 6$$

We can see that, in general, the values of criteria increase when we increase the number of cutpoints, say k. These results are to be expected since when we increase the number of cutpoints, we will get more information from the sample; thus the values of criteria should increase. We produce plots in figure 3.1 for the logistic distribution with D-optimality and the normal distributions with e_1-optimality to illustrate the changing criterion values with k changing from 3 to 6.

Returning to the tables 3.10 to 3.17, we now focus on the results on the double exponential and double reciprocal distributions. In both cases, when the criterion is e_1-optimality and the number of categories are three, we see that the criterion value reaches a maximum value of -1 when the value of z reaches 0. Thus, the three category case will reduce to the two category case. We already considered this case in particular in chapter two. We see that when the value of z is 0, the information matrix is singular. However,

Figure 3.1: Plots of criteria value vs. the number of categories k



we still have an optimal design. See Ford et al (1992)

In the case of the double exponential distribution, $k = 4$ and e_1-optimality, the criteria has a constant value of -1 because the information matrix is diagonal, the first entry being 1 (we will investigate this later on).

When $k = 5$, some criteria in the case of the double reciprocal distribution (D- and e_1-optimality) and some criteria in the case of the double exponential distribution (D-, A-, and e_1-optimality) reach their maximizing value when the two middle cutpoints coincide at 0. Thus, the five category case reduces to four category case.

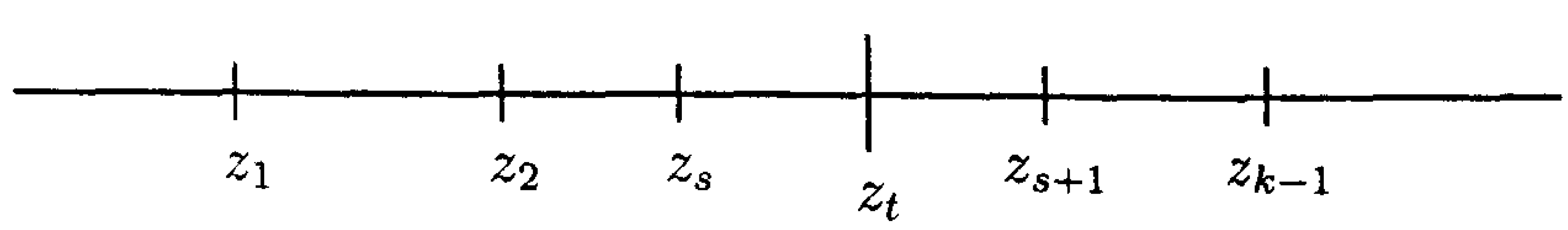
In the case of the double exponential distribution, when $k = 6$, four criteria, namely D-, A-, e_2- and E-optimality, have the same optimal cut-points, i.e. 1.02 and 2.61 while the optimal value of the criterion in the case of D-

optimality is -0.1981 and -1.2192 for the remaining cases. In the case of the double exponential distribution, when k=6 the e_1-optimal criterion has the same optimum as in the case of $k = 4$, the criterion always has the value of -1 for any value of z_1 and z_2.

Similar results appears for all criteria in the case of double reciprocal distribution.

In the following part, we will explain some of the results above by checking for an increase in the criterion values when the number of cutpoints (number of categories) increases by 1.

We assume that the cutpoint z_t is inserted between two other cutpoints z_s and z_{s+1}. We illustrate this situation by the following diagram:



We will compare the new information matrix (we call it New I_z) after inserting the cutpoint z_t and the old information matrix (we call it Old I_z) before inserting the cutpoint z_t.

In the previous chapter, we have constructed the following formula for the Fisher information matrix:

$$I_X(\underline{\gamma}) = \sum_{i=1}^{k} \frac{\partial \theta_i}{\partial \underline{\gamma}} \frac{1}{\theta_i} \left(\frac{\partial \theta_i}{\partial \underline{\gamma}} \right)^T,$$

where:

$$\frac{\partial \theta_i}{\partial \underline{\gamma}} = \begin{pmatrix} 1 & 1 \\ x_{i-1} & x_i \end{pmatrix} \begin{pmatrix} -f(z_{i-1}) \\ f(z_i) \end{pmatrix}.$$

And through the standardization:

$$I_Z(\underline{\gamma}) = \sum_{i=1}^{k} \frac{\partial \theta_i}{\partial \underline{\gamma}} \frac{1}{\theta_i} \left(\frac{\partial \theta_i}{\partial \underline{\gamma}} \right)^T,$$

where:

$$\frac{\partial \theta_i}{\partial \underline{\gamma}} = \begin{pmatrix} 1 & 1 \\ z_{i-1} & z_i \end{pmatrix} \begin{pmatrix} -f(z_{i-1}) \\ f(z_i) \end{pmatrix}.$$

Let:

$$\underline{v}_i = f(z_i) \begin{pmatrix} 1 \\ z_i \end{pmatrix}.$$

We have following formula for the Fisher information matrix:

$$I_{\underline{Z}}(\underline{\gamma}) = \sum_{i=1}^{k} \frac{1}{\theta_i} [\underline{v}_i - \underline{v}_{i-1}][\underline{v}_i - \underline{v}_{i-1}]^T,$$

where:

$$\theta_o = 0, \quad \theta_k = 1$$

$$z_o = -\infty, \quad z_k = \infty$$

$$F(z_o) = 0, \quad F(z_k) = 1$$

$$f(z_o) = 0, \quad f(z_k) = 0$$

$$\underline{v}_o = 0, \quad \underline{v}_k = 0$$

Now suppose cutpoint z_t is inserted between two cutpoints z_s and z_{s+1}

Note that:

$$z_s \leq z_t \leq z_{s+1}$$

$$\underline{v}_s = f(z_s) \begin{pmatrix} 1 \\ z_s \end{pmatrix}$$

$$\underline{v}_{s+1} = f(z_{s+1}) \begin{pmatrix} 1 \\ z_{s+1} \end{pmatrix}$$

$$\underline{v}_t = f(z_t) \begin{pmatrix} 1 \\ z_t \end{pmatrix}$$

$$\underline{v}_{s+1} - \underline{v}_s = (\underline{v}_{s+1} - \underline{v}_t) + (\underline{v}_t - \underline{v}_s)$$

$$\theta_{s+1} - \theta_s = (\theta_{s+1} - \theta_t) + (\theta_t - \theta_s)$$

where:

$$\theta_t = F(z_t) - F(z_s)$$

Let D be the difference between New I_z and Old I_z and note that $\theta_i = F(z_i) - F(z_{i-1})$, we have:

$$\begin{aligned} D \;=\; & \frac{1}{F(z_{s+1}) - F(z_t)} \left[\underline{v}_{s+1} - \underline{v}_t\right] \left[\underline{v}_{s+1} - \underline{v}_t\right]^T \\ & + \frac{1}{F(z_t) - F(z_s)} \left[\underline{v}_t - \underline{v}_s\right] \left[\underline{v}_t - \underline{v}_s\right]^T \\ & - \frac{1}{F(z_{s+1}) - F(z_s)} \left[\underline{v}_{s+1} - \underline{v}_s\right] \left[\underline{v}_{s+1} - \underline{v}_s\right]^T \end{aligned}$$

If z_s, z_t, z_{s+1} and $F(.)$ are symmetric about 0, it follows that:

$$\underline{v}_s = f(-z) \begin{pmatrix} 1 \\ -z \end{pmatrix}$$

$$\underline{v}_{s+1} = f(z) \begin{pmatrix} 1 \\ z \end{pmatrix}$$

$$\underline{v}_t = f(0) \begin{pmatrix} 1 \\ 0 \end{pmatrix}$$

$$F(-z) = 1 - F(z), \quad F(0) = 1/2$$

In this case, we insert the cutpoint 0 between two cutpoints z_s and z_{s+1}. So we have:

$$\begin{aligned} D &= \frac{1}{[F(z)-0.5]}\left[f(z)\begin{pmatrix}1\\z\end{pmatrix}-f(0)\begin{pmatrix}1\\0\end{pmatrix}\right]\left[f(z)\begin{pmatrix}1\\z\end{pmatrix}-f(0)\begin{pmatrix}1\\0\end{pmatrix}\right]^T \\ &+\frac{1}{[F(z)-0.5]}\left[f(0)\begin{pmatrix}1\\0\end{pmatrix}-f(z)\begin{pmatrix}1\\-z\end{pmatrix}\right]\left[f(0)\begin{pmatrix}1\\0\end{pmatrix}-f(z)\begin{pmatrix}1\\-z\end{pmatrix}\right]^T \\ &-\frac{1}{[2F(z)-1]}\left[f(z)\begin{pmatrix}1\\z\end{pmatrix}-f(z)\begin{pmatrix}1\\-z\end{pmatrix}\right]\left[f(z)\begin{pmatrix}1\\z\end{pmatrix}-f(z)\begin{pmatrix}1\\-z\end{pmatrix}\right]^T \end{aligned}$$

$$\begin{aligned} D &= \frac{1}{[2F(z)-1]}\left\{\begin{pmatrix}f(z)-f(0)\\zf(z)\end{pmatrix}\begin{pmatrix}f(z)-f(0)\\zf(z)\end{pmatrix}^T\right. \\ &\left.+2\begin{pmatrix}f(0)-f(z)\\zf(z)\end{pmatrix}\begin{pmatrix}f(0)-f(z)\\zf(z)\end{pmatrix}^T-\begin{pmatrix}0\\2zf(z)\end{pmatrix}\begin{pmatrix}0\\2zf(z)\end{pmatrix}^T\right\} \end{aligned}$$

After simplification, D has the form:

$$D=\frac{1}{[2F(z)-1]}\left\{\begin{pmatrix}4[f(0)-f(z)]^2 & 0\\ 0 & 4z^2f(z)\end{pmatrix}-\begin{pmatrix}0 & 0\\ 0 & 4z^2f(z)\end{pmatrix}\right\}$$

And finally:

$$D=\begin{pmatrix}\dfrac{4[f(0)-f(z)]^2}{2F(z)-1} & 0\\ 0 & 0\end{pmatrix} \tag{3.15}$$

The first diagonal entry of the matrix D is positive. This explains why all the criteria (except the e_2-criteria) increase when we add a cut-point of zero between two symmetric cut-points. In the case of the e_2-criteria, because the vector $e_2 = (0,1)$, the values of the criteria stay the same when adding a cut-point of zero between two symmetric cut-points.

In the general case, when z_s, z_t, z_{s+1} are not necessarily symmetric, we have the following formula for the matrix D:

$$\begin{aligned} D &= \frac{1}{F(z_{s+1}) - F(z_t)} \times \\ &\times \left[f(z_{s+1}) \begin{pmatrix} 1 \\ z_{s+1} \end{pmatrix} - f(z_t) \begin{pmatrix} 1 \\ z_t \end{pmatrix} \right] \left[f(z_{s+1}) \begin{pmatrix} 1 \\ z_{s+1} \end{pmatrix} - f(z_t) \begin{pmatrix} 1 \\ z_t \end{pmatrix} \right]^T \\ &+ \frac{1}{F(z_t) - F(z_s)} \times \\ &\times \left[f(z_t) \begin{pmatrix} 1 \\ z_t \end{pmatrix} - f(z_s) \begin{pmatrix} 1 \\ z_s \end{pmatrix} \right] \left[f(z_t) \begin{pmatrix} 1 \\ z_t \end{pmatrix} - f(z_s) \begin{pmatrix} 1 \\ z_s \end{pmatrix} \right]^T \\ &- \frac{1}{F(z_{s+1}) - F(z_s)} \times \\ &\left[f(z_{s+1}) \begin{pmatrix} 1 \\ z_{s+1} \end{pmatrix} - f(z_s) \begin{pmatrix} 1 \\ z_s \end{pmatrix} \right] \left[f(z_{s+1}) \begin{pmatrix} 1 \\ z_{s+1} \end{pmatrix} - f(z_s) \begin{pmatrix} 1 \\ z_s \end{pmatrix} \right]^T \end{aligned}$$

The matrix D will have the form:

$$D = \begin{pmatrix} a & b \\ b & c \end{pmatrix} \tag{3.16}$$

We will illustrate that the matrix D is a non-negative definite matrix. One of the conditions for the matrix D of the form 3.16 to be non-negative definite is that it satisfies:

$$\begin{cases} a & \geq 0 \\ c & \geq 0 \\ |D| & \geq 0 \end{cases}$$

Another necessary and sufficient condition for a matrix to be non-negative definite is that all of its eigenvalues are non-negative. We will use the second condition to check for the non-negative definite property of the matrix D. In order to do that, we will calculate the minimum eigenvalue of matrix D and show that it is non-negative.

The eigenvalues of matrix D (denoted by λ) are the solutions of the following equation:

$$\begin{aligned} |D - \lambda I| &= 0 \\ (a-\lambda)(c-\lambda) - b^2 &= 0 \\ \lambda^2 - (a+c)\lambda + (ac - b^2) &= 0 \end{aligned}$$

The solutions are:

$$\lambda_{1,2} = \frac{(a+c) \pm \sqrt{(a+c)^2 - 4(ac-b^2)}}{2}$$

$$\lambda_{1,2} = \frac{(a+c) \pm \sqrt{(a-c)^2 + 4b^2}}{2}$$

So the smaller solution (smaller eigenvalue of matrix D) is:

$$\lambda_2 = \frac{(a+c) - \sqrt{(a-c)^2 + 4b^2}}{2}$$

with a, b and c being functions of z_s, z_t and z_{s+1}.

Now we will find the minimum value of the smaller eigenvalue of D. If this minimum value is non-negative, it follows that all the eigenvalues of D are non-negative and D is also non-negative definite. To search for the minimum value of the smaller eigenvalue of matrix D, we use a multiplicative algorithm. For convenience, we transform the variables as follows:

Let $z_s = z_1$, $z_t = z_2$, $z_{s+1} = z_3$.

$P_1 = F(z_1)$

$P_2 = F(z_2) - F(z_1)$

$P_3 = F(z_3) - F(z_2)$.

$P_4 = 1 - F(z_3)$.

Note that $P_i \geq 0$ and $\sum P_i = 1$. Now the problem can be stated as follows:

Find (P_1, P_2, P_3, P_4) that minimizes the function:

$$\lambda_2 = \lambda(P_1, P_2, P_3, P_4) = \frac{(a+c) - \sqrt{(a-c)^2 + 4b^2}}{2}$$

with the constraints:

$$P_1, P_2, P_3, P_4 \geq 0$$

$$P_1 + P_2 + P_3 < 1$$

$$or\ P_1 + P_2 + P_3 + P_4 = 1$$

We cannot determine the minima of P_1, P_2, P_3, P_4 explicitly. Numerical techniques are needed. We use the following multiplicative algorithm: (see Mandal and Torsney (2000))

$$P_j^{(r+1)} = \frac{P_j^{(r)} m(x_j^{(r+1)})}{\sum_{i=1}^{k} P_i^{(r)} m(x_i^{(r)})}$$

in which:

- $m(.)$ is a positive increasing function (e.g. $m(z) = \Phi(\delta z)$), $\delta = 1$
- $x_j^{(r)} = d_j^{(r)} = \partial\lambda_2/\partial P_j \mid \underline{P} = \underline{P}^{(r)}$ or
- $x_j^{(r)} = F_j(r) = d_j^{(r)} - \sum P_i^{(r)} d_i^{(r)}$, which is a directional derivative of λ_2.

This algorithm has some important properties needed to be considered. We will explore this algorithm and such properties in more detail in the next chapter.

This algorithm is for finding the maximum value of a function. Since our purpose is to find the minimum value of the eigenvalue λ_2, we will find the maximum value of the function $-\lambda_2$. Let $\lambda_3 = -\lambda_2$. We will make the choice:

$$x_j^{(r)} = \left[\frac{\partial\lambda_3}{\partial P_j} - \sum_{i=1}^{4} P_i \frac{\partial\lambda_3}{\partial P_i}\right], \quad j = 1, 2, 3, 4 \tag{3.17}$$

So we need to calculate:

$$\frac{\partial \lambda_3}{\partial P_i} = \sum_{j=1}^{3} \frac{\partial \lambda_3}{\partial z_j}\frac{\partial z_j}{\partial P_i}, \quad i = 1,2,3 \tag{3.18}$$

$$\lambda_3 = -\frac{(a+c) - \sqrt{(a-c)^2 + 4b^2}}{2}$$

Here are the formula for $\partial\lambda_3/\partial z_j$:

$$\begin{aligned}
\frac{\partial \lambda_3}{\partial z_j} &= -\frac{1}{2}\left(\frac{\partial a}{\partial z_j} + \frac{\partial c}{\partial z_j}\right) \\
&\quad + \frac{1}{2\sqrt{(a-c)^2+4b^2}}\left[2a\frac{\partial a}{\partial z_j} - 2\left(a\frac{\partial c}{\partial z_j} + c\frac{\partial a}{\partial z_j}\right) + 2c\frac{\partial c}{\partial z_j} + 8b\frac{\partial b}{\partial z_j}\right]
\end{aligned}$$

And it follows that we need to calculate:

$$a = \frac{[f(z_3)-f(z_2)]^2}{F(z_3)-F(z_2)} + \frac{[f(z_2)-f(z_1)]^2}{F(z_2)-F(z_1)} - \frac{[f(z_3)-f(z_1)]^2}{F(z_3)-F(z_1)}$$

$$\begin{aligned}
\frac{\partial a}{\partial z_1} &= 2f'(z_1)\left[\frac{f(z_3)-f(z_1)}{F(z_3)-F(z_1)} - \frac{f(z_2)-f(z_1)}{F(z_2)-F(z_1)}\right] \\
&\quad + f(z_1)\left[\frac{(f(z_2)-f(z_1))^2}{(F(z_2)-F(z_1))^2} - \frac{(f(z_3)-f(z_1))^2}{(F(z_3)-F(z_1))^2}\right]
\end{aligned}$$

$$\begin{aligned}
\frac{\partial a}{\partial z_2} &= 2f'(z_2)\left[\frac{f(z_2)-f(z_1)}{F(z_2)-F(z_1)} - \frac{f(z_3)-f(z_2)}{F(z_3)-F(z_2)}\right] \\
&\quad + f(z_2)\left[\frac{(f(z_3)-f(z_2))^2}{(F(z_3)-F(z_2))^2} - \frac{(f(z_2)-f(z_1))^1}{(F(z_2)-F(z_1))^2}\right]
\end{aligned}$$

$$\begin{aligned}
\frac{\partial a}{\partial z_3} &= 2f'(z_3)\left[\frac{f(z_3)-f(z_2)}{F(z_3)-F(z_2)} - \frac{f(z_3)-f(z_1)}{F(z_3)-F(z_1)}\right] \\
&\quad + f(z_3)\left[\frac{(f(z_3)-f(z_1))^2}{(F(z_3)-F(z_1))^2} - \frac{(f(z_3)-f(z_2))^2}{(F(z_3)-F(z_2))^2}\right]
\end{aligned}$$

$$\begin{aligned}
b &= \frac{[f(z_3)-f(z_2)]\,[z_3f(z_3)-z_2f(z_2)]}{F(z_3)-F(z_2)} + \frac{[f(z_2)-f(z_1)]\,[z_2f(z_2)-z_1f(z_1)]}{F(z_2)-F(z_1)} \\
&\quad - \frac{[f(z_3)-f(z_1)]\,[z_3f(z_3)-z_1f(z_1)]}{F(z_3)-F(z_1)}
\end{aligned}$$

$$
\begin{aligned}
\frac{\partial b}{\partial z_1} &= \frac{-f'(z_1)\left[z_2 f(z_2) - z_1 f(z_1)\right] - \left[f(z_2) - f(z_1)\right]\left[f(z_1) + z_1 f'(z_1)\right]}{F(z_2) - F(z_1)} \\
&+ \frac{f'(z_1)\left[z_3 f(z_3) - z_1 f(z_1)\right] + \left[f(z_3) - f(z_1)\right]\left[f(z_1) + z_1 f'(z_1)\right]}{F(z_3) - F(z_1)} \\
&+ \frac{f(z_1)\left[f(z_2) - f(z_1)\right]\left[z_2 f(z_2) - z_1 f(z_1)\right]}{(F(z_2) - F(z_1))^2} \\
&- \frac{f(z_1)\left[f(z_3) - f(z_1)\right]\left[z_3 f(z_3) - z_1 f(z_1)\right]}{(F(z_3) - F(z_1))^2}
\end{aligned}
$$

$$
\begin{aligned}
\frac{\partial b}{\partial z_2} &= \frac{-f'(z_2)\left[z_3 f(z_3) - z_2 f(z_2)\right] - \left[f(z_3) - f(z_2)\right]\left[f(z_2) + z_2 f'(z_2)\right]}{F(z_3) - F(z_2)} \\
&+ \frac{f'(z_2)\left[z_2 f(z_2) - z_1 f(z_1)\right] + \left[f(z_2) - f(z_1)\right]\left[f(z_2) + z_2 f'(z_2)\right]}{F(z_2) - F(z_1)} \\
&+ \frac{f(z_2)\left[f(z_3) - f(z_2)\right]\left[z_3 f(z_3) - z_2 f(z_2)\right]}{(F(z_3) - F(z_2))^2} \\
&- \frac{f(z_2)\left[f(z_2) - f(z_1)\right]\left[z_2 f(z_2) - z_1 f(z_1)\right]}{(F(z_2) - F(z_1))^2}
\end{aligned}
$$

$$
\begin{aligned}
\frac{\partial b}{\partial z_3} &= \frac{-f'(z_3)\left[z_3 f(z_3) - z_2 f(z_2)\right] - \left[f(z_3) - f(z_2)\right]\left[f(z_3) + z_3 f'(z_3)\right]}{F(z_3) - F(z_2)} \\
&+ \frac{f'(z_3)\left[z_3 f(z_3) - z_1 f(z_1)\right] + \left[f(z_3) - f(z_1)\right]\left[f(z_3) + z_3 f'(z_3)\right]}{F(z_3) - F(z_1)} \\
&+ \frac{f(z_3)\left[f(z_3) - f(z_1)\right]\left[z_3 f(z_3) - z_1 f(z_1)\right]}{(F(z_3) - F(z_1))^2} \\
&- \frac{f(z_3)\left[f(z_3) - f(z_2)\right]\left[z_3 f(z_3) - z_2 f(z_2)\right]}{(F(z_3) - F(z_2))^2}
\end{aligned}
$$

$$
c = \frac{z_3 f(z_3) - z_2 f(z_2)]^2}{F(z_3) - F(z_2)} + \frac{[z_2 f(z_2) - z_1 f(z_1)]^2}{F(z_2) - F(z_1)} - \frac{[z_3 f(z_3) - z_1 f(z_1)]^2}{F(z_3) - F(z_1)}
$$

$$
\begin{aligned}
\frac{\partial c}{\partial z_1} &= \frac{-2\left[f(z_1) + z_1 f'(z_1)\right]\left[z_2 f(z_2) - z_1 f(z_1)\right]}{F(z_2) - F(z_1)} + \frac{f(z_1)\left[z_2 f(z_2) - z_1 f(z_1)\right]^2}{F(z_2) - F(z_1)} \\
&+ \frac{2\left[f(z_1) + z_1 f'(z_1)\right]\left[z_3 f(z_3) - z_1 f(z_1)\right]}{F(z_3) - F(z_1)} - \frac{f(z_1)\left[z_3 f(z_3) - z_1 f(z_1)\right]^2}{(F(z_3) - F(z_1))^2}
\end{aligned}
$$

$$\begin{aligned}\frac{\partial c}{\partial z_2} &= \frac{-2\left[f(z_2)+z_2f'(z_2)\right]\left[z_3f(z_3)-z_2f(z_2)\right]}{F(z_3)-F(z_2)}+\frac{f(z_2)\left[z_3f(z_3)-z_2f(z_2)\right]^2}{F(z_3)-F(z_2)}\\ &+\frac{2\left[f(z_2)+z_2f'(z_2)\right]\left[z_2f(z_2)-z_1f(z_1)\right]}{F(z_2)-F(z_1)}-\frac{f(z_2)\left[z_2f(z_2)-z_1f(z_1)\right]^2}{(F(z_2)-F(z_1))^2}\end{aligned}$$

$$\begin{aligned}\frac{\partial c}{\partial z_3} &= \frac{2\left[f(z_3)+z_3f'(z_3)\right]\left[z_3f(z_3)-z_2f(z_2)\right]}{F(z_3)-F(z_2)}-\frac{f(z_3)\left[z_3f(z_3)-z_2f(z_2)\right]^2}{F(z_3)-F(z_2)}\\ &-\frac{2\left[f(z_3)+z_3f'(z_3)\right]\left[z_3f(z_3)-z_1f(z_1)\right]}{F(z_3)-F(z_1)}-\frac{f(z_3)\left[z_3f(z_3)-z_1f(z_1)\right]^2}{(F(z_3)-F(z_1))^2}\end{aligned}$$

We also need to calculate $\frac{\partial z_j}{\partial P_i}$ where $j = 1,2,3$ and $i = 1,2,3,4$.

We know that:

$$F(z_1) = P_1 \rightarrow z_1 = F^{-1}(P_1)$$

$$\frac{\partial z_1}{\partial P_1} = \frac{1}{\partial F(P_1)/\partial P_1} = \frac{1}{f(P_1)}, \quad \frac{\partial z_1}{\partial P_2} = \frac{\partial z_1}{\partial P_3} = 0$$

$$F(z_2) - F(z_1) = P_2 \rightarrow z_2 = F^{-1}(P_1 + P_2)$$

$$\frac{\partial z_2}{\partial P_1} = \frac{\partial z_2}{\partial P_2} = \frac{1}{\partial F(P_1+P_2)/\partial P_1} = \frac{1}{\partial F(P_1+P_2)/\partial P_2} = \frac{1}{f(P_1+P_2)}, \quad \frac{\partial z_2}{\partial P_3} = 0$$

$$F(z_3) - F(z_2) = P_3 \rightarrow z_3 = F^{-1}(P_1 + P_2 + P_3)$$

$$\frac{\partial z_3}{\partial P_1} = \frac{\partial z_3}{\partial P_2} = \frac{\partial z_3}{\partial P_3} = \frac{1}{f(P_1+P_2+P_3)}$$

From the above formula, we can calculate $\frac{\partial \lambda_3}{\partial P_i}$ in 3.18. In order to calculate $\frac{\partial \lambda_3}{\partial P_4}$ in 3.17, we extend the form of λ_3 from the function of (P_1, P_2, P_3) to the function of (P_1, P_2, P_3, P_4) as follows:

$$\begin{aligned}\lambda_3 &= \lambda_3(P_1, P_2, P_3)\\ &= \frac{1}{4}[\lambda_3(P_1, P_2, P_3) + \lambda_3(1 - P_2 - P_3 - P_4, P_2, P_3)\\ &\quad +\lambda_3(P_1, 1 - P_1 - P_3 - P_4, P_3) + \lambda_3(P_1, P_2, 1 - P_1 - P_2 - P_4)]\\ &= \hat{\lambda}_3(P_1, P_2, P_3, P_4), \quad P_1 + P_2 + P_3 + P_4 = 1\end{aligned}$$

We have:

$$\frac{\partial \hat{\lambda}_3}{\partial P_j} = \frac{\partial \lambda_3}{\partial P_j} - \frac{1}{4}\sum_{j=1}^{3}\frac{\partial \lambda_3}{\partial P_j}, \quad j = 1,2,3 \tag{3.19}$$

$$\frac{\partial \hat{\lambda}_3}{\partial P_4} = 0 - \frac{1}{4}\sum_{j=1}^{3}\frac{\partial \lambda_3}{\partial P_j}, \quad j = 1,2,3 \tag{3.20}$$

It then follows that the directional derivatives of λ_3 and $\hat{\lambda}_3$ are identical namely

$$F_j^{(\lambda_3)} = F_j^{(\hat{\lambda}_3)} = \frac{\partial \lambda_3}{\partial P_j} - \sum_{i=1}^{4} P_i \frac{\partial \lambda_3}{\partial P_i}, \quad j = 1,2,3,4.$$

allowing for $\partial\lambda_3/\partial P_4 = 0$.

Now, we can start using the algorithm with the initial values for P_i, say $P_i^{(0)} = \frac{1}{4}$.

Using this algorithm, we see that all the minimum values of the smaller eigenvalue (denoted by λ) of the matrix D, whatever the distributions of the cutpoints, are always positive but approximately zero.

We quote the results from running the multiplicative algorithm for the case of the logistic distribution in the table 3.18 below. Note that in the table 3.18, the maximum values of λ_3 are negative, which means the minimum values of λ_2 are positive.

Table 3.18: Results: finding minimum eigenvalue using a multiplicative algorithm for logistic distribution

Iteration	Directional derivative	P_i	λ_3
1971	-0.000000023623027	P_1=0.2053	0
	0.000000013293094	P_2=0.7019	
	-0.000000035028102	P_3=0.0926	
	-0.000000065153065	P_4=0.0000	
1972	-0.000000028266256	P_1=0.2053	0
	0.000000013273539	P_2=0.7020	
	-0.000000035058472	P_3=0.0926	
	-0.000000028361215	P_4=0.0000	
1973	-0.000000028270197	P_1=0.2052	0
	0.000000013256412	P_2=0.7021	
	-0.000000035493534	P_3=0.0926	
	0.000000069396156	P_4=6.9497e-005	
1974	-0.0000000382741231	P_1=0.2052	0
	0.0000000813234841	P_2=0.7021	
	-0.0000000888081969	P_3=0.0926	
	0.0000000696381278	P_4=0.0000	

So both eigenvalues of the matrix D are non-negative. We now can conclude that matrix D is non-negative definite. Remember that D is the difference between the information matrix after inserting the cutpoint z_t between two other cutpoints z_s and z_{s+1} (New I_z) and the matrix before inserting this cutpoint. The matrix D is non-negative definite, which means if we insert a cut-point between 2 other cut-points, the criterion values always increase. This confirms the validity of our general results about the increasing tendency of the criterion values when the number of cutpoints increases.
In the next chapter we will investigate the multiplicative algorithm in more detail and the way of using it to find optimal cutpoints.

3.5 Contour plotting

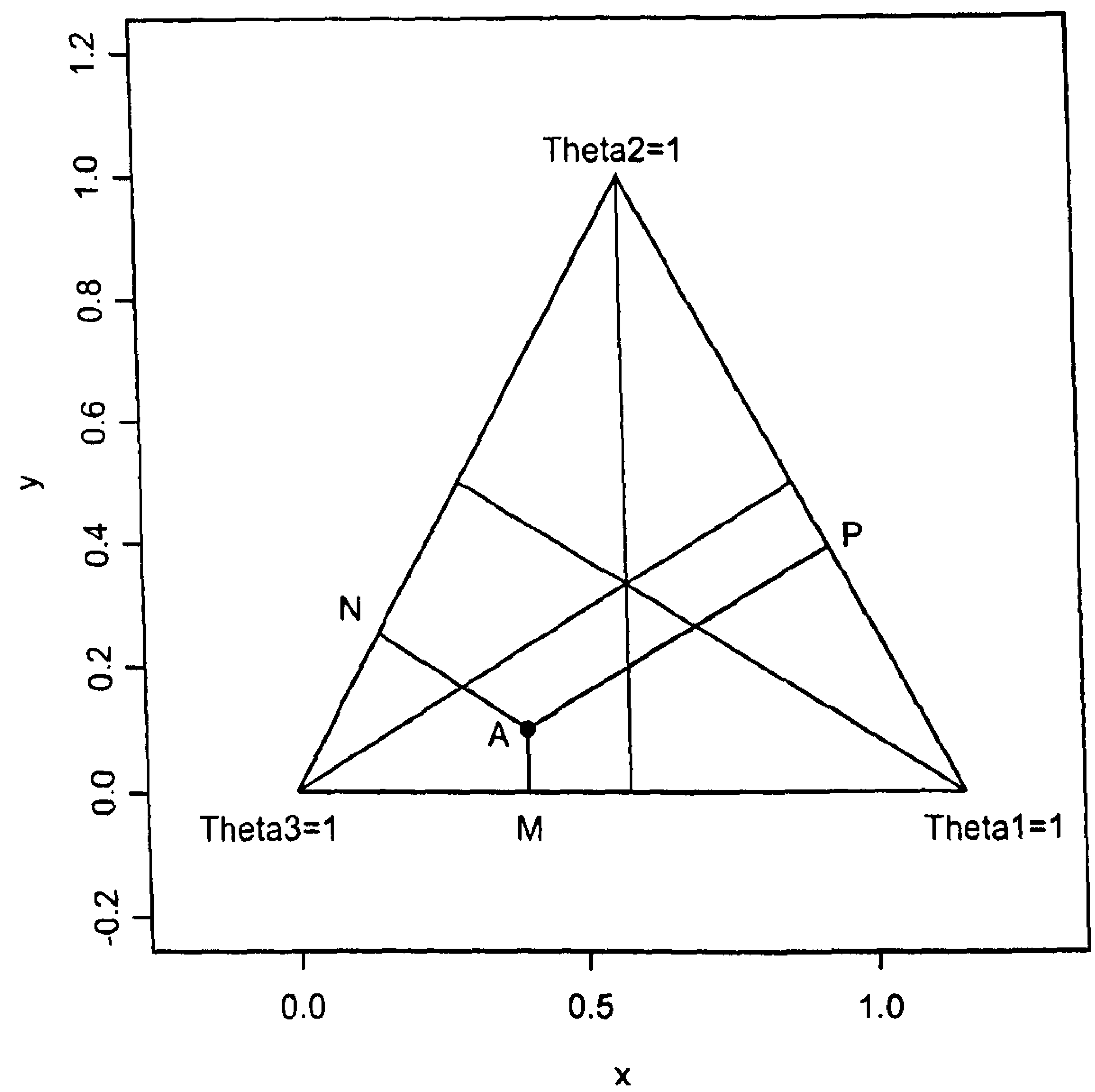
3.5.1 Introduction

We now produce contour plots in some situations. This allows us to relax the assumption of symmetry made in the above calculations.
We focus on investigating triangle plots and rectangular plots in three and four category cases. For the triangle plots, we have three cell probabilities $\theta_1, \theta_2, \theta_3$. So producing the plots is straight forward. We use rectangular plots, however, to deal with four cell probabilities, say $\theta_1, \theta_2, \theta_3$ and θ_4. We reduce the problem to two dimensions by imposing constraints on the cell probabilities. A sensible constraint is $\theta_1+\theta_2 = \theta_3+\theta_4 = 1/2$. This constraint makes sense if we want the current median to be a cutpoint.

3.5.2 Triangle and square contour plots

For three categories, we can use triangle plots to depict criterion values. The triangle contour plot is a way of dealing with the constraint that the three

Figure 3.2: The triangle contour plot



cell probabilities should sum to 1. In our case, we need to assess the criterion values with respect to 3 cell probabilities θ_1, θ_2 and θ_3. Figure 3.2 presents an equilateral triangle such that the sum of the perpendiculars of any point to the three sides equals one. At each vertex of the triangle, one cell probability equals one and the two remaining θ's equal zero. Every point (for instance point A) lying within the triangle represents a set of cell probabilities. From point A, if we draw the lines which are perpendicular to the respective sides of triangle, we have the measures for the cell probabilities. In this particular case, $AM = \theta_2, AN = \theta_1, AP = \theta_3$ and $AM + AN + AP = 1$.

Similarly in the four category case we use the rectangular plot. In our case, we impose the constraint on the cell probabilities that $\theta_1+\theta_2 = \theta_3+\theta_4 = 1/2$. So we can substitute for θ_2 in terms of θ_1 ($\theta_2 = 1 - \theta_1$) and θ_4 in terms of θ_3 ($\theta_4 = 1 - \theta_3$) leaving two free variables satisfying $0 < \theta_1 < 1/2$ and $0 < \theta_3 < 1/2$. So we plot criteria over the square defined by these ranges. Here are some contour plots we produce.

Figure 3.3: Contour plot: logistic distribution and D-opt

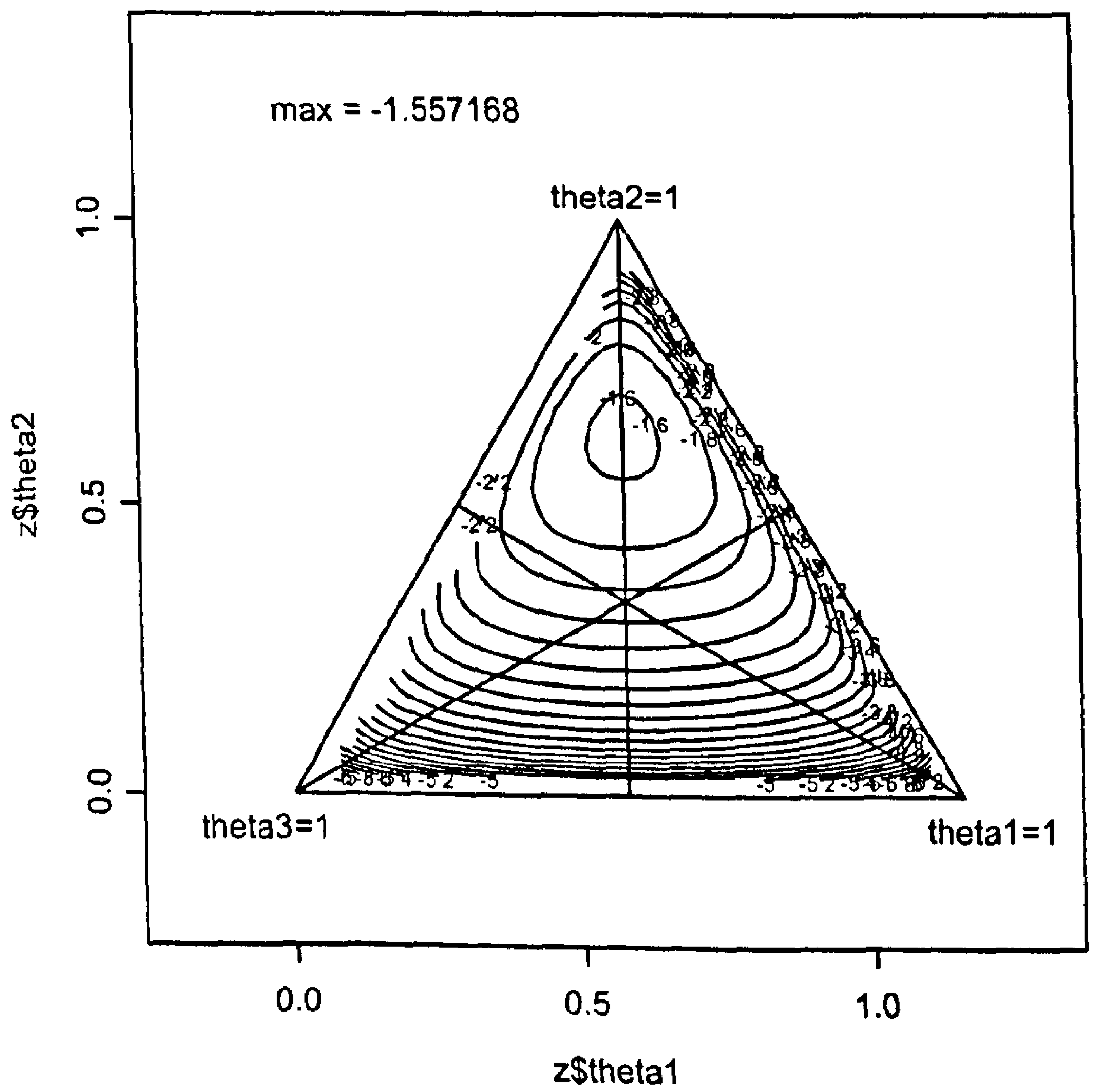


Figure 3.4: Contour plot: normal distribution and D-opt

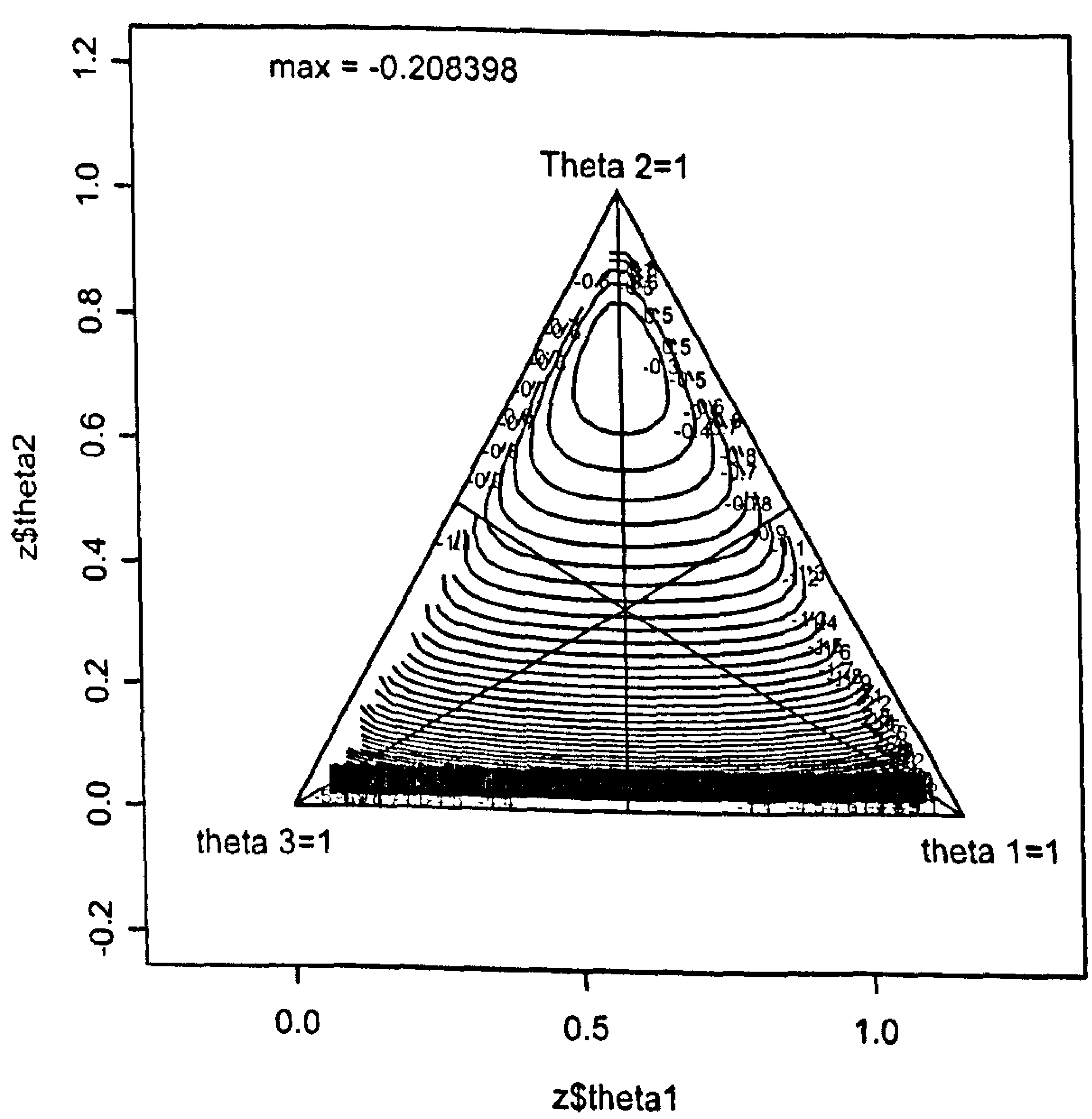


Figure 3.5: Contour plot: double-expo distribution and D-opt

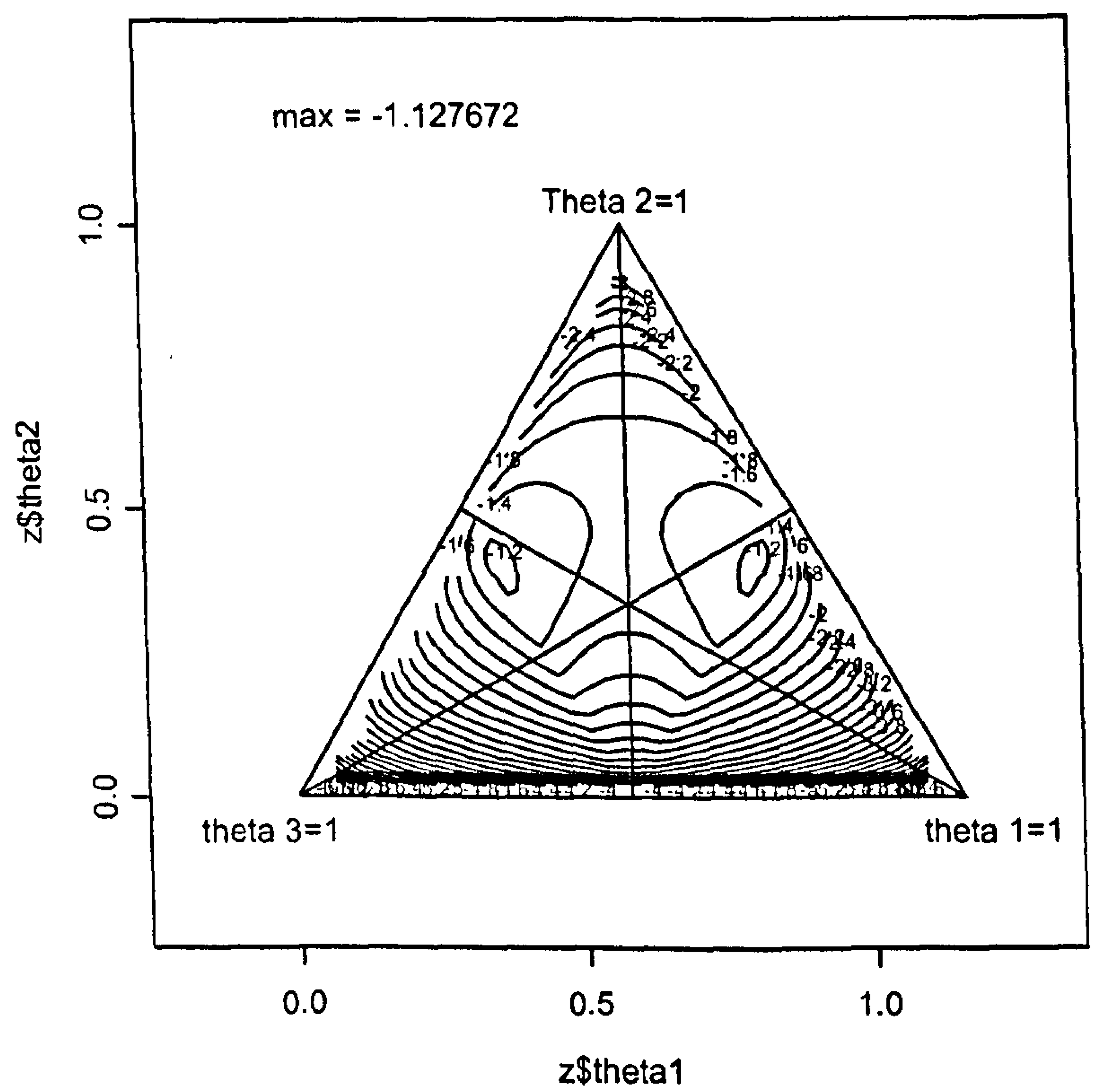


Figure 3.6: Contour plot: double-reciprocal distribution and D-opt

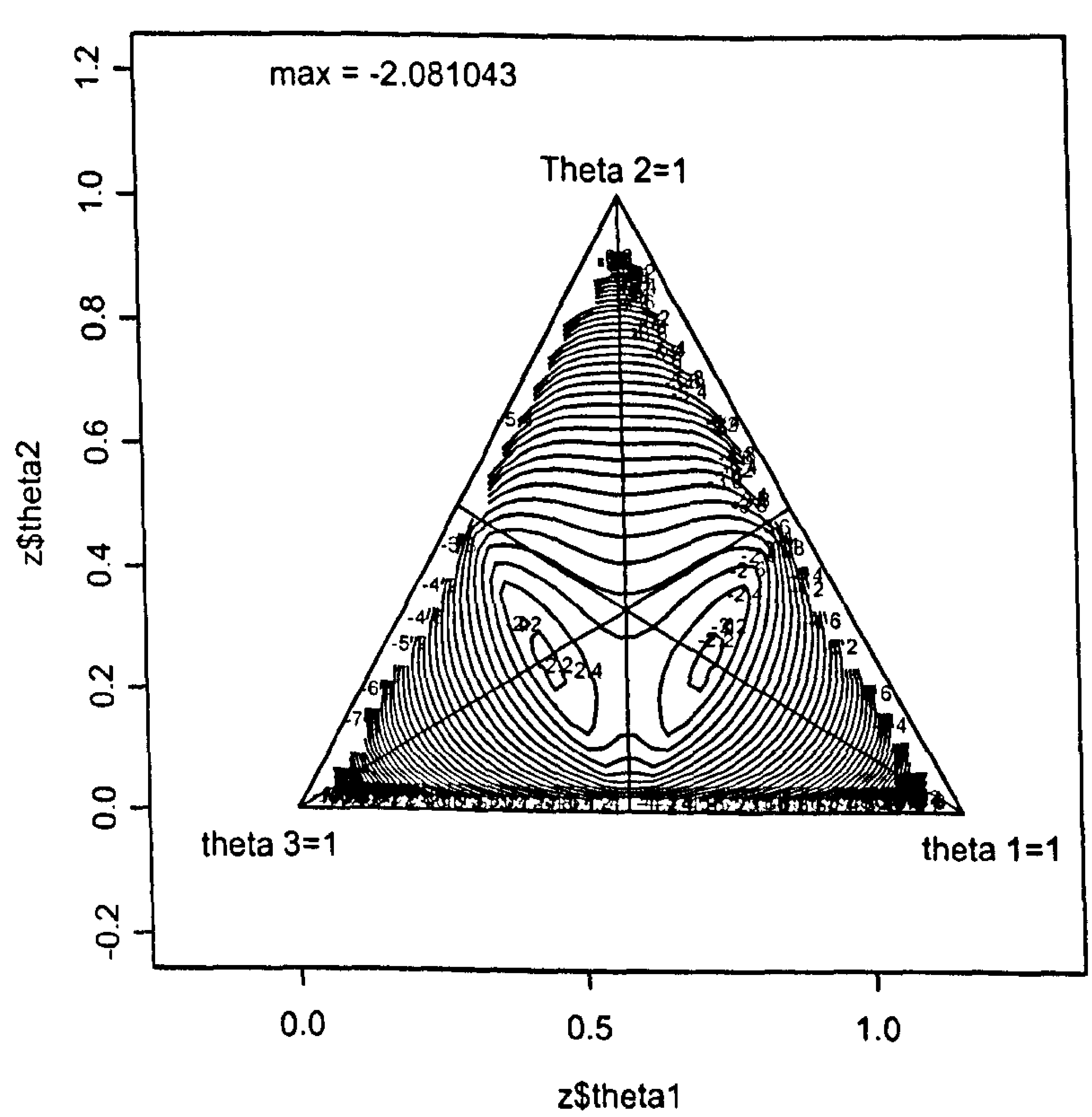


Figure 3.7: Contour plot: logistic distribution and A-opt

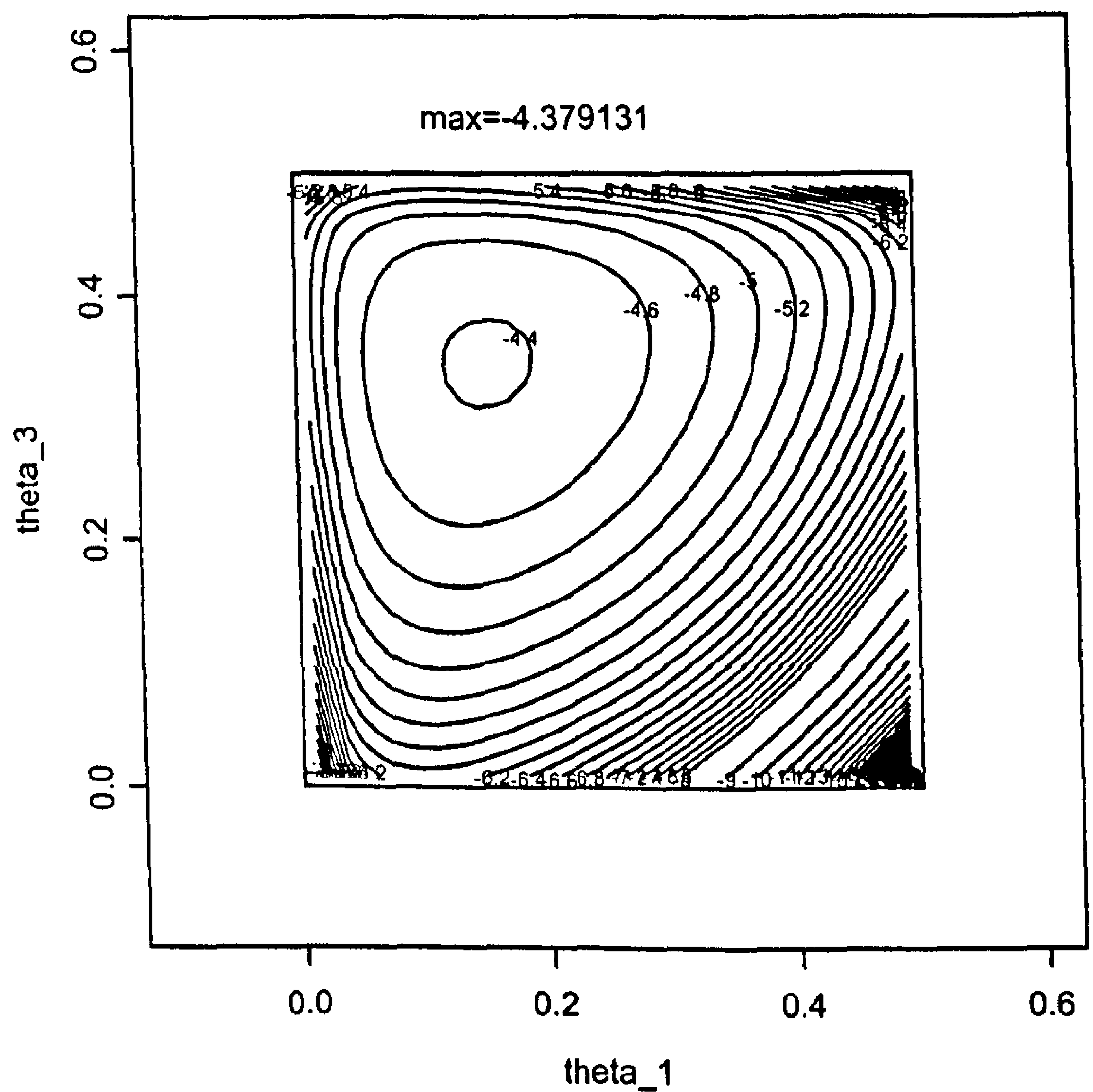


Figure 3.8: Contour plot: normal distribution and A-opt

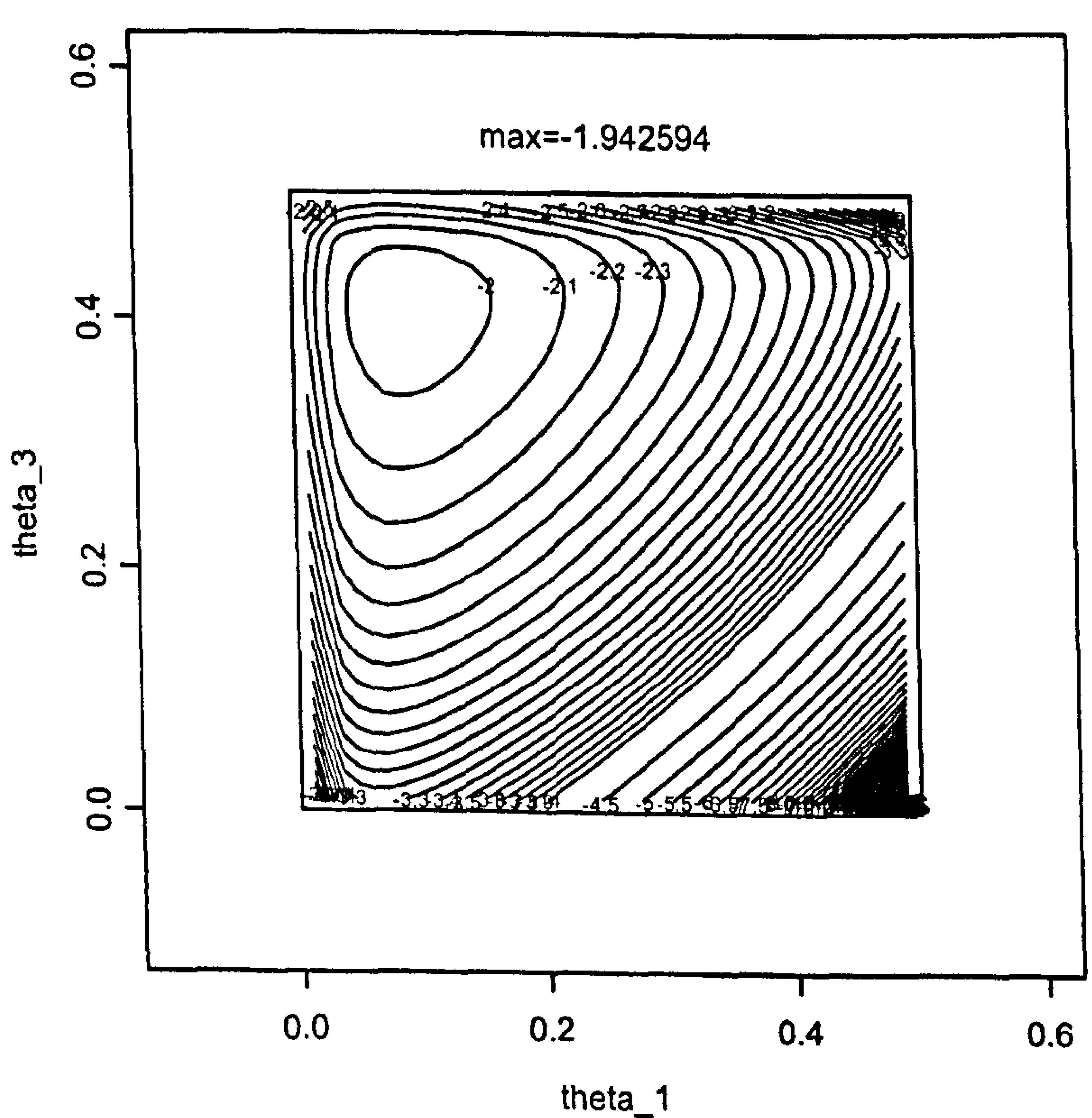


Figure 3.9: Contour plot: double-expo distribution and A-opt

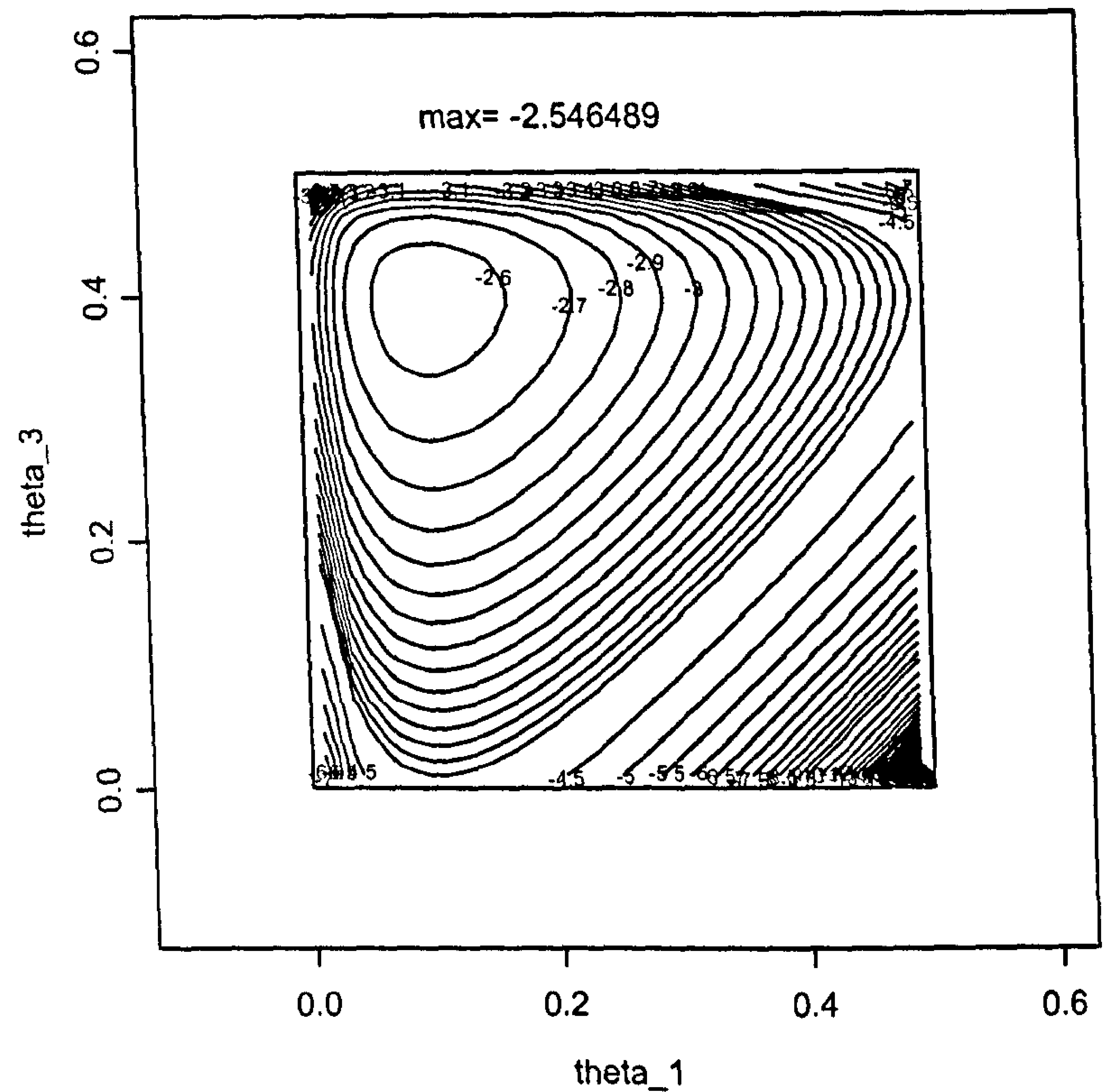
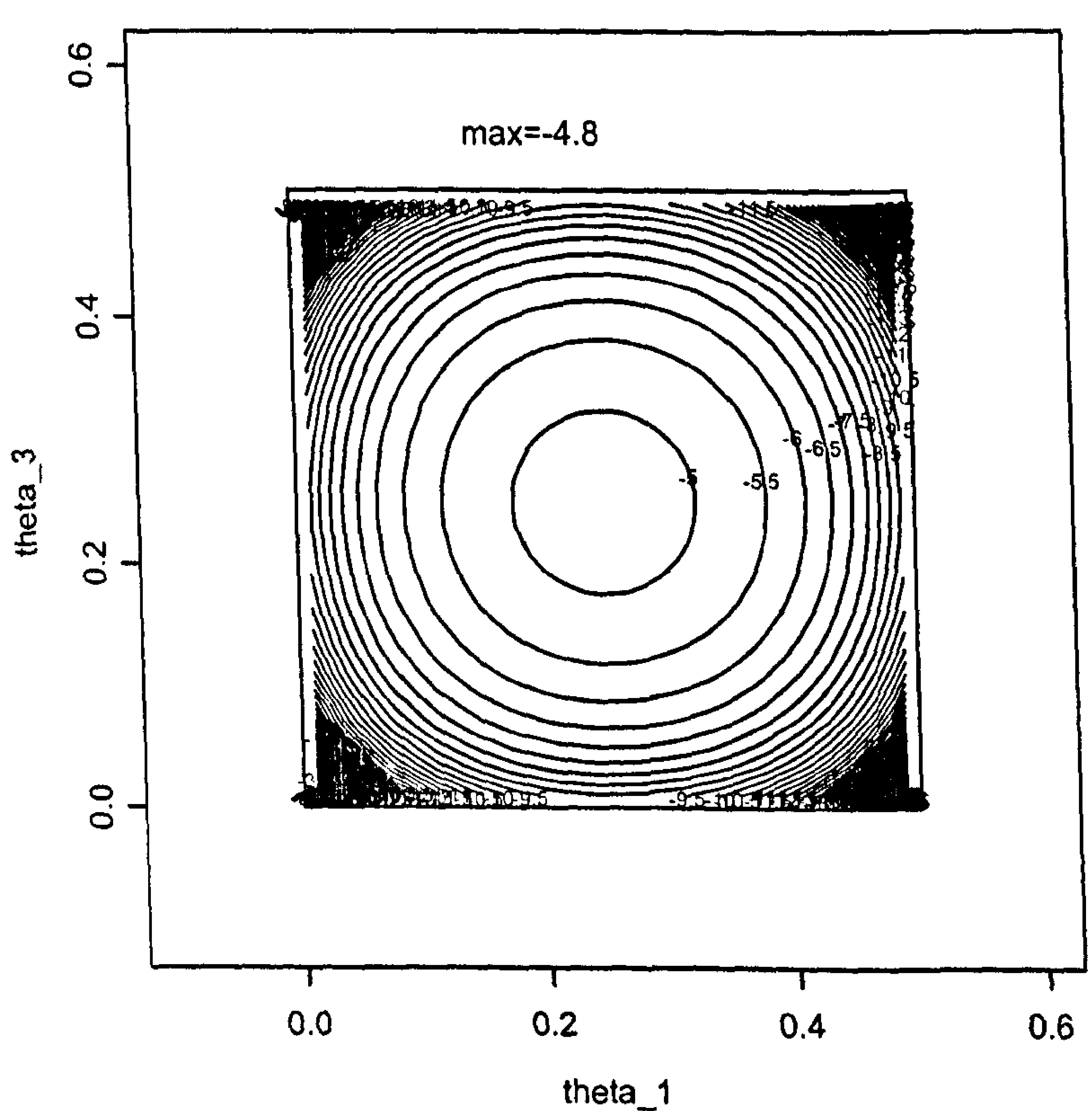


Figure 3.10: Contour plot: double-reciprocal distribution and A-opt



3.5.3 Some comments about the contour plots.

- Normally, the plots do not show the highest level of contours. However, we can roughly determine the relative positions of the optima on the plots.

- For the first two distributions where $k = 3$, the logistic and normal distribution, the results verify the results we obtained on using the search method. The optimal criteria and optimal cutpoints in the two methods are very similar. From the contour plots, we see that the optimal designs should lie on the vertical perpendicular of the equilateral triangular starting from the point ($\theta_2 = 1, \theta_1 = \theta_3 = 0$). The vertex positions of optima also confirm our assumption about the symmetry of the cutpoints, $\theta_1 = \theta_3$, i.e. the optimal results have the form $-z^*, z^*$. It was justifiable to search along perpendicular as we did.

- In the two other triangle contour plots for the symmetric double exponential and double reciprocal distributions, there are interesting results. The plots show that there are two optimal design points in two different positions. These two points are symmetrical with respect to the perpendicular from the top vertex of the triangle. For the double exponential, the optimal criterion value is -1.128, and the two optimal sets of cutpoints are (-1.609,0) and (0,1.609) and the two respective sets of cell probabilities are (0.1,0.4,0.5) and (0.5,0.4,0.1). For the double reciprocal, the optimal criterion value is -2.08, optimal cutpoints are (-1,0) and (0,1) and optimal cell probabilities are (0.25,0.25,0.5) and (0.5,0.25,0.25). We see that the optimal criterion values in these two cases are larger than the ones we found using the search method. The reason is that we assumed symmetric optimal cutpoints in the search method but in fact, by using the graphical approach, the optimal cutpoints are not symmetric.

- In the four-category case, for the rectangular contour plots for all four symmetric distributions, we can see that the optimal designs should lie on the diagonal line of the square, the line satisfying $\theta_1 + \theta_3 = 1/2$. So $\theta_2 = \theta_3$ and $\theta_1 = \theta_4$ and hence, this confirms that our previous assumption of symmetry of the cutpoints is valid

Here are some other contour plots for e_1- and e_2-optimality

Figure 3.11: The contour plot: k=3, 1 point, logistic dist and e_1-opt

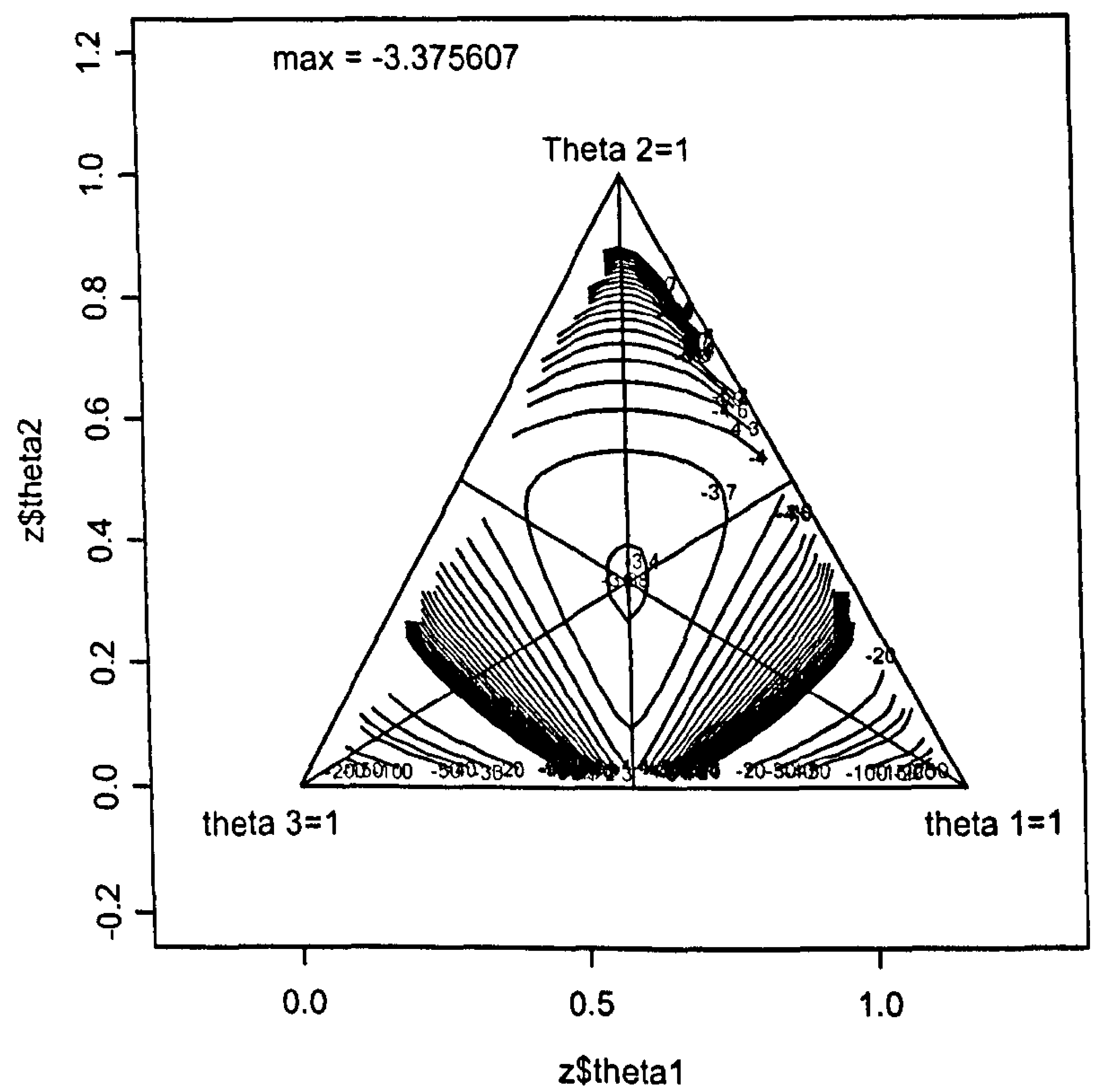


Figure 3.12: The contour plot: k=3, 1 point, normal dist and e_1-opt

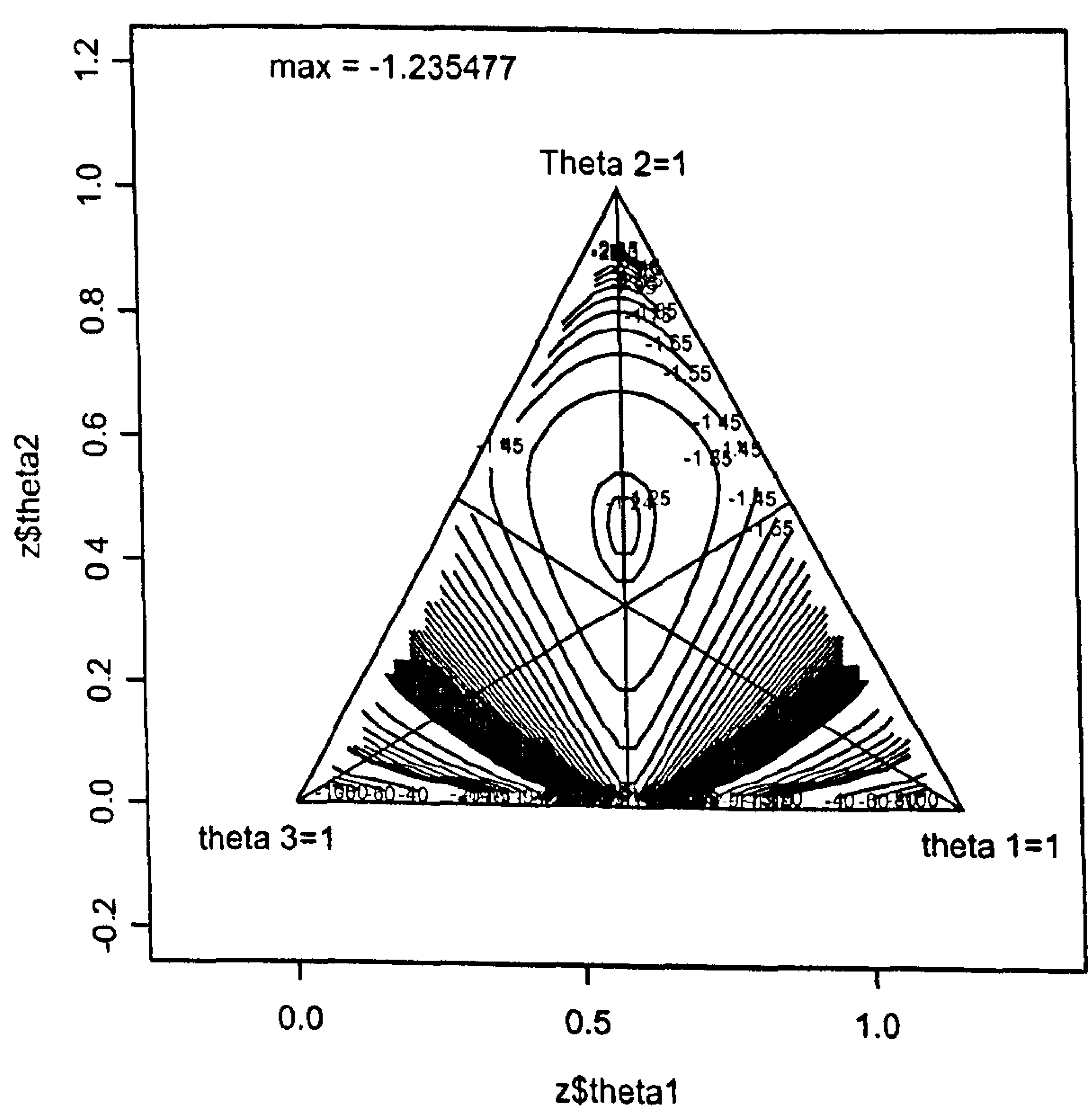


Figure 3.13: The contour plot: k=3, 1 point, double-expo dist and e_1-opt

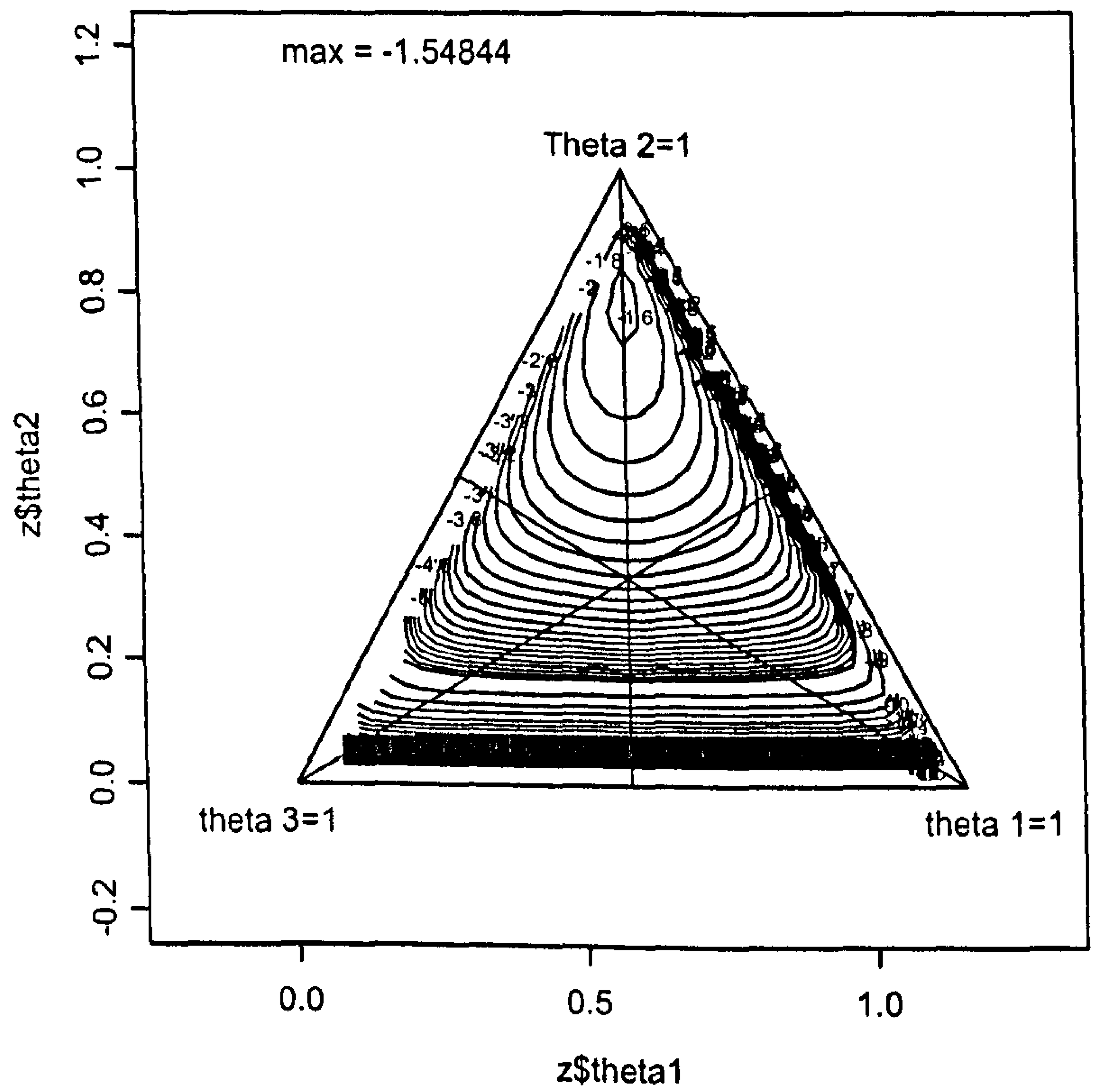


Figure 3.14: The contour plot: k=3, 1 point, double-reciprocal dist and e_1-opt

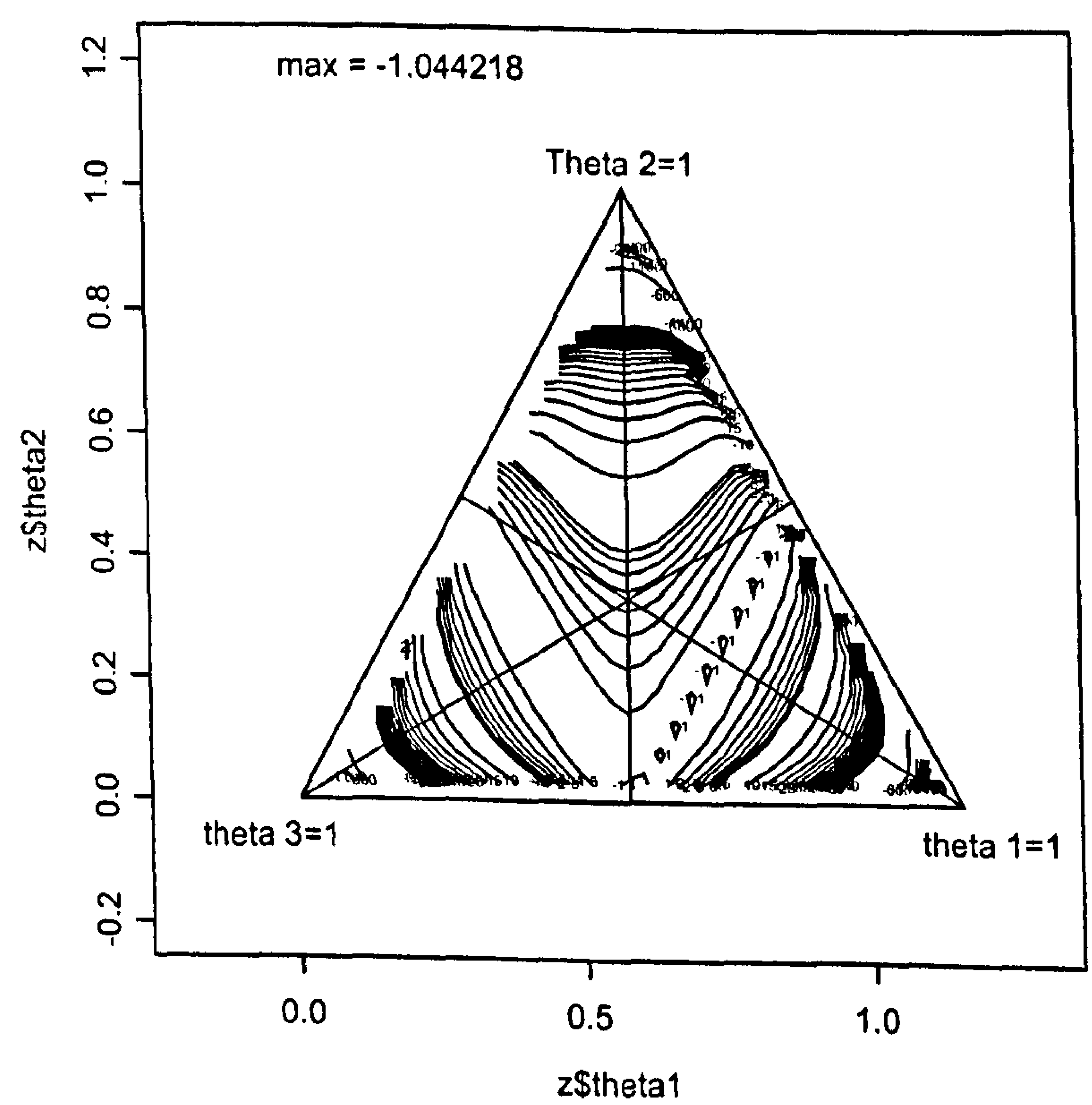


Figure 3.15: The contour plot: logistic distribution and e_2-opt

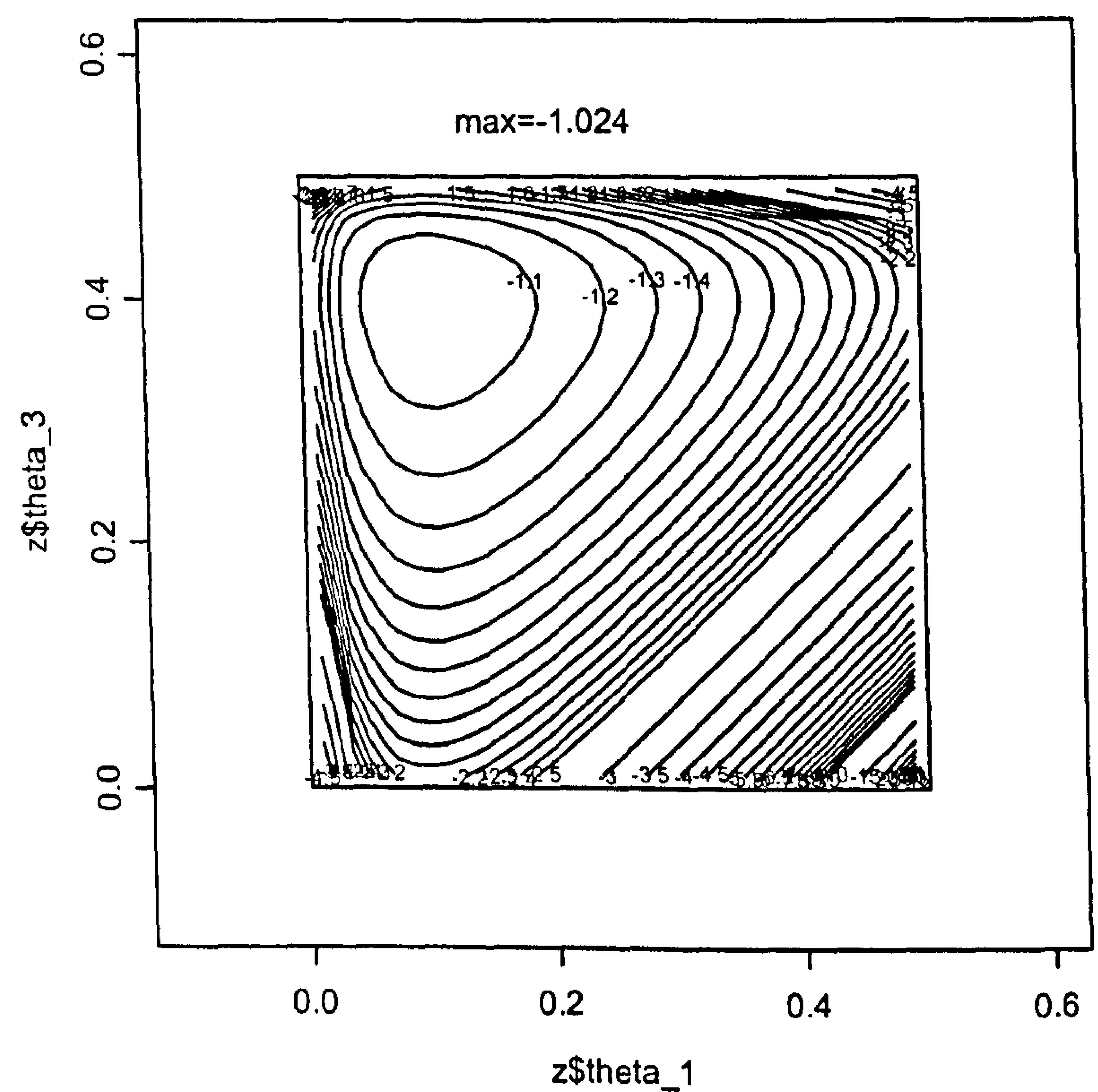


Figure 3.16: The contour plot: normal distribution and e_2-opt

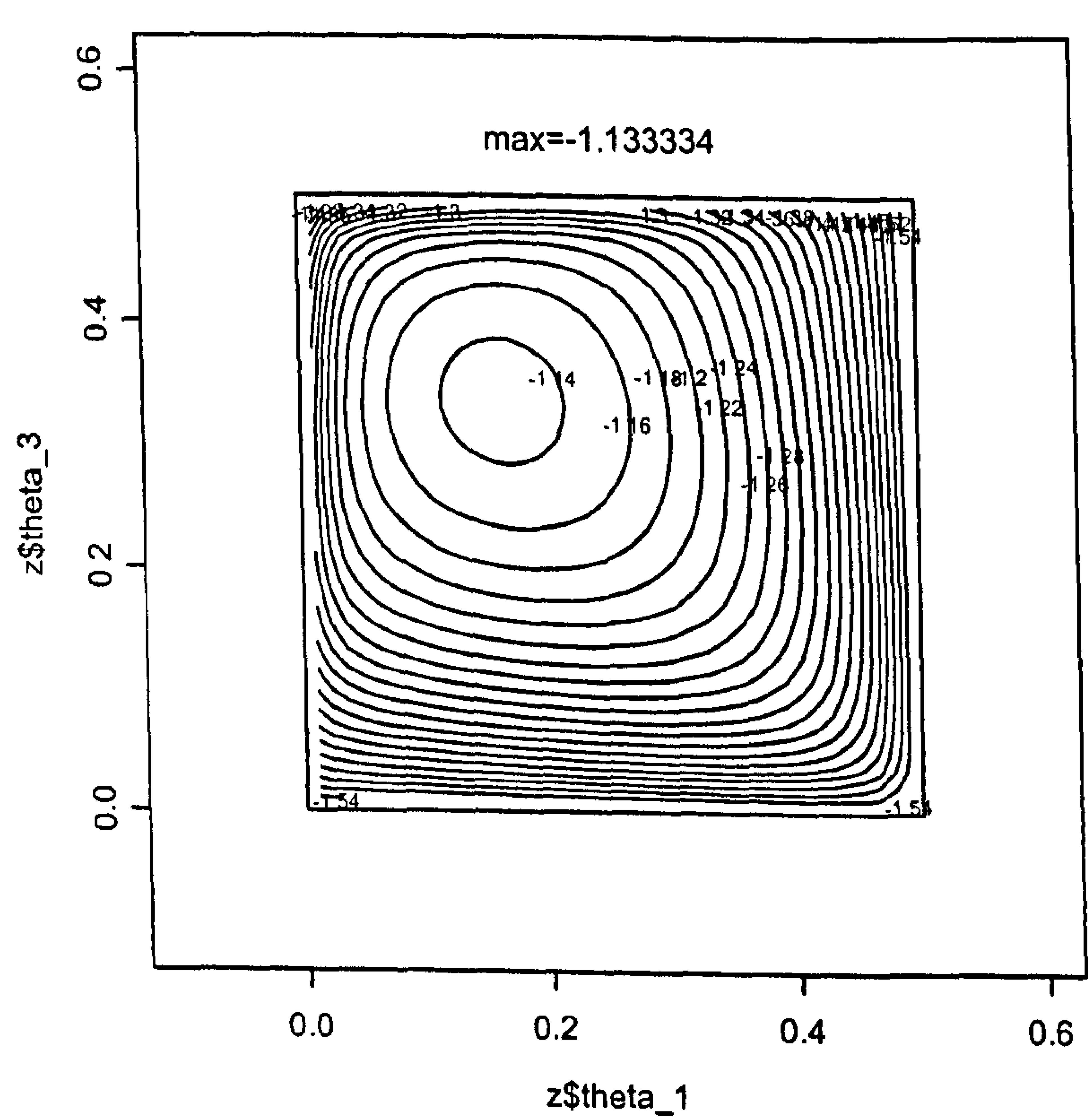


Figure 3.17: The contour plot: double-expo distribution and e_2-opt

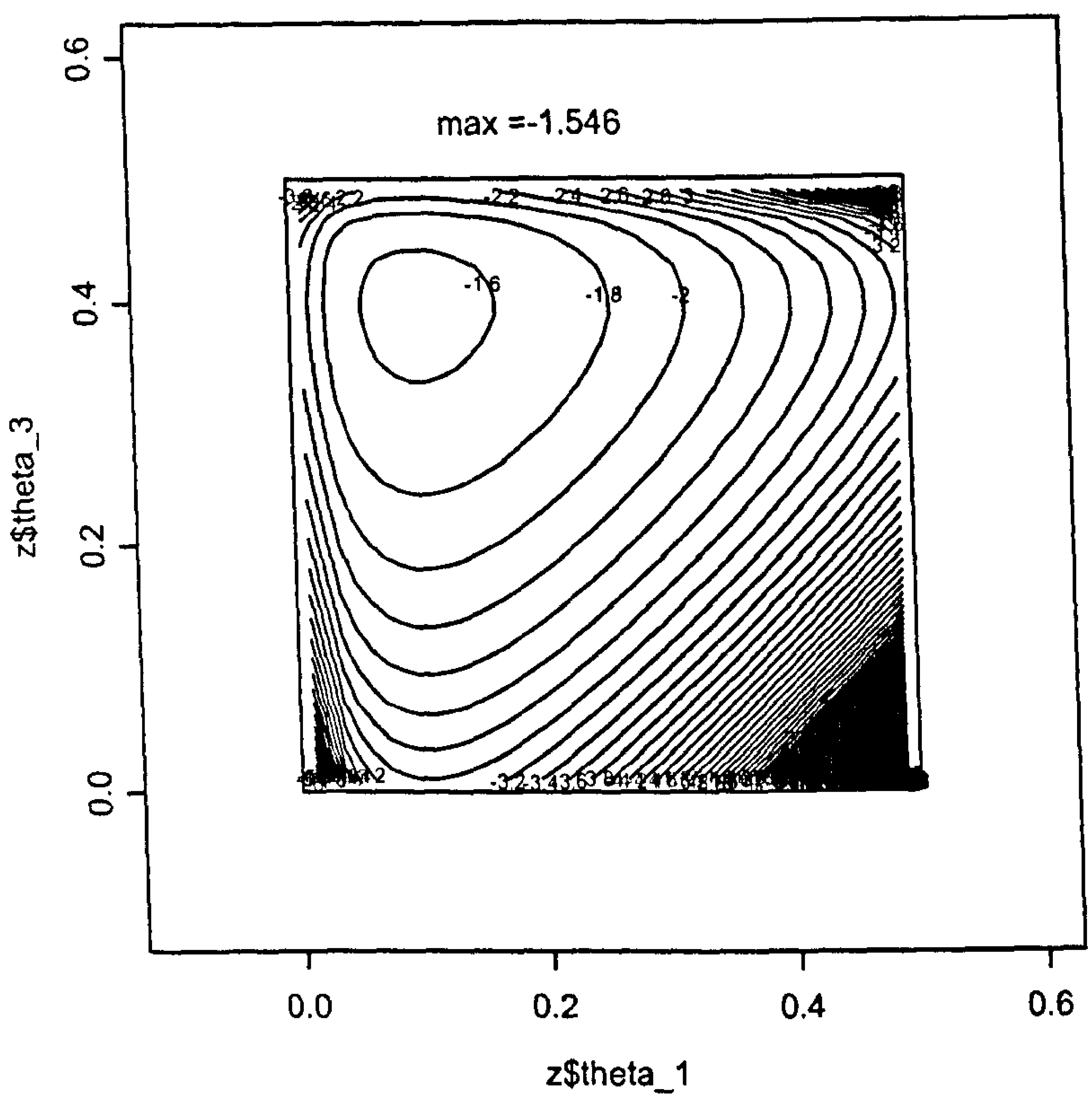
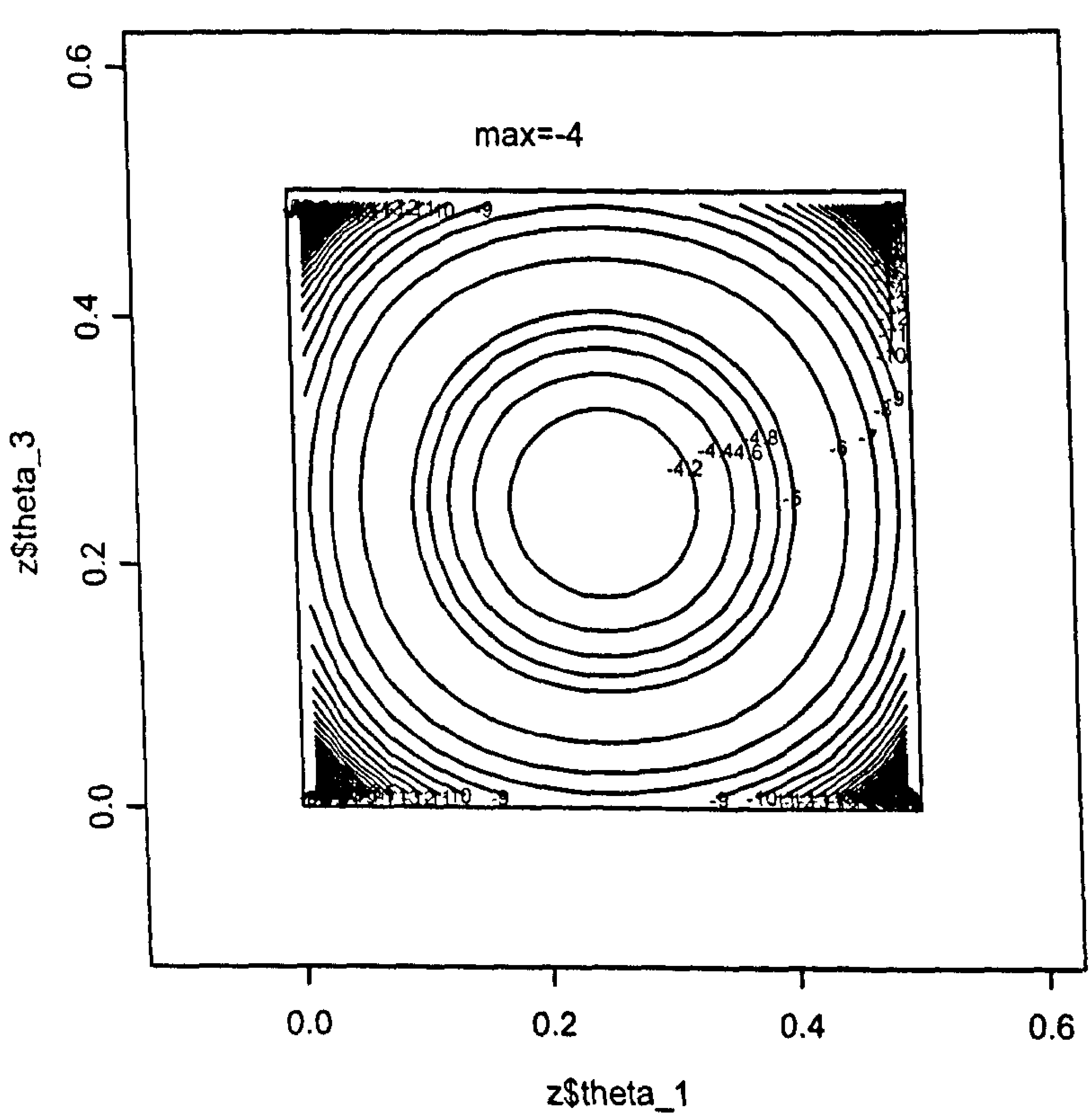


Figure 3.18: The contour plot: double-reciprocal distribution and e_2-opt



3.6 Some results for asymmetric distributions and three category case

In this section, we use the search method to find the optimal solutions for some asymmetric distributions, namely the complementary log-log and skewed logistic. Because the distributions are not symmetric, the cutpoints are not in the symmetric form as well. That is why, we consider the three category case only. The optimal cutpoints in the three category case have the following form:

$$\underset{\sim}{z}^* = (z_1^*, z_2^*)$$

The asymmetric distributions and their cdf, pdf functions are given in the table 3.19.

Table 3.19: Some asymmetric distributions considered

Case	Distribution	$F_i(z)$	$f_i(z)$
5	Complementary log-log	$1 - exp(-exp(z))$	$exp(z - exp(z))$
6-9	Skewed logit	$\{1 + exp(-z)\}^{-m}$	$m\{F_1(z)\}^{m-1} f_1(z)$
6	$m = 1/3$	...	...
7	$m = 2/3$	...	...
8	$m = 3/2$	...	...
9	$m = 3$	...	...

The tables of results and illustrated triangle plots are given below. We can see that the optimal cutpoints are not symmetric. The contour plots also illustrate this.

Table 3.20: Numerical results for D-optimality, k=3

Distribution	z_1^*	z_2^*	$F(z_1^*)$	$F(z_2^*)$	$\phi(z_1^*, z_2^*)$
Com. log-log	-1.3000	0.9500	0.2385	0.9246	-0.3994
Skewed logistic					
$m = 1/3$	-4.2300	0.4900	0.2429	0.8527	-2.5806
$m = 2/3$	-2.1300	1.1300	0.2242	0.8297	-1.9015
$m = 3/2$	-0.8800	1.8400	0.1587	0.8016	-1.2713
$m = 3$	-0.0200	2.4700	0.1212	0.7838	-0.9033

Table 3.21: Numerical results for A-optimality, k=3

Distribution	z_1^*	z_2^*	$F(z_1^*)$	$F(z_2^*)$	$\phi(z_1^*, z_2^*)$
Com. log-log	-1.1733	0.8938	0.2660	0.9132	-2.5909
Skewed logistic					
$m = 1/3$	-4.2400	0.1800	0.2421	0.8167	-10.894
$m = 2/3$	-1.5600	0.9100	0.3112	0.7981	-6.4078
$m = 3/2$	-0.7400	1.5700	0.1835	0.7531	-4.3012
$m = 3$	0.0500	2.9200	0.1346	0.8542	-4.1582

Table 3.22: Numerical results for e_1-optimality, k=3

Distribution	z_1^*	z_2^*	$F(z_1^*)$	$F(z_2^*)$	$\phi(z_1^*, z_2^*)$
Com. log-log	-0.6800	0.5400	0.3974	0.8202	-1.3601
Skewed logistic					
$m = 1/3$	-0.6300	0.8700	0.7030	0.8899	-8.0944
$m = 2/3$	-0.6700	0.7800	0.4857	0.7775	-4.5057
$m = 3/2$	-0.6900	0.5700	0.1930	0.5105	-2.7160
$m = 3$	-0.5800	0.3300	0.0462	0.1968	-2.5848

Table 3.23: Numerical results for e_2-optimality, k=3

Distribution	z_1^*	z_2^*	$F(z_1^*)$	$F(z_2^*)$	$\phi(z_1^*, z_2^*)$
Com. log-log	-1.6900	1.2900	0.1684	0.9735	-0.9278
Skewed logistic					
$m = 1/3$	-5.0500	1.6400	0.1853	0.9426	-1.2162
$m = 2/3$	-3.0000	1.9400	0.1310	0.9143	-1.0835
$m = 3/2$	-1.4600	2.4100	0.0818	0.8789	-0.9882
$m = 3$	-0.5000	2.9800	0.0538	0.8618	-0.9509

Figure 3.19: The contour plot: k=3, 1 point, comp-loglog dist and D-opt

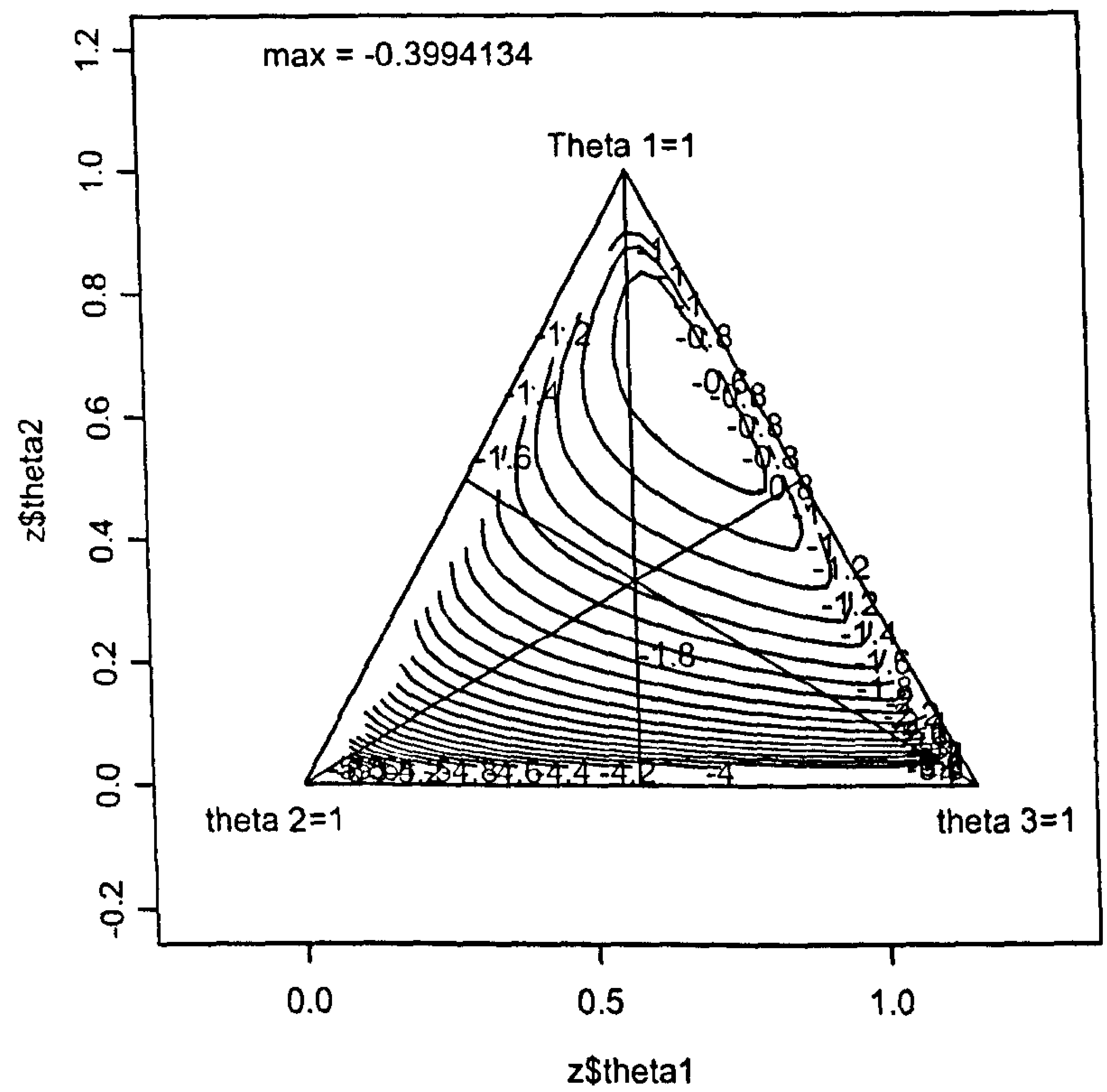


Figure 3.20: The contour plot: k=3, 1 point, skewed-logit dist (m=1/3) and D-opt

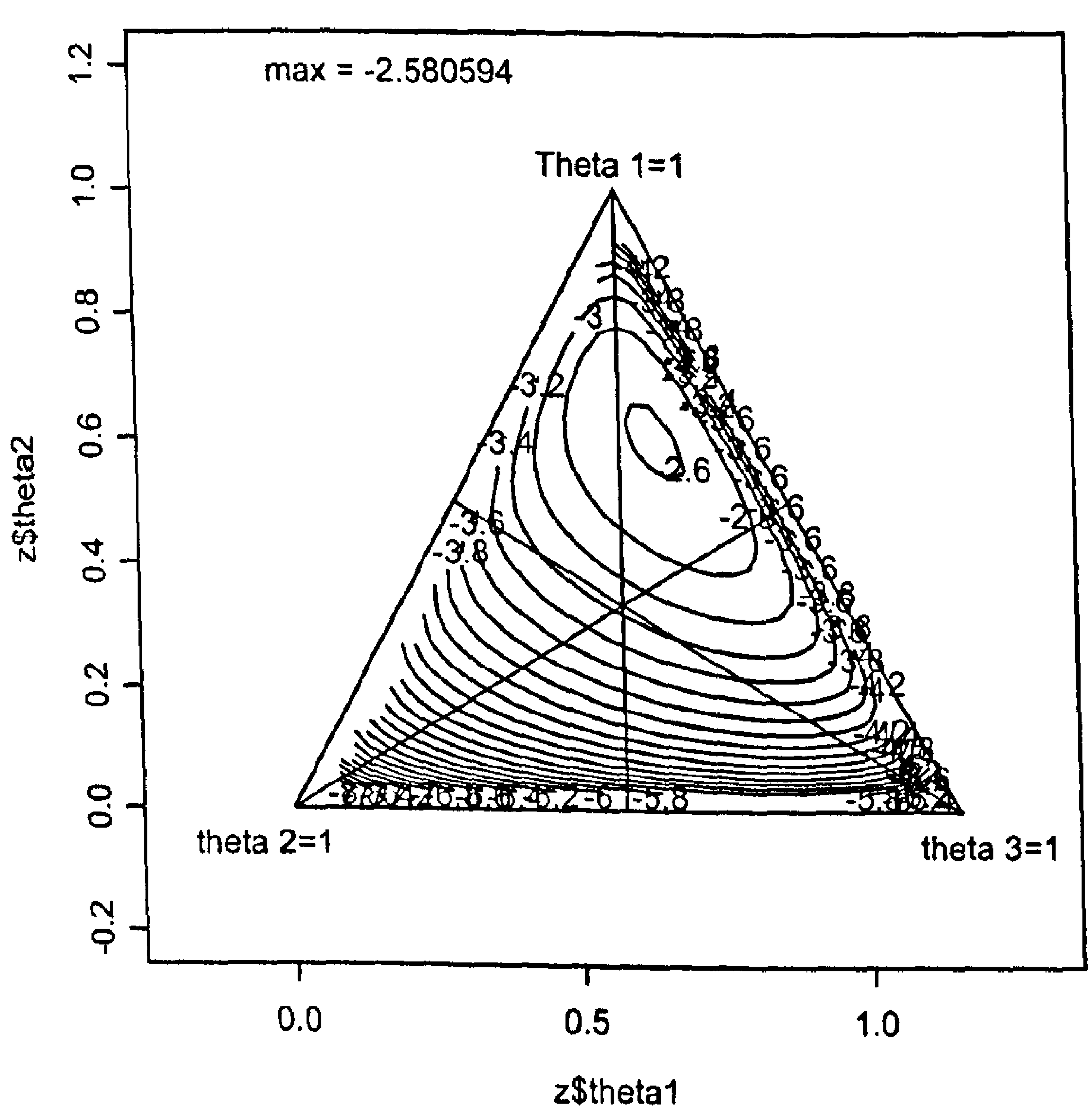


Figure 3.21: The contour plot: k=3, 1 point, comp-loglog dist and e_1-opt

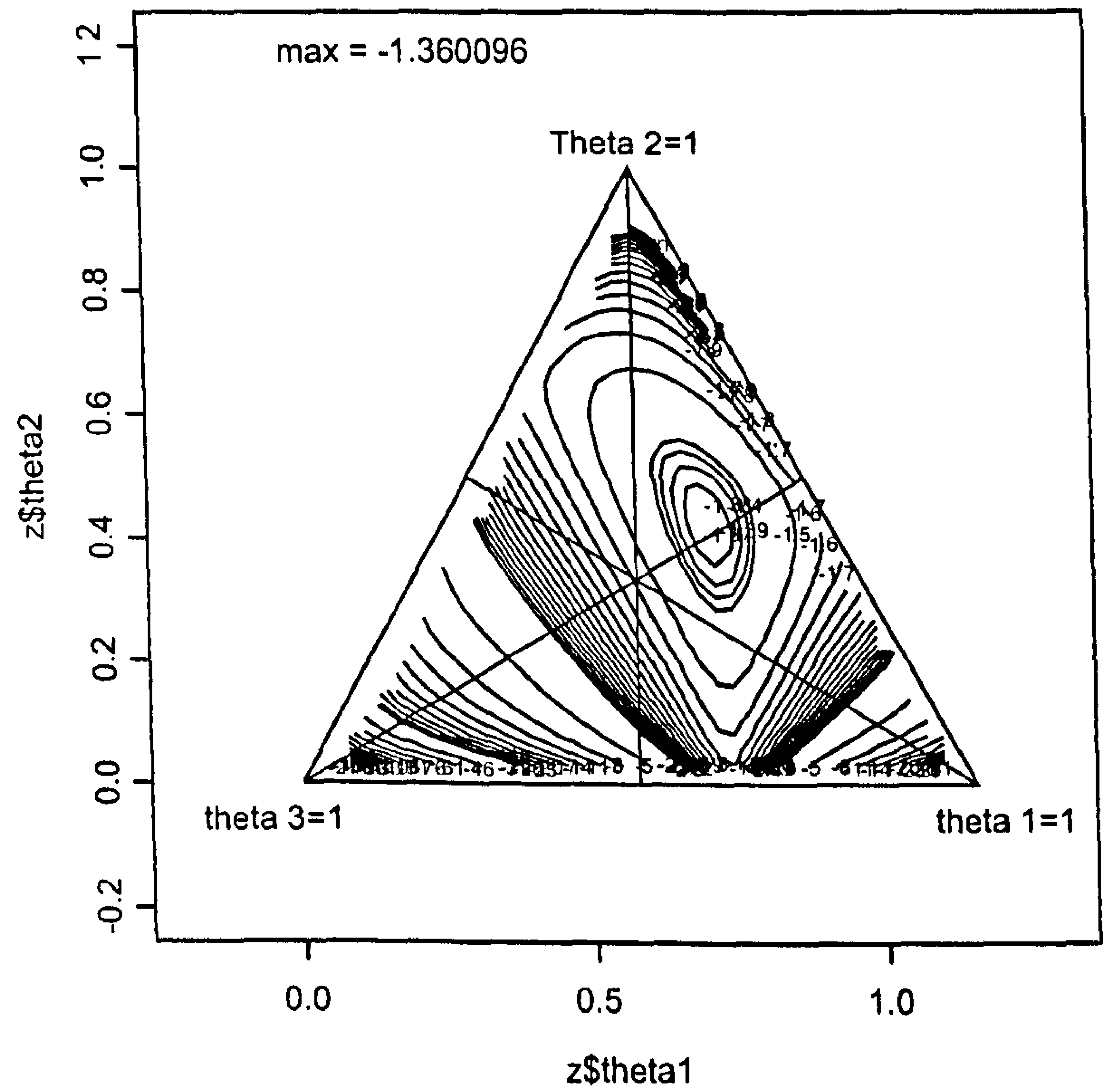
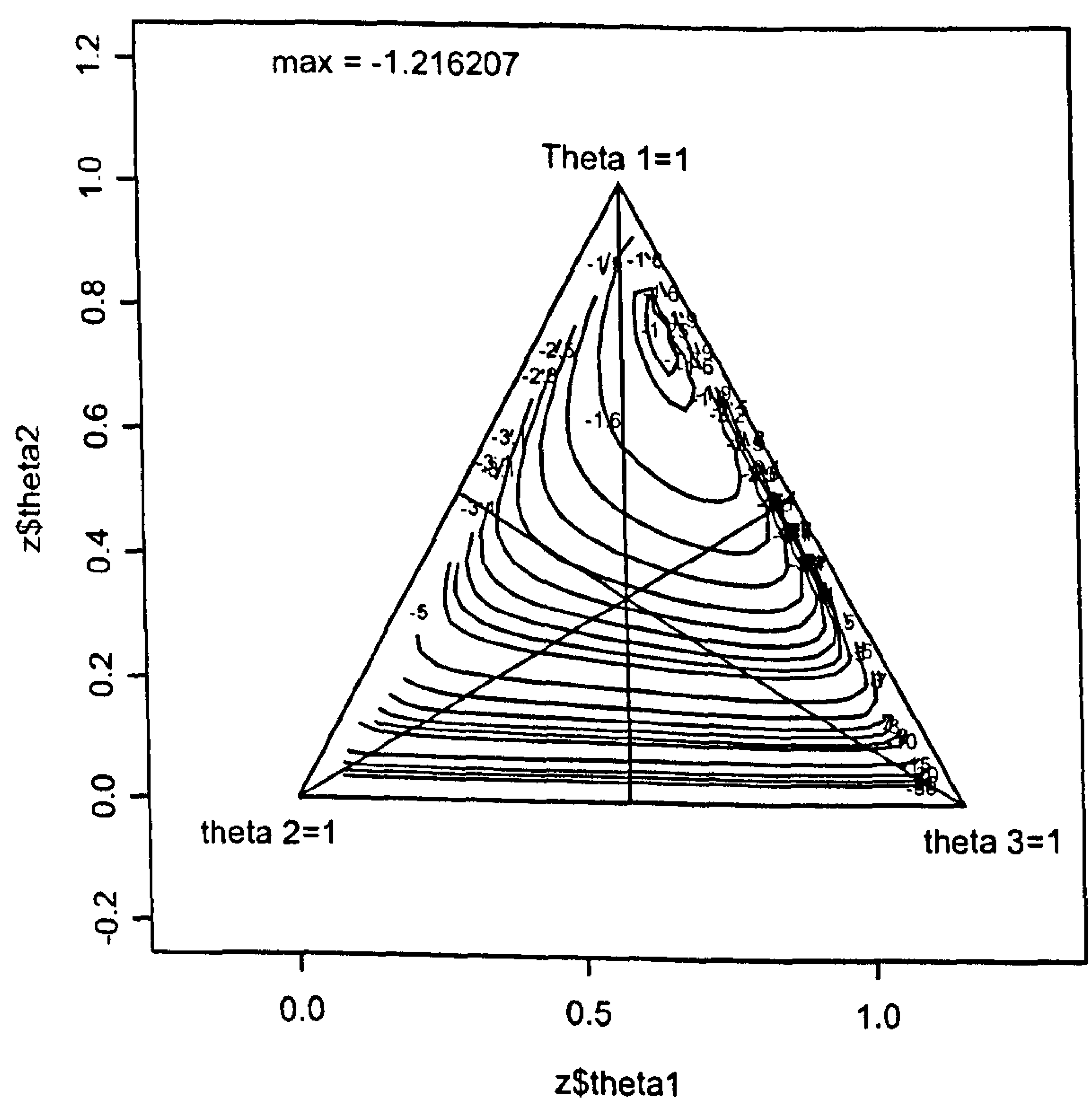


Figure 3.22: The contour plot: k=3, 1 point, skewed-logit dist (m=1/3) and e_2-opt



Chapter 4

A Multiplicative Algorithm for Finding Optimal One Point Designs

In the above chapter, we used search methods to find optimal results for 5 criteria, in the case of 3, 4, 5 and 6 categories and for 4 symmetric distributions. Search methods are limited to a small number of categories. In the case of symmetric distributions, by imposing the assumptions of symmetric cutpoints. we investigate up to the six category cases (i.e. up to two variable optimization problem). In asymmetric distribution cases, search methods are limited to the three category case. For a large number of categories and especially, asymmetric distributions, we need more sophisticated numerical optimization techniques. We need an algorithm. We already used a multiplicative algorithm in chapter three to find the minimum value of the smaller eigenvalue of the difference between two information matrices to check for its non-negative definitiveness. This algorithm will be introduced in more detail in this chapter with its properties and the way of using it in our particular problem (P1). At the end of the chapter, the results from using the algorithm will be presented for both symmetric and asymmetric distributions.

4.1 Introducing the algorithm for the problem (P1)

In constructing optimal design problems, explicit solutions are often not possible, except in some simple cases. Problem (P1), described in chapter one above, can not be solved analytically. That is why numerical techniques such as multiplicative algorithm noted above must be employed. This multiplicative algorithm has been devised for a constrained optimization problem (particularly for the design problem) which requires the calculation of an optimizing probability distribution.

4.1.1 A multiplicative algorithm

Problem (P1) states that we have to choose proportions p_i to maximize some criterion $\phi(p)$ subject to the constraints $p_i \geq 0, \quad \sum_{i=1}^{k} p_i = 1, \quad i = 1, 2, \ldots, k$. The criterion $\phi(p)$ can be a function ψ of the information matrix as we state in chapter one.

In our particular context, an example of problem (P1) turns out to be a transformation of the problem of determining the cut-points $z_1, z_2, \ldots, z_{k-1}$ optimally, namely the problem of determining the cell probabilities $\theta_1, \theta_2, \ldots, \theta_k$ to optimize a criterion (although in the first instance we only have an explicit dependence on $\theta_1, \theta_2, \ldots, \theta_{k-1}$). We can consider this objective problem as another version of problem (P1). This problem has two constraints on θ_i, namely $\theta_i \geq 0, \quad i = 1, 2, \ldots, k, \quad \sum_{i=1}^{k} \theta_i = 1$. An iteration that preserves these constraints and has some suitable properties is the following multiplicative algorithm. As introduced before in chapter three, we can describe this algorithm by the following formula:

$$\theta_j^{(r+1)} = \frac{\theta_j^{(r)} m(x_j^{(r+1)}, \delta)}{\sum_{i=1}^{k} \theta_i^{(r)} m(x_i^{(r)}, \delta)} \tag{4.1}$$

in which:

- δ is a positive free parameter.
- $m(z, \delta)$ is a positive increasing function of z for given δ (e.g. $m(z, \delta) = \Phi(\delta z)$).
- $x_j^{(r)} = d_j^{(r)} = \partial\psi/\partial\theta_j \mid \underline{\theta} = \underline{\theta}^{(r)}$ or
- $x_j^{(r)} = F_j^{(r)} = d_j^{(r)} - \sum \theta_i^{(r)} d_i^{(r)}$, the j^{th} vertex directional derivative.

This kind of iteration was first proposed by Torsney (1977), taking $x = d, m(d, \delta) = d^{\delta}$, with $\delta > 0$. This requires derivatives to be positive. Subsequent empirical studies include Silvey, Titterington and Torsney (1978), which is a study of the choice of δ when $m(d, \delta) = d^{\delta}, \delta > 0$. Torsney (1988) mainly considers $m(d, \delta) = e^{\delta d}$ in a variety of applications, for which one criterion $\phi(d, \delta)$ could have negative derivatives. Torsney and Alahmadi (1992) consider other choices of $m(d, \delta)$. Mandal and Torsney (2002a) consider systematic choices of $m(.,.)$.

Titterington (1976) describes a proof of monotonicity of $m(d, \delta) = d$ in the case of D-optimality. Torsney (1983) explores monoticity of particular values of δ for particular $\phi(p)$. Torsney (1983) also establishes a sufficient condition for monotocity of $m(d, \delta) = d^{\delta}, \delta = 1/(t+1)$ when the criterion $\phi(p)$ is homogeneous of degree $-t, t > 0$ with positive derivatives and proves this condition to hold in the case of linear design criteria such as c-optimality or A-optimality criteria when $t = 1$ so that $\delta = 1/2$.

Convergence results depend on properties of the criterion function $\phi(p)$, on the function $m(z,\delta)$ and on δ. In our case later on, we will consider some standardized criteria mentioned above, take $\delta = 1$ and $m(z,\delta)$ to be of the form of normal cdf function.

4.1.2 Properties of the algorithm

The multiplicative algorithm possesses the following properties considered by Torsney (1988), Torsney and Alahmadi (1992) and Mandal and Torsney (2002a).

1. $\theta^{(r)}$ is always feasible (i.e. $\theta^{(r)} \geq 0$, $\underline{1}^T \underline{\theta}^{(r)} = 1$).

2. $F_\phi\{\theta^{(r)}, \theta^{(r+1)}\} \geq 0$ with equality when the x_j corresponding to nonzero θ_j are equal (in this case $\theta^{(r+1)} = \theta^{(r)}$).

   Let $m(x) = m(x,\delta)$. Consider the equality case where x_j have a common value, say d. We have $x_j = d_j = d$ or $x_j = F_j = 0$. Thus, with $x = d$ or $x = 0$:

$$\begin{aligned} \theta_j^{(r+1)} &= \frac{\theta_j^{(r)} m(x_j)}{\sum_{i=1}^k \theta_i^{(r)} m(x_i)} \\ &= \frac{\theta_j^{(r)} m(x)}{m(x) \sum_{i=1}^k \theta_i^{(r)}} \\ &= \theta_j^{(r)} \end{aligned} \tag{4.2}$$

   Consider the case $x_j = d_j$ or $x_j = F_j$

   The inequality property (i.e. $F_\phi\{\theta^{(r)}, \theta^{(r+1)}\} > 0$) can be seen by letting a positive random variable X take the value $x_j = \dfrac{\partial \phi}{\partial \theta_j} = d_j$ with probability $\theta_j^{(r)}$.

Then:

$$F_\phi\{\theta^{(r)}, \theta^{(r+1)}\} = \frac{Cov\{X, m(X)\}}{E\{m(X)\}} \tag{4.3}$$

Proof:

$$\begin{aligned}
F_\phi\{\theta^{(r)}, \theta^{(r+1)}\} &= \left[\theta^{(r+1)} - \theta^{(r)}\right]^T \underline{d} \\
&= \sum_{i=1}^{k} \left[\theta_i^{(r+1)} - \theta_i^{(r)}\right] d_i \\
&= \sum_{i=1}^{k} \theta_i^{(r+1)} d_i - \sum_{i=1}^{k} \theta_i^{(r)} d_i \\
&= \frac{\sum_{i=1}^{k} \theta_i m(d_i) d_i}{\sum_{i=1}^{k} \theta_i m(d_i)} - \sum_{i=1}^{k} \theta_i d_i \qquad (4.4) \\
&= \frac{\left[\sum_{i=1}^{k} \theta_i m(d_i) d_i\right] - \left[\sum_{i=1}^{k} \theta_i d_i\right]\left[\sum_{i=1}^{k} \theta_i m(d_i)\right]}{\sum_{i=1}^{k} \theta_i m(d_i)} \qquad (4.5)
\end{aligned}$$

i.e.

$$F_\phi\{\theta^{(r)}, \theta^{(r+1)}\} = \frac{Cov\{X, m(X)\}}{E\{f(X)\}}$$

The argument then is that the covariance between X and $m(X)$ must be non-negative if $m(X)$ is increasing in X. Thus an increase in the criterion can be obtained by a partial but possibly not a full step from $\theta^{(r)}$ in the direction of $\theta^{(r+1)}$

3. An iterate $\theta^{(r)}$ is a fixed point of the iteration if the derivative $\partial\phi/\partial\theta_j^{(r)}$ corresponding to non-zero $\theta^{(r)}$ are all equal. This is a necessary but not a sufficient condition for $\theta^{(r)}$ to solve problem (P1).

4.2 Using the algorithm to find optimal one point designs

We now formulate the algorithm in our particular context.

In the formula for the algorithm, we will choose $x_j^{(r)}$ to be:

$$x_j = \frac{\partial \psi}{\partial \theta_j} - \sum_{i=1}^{k} \theta_i \frac{\partial \psi}{\partial \theta_i}, \quad j = 1, 2, \ldots, k. \tag{4.6}$$

Note that the function ψ is a criterion function as defined in problem (P1). We will use the function $m(x, \delta) = \Phi(\delta x)$, (the normal cdf function) because it is a symmetric function, has the range of value from 0 to 1 and the value of 0.5 when $x = 0$. In our case, we use $\delta = 1$. Now we need to determine the partial derivatives $\partial\psi/\partial\theta_i$

At the moment, we make clear our notation:

In the first instance, our criteria depend on $k-1$ cutpoints $(z_1, z_2, \ldots, z_{k-1})$.

Let:

$$\underline{z} = (z_1, z_2, \ldots, z_{k-1})$$

and

$$\underline{\tilde{\theta}} = (\theta_1, \theta_2, \ldots, \theta_{k-1}),$$

the first $k-1$ cell probabilities.

$\underline{\tilde{\theta}}$ and $\underline{z}$ are one to one related in the following way:

$\theta_1 = F(z_1) \rightarrow z_1 = F^{-1}(\theta_1)$

$\theta_2 = F(z_2) - F(z_1) \rightarrow F(z_2) = \theta_1 + \theta_2 \rightarrow z_2 = F^{-1}(\theta_1 + \theta_2)$

$\theta_3 = F(z_3) - F(z_2) \rightarrow F(z_3) = \theta_1 + \theta_2 + \theta_3 \rightarrow z_3 = F^{-1}(\theta_1 + \theta_2 + \theta_3)$

$\vdots$

$\theta_{k-1} = F(z_{k-1}) - F(z_{k-2}) \rightarrow F(z_{k-1}) = \theta_1 + \theta_2 + \ldots + \theta_{k-1} \rightarrow z_{k-1} = F^{-1}(\theta_1 + \theta_2 + \ldots + \theta_{k-1})$

Note that any criterion $\psi(\underline{z})$ can be transformed to a criterion, say $\tilde{\psi}(\underline{\tilde{\theta}}) = \psi\{\underline{z}(\underline{\tilde{\theta}})\}$ which explicitly depends on the first $k-1$ cell probabilities. It does not depend on θ_k:

$\theta_k = 1 - F(z_{k-1}) = 1 - \sum_{j=1}^{k-1} \theta_j$

In general we have $z_i = F^{-1}(\sum_{j=1}^{i} \theta_j)$, $\quad i = 1, 2, \ldots, k-1$.

Since:

$$\frac{\partial \tilde{\psi}(\underline{\tilde{\theta}})}{\partial \theta_i} = \frac{\partial \psi\{\underline{z}(\underline{\tilde{\theta}})\}}{\partial \theta_i} = \sum_{j=1}^{k-1} \frac{\partial \psi}{\partial z_j} \frac{\partial z_j}{\partial \theta_i}, \quad i = 1, 2, \ldots, k-1. \tag{4.7}$$

we have:

$$\frac{\partial \tilde{\psi}}{\partial \underline{\theta}} = A \frac{\partial \psi}{\partial \underline{z}}, \quad \underline{z} = (z_1, z_2, \ldots, z_{k-1}); \quad \underline{\theta} = (\theta_1, \theta_2, \ldots, \theta_{k-1}) \tag{4.8}$$

where:

$$A = \begin{pmatrix} \frac{\partial z_1}{\partial \theta_1} & \frac{\partial z_2}{\partial \theta_1} & \frac{\partial z_3}{\partial \theta_1} & \cdots & \frac{\partial z_{k-1}}{\partial \theta_1} \\ 0 & \frac{\partial z_2}{\partial \theta_2} & \frac{\partial z_3}{\partial \theta_2} & \cdots & \frac{\partial z_{k-1}}{\partial \theta_2} \\ 0 & 0 & \frac{\partial z_3}{\partial \theta_3} & \cdots & \frac{\partial z_{k-1}}{\partial \theta_3} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & \frac{\partial z_{k-1}}{\partial \theta_{k-1}} \end{pmatrix} \tag{4.9}$$

The computation of $\partial\psi/\partial\underline{z}$ depends on particular criteria. Below are the formulae for $\partial\psi/\partial\underline{z}$ for some criteria considered:

Assume that $\psi = \psi(I_z)$, where I_z is the Fisher information matrix.

1. D-optimality : $\psi(I_z) = log\, det(I_z) \;\rightarrow\; \frac{\partial \psi}{\partial z_j} = tr\left(I_z^{-1} \frac{\partial I_z}{\partial z_j}\right)$

2. A-optimality : $\psi(I_z) = -tr(I_z^{-1}) \;\rightarrow\; \frac{\partial \psi}{\partial z_j} = tr\left(I_z^{-2} \frac{\partial I_z}{\partial z_j}\right)$

3. e_1-optimality : $\psi(I_z) = -e_1^T I_z^{-1} e_1 \;\rightarrow\; \frac{\partial \psi}{\partial z_j} = e_1^T\, I_z^{-1} \frac{\partial I_z}{\partial z_j} I_z^{-1} e_1$

4. e_2-optimality : $\psi(I_z) = -e_2^T I_z^{-1} e_2 \;\rightarrow\; \frac{\partial\psi}{\partial z_j} = e_2^T\, I_z^{-1}\, \frac{\partial I_z}{\partial z_j}\, I_z^{-1}\, e_2.$

where: $e_1 = (1,0)^T$ and $e_2 = (0,1)^T$

We see that in all cases, $\frac{\partial\psi}{\partial z_j}$ depends on I_z^{-1} and $\frac{\partial I_z}{\partial z_j}$.
We need to calculate $\frac{\partial I_z}{\partial z_j}$.
From formula (3.14), we know:

$$I_z = Z D_f H D_\theta^{-1} H^T D_f Z^T$$

Using the product rule for derivatives, we have:

$$\begin{aligned}
\frac{\partial I_z}{\partial z_j} &= \left(\frac{\partial Z}{\partial z_j} D_f H D_\theta^{-1} H^T D_f Z^T\right) \\
&+ \left(Z \frac{\partial D_f}{\partial z_j} H D_\theta^{-1} H^T D_f Z^T\right) \\
&+ \left(Z D_f H \frac{\partial D_\theta^{-1}}{\partial z_j} H^T D_f Z^T\right) \\
&+ \left(Z D_f H D_\theta^{-1} H^T \frac{\partial D_f^T}{\partial z_j} Z^T\right) \\
&+ \left(Z D_f H D_\theta^{-1} H^T D_f \frac{\partial Z^T}{\partial z_j}\right)
\end{aligned}$$

where:

$$Z = \begin{pmatrix} 1 & 1 & \ldots & 1 & \ldots & 1 & 1 \\ z_1 & z_2 & \ldots & z_t & \ldots & z_{k-2} & z_{k-1} \end{pmatrix}$$

Then:

$$\frac{\partial Z}{\partial z_t} = \begin{pmatrix} 0 & 0 & \ldots & 0 & \ldots & 0 & 0 \\ 0 & 0 & \ldots & 1 & \ldots & 0 & 0 \end{pmatrix}$$

and:

$$D_f = diag\Big(f(z_1), f(z_2), \ldots, f(z_t), \ldots, f(z_{k-1})\Big)$$

Then:

$$\frac{\partial D_f}{\partial z_t} = diag\Big(0, 0, \ldots, f'(z_t), \ldots, 0\Big)$$

and:

$$\frac{\partial D_\theta^{-1}}{\partial z_j} = -D_\theta^{-2}\frac{\partial D_\theta}{\partial z_j} = -\frac{1}{D_\theta^2}\frac{\partial D_\theta}{\partial z_j}$$

$$\begin{aligned} D_\theta &= diag(\theta_1, \theta_2, \ldots, \theta_k) \\ &= diag\Big(F(z_1), F(z_2) - F(z_1), \ldots, 1 - F(z_{k-1})\Big) \end{aligned}$$

Then:

$$\begin{aligned} \frac{\partial D_\theta}{\partial z_1} &= f(z_1)\Big[diag(1, -1, 0, 0, \ldots, 0)\Big] \\ \frac{\partial D_\theta}{\partial z_2} &= f(z_2)\Big[diag(0, 1, -1, 0, \ldots, 0)\Big] \\ &\vdots \\ \frac{\partial D_\theta}{\partial z_{k-1}} &= f(z_{k-1})\Big[diag(0, 0, 0, 0, \ldots, 1, -1)\Big] \end{aligned}$$

Note:

- The matrix H does not depend on z
- $f'(z_j)$ is the derivative of $f(z_j)$ or second derivative of $F(z_j)$. They have formulae which depend on particular distributions of Z.

In the above formula, we have calculated $\frac{\partial \tilde{\psi}}{\partial \theta_i}$ for $i = 1, 2, \ldots, k-1$ since $\tilde{\psi}(\underline{\tilde{\theta}})$ only depended on these cell probabilities.

We now generate a function $\Psi(\theta_1, \theta_2, \ldots, \theta_k)$ with an explicit dependence on

θ_k.

$$\begin{aligned}
\Psi\Big(\theta_1,\theta_2,\ldots,\theta_k\Big) &= \frac{1}{k}\Bigg[\tilde{\psi}\Big(\theta_1,\theta_2,\ldots,\theta_{k-1}\Big) \\
&+ \tilde{\psi}\Big(1-\sum_{j=2}^{k}\theta_j,\theta_2,\ldots,\theta_{k-1}\Big) \\
&+ \tilde{\psi}\Big(\theta_1,1-\theta_1-\sum_{j=3}^{k}\theta_j,\ldots,\theta_{k-1}\Big) \\
&+ \ldots \\
&+ \tilde{\psi}\Big(\theta_1,\theta_2,\ldots,\theta_{k-2},1-\sum_{j=1}^{k-2}\theta_j-\theta_k\Big)\Bigg] \qquad (4.10)
\end{aligned}$$

Then:

$$\begin{aligned}
\frac{\partial\Psi}{\partial\theta_1} &= \frac{1}{k}\left[\frac{\partial\tilde{\psi}}{\partial\theta_1}+0+\left(\frac{\partial\tilde{\psi}}{\partial\theta_1}-\frac{\partial\tilde{\psi}}{\partial\theta_2}\right)+\ldots+\left(\frac{\partial\tilde{\psi}}{\partial\theta_1}-\frac{\partial\tilde{\psi}}{\partial\theta_{k-1}}\right)\right] \\
&= \frac{1}{k}\left[(k-1)\frac{\partial\tilde{\psi}}{\partial\theta_1}-\sum_{j=2}^{k-1}\frac{\partial\tilde{\psi}}{\partial\theta_j}\right] \\
&= \frac{1}{k}\left[k\frac{\partial\tilde{\psi}}{\partial\theta_1}-\sum_{j=1}^{k-1}\frac{\partial\tilde{\psi}}{\partial\theta_j}\right] \\
&= \frac{\partial\tilde{\psi}}{\partial\theta_1}-\frac{1}{k}\sum_{j=1}^{k-1}\frac{\partial\tilde{\psi}}{\partial\theta_j} \qquad (4.11)
\end{aligned}$$

Similarly:

$$\frac{\partial\Psi}{\partial\theta_j}=\frac{\partial\tilde{\psi}}{\partial\theta_j}-\frac{1}{k}\sum_{j=1}^{k-1}\frac{\partial\tilde{\psi}}{\partial\theta_j},\quad j=1,2,\ldots,k-1 \qquad (4.12)$$

and for θ_k:

$$\frac{\partial\Psi}{\partial\theta_k}=0-\frac{1}{k}\sum_{j=1}^{k-1}\frac{\partial\tilde{\psi}}{\partial\theta_j},\quad \text{since } \frac{\partial\tilde{\psi}}{\partial\theta_k}=0 \qquad (4.13)$$

Note that there is an alternative definition of $\Psi\Big(\theta_1, \theta_2, \ldots, \theta_k\Big)$ as follows:

$$\begin{aligned}
\Psi\Big(\theta_1, \theta_2, \ldots, \theta_k\Big) &= \theta_k \tilde{\psi}\Big(\theta_1, \theta_2, \ldots, \theta_{k-1}\Big) \\
&+ \theta_1 \tilde{\psi}\Big(1 - \sum_{j=2}^{k} \theta_j, \theta_2, \ldots, \theta_{k-1}\Big) \\
&+ \theta_2 \tilde{\psi}\Big(\theta_1, 1 - \theta_1 - \sum_{j=3}^{k} \theta_j, \ldots, \theta_{k-1}\Big) \\
&+ \ldots \\
&+ \theta_{k-1} \tilde{\psi}\Big(\theta_1, \theta_2, \ldots, \theta_{k-2}, 1 - \sum_{j=1}^{k-2} \theta_j - \theta_k\Big) \quad (4.14)
\end{aligned}$$

Using this definition, we recover the same formula for $\frac{\partial \Psi}{\partial \theta_j}$, $j = 1, 2, \ldots, k$

We note that:

$$\frac{\partial \Psi}{\partial \theta_j} = \frac{\partial \tilde{\psi}}{\partial \theta_j} - \mathit{constant}$$

In fact, $\tilde{\psi}$ and Ψ have the same F_j's, i.e. directional derivatives.

4.3 The results

In this section, we use the multiplicative algorithm first to verify the results obtained using the search method for symmetric distributions and asymmetric distributions (three category case), and secondly to find the optimal set of cutpoints and respective set of cell probabilities for asymmetric cases.

4.3.1 Symmetric distributions

In chapter three, using search methods for symmetric distributions, we assumed that the optimal cutpoints have symmetric forms. Using the algorithm, it is not necessary to impose this assumption. Arbitrary cutpoints will be used. However, if the initial cell probabilities are equal and the distribution is symmetric, then the cutpoints in all iterations are also symmetric.

We start the algorithm from both equal and unequal initial cell probabilities. We now construct the formula for the derivatives of the pdf function for some symmetric distributions.

- Logistic distribution.

$$F(z) = \frac{e^z}{1+e^z}$$

$$f(z) = F'(z) = F(z)[1 - F(z)]$$

$$f'(z) = F(z)[1 - F(z)][1 - 2F(z)]$$

- Normal distribution.

$$F(z) = \Phi(z) = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} e^{-u^2/2} du$$

$$f(z) = F'(z) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2}$$

$$f'(z) = -z\frac{1}{\sqrt{2\pi}} e^{-z^2/2} = -zf(z)$$

- Double exponential.

$$f(z) = \frac{1}{2} e^{-|z|}$$

$$f'(z) = \frac{-s}{2} e^{-|z|}$$

Note that: $s = sign(z)$ and $f'(0) = 0$

- Double reciprocal.

$$f(z) = \frac{1}{2}(1 + |z|)^{-2}$$

$$f'(z) = -s(1 + |z|)^{-3}$$

Note that: $s = sign(z)$ and $f'(0) = 0$

Running the algorithm to verify the results we obtained by using the search method, we see that for the four symmetric distributions, namely the logistic, normal, double-exponential and double reciprocal, the results are very consistent. The two sets of results are very similar. The algorithm converges very well (normally only fewer than 1000 iterations needed). However, the larger the number of categories is, the slower is the convergence of the algorithm. For example, in the case of the logistic distribution and D-optimality, we need about 300, 450, 700 and 960 iterations for the three, four, five and six category cases respectively for convergence. There are several elements which we can use to check for the convergence of the algorithm such as directional derivatives or the values of cutpoints and criteria. In our cases, we consider that the algorithm has converged if all the directional derivatives reach the values which are less than 10^{-6}.

We use either equal or unequal initial cell probabilities to start the algorithm. In the case of equal starting cell probabilities, the initial values $\theta_j^0 = 1/k$ for θ_j, $j = 1, 2, \ldots, k$ with k being the number of categories. In the unequal case, we choose the initial cell probabilities arbitrarily providing that they sum up to 1. For example, we can choose the initial set of cell probabilities (0.2, 0.3, 0.5) for the three category case or (0.1, 0.2, 0.3, 0.25, 0.15) for the five category case. We quote here the results we obtained by using the algorithm in the case of the logistic distribution and D-optimality for the number of categories running from three to six. In the tables below, the first column contains the number of iterations needed for convergence. The second column presents the directional derivatives in each iteration. The third and the last column show the cutpoints and criterion value in each iteration. Using both equal and unequal initial cell probabilities, we obtain the same optimal results but after different numbers of iterations. Note we explored other starting values but the algorithm always converges to same value.

Table 4.1: Solution: Logistic distribution, *D*-opt and k=3

Iteration	Directional derivatives	Cutpoints	Criterion value
316	-0.00000073	-1.46717329	-1.55667658
	0.00000043	1.46717329	
	-0.00000073		
317	-0.00000071	-1.46718044	-1.55667658
	0.00000042	1.46718044	
	-0.00000071		
318	-0.00000069	-1.46718740	-1.55667658
	0.00000041	1.46718740	
	-0.00000069		
319	-0.00000067	-1.46719418	-1.55667657
	0.00000040	1.46719418	
	-0.00000067		
320	-0.00000065	-1.46720078	-1.55667657
	0.00000039	1.46720078	
	-0.00000065		

Table 4.2: Solution: Logistic distribution, D-opt and k=4

Iteration	Directional derivatives	Cutpoints	Criterion value
430	-0.0000011	-1.97916827	-1.24833985
	0.00000035	0	
	0.00000035	1.97916827	
	-0.0000011		
431	-0.0000011	-1.97917821	-1.24833984
	0.00000034	0	
	0.00000034	1.97917821	
	-0.0000011		
432	-0.0000011	-1.97918797	-1.24833984
	0.00000034	0	
	0.00000034	1.97918797	
	-0.0000011		
433	-0.0000010	-1.97919756	-1.24833984
	0.00000033	0	
	0.00000033	1.97919756	
	-0.0000010		
434	-0.0000010	-1.97920698	-1.24833983
	0.00000033	0	
	0.00000033	1.97920698	
	-0.0000010		

Table 4.3: Solution: Logistic distribution, D-opt and k=5

Iteration	Directional derivatives	Cutpoints	Criterion value
699	-0.0000038	-2.50419943	-1.07096211
	-0.0000038	-0.84051687	
	0.0000058	0.84051687	
	-0.0000038	2.50419943	
	-0.0000038		
700	-0.0000038	-2.50423229	-1.07096193
	-0.0000038	-0.84056083	
	0.0000058	0.84056083	
	-0.0000038	2.50423229	
	-0.0000038		
701	-0.0000038	-2.50426492	-1.07096175
	-0.0000038	-0.84060450	
	0.0000058	-0.84060450	
	-0.0000038	2.50426492	
	-0.0000038		
702	-0.0000037	-2.50429732	-1.07096158
	-0.0000038	-0.84064787	
	-0.0000057	0.84064787	
	-0.0000038	2.50429732	
	-0.0000037		

Table 4.4: Solution: Logistic distribution, D-opt and k=6

Iteration	Directional derivatives	Cutpoints	Criterion value
961	-0.0000019	-2.90566517	-0.97876512
	-0.0000017	-1.33064179	
	0.0000012	0	
	0.0000012	1.33064179	
	-0.0000017	2.90566517	
	-0.0000019		
962	-0.0000019	-2.90568127	-0.97876511
	-0.0000017	-1.33065938	
	0.0000012	0	
	0.0000012	1.33065938	
	-0.0000017	2.90568127	
	-0.0000019		
963	-0.0000019	-2.90569727	-0.97876501
	-0.0000017	-1.33067685	
	0.0000012	0	
	0.0000012	1.33067685	
	-0.0000017	2.90569727	
	-0.0000019		
964	-0.0000019	-2.90571317	-0.97876507
	-0.0000017	-1.33069422	
	0.0000012	0	
	0.0000012	1.33069422	
	-0.0000017	2.90571317	
	-0.0000019		

We can see that the vertex directional derivatives in all cases are very close to zero. Thus, numerically, the optimality conditions are satisfied. These results also suggest that our assumption about the symmetry of the optimal cutpoints in the cases of symmetric distributions is reasonable. Using the algorithm for the case when the number of categories is larger than six, we also verify that the values of the criteria level off with k. For example, in the case of D-optimality and the logistic distribution, the values of criteria are -1.5567, -1.2483, -1.0714, -0.9784, -0.9198 and -0.8809 when k runs from 3 to 8 correspondingly.

4.3.2 Asymmetric distributions

We know that for asymmetric distributions, the optimal cutpoints are not in the symmetric form. Thus, when the number of cutpoints is large, it makes the number of variables too large for search methods (normally more than two). We can have a high dimensional problem. The use of multiplicative algorithm solves this difficulty. We now consider some asymmetric distributions, namely the complementary log-log and the skewed logistic distribution. Table 3.19 in chapter three introduces the formula for the cdf function $F(z)$ and pdf function $f(z)$ of these two asymmetric distributions. In order to carry out the algorithm, we need the derivative of pdf function $f'(z)$.

- Complementary log-log:

$$f(z) = exp[z - exp(z)]$$

$$f'(z) = [1 - exp(z)]f(z)$$

- Skewed logistic:

  Let $F_1(z), f_1(z)$ and $f_1'(z)$ be the cdf, pdf and first derivative of the pdf of z in the case of the logistic distribution. We have for the skewed logistic distribution:

$$F(z) = [F_1(z)]^m$$

$$f(z) = F'(z) = mf_1(z)[F_1(z)]^{m-1}$$

$$f'(z) = mf_1'(z)[F_1(z)]^{m-1} + m(m-1)[f_1(z)]^2[F_1(z)]^{m-1}$$

where $m > 0$. We choose the values of $m = 1/3, 2/3, 3/2$ and 3.

Tables 4.5 to 4.24 below report the results obtained by using the multiplicative algorithm for two asymmetric distributions and four criteria D-, A-, e_1- and e_2-optimality, the number of categories considered runs from three to six. In these tables, we quote the results on the optimal cutpoint z^*, the cdf value $F(z^*)$ at z^* and the optimal criterion value $\phi^*(z^*)$.

Some comments:

- As in the symmetric case, the algorithm converges very well for both distributions, for the four criteria and for the number of categories running from three to six. The number of iterations needed for convergence is almost the same as in the case of symmetric distributions. (We already quoted the results for the case of the logistic distribution and D-optimality). Again, when the number of categories increases, the speed of convergence of the algorithm decreases. We also start the algorithm with both equal initial cell probabilities and unequal arbitrary initial cell probabilities provided that they sum to 1. We also obtain the same solutions but with different numbers of iterations needed. Vertex directional derivatives are also very close to zero. However, with the same numbers of the iterations, the directional derivative in the case of asymmetric distributions are normally larger than those in the case of symmetric distributions. For example, after about 950 iterations, we obtain directional derivatives less than 10^{-5} in the case of the logistic distribution, D-optimality and 6 categories but these values are about less than 10^{-4} for the case of the skewed logistic distribution, D-optimality and 6 categories. However, they still satisfy our convergence requirement (all directional derivatives reach the values which are less than 10^{-4}). As we expected, the optimal cutpoints in the asymmetrical cases are not symmetric.

- For the three category case, the results obtained by using the algorithm are very similar to those obtained by the search method.
- The criterion value in each case also increases but levels off when the the number of categories increases.

Table 4.5: The results: D-optimality and complementary log-log distribution

k		$z_1^*, F(z_1^*)$	$z_2^*, F(z_2^*)$	$z_3^*, F(z_3^*)$	$z_4^*, F(z_4^*)$	$z_5^*, F(z_5^*)$	$\phi^*(z^*)$
3	z^*	-1.2978	0.9580				-0.3989
	$F(z^*)$	0.2389	0.9262				
4	z^*	-1.5561	0.3350	1.2270			-0.0872
	$F(z^*)$	0.1901	0.7529	0.9670			
5	z^*	-2.2591	-0.6688	0.6724	1.3508		0.1172
	$F(z^*)$	0.0991	0.4008	0.8590	0.9789		
6	z^*	-2.5610	-1.0083	0.2041	0.9447	1.4837	0.2194
	$F(z^*)$	0.0743	0.3056	0.7066	0.9236	0.9878	

Table 4.6: The results: A-optimality and complementary log-log distribution

k		$z_1^*, F(z_1^*)$	$z_2^*, F(z_2^*)$	$z_3^*, F(z_3^*)$	$z_4^*, F(z_4^*)$	$z_5^*, F(z_5^*)$	$\phi^*(z^*)$
3	z^*	-1.1726	0.8945				-2.5854
	$F(z^*)$	0.2662	0.9133				
4	z^*	-1.8183	-0.1945	1.0678			-2.1999
	$F(z^*)$	0.1498	0.5609	0.9454			
5	z^*	-2.2158	-0.6880	0.4949	1.2755		-2.0220
	$F(z^*)$	0.1033	0.3950	0.8061	0.9721		
6	z^*	-2.6414	-1.1408	-0.0993	0.7737	1.3962	-1.9297
	$F(z^*)$	0.0687	0.2735	0.5956	0.8855	0.9824	

Table 4.7: The results: e_1-optimality and complementary log-log distribution

k		$z_1^*, F(z_1^*)$	$z_2^*, F(z_2^*)$	$z_3^*, F(z_3^*)$	$z_4^*, F(z_4^*)$	$z_5^*, F(z_5^*)$	$\phi^*(z^*)$
3	z^*	-0.6792	0.5400				-1.3593
	$F(z^*)$	0.3977	0.8204				
4	z^*	-1.1839	-0.0069	0.8166			-1.2510
	$F(z^*)$	0.2636	0.6295	0.8959			
5	z^*	-1.6033	-0.4159	0.3404	0.9955		-1.2017
	$F(z^*)$	0.1822	0.4829	0.7547	0.9332		
6	z^*	-1.9686	-0.7551	-0.0160	0.5653	1.1248	-1.1747
	$F(z^*)$	0.1303	0.3749	0.6261	0.8279	0.9540	

Table 4.8: The results: e_2-optimality and complementary log-log distribution

k		$z_1^*, F(z_1^*)$	$z_2^*, F(z_2^*)$	$z_3^*, F(z_3^*)$	$z_4^*, F(z_4^*)$	$z_5^*, F(z_5^*)$	$\phi^*(z^*)$
3	z^*	-1.6903	1.2909				-0.9163
	$F(z^*)$	0.1684	0.9736				
4	z^*	-2.7345	-1.1762	1.2291			-0.7954
	$F(z^*)$	0.0628	0.2654	0.9454			
5	z^*	-2.2158	-1.1644	0.0603	1.2431		-0.7952
	$F(z^*)$	0.0639	0.2680	0.6543	0.9687		
6	z^*	-3.4835	-1.9499	-0.9253	1.0572	1.5647	-0.6961
	$F(z^*)$	0.0302	0.1326	0.3272	0.9437	0.9916	

Table 4.9: The results: D-optimality and Skewed logistic distribution, m=1/3

k		$z_1^*, F(z_1^*)$	$z_2^*, F(z_2^*)$	$z_3^*, F(z_3^*)$	$z_4^*, F(z_4^*)$	$z_5^*, F(z_5^*)$	$\phi^*(z^*)$
3	z^*	-4.1235	0.4906				-2.5781
	$F(z^*)$	0.2516	0.8528				
4	z^*	-5.6530	-1.2682	1.0690			-2.2484
	$F(z^*)$	0.1517	0.6032	0.9062			
5	z^*	-7.3804	-2.8261	-0.2490	1.5606		-2.0669
	$F(z^*)$	0.0854	0.3824	0.7594	0.9384		
6	z^*	-8.6235	-3.9515	-1.3136	0.3378	1.9763	-1.9704
	$F(z^*)$	0.0564	0.2661	0.5961	0.8357	0.9576	

Table 4.10: The results: D-optimality and Skewed logistic distribution, m=2/3

k		$z_1^*, F(z_1^*)$	$z_2^*, F(z_2^*)$	$z_3^*, F(z_3^*)$	$z_4^*, F(z_4^*)$	$z_5^*, F(z_5^*)$	$\phi^*(z^*)$
3	z^*	-2.1719	1.1124				-1.8984
	$F(z^*)$	0.2187	0.8273				
4	z^*	-2.8877	-0.4218	1.6471			-1.5847
	$F(z^*)$	0.1406	0.5393	0.8892			
5	z^*	-3.6671	-1.4178	0.4616	2.1622		-1.4081
	$F(z^*)$	0.0853	0.3362	0.7219	0.9299		
6	z^*	-4.2792	-2.0461	-0.4420	0.9714	2.5632	-1.3150
	$F(z^*)$	0.0571	0.2357	0.5349	0.8073	0.9517	

Table 4.11: The results: D-optimality and Skewed logistic distribution, m=3/2

k		$z_1^*, F(z_1^*)$	$z_2^*, F(z_2^*)$	$z_3^*, F(z_3^*)$	$z_4^*, F(z_4^*)$	$z_5^*, F(z_5^*)$	$\phi^*(z^*)$
3	z^*	-0.8826	1.8285				-1.2688
	$F(z^*)$	0.1582	0.7997				
4	z^*	-1.2870	0.3837	2.3103			-0.9631
	$F(z^*)$	0.1006	0.4587	0.8676			
5	z^*	-1.6777	-0.3578	1.2204	2.8575		-0.7826
	$F(z^*)$	0.0624	0.2639	0.6784	0.9196		
6	z^*	-1.9625	-0.7593	0.4052	1.6880	3.2526	-0.6901
	$F(z^*)$	0.0432	0.1800	0.4646	0.7753	0.9446	

Table 4.12: The results: D-optimality and Skewed logistic distribution, m=3

k		$z_1^*, F(z_1^*)$	$z_2^*, F(z_2^*)$	$z_3^*, F(z_3^*)$	$z_4^*, F(z_4^*)$	$z_5^*, F(z_5^*)$	$\phi^*(z^*)$
3	z^*	-0.0214	2.4663				-0.9032
	$F(z^*)$	0.1210	0.7831				
4	z^*	-0.3433	0.9923	2.8813			-0.5984
	$F(z^*)$	0.0714	0.3882	0.8490			
5	z^*	-0.6052	0.3951	1.8589	3.4719		-0.4103
	$F(z^*)$	0.0440	0.2133	0.6476	0.9123		
6	z^*	-0.8016	0.0678	1.0530	2.2917	3.8482	-0.3157
	$F(z^*)$	0.0296	0.1381	0.4074	0.7490	0.9386	

Table 4.13: The results: A-optimality and Skewed logistic distribution, m=1/3

k		$z_1^*, F(z_1^*)$	$z_2^*, F(z_2^*)$	$z_3^*, F(z_3^*)$	$z_4^*, F(z_4^*)$	$z_5^*, F(z_5^*)$	$\phi^*(z^*)$
3	z^*	-4.2423	0.1658				-10.8904
	$F(z^*)$	0.2419	0.8150				
4	z^*	-4.9267	-0.6885	0.9521			-9.3333
	$F(z^*)$	0.1930	0.6940	0.8969			
5	z^*	-6.0457	-1.2725	0.0730	1.4523		-8.8322
	$F(z^*)$	0.1331	0.6026	0.8032	0.9323		
6	z^*	-6.7469	-1.9361	-0.5313	0.5045	1.8026	-8.6051
	$F(z^*)$	0.1054	0.5014	0.7180	0.8543	0.9504	

Table 4.14: The results: A-optimality and Skewed logistic distribution, m=2/3

k		$z_1^*, F(z_1^*)$	$z_2^*, F(z_2^*)$	$z_3^*, F(z_3^*)$	$z_4^*, F(z_4^*)$	$z_5^*, F(z_5^*)$	$\phi^*(z^*)$
3	z^*	-1.5689	0.9058				-6.4039
	$F(z^*)$	0.3097	0.7974				
4	z^*	-2.3764	-0.1956	1.4310			-5.5956
	$F(z^*)$	0.1933	0.5883	0.8668			
5	z^*	-3.0432	-0.7595	-0.0392	1.2319		-5.3884
	$F(z^*)$	0.1274	0.4666	0.6216	0.8431		
6	z^*	-3.4809	-1.3728	-0.2101	0.8167	2.2002	-5.1308
	$F(z^*)$	0.0962	0.3444	0.5851	0.7835	0.9323	

Table 4.15: The results: A-optimality and Skewed logistic distribution, m=3/2

k		$z_1^*, F(z_1^*)$	$z_2^*, F(z_2^*)$	$z_3^*, F(z_3^*)$	$z_4^*, F(z_4^*)$	$z_5^*, F(z_5^*)$	$\phi^*(z^*)$
3	z^*	0.9155	1.5637				-4.3005
	$F(z^*)$	0.1850	0.7519				
4	z^*	-1.2399	0.1466	2.0578			-3.6925
	$F(z^*)$	0.1063	0.3930	0.8350			
5	z^*	-1.5745	-0.3813	0.7801	2.5119		-3.4572
	$F(z^*)$	0.0710	0.2585	0.5678	0.8896		
6	z^*	-1.7863	-0.6706	0.2349	1.2422	2.8031	-3.3370
	$F(z^*)$	0.0543	0.1968	0.4173	0.6835	0.9155	

Table 4.16: The results: A-optimality and Skewed logistic distribution, m=3

k		$z_1^*, F(z_1^*)$	$z_2^*, F(z_2^*)$	$z_3^*, F(z_3^*)$	$z_4^*, F(z_4^*)$	$z_5^*, F(z_5^*)$	$\phi^*(z^*)$
3	z^*	0.0459	2.9147				-4.1460
	$F(z^*)$	0.1338	0.8535				
4	z^*	0.0462	0.5020	2.9847			-3.4499
	$F(z^*)$	0.0535	0.2417	0.8624			
5	z^*	-0.8067	0.0095	0.8642	3.1128		-3.2221
	$F(z^*)$	0.0293	0.1267	0.3482	0.8776		
6	z^*	-0.8571	-0.0721	0.6985	1.9713	3.7204	-3.0897
	$F(z^*)$	0.0264	0.1119	0.2979	0.6762	0.9307	

Table 4.17: The results: e_1-optimality and Skewed logistic distribution, m=1/3

k		$z_1^*, F(z_1^*)$	$z_2^*, F(z_2^*)$	$z_3^*, F(z_3^*)$	$z_4^*, F(z_4^*)$	$z_5^*, F(z_5^*)$	$\phi^*(z^*)$
3	z^*	-0.6328	0.8651				-8.0724
	$F(z^*)$	0.7026	0.8894				
4	z^*	-0.9595	0.1652	1.3993			-7.7214
	$F(z^*)$	0.6518	0.8149	0.9291			
5	z^*	-4.2743	-1.0737	0.0923	1.3309		-7.7151
	$F(z^*)$	0.2394	0.6338	0.8057	0.9248		
6	z^*	-5.5820	-1.4499	-0.3954	0.4993	1.6386	-7.5401
	$F(z^*)$	0.1553	0.5749	0.7382	0.8537	0.9425	

Table 4.18: The results: e_1-optimality and Skewed logistic distribution, m=2/3

k		$z_1^*, F(z_1^*)$	$z_2^*, F(z_2^*)$	$z_3^*, F(z_3^*)$	$z_4^*, F(z_4^*)$	$z_5^*, F(z_5^*)$	$\phi^*(z^*)$
3	z^*	-0.6713	0.7790				-4.5050
	$F(z^*)$	0.4854	0.4854				
4	z^*	-1.0444	0.0803	1.2487			-4.2887
	$F(z^*)$	0.4076	0.6467	0.8452			
5	z^*	-1.2957	-0.3144	0.5288	1.5944		-4.1966
	$F(z^*)$	0.3587	0.5626	0.7342	0.8840		
6	z^*	-1.9301	-0.8364	-0.0492	0.7280	1.7744	-4.1582
	$F(z^*)$	0.2523	0.4504	0.6195	0.7690	0.8537	

Table 4.19: The results: e_1-optimality and Skewed logistic distribution, m=3/2

k		$z_1^*, F(z_1^*)$	$z_2^*, F(z_2^*)$	$z_3^*, F(z_3^*)$	$z_4^*, F(z_4^*)$	$z_5^*, F(z_5^*)$	$\phi^*(z^*)$
3	z^*	-0.6924	0.5731				-2.7141
	$F(z^*)$	0.1925	0.5114				
4	z^*	-1.1116	-0.0927	0.8927			-2.5627
	$F(z^*)$	0.1231	0.3292	0.5975			
5	z^*	-1.4172	-0.4949	0.2511	1.1069		-2.4974
	$F(z^*)$	0.0861	0.2330	0.4218	0.6515		
6	z^*	-1.4493	-0.5501	0.1250	0.7931	1.6680	-2.4733
	$F(z^*)$	0.0828	0.2212	0.3871	0.5713	0.7716	

Table 4.20: The results: e_1-optimality and Skewed logistic distribution, m=3

k		$z_1^*, F(z_1^*)$	$z_2^*, F(z_2^*)$	$z_3^*, F(z_3^*)$	$z_4^*, F(z_4^*)$	$z_5^*, F(z_5^*)$	$\phi^*(z^*)$
3	z^*	-0.5746	0.3335				-2.5671
	$F(z^*)$	0.0467	0.1977				
4	z^*	0.4025	0.5682	4.4694			-2.4688
	$F(z^*)$	0.2152	0.2601	0.9664			
5	z^*	-0.7206	0.0683	0.6882	2.8373		-2.3599
	$F(z^*)$	0.0350	0.1382	0.2948	0.8429		
6	z^*	-0.9552	-0.2510	0.3489	1.0573	3.7398	-2.2563
	$F(z^*)$	0.0214	0.0837	0.2015	0.4088	0.9319	

Table 4.21: The results: e_2-optimality and Skewed logistic distribution, m=1/3

k		$z_1^*, F(z_1^*)$	$z_2^*, F(z_2^*)$	$z_3^*, F(z_3^*)$	$z_4^*, F(z_4^*)$	$z_5^*, F(z_5^*)$	$\phi^*(z^*)$
3	z^*	-5.0511	1.6401				-1.2073
	$F(z^*)$	0.1852	0.9426				
4	z^*	-8.7098	-4.0279	1.4910			-1.0316
	$F(z^*)$	0.0548	0.2596	0.9345			
5	z^*	-8.4530	-3.7849	-0.6895	1.6124		-1.0278
	$F(z^*)$	0.0597	0.2810	0.6939	0.9411		
6	z^*	-10.254	-5.5174	-2.8107	0.8419	2.4679	-0.9140
	$F(z^*)$	0.0327	0.1587	0.3842	0.8874	0.9732	

Table 4.22: The results: e_2-optimality and Skewed logistic distribution, m=2/3

k		$z_1^*, F(z_1^*)$	$z_2^*, F(z_2^*)$	$z_3^*, F(z_3^*)$	$z_4^*, F(z_4^*)$	$z_5^*, F(z_5^*)$	$\phi^*(z^*)$
3	z^*	-3.0015	1.9389				-1.0765
	$F(z^*)$	0.1308	0.9142				
4	z^*	-4.5859	-2.3286	1.7613			-0.9584
	$F(z^*)$	0.0467	0.1990	0.8996			
5	z^*	-4.4967	-2.2443	1.3549	2.9381		-0.8703
	$F(z^*)$	0.0495	0.2094	0.8581	0.9661		
6	z^*	-5.4823	-3.2023	-1.7177	1.1927	2.7936	-0.8378
	$F(z^*)$	0.0257	0.1151	0.2850	0.8380	0.9611	

Table 4.23: The results: e_2-optimality and Skewed logistic distribution, m=3/2

k		$z_1^*, F(z_1^*)$	$z_2^*, F(z_2^*)$	$z_3^*, F(z_3^*)$	$z_4^*, F(z_4^*)$	$z_5^*, F(z_5^*)$	$\phi^*(z^*)$
3	z^*	-1.4597	2.40945				-0.9851
	$F(z^*)$	0.0818	0.8789				
4	z^*	-1.3354	0.6859	2.5600			-0.9727
	$F(z^*)$	0.0950	0.5423	0.8943			
5	z^*	-1.6517	-0.1955	1.2554	2.8824		-0.8625
	$F(z^*)$	0.0645	0.3031	0.6865	0.9215		
6	z^*	-2.0663	-0.8380	0.6497	1.9416	3.5010	-0.7939
	$F(z^*)$	0.0376	0.1659	0.5324	0.8178	0.9563	

Table 4.24: The results: e_2-optimality and Skewed logistic distribution, m=3

k		$z_1^*, F(z_1^*)$	$z_2^*, F(z_2^*)$	$z_3^*, F(z_3^*)$	$z_4^*, F(z_4^*)$	$z_5^*, F(z_5^*)$	$\phi^*(z^*)$
3	z^*	-0.4995	2.9747				-0.9479
	$F(z^*)$	0.0538	0.8612				
4	z^*	-0.4222	2.5542	4.1155			-0.8365
	$F(z^*)$	0.0620	0.7988	0.9526			
5	z^*	-0.9919	-0.1645	2.4116	3.9738		-0.7637
	$F(z^*)$	0.0198	0.0966	0.7728	0.9456		
6	z^*	-0.9073	-0.0379	1.8389	2.9448	4.4761	-0.7379
	$F(z^*)$	0.0237	0.1180	0.6423	0.8574	0.9666	

Chapter 5

Multiple Design Points

So far, we have considered only one point designs, under which we offer all the respondents the same set of cutpoints. In practice, several sets may be needed. In this chapter, we consider the case where we offer respondents one of several sets of cutpoints. The proportion of times they are used being determined by a set of weights to be chosen optimally. In this case, respondents will be divided into several groups and each group will be offered a common set of cutpoints. So, the optimal design problem turns out to be determining the sets of cutpoints and the respective weights optimally. We call this situation a multiple design point case. Actually, in chapter two, we already mentioned the concept of multiple design points when we reviewed the case of two categories. In this case, to ensure estimation of both parameters α and β for any distribution and most criteria, we need at least two distinct support points. We now introduce the problem of multiple design points in which two cases will be considered: multiple design points with equal weights, and with arbitrary weights. Then, the methods of finding optimal designs will be presented along with some results and contour plots. Finally, we will summarize both cases, the one point design case and multiple design point case in respect of choosing the number of design points and the number of categories.

5.1 The problem and notations

We will extend the notation from the one design point problem to the multiple design point problem and extend the formula for the expected information matrix.

5.1.1 The problem

Suppose that we offer respondents one of I sets of cutpoints, indexed by i, each consisting of $k-1$ cutpoints. We denote these sets by $\underline{z}^{(i)}$, $i = 1, 2, \ldots, I$ and the corresponding cell probabilities by $\underline{\theta}^{(i)}$ so that:

$$\underline{z}^{(i)} = \{z_1^{(i)}, z_2^{(i)}, \ldots z_{k-1}^{(i)}\},$$

$$\underline{\theta}^{(i)} = \{\theta_1^{(i)}, \theta_2^{(i)}, \ldots \theta_k^{(i)}\}, \quad \sum_{j=1}^{k} \theta_j^{(i)} = 1.$$

The proportion of observations to be taken at each set is called the design weights and denoted by p_i. Hence: $\sum_{i=1}^{I} p_i = 1$

So, our problem turns out to be:

$$\textbf{Choose } p_i, \underline{\theta}^{(i)} \textbf{ optimally}$$

$$\textbf{subjected to: } p_i \geq 0, \quad \sum_{i=1}^{I} p_i = 1,$$

$$\underline{\theta}^{(i)} \geq 0, \quad \sum_{j=1}^{k} \theta_j^{(i)} = 1, \quad i = 1, 2, \ldots, I.$$

This problem is called problem (P2). This is an optimization problem with respect to $I+1$ distributions or the problem of optimizing the function $\phi(\underline{p}, \underline{\theta}^{(1)}, \underline{\theta}^{(2)}, \ldots, \underline{\theta}^{(I)})$.

5.1.2 The expected information matrix

We denote by $I_{\underline{z}^{(i)}}$ the information matrix at $\underline{z}^{(i)}$ or by $I_{\underline{\theta}^{(i)}}$ the information matrix at $\underline{\theta}^{(i)}$. Then, the expected per observation information matrix of this design problem is:

$$M(p) = \sum_{i=1}^{I} p_i I_{\underline{\theta}^{(i)}}. \tag{5.1}$$

Our optimizing function will be a function of the expected information matrix, i.e.

$$\phi(\underline{p}, \underline{\theta}^{(1)}, \underline{\theta}^{(2)}, \ldots, \underline{\theta}^{(I)}) = \psi\{M(p)\}.$$

Note that the multiple design point problem is an extension of the one point design problem. So, in principle, the methods we use in the one design point case to find an optimal design can be applied to the case of multiple design points. However, in the multiple design point problem, there normally are many variables of interest, the use of the search method sometimes may not be relevant. In the following sections, we consider two cases, one when we assume equal design weights with (for simplicity) constraints imposed on the cell probabilities and a second one with arbitrary design weights and no constraints on cell probabilities. In the first case, we still use the search method and a graphical approach to find optimal solutions. In the second case, we will focus on using the multiplicative algorithm.

5.2 Multiple point designs with constraints and equal design weighting

We first assume that the optimal weights are equal, i.e. $p_i^* = 1/I$ for all sets of cutpoints. We set up the new design points by using a single set of cell probability values as in the one point design case but using permutations of them, to define different cutpoint sets. For example, in the three category case, to set up the two design points, the first design point will be $(\theta_1, \theta_2, \theta_3)$ and for symmetry purposes the second one will be $(\theta_3, \theta_2, \theta_1)$. We will focus on the cases of three and four categories. In the three category case, the number of design points that will be considered is 2, 3 and 6. For the four category case, there will be 2, 4 and 8 design points. For higher numbers of categories, the same method of investigation still applies.

5.2.1 The case of three categories

In this case, we have three cell probability values, say $(\theta_1, \theta_2, \theta_3)$ for a one point design. For 2, 3 and 6 point designs, we consider the following designs from which to construct optimal designs.

- Design 3A: one design point

$$(\theta_1, \theta_2, \theta_3)$$

- Design 3B: two design points

$$(\theta_1, \theta_2, \theta_3), (\theta_3, \theta_2, \theta_1)$$

- Design 3C: three design points

$$(\theta_1, \theta_2, \theta_3), (\theta_2, \theta_3, \theta_1), (\theta_3, \theta_1, \theta_2)$$

- Design 3D: six design points

$$(\theta_1, \theta_2, \theta_3), (\theta_2, \theta_3, \theta_1), (\theta_3, \theta_1, \theta_2)$$

$$(\theta_3, \theta_2, \theta_1), (\theta_1, \theta_3, \theta_2), (\theta_2, \theta_1, \theta_3)$$

We have used the concept of Latin squares to construct designs 3C and 3D. The design weights for case 1 through case 4 above are 1, 1/2, 1/3 and 1/6 respectively.

5.2.2 The case of four categories

In this case, the cell probability values are $(\theta_1, \theta_2, \theta_3, \theta_4)$. For the purpose of producing contour plots and avoiding too many combinations of cell probabilities, we impose the following constraint on them:

$$\theta_1 + \theta_2 = \theta_3 + \theta_4 = 1/2$$

Here are the designs considered for the case of four categories:

- Design 4A: one design point

$$(\theta_1, \theta_2, \theta_3, \theta_4)$$

- Design 4B: two design points

$$(\theta_1, \theta_2, \theta_3, \theta_4), (\theta_2, \theta_1, \theta_4, \theta_3)$$

- Design 4C: four design points

$$(\theta_1, \theta_2, \theta_3, \theta_4), (\theta_2, \theta_1, \theta_3, \theta_4)$$

$$(\theta_1, \theta_2, \theta_4, \theta_3), (\theta_2, \theta_1, \theta_4, \theta_3)$$

- Design 4D: eight design points

$$(\theta_1, \theta_2, \theta_3, \theta_4), (\theta_2, \theta_1, \theta_3, \theta_4)$$

$$(\theta_1, \theta_2, \theta_4, \theta_3), (\theta_2, \theta_1, \theta_4, \theta_3)$$

$$(\theta_3, \theta_4, \theta_1, \theta_2), (\theta_3, \theta_4, \theta_2, \theta_1)$$

$$(\theta_4, \theta_3, \theta_1, \theta_2), (\theta_4, \theta_3, \theta_2, \theta_1)$$

The design weights in this case will be 1, 1/2, 1/4 and 1/8 respectively.

5.2.3 Graphical approach

Since we assume equality for the weights p_i and impose constraints on the θ_j, $j = 1, 2, 3, 4$ (in the four category case), we can use a graphical approach to find optimal solutions. Given the manner above of setting up the cell probabilities for the design points, the number of free variables is the same as in the one design point case.

Design 3A and 4A above are one point designs. We already had results for these using the search method, graphical approach and a multiplicative algorithm in previous chapters. The remaining designs are multiple point designs. We are focusing on four criterion functions, namely D-, A-, e_1- and e_2-optimality and four symmetric distributions for the variables, namely the logistic, normal, double-exponential, double-reciprocal. For the higher number of categories and asymmetric distributions, we will use the algorithm in the next section. Note that in the multiple design point case, we have I sets of cutpoints (as defined above, I is the number of design points). In order to complete the calculation of the expected information matrix $M(p)$, we treat each set of cutpoints as one in the one point designs, i.e. we calculate the information matrix for each set using formula 3.14 and then calculate the expected information matrix using formula 5.1.

5.2.4 Some results

For the purpose of determining which case is better in terms of criterion values (among the different number of design point cases), we summarize in tables 5.1 to 5.8 the optimal criterion values for designs 3A to 3D for the three category case and for designs 4A to 4D for the four category case for four symmetric distributions. We also quote the optimal set (or sets) of cell probabilities and optimal cutpoints for the case of D-optimality and the logistic distribution and for the case of A-optimality and the normal distribution to see differences amongst cases. The contour plots in figures 5.1 to 5.3 illustrate the results and show the positions of the optimal design points (or θ-values) in the above cases.

Table 5.1: The criterion values of multiple point designs: k=3, logistic distribution

Design	D-optimality	A-optimality	e_1-optimality	e_2-optimality
3A	-1.5572	-5.0182	-3.3756	-1.0226
3B	-1.5572	-5.0182	-3.3756	-1.0226
3C	-2.0560	-5.7166	-3.3753	-1.5574
3D	-2.0560	-5.7166	-3.3753	-1.5574

Table 5.2: The criterion values of multiple point designs: k=3, normal distribution

Design	D-optimality	A-optimality	e_1-optimality	e_2-optimality
3A	-0.2070	-2.2801	-1.2348	-0.7666
3B	-0.2070	-2.2801	-1.2348	-0.7666
3C	-0.7962	-2.9905	-1.2607	-1.1316
3D	-0.7962	-2.9905	-1.2607	-1.1316

Table 5.3: The criterion values of multiple point designs: k=3, double exponential distribution

Design	D-optimality	A-optimality	e_1-optimality	e_2-optimality
3A	-1.1277	-3.9971	-1.0000	-1.5484
3B	-1.1277	-3.9971	-1.0000	-1.5484
3C	-1.3208	-4.3536	-1.0000	-2.3300
3D	-1.3208	-4.3536	-1.0000	-2.3300

Table 5.4: The criterion values of multiple point designs: k=3, double reciprocal distribution

Design	D-optimality	A-optimality	e_1-optimality	e_2-optimality
3A	-2.0710	-7.6493	-1.0442	-4.0115
3B	-2.0743	-7.3344	-0.9612	-4.0040
3C	-2.0703	-7.1551	-1.0955	-4.5022
3D	-2.0760	-7.1551	-1.0927	-4.5022

Table 5.5: The criterion values of multiple point designs: k=4, logistic distribution

Design	D-optimality	A-optimality	e_1-optimality	e_2-optimality
4A	-1.2483	-4.3788	-3.2000	-1.0226
4B	-1.5504	-4.6729	-3.2000	-1.4729
4C	-1.5504	-4.6729	-3.2000	-1.4729
4D	-1.5504	-4.6729	-3.2000	-1.4729

Table 5.6: The criterion values of multiple point designs: k=4, normal distribution

Design	D-optimality	A-optimality	e_1-optimality	e_2-optimality
4A	0.1001	-1.9425	-1.1331	-0.7666
4B	-0.4578	-2.5211	-1.1333	-1.3266
4C	-0.4578	-2.5211	-1.1333	-1.3266
4D	-0.4578	-2.5211	-1.1333	-1.3266

Table 5.7: The criterion values of multiple point designs: k=4, double exponential distribution

Design	D-optimality	A-optimality	e_1-optimality	e_2-optimality
4A	-0.4359	-2.5464	-1.0000	-1.5464
4B	-0.7330	-3.0813	-1.0000	-2.0813
4C	-0.7330	-3.0813	-1.0000	-2.0813
4D	-0.7330	-3.0813	-1.0000	-2.0813

Table 5.8: The criterion values of multiple point designs: k=4, double reciprocal distribution

Design	D-optimality	A-optimality	e_1-optimality	e_2-optimality
4A	-1.1631	-4.8000	-0.8000	-4.0000
4B	-1.1631	-4.8000	-0.8000	-4.0000
4C	-1.1631	-4.8000	-0.8000	-4.0000
4D	-1.1631	-4.8000	-0.8000	-4.0000

Some comments:

We discuss tables 5.1 to 5.8. To begin, for the three category case, the optimal criterion values for one and two point designs are almost the same. Similarly, the optimal criterion values for three and six points designs are the same. Look at designs 3A to 3D, we see that design 3A and 3B, 3C and 3D coincide if $\theta_1 = \theta_3$. We also see that compared to three or six point designs, one or two point designs are better in terms of optimal criterion values.
For the four category case, one point design is the best while two, four and eight point designs are the same in terms of optimal criterion values.
There are some exceptions. In the three category case, for the case of the double reciprocal distribution where two point designs seem to be better for D-, e_1- and e_2-optimality and the three point design is the best for A-optimality (one point design is worst). For e_1-optimality, the logistic and double-exponential distribution, the optimal criterion values stay the same in all four cases, namely one, two, three and six point designs. Similarly in the four category case, for the logistic and double exponential distribution, the e_1-optimality criterion values are the same for all four cases, namely one, two, four and eight point designs. Also, for the case of the double reciprocal distribution, the criterion values are unchanged when the number of design points increases from two to eight for all criteria considered. We now show details of some of the optimizing designs.

Table 5.9: The optimal cell probabilities and optimal cutpoints: k=3, logistic distribution and D-optimality

Design	Alternative θ_i^*'s	Corresponding cutpoints	Optimal criteria
3A	(0.18,0.64,0.18)	(-1.5163,1.5163)	-1.5572
3B	(0.18,0.64,0.18)	(-1.5163,1.5163)	-1.5572
3C	(0.56,0.22,0.22)	(0.2411,1.2656)	-2.0559
	(0.22,0.56,0.22)	(-1.2656,1.2656)	
	(0.22,0.22,0.56)	(-1.2656,-0.2411)	
3D	(0.56,0.22,0.22)	(0.2411,1.2656)	-2.0559
	(0.22,0.56,0.22)	(-1.2656,1.2656)	
	(0.22,0.22,0.56)	(-1.2656,-0.2411)	

Table 5.10: The optimal cell probabilities and optimal cutpoints: k=4, logistic distribution and D-optimality

Design	Alternative θ_i^*'s	Corresponding cutpoints	Optimal criteria
4A	(0.12,0.38,0.38,0.12)	(-1.9924,0,1.9924)	-1.2484
4B	(0.25,0.25,0.25,0.25)	(-1.0986,0,1.0986)	-1.5504
4C	(0.25,0.25,0.25,0.25)	(-1.0986,0,1.0986)	-1.5504
4D	(0.25,0.25,0.25,0.25)	(-1.0986,0,1.0986)	-1.5504

Table 5.11: The optimal cell probabilities and optimal cutpoints: k=3, normal distribution and A-optimality

Design	Alternative θ_i^*'s	Corresponding cutpoints	Optimal criteria
3A	(0.15,0.70,0.15)	(-1.0364,1.0364)	-2.2801
3B	(0.15,0.70,0.15)	(-1.0364,1.0364)	-2.2801
3C	(0.15,0.70,0.15)	(-1.0364,1.0364)	-2.9905
	(0.15,0.15,0.70)	(-1.0364,-0.5244)	
	(0.70,0.15,0.15)	(0.5244,1.0364)	
3D	(0.15,0.70,0.15)	(-1.0364,1.0364)	-2.9905
	(0.15,0.15,0.70)	(-1.0364,-0.5244)	
	(0.70,0.15,0.15)	(0.5244,1.0364)	

Table 5.12: The optimal cell probabilities and optimal cutpoints: k=4, normal distribution and A-optimality

Design	Alternative θ_i^*'s	Corresponding cutpoints	Optimal criteria
4A	(0.09,0.41,0.41,0.09)	(-1.3408,0,1.3408)	-1.8425
4B	(0.19,0.31,0.31,0.19)	(-0.8779,0,1.8779)	-2.5211
	(0.19,0.31,0.19,0.31)	(-0.8779,0,0.4959)	
	(0.31,0.19,0.31,0.19)	(-0.4959,0,0.8779)	
	(0.31,0.19,0.19,0.31)	(-0.4959,0,0.4959)	
4C	(0.19,0.31,0.31,0.19)	(-0.8779,0,1.8779)	-2.5211
	(0.19,0.31,0.19,0.31)	(-0.8779,0,0.4959)	
	(0.31,0.19,0.31,0.19)	(-0.4959,0,0.8779)	
	(0.31,0.19,0.19,0.31)	(-0.4959,0,0.4959)	
4D	(0.19,0.31,0.31,0.19)	(-0.8779,0,1.8779)	-2.5211
	(0.19,0.31,0.19,0.31)	(-0.8779,0,0.4959)	
	(0.31,0.19,0.31,0.19)	(-0.4959,0,0.8779)	
	(0.31,0.19,0.19,0.31)	(-0.4959,0,0.4959)	

Figure 5.1: The contour of criterion values versus three cell probabilities (Logistic distribution and D-optimality)

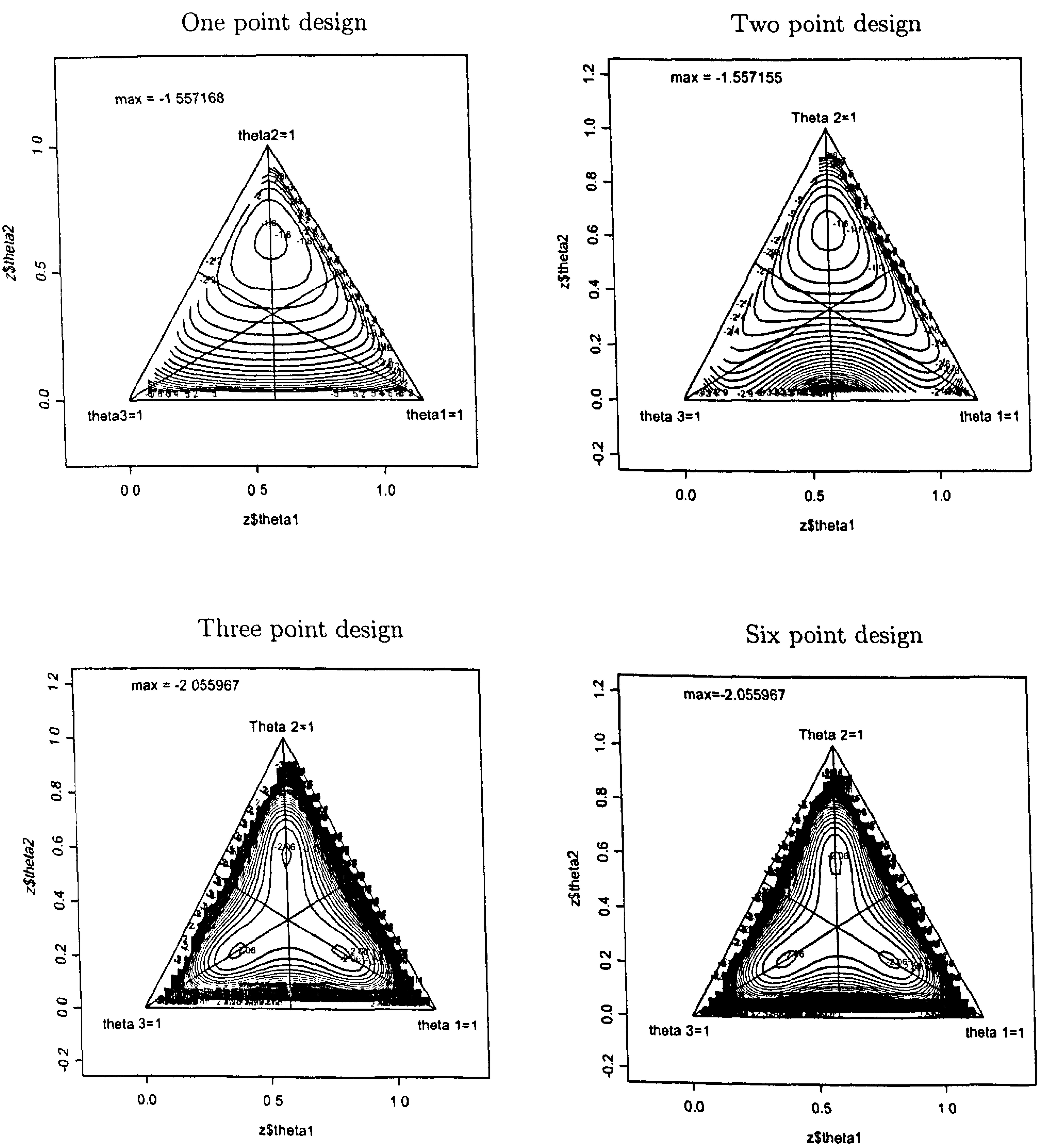


Figure 5.2: The contour of criterion values versus four cell probabilities (Logistic distribution and D-optimality)

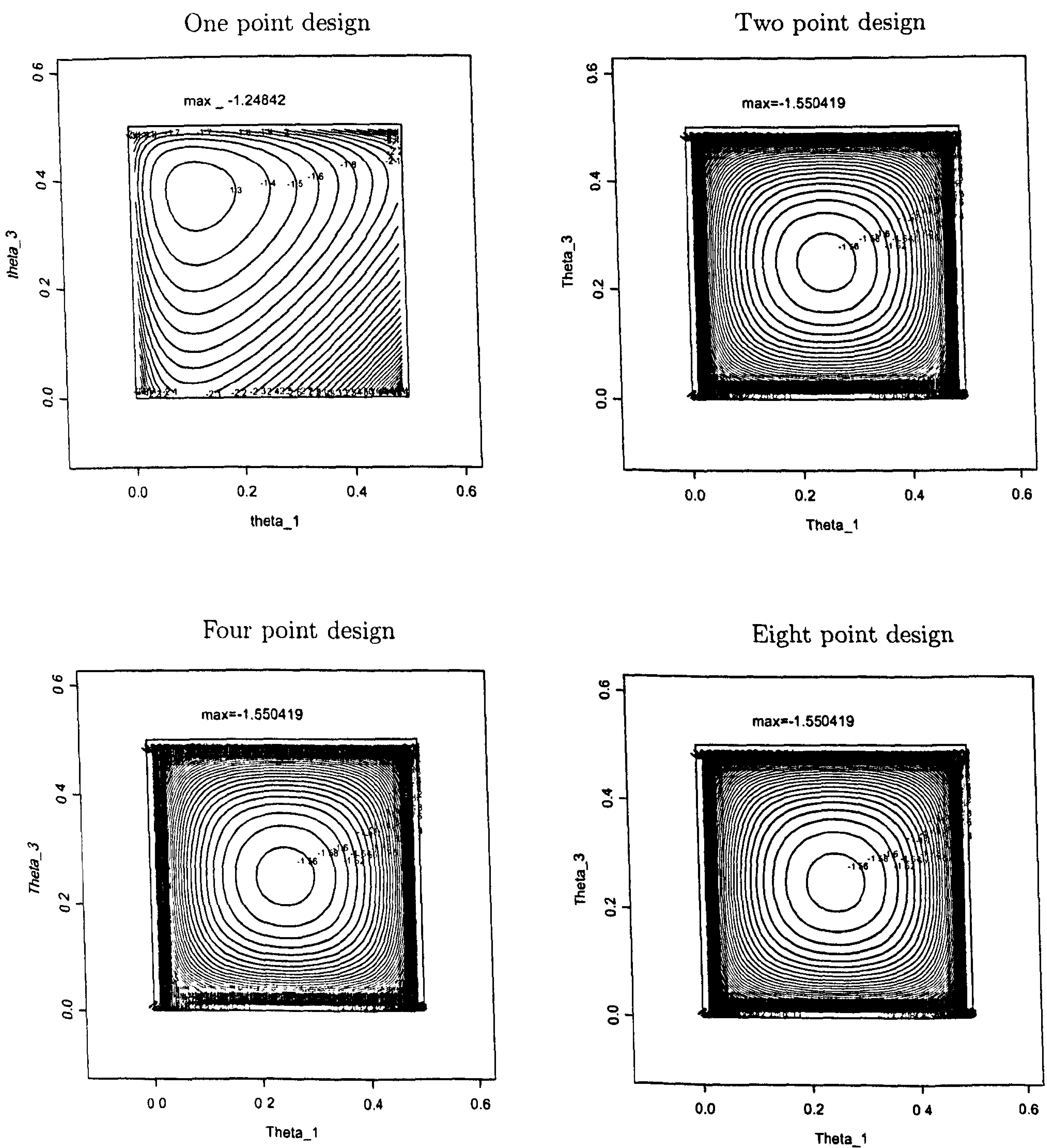
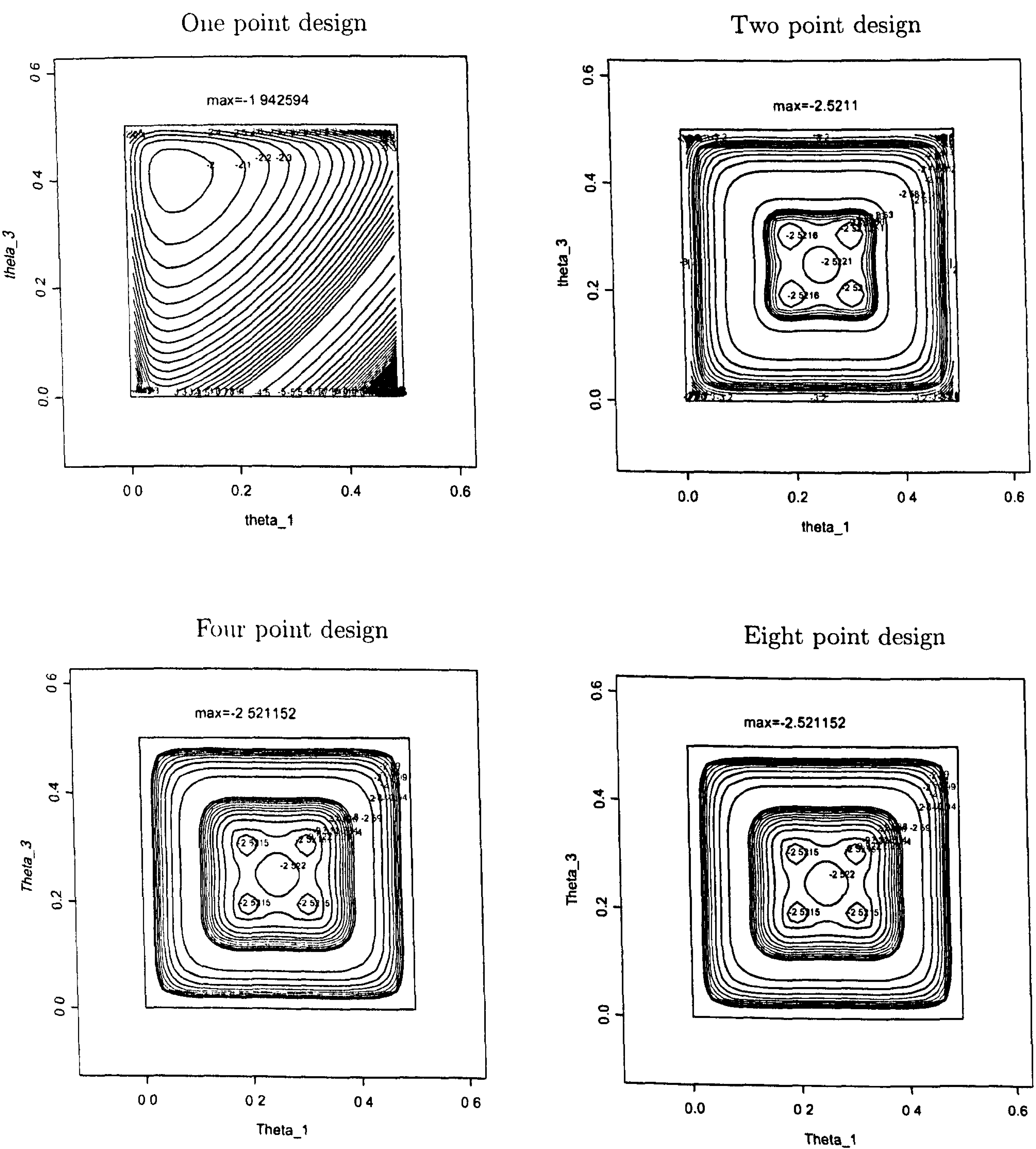


Figure 5.3: The contour of criterion values versus four cell probabilities (Normal distribution and A-optimality)



We can consider the results in more details by looking at tables 5.9 to 5.12 and the contour plots 5.1 to 5.3 which results for D-optimality combined with the logistic distribution and A-optimality combined with the normal distribution. For the three category case there is in both examples, a unique optimizing design of the form 3A and 3B and these coincide; since for 3B, $\theta_1 = \theta_3$. Hence criterion value is the same. We have a similar result in respect of the best 3C and 3D designs. These share a unique optimizing design, but (not surprisingly) with three possible sets of optimizing θ-values, under each solution two cell probabilities are equal. Hence the six point design of 3D reduces to the three point design of 3C.

The results are different in the four category case. For the logistic distribution and D-optimality, the best 4B, 4C and 4D designs coincide since all optimal θ-values are equal. In other words, the best 4B, 4C and 4D designs reduce to a one point optimal design with equal cell probabilities at the optimum. For the normal distribution and A-optimality, there are four distinct solutions, each 'equivalent' to $\theta_2 = \theta_3$ (and hence $\theta_1 = \theta_4$) for designs 4B, 4C, 4D. The implications of this are that there are two optimal designs of the form 4B, one given by $(\theta_1, \theta_2, \theta_3, \theta_4) = (0.19, 0.31, 0.19, 0.31)$ or $(0.31, 0.19, 0.31, 0.19)$ and the other by $(\theta_2, \theta_1, \theta_4, \theta_3) = (0.19, 0.31, 0.31, 0.19)$ or $(0.31, 0.19, 0.19, 0.31)$. In contrast, there is a common unique optimizing design for cases 4C, 4D, the eight points of the latter reducing to four distinct points.

In the following section, we will consider the results of the multiple point designs when there are no constraints on weights and cell probabilities.

5.3 Multiple point designs with arbitrary weights and no constraints

In this section, we investigate multiple point designs without any assumptions or constraints on the weights p_i and cell probabilities θ_j. For example, in the case of two point designs with three cell probabilities at each point, we have two cutpoints in each design point and two weights. In the case where the number of variables considered is large, more relevant techniques for finding optimal solutions should be employed. We will use the extension of the multiplicative algorithm as introduced the in previous chapter for this purpose. For problem (P2), we apply the iteration for both sets of cell probabilities and the design weights.

5.3.1 Multiplicative algorithm for multiple point designs

Here are the formula of the multiplicative algorithm for finding optimal p^* and θ^*:

$$p_i^{(r+1)} = \frac{p_i^{(r)} m_p\{F_i(p^{(r)}), \delta\}}{\sum_{j=1}^{I} p_j^{(r)} m_p\{F_j(p^{(r)}), \delta\}} \tag{5.2}$$

$$\theta_j^{(i)(r+1)} = \frac{\theta_j^{(i)(r)} m_i\{F_j^{(i)}(\underline{\theta}^{(i)(r)}), \delta\}}{\sum_{t=1}^{k} \theta_t^{(i)(r)} m_i\{F_t^{(i)}(\underline{\theta}^{(i)(r)}), \delta\}} \tag{5.3}$$

where:

- $F_i(p)$ is the i^{th} directional derivatives w.r.t p_i
- $F_j^{(i)}(\underline{\theta}^{(i)})$ is the j^{th} directional derivatives w.r.t $\theta_j^{(i)}$

- $m_p(z, \delta)$ and $m_i(z, \delta)$ are once again positive increasing functions of z for given positive δ. δ is one of the free parameters that these two functions may depend on. In our case, we use the normal cdf function for both $m_p(z, \delta)$ and $m_i(z, \delta)$ and δ=1, i.e. $m_i(z, \delta) = \Phi(\delta z)$.

The procedure for calculation of directional derivatives is exactly the same as in the case of one point designs but remember that our criterion functions now are functions of the p_i and $\theta_j^{(i)}$, $i = 1, 2, \ldots, I$. We are still using four criteria: D-, A-, e_1- and e_2-optimality.

In particular, if the criterion function is:

$$\phi(\underline{p}, \underline{\theta}^{(1)}, \underline{\theta}^{(2)}, \ldots, \underline{\theta}^{(I)}) = \psi\{M(p)\}.$$

Then:

$$F_i(p) = \frac{\partial \psi}{\partial p_i} - \sum_{l=1}^{I} p_l \frac{\partial \psi}{\partial p_l}, \quad i = 1, 2, \ldots, I,$$

$$F_j^{(i)}(\underline{\theta}^{(i)}) = \frac{\partial \psi}{\partial \theta_j} - \sum_{t=1}^{k} \theta_t \frac{\partial \psi}{\partial \theta_t}, \quad j = 1, 2, \ldots, k.$$

All the procedures for calculation of these are exactly the same as in the case of one point designs. However, for applying the multiplicative algorithm for multiple point designs, we also need to calculate derivatives of criteria with respect to the p_i. We summarize the formulae for $\dfrac{\partial \psi}{\partial z_j^{(i)}}$ and $\dfrac{\partial \psi}{\partial p_i}$ in the following table:

Table 5.13: The derivatives of $\psi\{M(p)\}$ w.r.t. $z_j^{(i)}$ and p_i

Criteria	Criterion function $M(p)$	$\frac{\partial\psi}{\partial z_j^{(i)}}$	$\frac{\partial\psi}{\partial p_i}$
D-optimality	$log\, det\{M(p)\}$	$p_i\, tr\left(M^{-1}(p)\frac{\partial I_{\underline{z}^{(i)}}}{\partial z_j^{(i)}}\right)$	$tr(M^{-1}(p))I_{\underline{z}^{(i)}}$
A-optimality	$-tr\{M^{-1}(p)\}$	$p_i\, tr\left(M^{-2}(p)\frac{\partial I_{\underline{z}^{(i)}}}{\partial z_j^{(i)}}\right)$	$tr(M^{-2}(p))I_{\underline{z}^{(i)}}$
e_1-optimality	$-e_1^T\, M^{-1}(p)\, e_1$	$p_i\, e_1^T\, M^{-1}(p)\, \frac{\partial I_{\underline{z}^{(i)}}}{\partial z_j^{(i)}} M^{-1}(p)\, e_1$	$e_1^T\, M^{-1}(p)\, I_{\underline{z}^{(i)}}\, M^{-1}(p)\, e_1$
e_2-optimality	$-e_2^T\, M^{-1}(p)\, e_2$	$p_i\, e_2^T M^{-1}(p)\, \frac{\partial I_{\underline{z}^{(i)}}}{\partial z_j^{(i)}}\, M^{-1}(p)\, e_2$	$e_2^T\, M^{-1}(p)\, I_{\underline{z}^{(i)}}\, M^{-1}(p)\, e_2$

$e_1 = (1,0)^T$, $e_2 = (0,1)^T$, $M(p) = \sum_{i=1}^{I} p_i\, I_{\underline{z}^{(i)}}$

5.3.2 Some results

In the two tables below, we summarize the results obtained by using the multiplicative algorithm in the cases of two point and three point designs for one symmetric distribution (logistic, table 5.10) and one asymmetric distribution (skewed logistic, table 5.11). In both cases, we use the D-criterion. We run the algorithm for the four cases where the number of categories is 3, 4, 5 and 6. We use arbitrary initial values for the sets of cell probabilities and the design weights. The reason for using these arbitrary initial values is that we need to start from distinct design points. If the cell probabilities were initially equal for all design points, there would be only one distinct design point and the algorithm would not change this. For example, in the case of two design points and three categories, we use two sets of values $\{(0.3,\ 0.4,\ 0.3);(0.2,\ 0.5,\ 0.3)\}$ for the initial cell probabilities and a set of values (0.4, 0.6) for the initial weights (although these could have been equal). In all cases, the higher the number of categories is, the more slowly the algorithm converges. For instance, in the case of the logistic distribution and D-optimality, for the two point design case, the number of iterations needed for convergence (as in chapter 4, we define the the algorithm to have converged whenever the respective directional derivatives are very close to zero) are about 420, 565, 743, 1034 for the three, four, five and six category case in turn. For the three point design case, the number of iterations needed are 940, 1400, 1900 and 2690. All of the directional derivative values are less than 10^{-6}.

The results show that the optimal two point designs and three point designs are almost the same in terms of optimal cut points and criterion values. In general, the algorithm tends to converge to the same optimal cutpoints whatever the number of design points is. If we compare with the results from the one point design case, the multiple point designs are not better. The optimal cutpoints and the optimal criterion values do not increase when we move from a one point to a three point design. In consequence, the design weights do not matter and we do not need to use multiple design points. Our assumption about the symmetry of cutpoint sets in the case of symmetric distributions proves to be justified. In tables 5.14 and 5.15, we present the results of optimal cutpoints (two sets and three sets for the two point design and three point design cases), respective optimal weights and optimal criterion values. In all but one case (three design points for $k = 6$ in table 5.15) all design points are approximately equal suggesting one design point.

Table 5.14: The results of two point and three point designs for D-optimality and logistic distribution

k	2 design points		3 design points	
	Cutpoints	Weights; $\phi^*(.)$	Cutpoints	Weights: $\phi^*(.)$
3	(-1.555, 1.555) (-1.555, 1.555)	(0.376, 0.624) $\phi^*(.)$=-1.557	(-1.466, 1.462) (-1.462, 1.466) (-1.468, 1.468)	(0.249, 0.249, 0.502) $\phi^*(.)$=-1.557
4	(-1.978, 0, 1.978) (-1.978, 0, 1.978)	(0.454, 0.546) $\phi^*(.)$=-1.248	(-1.984, 0, 1.983) (-1.924, 0, 1.921) (-1.943, 0, 1.931)	(0.585, 0.246, 0.169) $\phi^*(.)$=-1.248
5	(-2.565, -0.861, 0.854, 2.533) (-2.611, -0.851, 0.872, 2.512)	(0.374, 0.626) $\phi^*(.)$=-1.071	(-2.491, -0.823, 0.868, 2.528) (-2.502, -0.837, 0.847, 2.511) (-2.554, -0.843, 0.850, 2.661)	(0.236, 0.249, 0.515) $\phi^*(.)$=-1.070
6	(-2.895, -1.319, 0, 1.319, 2.895) (-2.839, -1.261, 0, 1.252, 2.834)	(0.458, 0.542) $\phi^*(.)$=-0.978	(-2.734, -1.257, 0, 1.232, 2.754) (-2.644, -1.230, 0, 1.496, 2.446) (-2.849, -1.458, 0, 1.410 2.875)	(0.412, 0.297, 0.291) $\phi^*(.)$=-0.981

Table 5.15: The results of two point and three point designs for D-optimality and skewed logistic distribution

k	2 design points		3 design points	
	Cutpoints	Weights: $\phi^*(.)$	Cutpoints	Weights; $\phi^*(.)$
3	(-4.123, 0.490) (-4.115, 0.491)	(0.614, 0.386) $\phi^*(.)$=-2.578	(-4.102, 0.493) (-4.078, 0.495) (-4.129, 0.4915)	(0.278. 0.287, 0.435) $\phi^*(.)$=-2.578
4	(-5.782, -1.381, 1.029) (-5.530, -1.162, 1.107)	(0.623 0.377) $\phi^*(.)$=-2.248	(-5.772, -1.371, 1.026) (-5.449, -1.372, 1.022) (-5.812, -1.429, 1.007)	(0.452, 0.214, 0.334) $\phi^*(.)$=-2.249
5	(-7.332, -2.794, -0.243, 1.564) (-7.286, -2.772, -0.251, 1.559)	(0.553, 0.447) $\phi^*(.)$=-2.066	(-7.809, -3.299, -0.490, 1.416) (-6.586, -2.520, -0.128, 1.632) (-7.181, -2.806, -0.266, 1.550)	(0.374, 0.278 0.348) $\phi^*(.)$=-2.069
6	(-8.961, -4.276, -1.581, 0.207, 1.872) (-8.734, -4.154, -1.769, 0.116, 1.804)	(0.561, 0.439) $\phi^*(.)$=-1.972	(-9.023, -4.448, -1.773, 0.083, 1.783) (-7.722, -3.429, -1.231, 0.366, 1.988) (-7.726, -3.429, -1.238, 0.363, 1.988)	(0.402, 0.299, 0.299) $\phi^*(.)$=-1.973

5.3.3 The case of unequal number of cutpoints across design points

In the previous cases, we assume the number of cutpoints (the number of categories) in each design point are the same. In this section, we will investigate the case where these numbers are different.

We consider that we have a two point design problem, one point with three categories and another point with four categories. So, the optimal cell probabilities will have the form:

$$\{(\theta_1^*, \theta_2^*, \theta_3^*), (\theta_4^*, \theta_5^*, \theta_6^*, \theta_7^*)\}.$$

The sets of optimal cutpoints are:

$$\{(z_1^*, z_2^*), (z_3^*, z_4^*, z_5^*)\}.$$

The optimal weights are:

$$(p_1^*, p_2^*).$$

Similarly, if we consider a three point design with one point with three categories, one point with four categories and the remaining one with five categories, the sets of optimal cell probabilities, cutpoints and optimal weights respectively are:

$$\{(\theta_1^*, \theta_2^*, \theta_3^*), (\theta_4^*, \theta_5^*, \theta_6^*, \theta_7^*), (\theta_8^*, \theta_9^*, \theta_{10}^*, \theta_{11}^*, \theta_{12}^*)\}$$

and:

$$\{(z_1^*, z_2^*), (z_3^*, z_4^*, z_5^*), (z_6^*, z_7^*, z_8^*, z_9^*)\}$$

and:

$$(p_1^*, p_2^*, p_3^*).$$

We can use the multiplicative algorithm to find optimal solutions. The tables below show the results for two examples of these kinds of designs for the case of D-optimality. For the logistic distribution, we consider the case of three and four categories and for the skewed logistic distribution the cases of four and five categories.

Table 5.16: 2 points: three and four categories, logistic distribution, D-optimality

Point order and weights	Directional derivatives	Cutpoints and weights	Criterion values
Point 1	0.000000016	-1.47569480	-1.24959796
	-0.000000033	1.47569480	
	0.000000016		
Point 2	0.0000000099	-1.97968464	
	-0.0000000099	0	
	-0.0000000099	1.97968464	
	0.0000000099		
Weights	0.000000099	0.00040196	
	0.000000099	0.99959804	

Table 5.17: 3 points: three, four and five categories, skewed logistic distribution, D-optimality

Point order and weights	Directional derivatives	Cutpoints and weights	Criterion values
Point 1	0.00000032	-3.83569480	-2.06956878
	0.00000023	0.25512450	
	-0.00000020		
Point 2	0.00000054	-4.43392095	
	-0.00000011	-1.72071692	
	-0.00000075	0.30690745	
	0.00000013		
Point 3	0.0000000068	-7.37536562	
	0.0000000015	-2.82211890	
	-0.0000000093	-0.24585599	
	-0.0000000096	1.56297317	
	-0.0000000035		
Weights	0.0000000018	0.00040196	
	-0.0000000018	0.00021454	
	-0.0000000018	0.99938350	

Some comments:

The algorithm converges very well in both cases. We need about 1250 iterations in the three and four category case and 2110 iterations in the three, four and five category case for the convergence (the directional derivatives in all cases are less than 10^{-6} of the algorithm. We see that if the number of cutpoints are not equal in each point, the optimal criteria are always the same as the values that we achieve for the one design point case, when the design point has the largest number of cutpoints. For example, in the above tables, the two point design with three and four categories, the optimal criteria and the respective optimal cutpoints are the same as those for the case of four categories alone, i.e. -1.245 for optimal criterion value and (-1.98,0, 1.98) for the optimal cutpoints. Similarly, in the case of the three point design with three, four and five categories, we attain the optimal solution of the one point five category case design. These results are confirmed because the optimal weights put unit weight to the point with the highest number of categories. In the examples above, weights are almost one for the four category design point (case 1) and five category design point (case 2). Weights are almost zero for the remaining design points. Such results are to be expected since we know that the optimal criterion values increase when we increase the number of cutpoints. So, if we consider offering respondents either three category bids or four category bids with respective design weights to be chosen optimally, it is reasonable to expect that the four category bids will dominate in terms of design weight and the optimal design will be the optimal design of the four category case with weight 1 (weight for four category bid is one). So, it would seem that in general it is not necessary to consider the case where the number of cutpoints (or categories) are not equal across design points.

5.4 Choosing the number of design points

Our main purpose is to identify the optimal number of design points and the optimal values for their cutpoints or cell probabilities. We have already used various methods to find optimal solutions such as search methods, graphical approaches and a multiplicative algorithm. The optimal solutions depend on the criterion under consideration. From all the results we have, we can make some statements on how to choose optimal solutions as follows.

- In general, the results obtained from the methods listed above are very consistent. Although we sometimes imposed some constraints on the cutpoints and design points in the search method and the graphical approach, the results obtained are verified by using the algorithm without any assumptions.

- The general tendency in the results is that when we increase the number of cutpoints, the optimal criteria initially increase but then level off. So in practice, using four or five categories seems to be suitable.

- In most cases, multiple point designs are not better than one point designs in term of the values of criteria, i.e. the criterion values either stay the same or decrease. So, using one point designs for the survey is enough. There is an exception in the case of the double reciprocal distribution where a two point design is better than a one point design. In this case, we would be better to use two point designs.

Chapter 6

The Bivariate Approach

In this chapter, we continue to identify optimal cutpoints and cell probabilities by a different method called the bivariate approach. We first summarize this method in the literature. Then we consider the problem of the bivariate approach by introducing the formula for the Fisher information matrix. We also investigate the particular situations where sets of cutpoints will be offered and the method of searching for optimal results. Finally, the results by this method will be compared with previous cases and conclusions reached.

6.1 Introduction

In previous chapters, we already considered cases where there is only one variable of interest. We call these cases univariate approaches. In our particular context, the univariate approach corresponds to the case where we offer respondents a set of cutpoints (one point design) or one of a set of cutpoints (multiple point design). In the case of multiple point designs, the sets of cutpoints are distinct and we have respective weights for each design point. Now, assume that we want to consider two variables simultaneously and these two variables are related to each other. In our context, the way of dealing with two variables is called the bivariate approach.

In chapter 2, we introduced the concept of Contingent Valuation (CV) studies, the aim of which is to estimate a population mean or total willingness to pay (WTP) for some non-market commodity. In particular, we described dichotomous choice of single bound CV studies in which respondents are offered to which they respond 'YES' if their WTP is larger than bid and otherwise 'NO'. An extension of the single bound CV study is the double bound CV study under which a second bound is offered to each respondent, higher if the first bid answer is 'YES', lower otherwise. Thus the bid is a middle one and three together are cutpoints defining four categories of WTP values. Optimal choice for them can be determined using the results of chapter 3 assuming WTP has a c.d.f of the form $F(\frac{WTP-\mu}{\sigma})$. We describe this approach as univariate approach. Note that this sequential approach is not natural for other context such as estimating mean income.

However, the sequential nature of the above process has led to what we call the bivariate approach. This allows a change in the (marginal) distribution of WTP between the two bids. In keeping with this we denote by WTP_1 and WTP_2 the willingness to pay of the respondent at the first and second bid respectively. We need a joint distribution for WTP_1, WTP_2. A popular assumption has seen bivariate normal or log-normal distribution. In the literature, several authors have done some work on this topic.

Cameron and Quiggin (1994) propose the use of a bivariate probit (or normal) contingent valuation model when respondents are offered a follow-up bid to an initial contingent valuation question. They adopt several competing specifications based on the bivariate probit, and compare them to the double bound model to analyze CV data. The distinction between single bound and double bound is that in the first case, only one bid is offered to each individual while in the latter case, a second bid is offered, higher than the first if the answer to that was 'YES', and lower otherwise. Alberini (1994) compares the associated double bound models and recently proposed bivariate models of WTP. She carries out some Monte Carlo simulations to show that the double bound estimates of mean or median WTP can be surprisingly robust to departure from the true, bivariate model, and that the double bound model is often superior to the bivariate model in terms of the mean square error of the estimates. In another paper Alberini (1995) finds optimal designs for discrete choice Contingent Valuation surveys using various models including the bivariate one. In the double bound context, she assumes that the two willingness to pay values (WTP_1) and (WTP_2) (or $log(WTP_1)$ and $log(WTP_2)$) have a bivariate normal distribution with a common location parameter (μ), scale parameter (σ) and a correlation coefficient (ρ). Dependent on the values of ρ ($0 \leq |\rho| \leq 1$), she constructs two two-point designs given the first bid answer. She assumed the first bid to be median WTP. If the first bid answer is 'YES', she calculates a design weight λ for a higher second bid and $1 - \lambda$ for a lower second bid; these standardized bids and λ being determined optimally. Similarly, if the first bid answer is 'NO', the design weight for the higher second bid is $1 - \lambda$ and λ is the design weight for the lower second bid. We use the same notation λ in both cases (Alberini denotes them by λ^{UP} and λ^{DN}) as optimal values are equal because of the symmetry of the bivariate normal distribution, and the choice of the first bid as being the median. The two designs differ only by a change of sign.

Later on, the bivariate approach is further pursued by Alberini, Carson and Kanninen (1997). Using bivariate probit specifications, they model different behavioral patterns induced by the reiteration of the elicitation question after the first answer has been obtained. Such different behavioral hypotheses are then tested as competing models by means of standard specification tests. One very common difficulty when we assume the bivariate normal for the joint distribution of two WTPs is that this assumption is not supported by the data and this may lead to biased estimates. In our case, there is another difficulty; namely in the computation of bivariate normality distribution function. We therefore extend our analysis to alternative bivariate models, namely Copula models which are characterized by a great flexibility in the distributional shape of their marginals, and in their dependence structure. In our case, we will focus on using the Plackett Copula which we will mention later on.

6.2 Construction of the problem

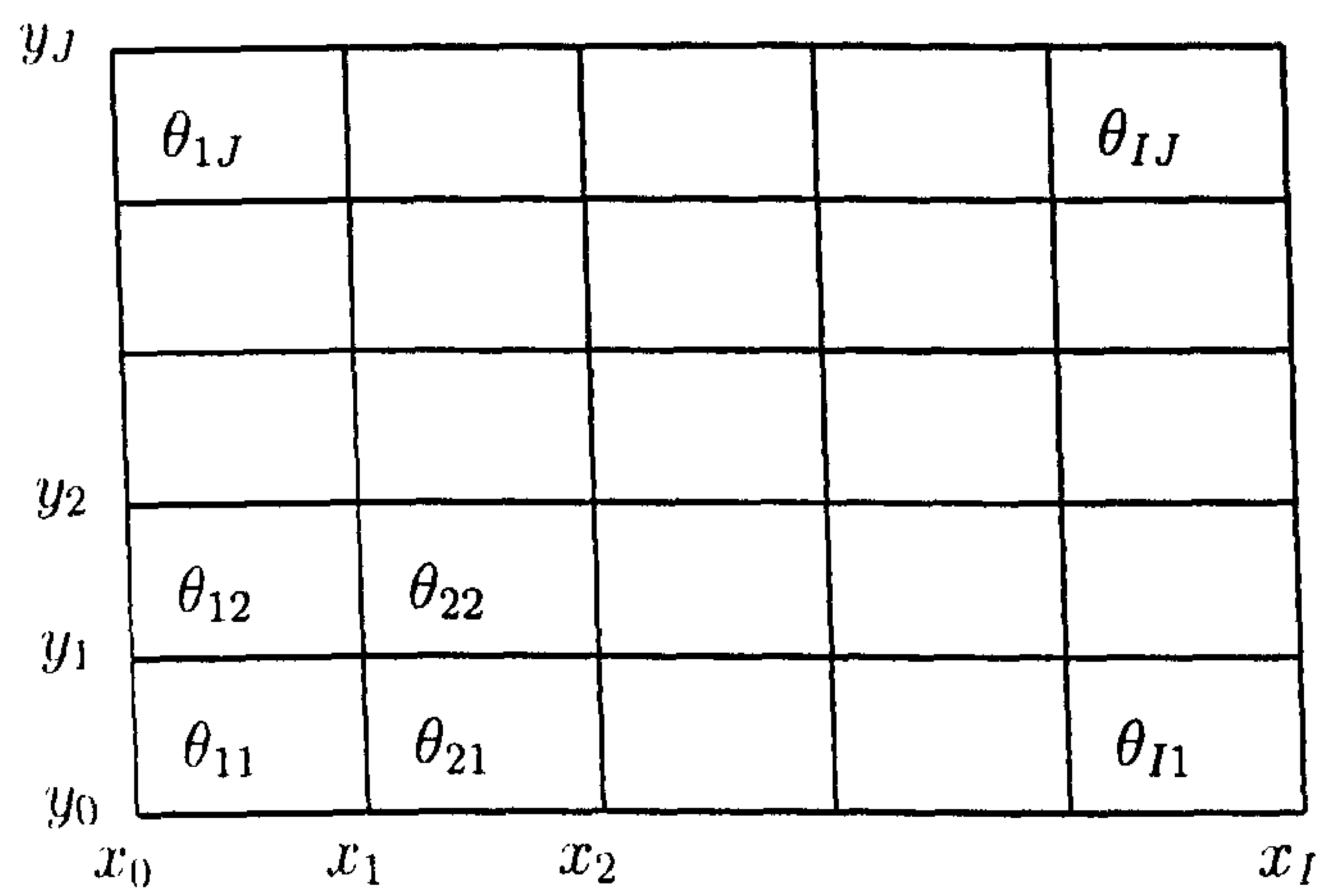
Suppose that in a survey (as in Contingent Valuation studies or market research), we are interested in two aspects of the population which are specified by two random variables X and Y, $X \in [C_1, D_1]$; $Y \in [C_2, D_2]$. In other words, we are concerned about two dimensions of the subjects and these two dimensions are characterized by two random variables X and Y. As in the univariate case, the questions are categorical with the recorded cutpoints. X has I categories and Y has J categories. So, the respective numbers of cutpoints are $I-1$ for variable X and $J-1$ for variable Y.

$$\underline{X} = (x_1, x_2, \ldots, x_{I-1}); \quad \underline{Y} = (y_1, y_2, \ldots, y_{J-1}).$$

We have a two-variable problem. The cell probability θ_{ij} is the probability that the first response falls between two cutpoints x_{i-1} and x_i and the second response falls between two cutpoints y_{j-1} and y_j:

$$\theta_{ij} = P(x_{i-1} \leq X \leq x_i, y_{j-1} \leq Y \leq y_j).$$

We can use the following diagram to depict the relationships between the variables and cell probabilities.



$$x_0 = C_1, x_I = D_1; \ \ y_0 = C_2, y_J = D_2.$$

We assume that the two responses from the respondents are related and are characterized by the joint distribution function between X and Y, namely $F(x_i, y_j)$. We further assume that the two variables X and Y have the same standardized marginal distribution function but their location and scale parameters may be the same or different. For generality, we denote these parameters as μ_X, σ_X and μ_Y, σ_Y, respectively. We can standardize the variables X and Y as follows:

$$U = \frac{X - \mu_X}{\sigma_X} = \alpha_X + \beta_X X, \tag{6.1}$$

$$V = \frac{Y - \mu_Y}{\sigma_Y} = \alpha_Y + \beta_Y Y, \tag{6.2}$$

where:

$$\alpha_X = -\frac{\mu_X}{\sigma_X}, \ \beta_X = \frac{1}{\sigma_X}; \ \alpha_Y = -\frac{\mu_Y}{\sigma_Y}, \ \beta_Y = \frac{1}{\sigma_Y}$$

and the design spaces are $[A_1, B_1]$ and $[A_2, B_2]$ for U and V respectively,

$$A_1 = \frac{C_1 - \mu_X}{\sigma_X}, \quad B_1 = \frac{D_1 - \mu_X}{\sigma_X},$$

$$A_2 = \frac{C_2 - \mu_Y}{\sigma_Y}, \quad B_2 = \frac{D_2 - \mu_Y}{\sigma_Y}.$$

For convenience, we denote:

$$\underline{\gamma} = (\gamma_1, \gamma_2, \gamma_3, \gamma_4)^T = (\alpha_X, \beta_X, \alpha_Y, \beta_Y)^T.$$

If U and V have joint distribution $F_2(u, v; \psi)$ where ψ is the measure of association between U and V, then X and Y have joint distribution

$$\begin{aligned} F_2(x, y, \psi) &= F_2\left\{\frac{x-\mu_X}{\sigma_X}, \frac{y-\mu_Y}{\sigma_Y}; \psi\right\} \\ &= F_2(\alpha_X + \beta_X x, \alpha_Y + \beta_Y y; \psi) \\ &= F_2(u, v; \psi). \end{aligned} \tag{6.3}$$

$F_2(u, v; \psi)$ is a standardized bivariate cdf model. As mentioned above, in the literature, many authors use the bivariate probit model. In our case, we focus on using a copula form of the function $F_2(u, v; \psi)$. A copula has the form:

$$F_2(u, v; \psi) = H\{F_1(u), F_1(v); \psi\} \tag{6.4}$$

where $F_1(u)$ and $F_1(v)$ are standardized marginal cdf's of U and V respectively.

The Plackett Copula will be used as the joint distribution function between the two variables X and Y.

6.2.1 Concept of copula and Plackett copula

In this section, we will briefly introduce the concept of copula and the Plackett copula. A class of functions called copulas was first introduced by A. Sklar in 1959, when answering a question raised by M. Fréchet about the relationship between a multidimensional probability function and its lower dimensional margins. These new functions are restrictions to $[0,1]^2$ of bivariate distribution functions whose margins are uniform in $[0,1]$. In short, Sklar showed that if H is a bivariate distribution function with margins $F(x)$ and $G(y)$, then there exists a copula C such that $H(x,y) = C\{F(x), G(y)\}$.
At the beginning, copulas were mainly used in the development of the theory of probabilistic metric spaces. Later, they were of interest for defining nonparametric measures of the dependence between random variables, and since then, they have began to play an important role in probability and mathematical statistics.
Here is the definition of copula for the bivariate case:
Definition: A copula is a function $C : [0,1]^2 \to [0,1]$ which satisfies:

1. For every u, v in $[0,1]$, $C(u,0) = C(0,v) = 0$ and $C(u,1) = u$ and $C(1,v) = v$.

2. For every u_1, u_2, v_1, v_2 in $[0,1]$ such that $u_1 \leq u_2$ and $v_1 \leq v_2$, we have:

$$C(u_2,v_2) - C(u_2,v_1) - C(u_1,v_2) + C(u_1,v_1) \geq 0$$

One of the copula families that we use throughout this chapter is the Plackett family. We can define this copula as follows:

For $\psi > 0$, the members of the Plackett family are defined by:

$$\begin{aligned} F_2(u_i, v_j) &= H\{F_1(u_i), F_1(v_j)\} = H(F_u, F_v) \\ &= \frac{[1 + (\psi - 1)(F_u + F_v)] - \sqrt{[1 + (\psi - 1)(F_u + F_v)]^2 - 4\psi(\psi - 1)F_u F_v}}{2(\psi - 1)} \end{aligned}$$

where for us, $F_u = F_1(u_i)$ and $F_v = F_1(v_j)$ and ψ is the measure of association. From now on, we denote $F_1(.)$ as the marginal distribution function and $F_2(u, v)$ the joint distribution function.

Let: $L = \{1 + (\psi - 1)[F_1(u_1) + F_1(v_2)]\}$, the above formula is as follows:

$$H(F_u, F_v) = \frac{L - \sqrt{L^2 - 4\psi(\psi - 1)F_u F_v}}{2(\psi - 1)} \tag{6.5}$$

There are three cases depending on value of ψ:

- $\psi = 1$: X and Y are independent.
- $\psi > 1$: There is positive association between X and Y.
- $\psi < 1$: There is negative association between X and Y.

Note that in the limit as $\psi \to 1$, using *l'Hopital's* rule, we have:

$$H(F_u, F_v) \to F_1(u_i)F_1(v_j)$$

With the definition of the copula $F_2(u_i, v_j)$, we can calculate the cell probability θ_{ij} in the following way:

$$\theta_{ij} = F_2(u_i, v_j) - F_2(u_{i-1}, v_j) - F_2(u_i, v_{j-1}) + F_2(u_{i-1}, v_{j-1}) \tag{6.6}$$

with:

$$F_2(u_0, v_j) = F_2(u_i, v_0) = 0$$

$$F_2(u_I, v_j) = F_1(v_j), \quad F_2(u_i, v_J) = F_1(u_i) \text{ and } F_2(u_I, v_J) = 1$$

6.2.2 Fisher information matrix

In the univariate case, the underlying model is a one dimensional multinomial distribution with cell probabilities θ_i. In the bivariate case, our model is also multinomial but with two dimensions. We will construct the formula for the Fisher information matrix. This procedure is similar to the procedure we used in the univariate case.

Denote:

$$\theta_{ij} = P(x_{i-1} \leq X \leq x_i, y_{j-1} \leq Y \leq y_j), \quad i = 1, 2, \ldots, I; \quad j = 1, 2, \ldots, J$$

and:

$$\underline{\theta} = (\theta_{11}, \theta_{12}, \ldots, \theta_{1I}, \ldots \theta_{I1}, \theta_{I2}, \ldots \theta_{IJ}), \quad \underline{\theta} = \underline{\theta}(\underline{\gamma}), \quad \sum_{i=1}^{I} \sum_{j=1}^{J} \theta_{ij} = 1,$$

where $\underline{\gamma}$ is a vector of parameters to be estimated. (In our case, $\underline{\gamma} = (\gamma_1, \gamma_2, \gamma_3, \gamma_4)^T = (\alpha_X, \beta_X, \alpha_Y, \beta_Y)^T$ or $\underline{\gamma} = (\alpha, \beta)^T$).

Now let:

$$\underline{U} = (u_{11}, u_{12}, \ldots, u_{1I}, \ldots u_{I1}, u_{I2}, \ldots u_{IJ})$$

where:

$$U_{ij} = \begin{cases} 1 & : \text{ if } \ (x_{i-1} \leq X \leq x_i) \text{ and } (y_{j-1} \leq X \leq y_j) \\ 0 & : \text{ if } \ \text{otherwise} \end{cases}$$

Then, $\underline{U}$ has a multinomial distribution.

$$\underline{U} \sim \mathcal{M}(1, \underline{\theta}). \tag{6.7}$$

$$\begin{aligned} E(U_{ij}) &= \theta_{ij}, \\ Var(U_{ij}) &= \theta_{ij}[1 - \theta_{ij}], \\ Cov(U_{ij}, U_{rs}) &= -\theta_{ij}\theta_{rs}. \end{aligned}$$

The covariance matrix is:

$$Cov(\underline{U}) = D_{\underline{\theta}} - \underline{\theta}\underline{\theta}^T$$

where $D_{\underline{\theta}} = diag(\underline{\theta})$, $\underline{\theta}$ is defined above.

The Log-likelihood function is:

$$\sum_{i=1}^{I}\sum_{j=1}^{J} U_{ij} ln(\theta_{ij})$$

Extending the formula 3.7 for Fisher information matrix in the univariate case, we have the general formula for Fisher information matrix in bivariate case as follows:

$$I_{X,Y}(\underline{\gamma}) = \sum_{i=1}^{I}\sum_{j=1}^{J} \left(\frac{\partial\theta_{ij}}{\partial\underline{\gamma}}\right) \frac{1}{\theta_{ij}} \left(\frac{\partial\theta_{ij}}{\partial\underline{\gamma}}\right)^T \quad (6.8)$$

6.2.3 Design objectives

As in the univariate case, our objective is to choose the cutpoints u_i and v_j optimally in order to obtain good estimation of some aspects of the parameters in the models 6.1 and 6.2. Depending on the objective of having good estimation of each parameter alone or good estimation of all parameters simultaneously, we can construct the following criteria.

1. Efficient estimation of each parameter alone:

For instance, if we want to have efficient estimation of parameter μ_X, we will minimize $Var(\hat{\mu}_X)$ where $\hat{\mu}_X$ is the estimator of μ_X.

We know that:

$$\mu_X = -\alpha_X/\beta_X \Rightarrow \hat{\mu}_X = -\hat{\alpha}_X/\hat{\beta}_X \quad \text{and} \quad Var(\hat{\mu}_X) \cong Var(\underline{c}_1^T\hat{\underline{\gamma}}),$$

where:

$$\underline{c}_1 = \frac{\partial\mu_X}{\partial\underline{\gamma}} \propto \frac{-(1, \mu_X, 0, 0)^T}{\beta_X}.$$

Through standardization, we have:

$$\underline{c}_1^T I_{X,Y}(\underline{\gamma})\underline{c}_1 = \underline{c}_{B_1}^T I_{U,V}(\underline{\gamma})\underline{c}_{B_1},$$

where

$$\underline{c}_{B_1} = B\underline{c}_1,$$

in which B is the matrix

$$B = \begin{pmatrix} 1 & 0 & 0 & 0 \\ \alpha_X & \beta_X & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & \alpha_Y & \beta_Y \end{pmatrix}.$$

So the initial form of a c-optimal criterion, as a function of $I_{X,Y}(\underline{\gamma})$, is transformed to another c-optimal criterion as a function of $I_{U,V}(\underline{\gamma})$ which is the standardized version of the Fisher information matrix with respect to the variables U and V.
In this particular case, as $\underline{c}_1 = \dfrac{-(1,\mu_X,0,0)^T}{\beta_X} \Rightarrow \underline{c}_{B_1} = B\underline{c}_1 = (\dfrac{-1}{\beta_X},0,0,0)^T$
So, in order to have efficient estimation of the parameter μ_X, we minimize $\underline{c}_{B_1}^T I_{U,V}^{-1}(\underline{\gamma})\underline{c}_{B_1}$ which is equivalent to minimizing
$(1,0,0,0)I_{U,V}^{-1}(\underline{\gamma})(1,0,0,0)^T$ i.e. $\underline{c}_{B_1} \propto \underline{e}_1 = (1,0,0,0)^T$.
We have the following definition:

- e_1-**optimality**: A design is called e_1-optimal if it maximizes the value of the function:

  $$-\underline{e}_1^T I_{U,V}^{-1}(\underline{\gamma})\underline{e}_1,$$

  where $\underline{e}_1 = (1,0,0,0)$.

Similarly, if we want to have efficient estimation of either parameter σ_X or μ_Y or σ_Y we minimize $Var(\hat{\sigma}_X)$ or $Var(\hat{\mu}_Y)$ or $Var(\hat{\sigma}_Y)$ respectively . Using the same manner as above, we have:

$$\underline{c}_2 = \frac{\partial \sigma_X}{\partial \underline{\gamma}} \propto \frac{-(0,1,0,0)^T}{\beta_X^2}$$
$$\underline{c}_3 = \frac{\partial \mu_Y}{\partial \underline{\gamma}} \propto \frac{-(0,0,1,\mu_Y)^T}{\beta_Y}$$
$$\underline{c}_4 = \frac{\partial \sigma_Y}{\partial \underline{\gamma}} \propto \frac{-(0,0,0,1)^T}{\beta_Y^2}$$

and:

$$\underline{c}_{B_2} = B\underline{c}_2 = (0, -\alpha_X, 0, 0)^T$$
$$\underline{c}_{B_3} = B\underline{c}_3 = (0, 0, \frac{-1}{\beta_X}, 0, 0, 0)^T$$
$$\underline{c}_{B_4} = B\underline{c}_4 = (0, 0, 0, -\alpha_Y)^T$$

We also easily see that:

$$\underline{c}_{B_2} \propto \underline{e}_2 = (0,1,0,0)^T$$
$$\underline{c}_{B_3} \propto \underline{e}_3 = (0,0,1,0)^T$$
$$\underline{c}_{B_4} \propto \underline{e}_4 = (0,0,0,1)^T$$

Thus, we have following definitions:

- e_2-**optimality**: A design is called e_2-optimal if it maximizes the value of the function:

$$-\underline{e}_2^T I_{U,V}^{-1}(\underline{\gamma})\underline{e}_2,$$

where $\underline{e}_1 = (0,1,0,0)$.

- e_3-**optimality**: A design is called e_3-optimal if it maximizes the value of the function:

$$-\underline{e}_3^T I_{U,V}^{-1}(\underline{\gamma})\underline{e}_3,$$

where $\underline{e}_1 = (0,0,1,0)$.

- e_4-**optimality**: A design is called e_4-optimal if it maximizes the value of the function:

$$-\underline{e}_4^T I_{U,V}^{-1}(\underline{\gamma})\underline{e}_4,$$

where $\underline{e}_1 = (0, 0, 0, 1)$.

2. Efficient estimation of all parameters:

We wish to make $Cov(\hat{\underline{\gamma}})$ 'small'. As in the univariate case in chapter two, we can use either D-optimality or A-optimality.

- D-**optimality** : $maximize\{log\, det(I_{U,V})\}$
- A-**optimality** : $maximize\{-tr\,(I_{U,V}^{-1})\}$

Possible marginal distributions for variables X and Y are symmetric distributions such as the logistic, normal, double exponential and double reciprocal and asymmetric distributions such as the complementary log-log and skewed logistic. We will investigate these in detail by dividing into two main cases. The first one is a two parameter case when we assume that the location and scale parameters of X and Y are the same. The second case is the four parameter case when we assume these parameters are different. We will come back to these cases later on.

6.3 Case 1: The two marginal distributions are identical in their parameters

The assumed distributions for two variables X and Y are the same and can be in the form of symmetric distributions such as logistic, normal, double exponential and double reciprocal or asymmetric distributions such as complementary log-log and skewed logistic.

6.3.1 Models and the Fisher information matrix

In this case, we assume in the model 6.1 and 6.2 that $\alpha_X = \alpha_Y = \alpha$ and $\beta_X = \beta_Y = \beta$. Our models are then two parameter ones under which:

$$u_i = \alpha + \beta x_i \tag{6.9}$$

$$v_j = \alpha + \beta y_j \tag{6.10}$$

The Fisher information matrix 6.8 can be written in the form:

$$I(\alpha, \beta / u_i, v_j) = \sum_{i=1}^{I} \sum_{j=1}^{J} \frac{1}{\theta_{ij}} \begin{bmatrix} \left(\frac{\partial \theta_{ij}}{\partial \alpha}\right)^2 & \frac{\partial \theta_{ij}}{\partial \alpha} \frac{\partial \theta_{ij}}{\partial \beta} \\ \frac{\partial \theta_{ij}}{\partial \alpha} \frac{\partial \theta_{ij}}{\partial \beta} & \left(\frac{\partial \theta_{ij}}{\partial \beta}\right)^2 \end{bmatrix} \tag{6.11}$$

In order to compute the Fisher information matrix, we need to have the derivatives of θ_{ij} with respect to the parameters α and β. We now demonstrate the calculation procedure by using some particular cases.

6.3.2 Symmetric distribution cases

We first focus on the symmetric marginal distributions, say logistic, normal, double exponential and double reciprocal distributions. Because of the symmetry, we also assume that the forms of the cutpoints in each dimension are symmetrical. The number of the cutpoints in each dimension may or may not be equal. The sets of cutpoints in each dimension can be the same or different. Since the model has two parameters, to ensure estimation of both parameters, we need at least one cutpoint in each dimension. We use the notation $\{(.);(.)\}$ to present the two sets of cutpoints in two dimensions where the first cutpoint or set of cutpoints is enclosed in the first round parentheses.

With this notation, we can write the form for the most simple case, which has one cutpoint in each direction, as follows:

$$\{(x);(y)\}$$

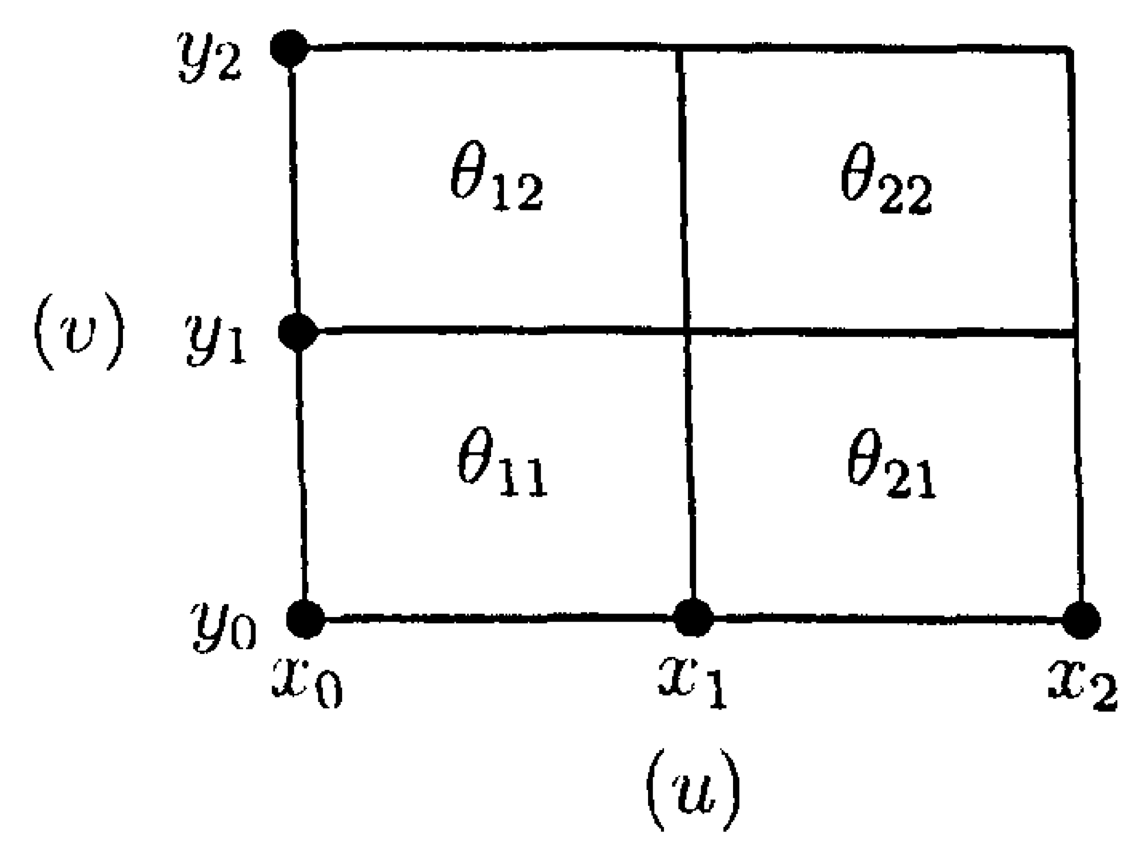
If we transform from x and y to the u and v through the standardization, we have the form: $\{(u);(v)\}$.

Similarly, we can extend to other cases and here are some we will consider:

1. $\{(x);(y)\} \rightarrow \{(u);(v)\}$
2. $\{(0);(-y,y)\} \rightarrow \{(0);(-v,v)\}$
3. $\{(0);(-y,0,y)\} \rightarrow \{(0);(-v,0,v)\}$
4. $\{(-x,x);(-x,x)\} \rightarrow \{(-u,u);(-u,u)\}$
5. $\{(-x,x);(-x,0,x)\} \rightarrow \{(-u,u);(-u,0,u)\}$
6. $\{(-x,x);(-y,0,y)\} \rightarrow \{(-u,u);(-v,0,v)\}$
7. $\{(-x,x);(-y,y)\} \rightarrow \{(-u,u);(-v,v)\}$
8. $\{(-x,0,x);(-x,0,x)\} \rightarrow \{(-u,0,u);(-u,0,u)\}$
9. $\{(-x,0,x);(-y,0,y)\} \rightarrow \{(-u,0,u);(-v,0,v)\}$
10. $\{(-x_2,-x_1,x_1,x_2);(-x_2,-x_1,x_1,x_2)\}$
    $\rightarrow \{(-u_2,-u_1,u_1,u_2);(-u_2,-u_1,u_1,u_2)\}$
11. $\{(-x_2,-x_1,0,x_1,x_2);(-x_2,-x_1,0,x_1,x_2)\}$
    $\rightarrow \{(-u_2,-u_1,0,u_1,u_2);(-u_2,-u_1,0,u_1,u_2)\}$

In all the cases above, we are generating a one or two variable optimization problem. We will use search methods to find optimal solutions. Since search methods are used, we limit the number of variables to two. Then the number of cutpoints in each dimension can vary from one to five (including cutpoint 0).

We now consider the first case above to demonstrate the computational procedure. We use the square diagram below to describe the case.



In this case, we have only one cutpoint for each dimension. This is a two variable and two parameter problem.

If the design space for X and Y is from $-\infty$ to ∞, in this case we have $x_0 = y_0 = -\infty$ and $x_2 = y_2 = \infty$. Through the standardized transformation, we have new variables:

$$u = \alpha + \beta x_1 \text{ and } v = \alpha + \beta y_1$$

Using formula 6.6, we have:

$$\begin{aligned}
\theta_{11} &= F_2(u, v) \\
\theta_{12} &= F_1(u) - F_2(u, v) \\
\theta_{21} &= F_1(v) - F_2(u, v) \\
\theta_{22} &= 1 - F_1(u) - F_1(v) + F_2(u, v)
\end{aligned}$$

Here $F_1(.)$ is the marginal distribution function and $F_2(..)$ is the joint distribution function.

We now calculate the derivatives.

$$\begin{aligned}
\frac{\partial\theta_{11}}{\partial\alpha} &= \frac{\partial F_2(u,v)}{\partial u} + \frac{\partial F_2(u,v)}{\partial v} \\
\frac{\partial\theta_{11}}{\partial\beta} &= x_1\frac{\partial F_2(u,v)}{\partial u} + y_1\frac{\partial F_2(u,v)}{\partial v} \\
\frac{\partial\theta_{12}}{\partial\alpha} &= f(u) - \frac{\partial\theta_{11}}{\partial\alpha} \\
\frac{\partial\theta_{12}}{\partial\beta} &= x_1 f(u) - \frac{\partial\theta_{11}}{\partial\beta} \\
\frac{\partial\theta_{21}}{\partial\alpha} &= f(v) - \frac{\partial\theta_{11}}{\partial\alpha} \\
\frac{\partial\theta_{21}}{\partial\beta} &= y_1 f(v) - \frac{\partial\theta_{11}}{\partial\beta} \\
\frac{\partial\theta_{22}}{\partial\alpha} &= -[f(u) + f(v)] + \frac{\partial\theta_{11}}{\partial\alpha} \\
\frac{\partial\theta_{22}}{\partial\beta} &= -[x_1 f(u) + y_1 f(v)] + \frac{\partial\theta_{11}}{\partial\beta}
\end{aligned}$$

where $f(.)$ is the density function: $f(.) = F'(.)$.

We know that:

$$\begin{pmatrix} 1 \\ u_i \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ \alpha & \beta \end{pmatrix}\begin{pmatrix} 1 \\ x_i \end{pmatrix} = B\begin{pmatrix} 1 \\ x_i \end{pmatrix} \rightarrow \begin{pmatrix} 1 \\ x_i \end{pmatrix} = B^{-1}\begin{pmatrix} 1 \\ u_i \end{pmatrix}$$

Then:

$$I(\alpha,\beta|x,y) = BI(u,v)B^T$$

So, through the standardization, we transform from variables X and Y to U and V and evaluate the information matrix $I(u,v)$ or functions of it which are independent of the parameters α and β.

In the above formula, we can replace x and y by u and v without any loss of generality.

In our case, we use the Plackett copula as the joint distribution function between the two variables X and Y or between U and V. We know that:

$$F_2(u,v) = H\{F_1(u), F_1(v)\} = H(F_u, F_v)$$

Then:

$$\begin{aligned}\frac{\partial F_2(u,v)}{\partial u} &= f(u)\frac{\partial H\{F_1(u),F_1(v)\}}{\partial F_1(u)} = f(u)\frac{\partial H(F_u,F_v)}{\partial F_u}\\ \frac{\partial F_2(u,v)}{\partial v} &= f(v)\frac{\partial H\{F_1(u),F_1(v)\}}{\partial F_1(v)} = f(v)\frac{\partial H(F_u,F_v)}{\partial F_v}\end{aligned}$$

From formula 6.5, let $P = \sqrt{L^2 - 4\psi(\psi-1)F_uF_v}$. We calculate the derivatives as:

$$\begin{aligned}\frac{\partial H(F_u,F_v)}{\partial F_u} &= \frac{(\psi-1) - \dfrac{\frac{1}{2}[2(\psi-1)L - 4\psi(\psi-1)F_v]}{P}}{2(\psi-1)}\\ &= \frac{1}{2}\left[1 - \frac{L-2\psi F_v}{P}\right] = \frac{1}{2}\left[1 + \frac{2\psi F_v - L}{P}\right]\end{aligned}$$

Or:

$$\frac{\partial H\{F_1(u),F_1(v)\}}{\partial F_1(u)} = \frac{1}{2}\left[1 + \frac{(\psi+1)F_1(v) - (\psi-1)F_1(u) - 1}{P}\right] \tag{6.12}$$

Similarly:

$$\frac{\partial H\{F_1(u),F_1(v)\}}{\partial F_1(v)} = \frac{1}{2}\left[1 + \frac{(\psi+1)F_1(u) - (\psi-1)F_1(v) - 1}{P}\right] \tag{6.13}$$

We are still using search methods to find the optimal solutions. We start the calculation procedure from values for u and v. Then, given our assumed distribution for the variables, we compute the values of cell probabilities θ_{ij} and the derivatives of θ_{ij} w.r.t. parameters. The next step is to calculate the

Fisher information matrix and evaluate the values for the chosen criterion function. Note that the optimal designs will depend on the values of the coefficient of association ψ.

6.3.3 Some results.

In tables 6.1 and 6.2 below, we report the D-optimal solutions in respect of the logistic distribution for the four cases (1), (4), (6) and (8) above, namely $\{(u);(v)\}$, $\{(-u,u);(-u,u)\}$,
$\{(-u,0,u);(-u,0,u)\}$ and $\{(-u,u);(-v,0,v)\}$. For each case, we change the value of the coefficient of association ψ from 0.001 (very low negative association) to 100 (very high positive association). The optimal cutpoints and optimal criterion values are also reported. Note that after standardization, cases (4) and (6) are one-variable optimizing problems but cases (1) and (8) are two-variable optimizing problems.

Comments

We can see that when the parameter ψ increases, the results do not change significantly in terms of optimal cutpoints and optimal criterion values. For example, in the case of $\{(u);(v)\}$, the optimal u^* and v^* are symmetrical and slightly change from the smallest value of -1.55 to the biggest value of -1.37. So, we can say that the optimal cutpoints are not very sensitive to the value of ψ.

When the parameter ψ approaches 1, we see that the optimal criterion values obtained are similar to the ones we found in the respective univariate cases. For instance, in the case $\{(-u,u);(-v,0,v)\}$, when $\psi = 0.999$, $u^* = 1.52$,

$v^* = 2.06$. These values are similar to the optimal cutpoints in the three and four category cases respectively. We will investigate the independent case ($\psi = 1$) to verify this result.

We also found similar results in terms of optimal cutpoints and optimal criterion values between the following pairs of cases:

$$\{(-u, u); (-v, v)\} \text{ and } \{(-u, u); (-u, u)\}$$

$$\{(-u, 0, u); (-v, 0, v)\} \text{ and } \{(-u, 0, u); (-u, 0, u)\}$$

Table 6.1: D-optimal cutpoints for: logistic distribution; bivariate approach; two parameters; cases (1) and (4).

ψ	Form$\{(u);(v)\}$		Form $\{(-u,u);(-u,u)\}$	
	(u^*,v^*)	$\Phi^*(u^*,v^*)$	u^*	$\Phi^*(u^*)$
0.001	(-1.3700, 1.3700)	0.2969	1.1200	2.0450
0.010	(-1.3900, 1.3900)	-0.6946	1.1000	1.0895
0.100	(-1.4700, 1.4700)	-1.4005	1.1900	0.3695
0.200	(-1.5000, 1.5000)	-1.5168	1.2700	0.1984
0.300	(-1.5200, 1.5200)	-1.5610	1.3300	0.1053
0.400	(-1.5300, 1.5300)	-1.5825	1.3700	0.0410
0.500	(-1.5300, 1.5300)	-1.5940	1.3900	-0.0087
0.600	(-1.5400, 1.5400)	-1.6003	1.4200	-0.0498
0.700	(-1.5400, 1.5400)	-1.6040	1.4300	-0.0851
0.800	(-1.5500, 1.5500)	-1.6060	1.4500	-0.1164
0.900	(-1.5400, 1.5400)	-1.6068	1.4600	-0.1446
0.990	(-1.5400, 1.5400)	-1.6070	1.4700	-0.1679
0.999	(-1.5400, 1.5400)	-1.6070	1.4700	-0.1701
2.000	(-1.5500, 1.5500)	-1.6005	1.5100	-0.3551
3.000	(-1.5300, 1.5300)	-1.5919	1.5200	-0.4738
4.000	(-1.5200, 1.5200)	-1.5862	1.5200	-0.5606
5.000	(-1.5100, 1.5100)	-1.5821	1.5200	-0.6280
10.00	(-1.4900, 1.4900)	-1.5716	1.5100	-0.8292
20.00	(-1.4800, 1.4800)	-1.5648	1.5000	-1.0051
50.00	(-1.4700, 1.4700)	-1.5601	1.4900	-1.1869
100.0	(-1.4700, 1.4700)	-1.5584	1.4300	-1.2897

Table 6.2: D-optimal cutpoints for: logistic distribution; bivariate approach; two parameters; cases (6) and (8).

ψ	Form $\{(-u,u);(-v,0,v)\}$		Form $\{(-u,0,u);(-u,0,u)\}$	
	(u^*,v^*)	$\Phi^*(u^*,v^*)$	u^*	$\Phi^*(u^*)$
0.001	(0.1100, 2.1800)	2.3819	1.3900	2.7065
0.010	(0.3300, 1.8800)	1.4279	1.5200	1.1985
0.100	(0.9200, 1.8900)	0.6568	1.6700	0.9006
0.200	(1.1300, 1.9400)	0.4344	1.8100	0.6520
0.300	(1.2400, 1.9500)	0.3143	1.8800	0.5141
0.400	(1.3600, 1.9700)	0.2345	1.9100	0.4199
0.500	(1.4200, 2.0000)	0.1739	1.9400	0.3489
0.600	(1.4500, 2.0100)	0.1257	1.9500	0.2921
0.700	(1.4700, 2.0200)	0.0846	1.9600	0.2449
0.800	(1.4900, 2.0400)	0.0488	1.9700	0.2045
0.900	(1.5000, 2.0400)	0.0177	1.9700	0.1692
0.990	(1.5100, 2.0500)	-0.0076	1.9800	0.1409
0.999	(1.5200, 2.0600)	-0.0100	1.9800	0.1382
2.000	(1.5700, 2.0900)	-0.2011	2.0100	-0.0631
3.000	(1.6000, 2.0900)	-0.3147	2.0300	-0.1776
4.000	(1.6000, 2.1000)	-0.3948	2.0300	-0.2576
5.000	(1.6000, 2.1100)	-0.4553	2.0400	-0.3188
10.00	(1.5800, 2.2000)	-0.6296	2.0300	-0.5019
20.00	(1.4100, 2.4000)	-0.7687	2.0200	-0.6683
50.00	(1.3200, 2.5700)	-0.8833	2.0100	-0.8502
100.0	(1.2500, 2.7000)	-0.9289	2.2000	-0.9557

6.3.4 Special case: $\psi = 1$.

Consider the case $\psi = 1$, i.e. the two variables X and Y are independent.

Denote:

$$\theta_{i.} = P(x_{i-1} \leq X \leq x_i) = \sum_{j=1}^{J} \theta_{ij}$$

and:

$$\theta_{.j} = P(y_{j-1} \leq Y \leq y_i) = \sum_{i=1}^{I} \theta_{ij},$$

$$\sum_{i=1}^{I} \theta_{i.} = \sum_{j=1}^{J} \theta_{.j} = 1.$$

$\theta_{i.}$ and $\theta_{.j}$ are in turn the cell probabilities if we consider the two variables X and Y as the two separate univariate cases. Since X and Y are independent, we have:

$$\theta_{ij} = \theta_{i.} \times \theta_{.j}. \tag{6.14}$$

$\theta_{i.}$ and $\theta_{.j}$ depend on the two parameters α and β. So

$$\frac{\partial \theta_{ij}}{\partial \gamma} = \frac{\partial (\theta_{i.} \theta_{.j})}{\partial \gamma} = \left(\frac{\partial \theta_{i.}}{\partial \gamma}\right)\theta_{.j} + \theta_{i.}\left(\frac{\partial \theta_{.j}}{\partial \gamma}\right), \tag{6.15}$$

where γ can be either α or β.

From formula 6.11, we can find the elements of the Fisher information matrix.

First, the two diagonal elements are (for $\gamma = (\alpha, \beta)$):

$$
\begin{aligned}
I_{\gamma\gamma} &= \sum_{i=1}^{I}\sum_{j=1}^{J}\frac{1}{\theta_{ij}}\left(\frac{\partial\theta_{ij}}{\partial\gamma}\right)^2 \\
&= \sum_{i=1}^{I}\sum_{j=1}^{J}\frac{1}{\theta_{i.}\theta_{.j}}\left[\left(\frac{\partial\theta_{i.}}{\partial\gamma}\right)\theta_{.j} + \theta_{i.}\left(\frac{\partial\theta_{.j}}{\partial\gamma}\right)\right]^2 \\
&= \sum_{i=1}^{I}\sum_{j=1}^{J}\frac{1}{\theta_{i.}\theta_{.j}}\left[\left(\frac{\partial\theta_{i.}}{\partial\gamma}\right)^2(\theta_{.j})^2 + 2\left(\frac{\partial\theta_{i.}}{\partial\gamma}\frac{\partial\theta_{.j}}{\partial\gamma}\right)(\theta_{i.}\theta_{.j}) + (\theta_{i.})^2\left(\frac{\partial\theta_{.j}}{\partial\gamma}\right)^2\right] \\
&= \sum_{i=1}^{I}\sum_{j=1}^{J}\frac{1}{\theta_{i.}}\left(\frac{\partial\theta_{i.}}{\partial\gamma}\right)^2\theta_{.j} + 2\sum_{i=1}^{I}\sum_{j=1}^{J}\frac{\partial\theta_{i.}}{\partial\gamma}\frac{\partial\theta_{.j}}{\partial\gamma} + \sum_{i=1}^{I}\sum_{j=1}^{J}\left(\theta_{i.}\frac{\partial\theta_{i.}}{\partial\gamma}\right)^2\theta_{.j}\frac{1}{\theta_{.j}} \\
&= \left[\sum_{i=1}^{I}\frac{1}{\theta_{i.}}\left(\frac{\partial\theta_{i.}}{\partial\gamma}\right)^2\right]\left(\sum_{j=1}^{J}\theta_{.j}\right) + 2\left[\sum_{i=1}^{I}\frac{\partial\theta_{i.}}{\partial\gamma}\right]\left[\sum_{j=1}^{J}\frac{\partial\theta_{.j}}{\partial\gamma}\right] \\
&\qquad + \left(\sum_{i=1}^{I}\theta_{i.}\right)\left[\sum_{j=1}^{J}\frac{1}{\theta_{.j}}\left(\frac{\partial\theta_{.j}}{\partial\gamma}\right)^2\right] \\
&= \sum_{i=1}^{I}\frac{1}{\theta_{i}}\left(\frac{\partial\theta_{i.}}{\partial\gamma}\right)^2 + 0 + \sum_{j=1}^{J}\frac{1}{\theta_{.j}}\left(\frac{\partial\theta_{.j}}{\partial\gamma}\right)^2
\end{aligned}
\tag{6.16}
$$

Since $\sum_{i=1}^{I}\theta_{i.} = \sum_{j=1}^{J}\theta_{.j} = 1$, then $\sum_{i=1}^{I}\frac{\partial\theta_{i.}}{\partial\gamma} = \sum_{j=1}^{J}\frac{\partial\theta_{.j}}{\partial\gamma} = 0$

The two off-diagonal elements are identical given symmetry of I, and can be found as follows:

$$
\begin{aligned}
I_{\alpha\beta} &= \sum_{i=1}^{I}\sum_{j=1}^{J}\frac{1}{\theta_{ij}}\left(\frac{\partial\theta_{ij}}{\partial\alpha}\right)\left(\frac{\partial\theta_{ij}}{\partial\beta}\right)^{T} \\
&= \sum_{i=1}^{I}\sum_{j=1}^{J}\frac{1}{\theta_{i.}\theta_{.j}}\left[\left(\frac{\partial\theta_{i}}{\partial\alpha}\right)\theta_{.j}+\theta_{i.}\left(\frac{\partial\theta_{.j}}{\partial\alpha}\right)\right]\left[\left(\frac{\partial\theta_{i.}}{\partial\beta}\right)\theta_{.j}+\theta_{i.}\left(\frac{\partial\theta_{.j}}{\partial\beta}\right)\right] \\
&= \sum_{i=1}^{I}\sum_{j=1}^{J}\frac{1}{\theta_{i.}\theta_{.j}}\left(\frac{\partial\theta_{i.}}{\partial\alpha}\right)\left(\frac{\partial\theta_{i.}}{\partial\beta}\right)(\theta_{.j})^2+\sum_{i=1}^{I}\sum_{j=1}^{J}\frac{1}{\theta_{i.}\theta_{.j}}\theta_{i.}\theta_{.j}\left(\frac{\partial\theta_{i.}}{\partial\alpha}\frac{\partial\theta_{.j}}{\partial\beta}+\frac{\partial\theta_{i.}}{\partial\beta}\frac{\partial\theta_{.j}}{\partial\alpha}\right) \\
&\quad +\sum_{i=1}^{I}\sum_{j=1}^{J}\frac{1}{\theta_{i.}\theta_{.j}}(\theta_{i.})^2\left(\frac{\partial\theta_{.j}}{\partial\alpha}\right)\left(\frac{\partial\theta_{.j}}{\partial\beta}\right) \\
&= \left(\sum_{i=1}^{I}\frac{1}{\theta_{i.}}\frac{\partial\theta_{i.}}{\partial\alpha}\frac{\partial\theta_{i.}}{\partial\beta}\right)\left(\sum_{j=1}^{J}\theta_{.j}\right)+\sum_{i=1}^{I}\frac{\partial\theta_{i.}}{\partial\alpha}\sum_{j=1}^{J}\frac{\partial\theta_{.j}}{\partial\beta}+\sum_{i=1}^{I}\frac{\partial\theta_{i.}}{\partial\beta}\sum_{j=1}^{J}\frac{\partial\theta_{.j}}{\partial\alpha} \\
&\quad +\left(\sum_{i=1}^{I}\theta_{i.}\right)\left(\sum_{j=1}^{J}\frac{1}{\theta_{.j}}\frac{\partial\theta_{.j}}{\partial\alpha}\frac{\partial\theta_{.j}}{\partial\beta}\right) \\
&= \sum_{i=1}^{I}\frac{1}{\theta_{i.}}\frac{\partial\theta_{i.}}{\partial\alpha}\frac{\partial\theta_{i.}}{\partial\beta}+0+0+\sum_{j=1}^{J}\frac{1}{\theta_{.j}}\frac{\partial\theta_{.j}}{\partial\alpha}\frac{\partial\theta_{.j}}{\partial\beta} \qquad (6.17)
\end{aligned}
$$

So, the Fisher information matrix in this case is of the form:

$$
I(\alpha,\beta/u,v)=\begin{bmatrix}
\sum_{i=1}^{I}\frac{1}{\theta_{i.}}\left(\frac{\partial\theta_{i.}}{\partial\alpha}\right)^2+\sum_{j=1}^{J}\frac{1}{\theta_{.j}}\left(\frac{\partial\theta_{.j}}{\partial\alpha}\right)^2 & \sum_{i=1}^{I}\frac{1}{\theta_{i.}}\frac{\partial\theta_{i.}}{\partial\alpha}\frac{\partial\theta_{i.}}{\partial\beta}+\sum_{j=1}^{J}\frac{1}{\theta_{.j}}\frac{\partial\theta_{.j}}{\partial\alpha}\frac{\partial\theta_{.j}}{\partial\beta} \\
\sum_{i=1}^{I}\frac{1}{\theta_{i.}}\frac{\partial\theta_{i.}}{\partial\alpha}\frac{\partial\theta_{i.}}{\partial\beta}+\sum_{j=1}^{J}\frac{1}{\theta_{.j}}\frac{\partial\theta_{.j}}{\partial\alpha}\frac{\partial\theta_{.j}}{\partial\beta} & \sum_{i=1}^{I}\frac{1}{\theta_{i.}}\left(\frac{\partial\theta_{i.}}{\partial\beta}\right)^2+\sum_{j=1}^{J}\frac{1}{\theta_{.j}}\left(\frac{\partial\theta_{.j}}{\partial\beta}\right)^2
\end{bmatrix} \qquad (6.18)
$$

Form 6.18 reveals an expected result namely that in the two parameter case of a bivariate model, if the two variables X and Y are independent ($\psi = 1$), the Fisher information matrix is the sum of two Fisher information matrices of X and Y if we consider them as separate univariate cases, i.e.

$$I_{X,Y} = I_X + I_Y$$

or:

$$I(\alpha, \beta | u_i, v_j) = I(\alpha, \beta | u_i) + I(\alpha, \beta | v_j) \tag{6.19}$$

This structure of the Fisher information matrix explains the similarity of results, in terms of optimal cutpoints and criterion values, we found between the bivariate and the respective univariate cases for ψ close to 1.

6.3.5 Asymmetric distribution cases.

In the asymmetrical case, the cutpoints in each dimension may not be symmetric or zero may not be a cutpoint. As mentioned above, because we use a search method to find the optimal solution, we need to limit the search to at most a two variable optimization problem. In addition to the condition that there are at least two cutpoints in each case to ensure the estimation of two parameters, we consider the following scenarios for the asymmetrical distribution case.

1. $\{(x); (y)\} \rightarrow \{(u); (v)\}$

2. $\{(x, y); (x, y)\} \rightarrow \{(u, v); (u, v)\}$

Table 6.3: D-optimal cutpoints for: complementary log-log distribution; bivariate approach; two parameters.

ψ	Form$\{(u);(v)\}$		Form $\{(u,v);(u,v)\}$	
	(u^*,v^*)	$\Phi^*(u^*,v^*)$	(u^*,u^*)	$\Phi^*(u^*,v^*)$
0.001	(0.5900, -1.7100)	1.2729	(-0.3700, 1.1000)	3.5172
0.010	(0.6200, -1.7100)	0.3001	(-0.3800, 1.0900)	2.4461
0.100	(0.8500, -1.4800)	-0.3027	(-0.6000, 1.0200)	1.5453
0.200	(0.9300, -1.3800)	-0.3719	(-0.8200, 0.9900)	1.3476
0.300	(0.9600, -1.3500)	-0.3970	(-0.9800, 0.9700)	1.2483
0.400	(0.9700, -1.3400)	-0.4091	(-1.0700, 0.9600)	1.1833
0.500	(0.9700, -1.3400)	-0.4155	(-1.1400, 0.9600)	1.1352
0.600	(0.9800, -1.3300)	-0.4192	(-1.1900, 0.9600)	1.0966
0.700	(0.9800, -1.3400)	-0.4212	(-1.2300, 0.9600)	1.0641
0.800	(0.9800, -1.3400)	-0.4222	(-1.2500, 0.9600)	1.0357
0.900	(0.9800, -1.3400)	-0.4227	(-1.2800, 0.9600)	1.0104
0.990	(0.9800, -1.3400)	-0.4229	(-1.2900, 0.9600)	0.9896
0.999	(0.9800, -1.3400)	-0.4229	(-1.3000, 0.9600)	0.9876
2.000	(0.9800, -1.3300)	-0.4190	(-1.3700, 0.9800)	0.8236
3.000	(0.9700, -1.3300)	-0.4150	(-1.3800, 1.0000)	0.7175
4.000	(0.9700, -1.3200)	-0.4123	(-1.3800, 1.0100)	0.6391
5.000	(0.9700, -1.3200)	-0.4103	(-1.3700, 1.0200)	0.5773
10.00	(0.9600, -1.3100)	-0.4054	(-1.3400, 1.0500)	0.3873
20.00	(0.9600, -1.3100)	-0.4024	(-1.3100, 1.0600)	0.2113
50.00	(0.9600, -1.3000)	-0.4003	(-1.2900, 1.0500)	0.0177
100.0	(0.9600, -1.3000)	-0.3997	(-1.2900, 1.0300)	-0.0943

6.4 Case 2: The two marginal distributions differ in their parameters

Now we consider the case when the parameters in the models of the two dimensions are different.

6.4.1 Model and the Fisher information matrix

If the parameters of the two marginal distributions of X and Y are different, i.e. $\alpha_X \neq \alpha_Y$ and $\beta_X \neq \beta_Y$, we have a four parameter model. Denote $\alpha_X = \alpha_1$, $\alpha_Y = \alpha_2$, $\beta_X = \beta_1$, $\beta_Y = \beta_2$, the standardized design variables are:

$$u_i = \alpha_1 + \beta_1 x_i \tag{6.20}$$

$$v_j = \alpha_2 + \beta_2 y_j \tag{6.21}$$

Let $\underline{\gamma} = (\alpha_1, \beta_1, \alpha_2, \beta_2)^T$. The Fisher information matrix 6.8 is of the form:

$$I(\underline{\gamma}/u_i, v_j) = \sum_{i=1}^{I}\sum_{j=1}^{J}\frac{1}{\theta_{ij}}\begin{bmatrix} \left(\frac{\partial\theta_{ij}}{\partial\alpha_1}\right)^2 & \frac{\partial\theta_{ij}}{\partial\alpha_1}\frac{\partial\theta_{ij}}{\partial\beta_1} & \frac{\partial\theta_{ij}}{\partial\alpha_1}\frac{\partial\theta_{ij}}{\partial\alpha_2} & \frac{\partial\theta_{ij}}{\partial\alpha_1}\frac{\partial\theta_{ij}}{\partial\beta_2} \\ \frac{\partial\theta_{ij}}{\partial\alpha_1}\frac{\partial\theta_{ij}}{\partial\beta_1} & \left(\frac{\partial\theta_{ij}}{\partial\beta_1}\right)^2 & \frac{\partial\theta_{ij}}{\partial\alpha_2}\frac{\partial\theta_{ij}}{\partial\beta_1} & \frac{\partial\theta_{ij}}{\partial\beta_1}\frac{\partial\theta_{ij}}{\partial\beta_2} \\ \frac{\partial\theta_{ij}}{\partial\alpha_1}\frac{\partial\theta_{ij}}{\partial\alpha_2} & \frac{\partial\theta_{ij}}{\partial\alpha_2}\frac{\partial\theta_{ij}}{\partial\beta_1} & \left(\frac{\partial\theta_{ij}}{\partial\alpha_2}\right)^2 & \frac{\partial\theta_{ij}}{\partial\alpha_2}\frac{\partial\theta_{ij}}{\partial\beta_2} \\ \frac{\partial\theta_{ij}}{\partial\alpha_1}\frac{\partial\theta_{ij}}{\partial\beta_2} & \frac{\partial\theta_{ij}}{\partial\beta_1}\frac{\partial\theta_{ij}}{\partial\beta_2} & \frac{\partial\theta_{ij}}{\partial\alpha_2}\frac{\partial\theta_{ij}}{\partial\beta_2} & \left(\frac{\partial\theta_{ij}}{\partial\beta_2}\right)^2 \end{bmatrix} \tag{6.22}$$

In this case, we have four parameters. In order to ensure the estimation of all the parameters, we need a total of at least four distinguished cutpoints in the two dimensions. We consider the following scenarios.

1. $\{(0); (-y, 0, y)\} \rightarrow \{(0); (-v, 0, v)\}$

2. $\{(-x, x); (-x, x)\} \rightarrow \{(-u, u); (-u, u)\}$

3. $\{(-x, x); (-x, 0, x)\} \rightarrow \{(-u, u); (-u, 0, u)\}$

4. $\{(-x,0,x);(-x,0,x)\} \rightarrow \{(-u,0,u);(-u,0,u)\}$

5. $\{(-x,x);(-y,y)\} \rightarrow \{(-u,u);(-v,v)\}$

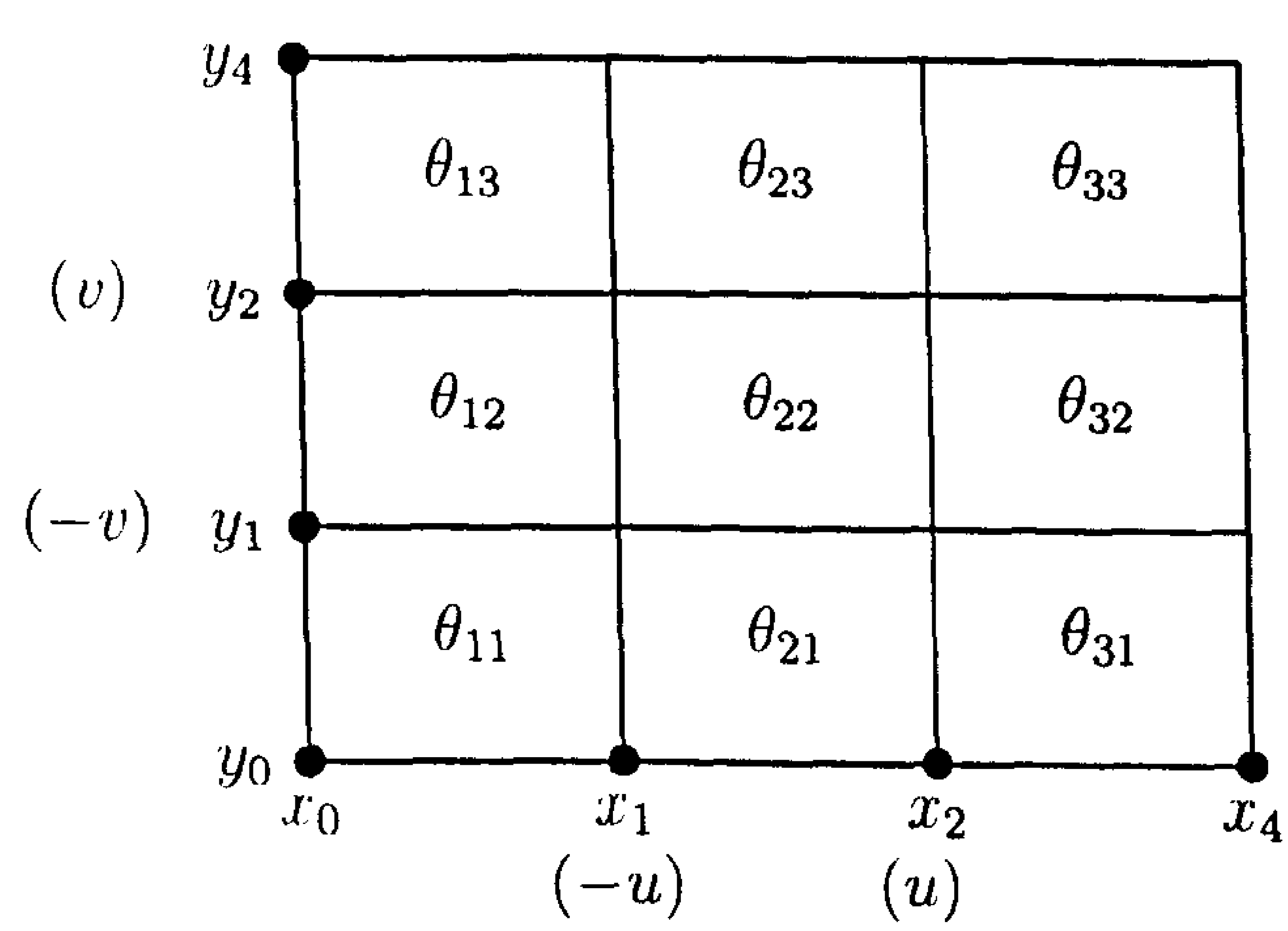
6. $\{(-x,x);(-y,0,y)\} \rightarrow \{(-u,u);(-v,0,v)\}$

7. $\{(-x,0,x);(-y,0,y)\} \rightarrow \{(-u,0,u);(-v,0,v)\}$

8. $\{(-x_2,-x_1,x_1,x_2);(-x_2,-x_1,x_1,x_2)\}$
   $\rightarrow \{(-u_2,-u_1,u_1,u_2);(-u_2,-u_1,u_1,u_2)\}$

9. $\{(-x_2,-x_1,0,x_1,x_2);(-x_2,-x_1,0,x_1,x_2)\}$
   $\rightarrow \{(-u_2,-u_1,0,u_1,u_2);(-u_2,-u_1,0,u_1,u_2)\}$

The calculation procedure is slightly different from the case of the two parameter model in the sense that u depends only on α_1 and β_1, v depend only on α_2 and β_2, so the derivative of θ_{ij} w.r.t the parameters are different compared with the two parameter case. We use the case $\{(-u,u);(-v,v)\}$ as an illustration for the calculation procedure.



The cell probabilities:

$$\begin{aligned}
\theta_{11} &= F_2(-u,-v)\\
\theta_{12} &= F_2(-u,v) - F_2(-u,-v)\\
\theta_{13} &= F_1(-u) - F_2(-u,v)\\
\theta_{21} &= F_2(u,-v) - F_2(-u,-v)\\
\theta_{22} &= F_2(u,v) - F_2(-u,v) - F_2(u,-v) + F_2(-u,-v)\\
\theta_{23} &= F_1(u) - F_1(-v) - F_2(u,v) + F_2(-u,v)\\
\theta_{31} &= F_1(-v) - F_2(u,-v)\\
\theta_{32} &= F_1(v) - F_1(-v) - F_2(u,v) + F_2(u,-v)\\
\theta_{33} &= 1 - F_1(v) - F_1(u) + F_2(u,v)
\end{aligned}$$

The derivatives of θ_{ij} w.r.t the parameters are:

$$\begin{aligned}
\frac{\partial\theta_{11}}{\partial\alpha_1} &= \frac{\partial F_2(-u,-v)}{\partial(-u)}\\
\frac{\partial\theta_{11}}{\partial\alpha_2} &= \frac{\partial F_2(-u,-v)}{\partial(-v)}\\
\frac{\partial\theta_{11}}{\partial\beta_1} &= x_1\frac{\partial F_2(-u,-v)}{\partial(-u)}\\
\frac{\partial\theta_{11}}{\partial\beta_2} &= y_1\frac{\partial F_2(-u,-v)}{\partial(-v)}
\end{aligned}$$

$$\begin{aligned}
\frac{\partial\theta_{12}}{\partial\alpha_1} &= \frac{\partial F_2(-u,v)}{\partial(-u)} - \frac{\partial\theta_{11}}{\partial\alpha_1}\\
\frac{\partial\theta_{12}}{\partial\alpha_2} &= \frac{\partial F_2(-u,v)}{\partial(v)} - \frac{\partial\theta_{11}}{\partial\alpha_2}\\
\frac{\partial\theta_{12}}{\partial\beta_1} &= x_1\frac{\partial F_2(-u,v)}{\partial(-u)} - \frac{\partial\theta_{11}}{\partial\beta_1}\\
\frac{\partial\theta_{12}}{\partial\beta_2} &= y_2\frac{\partial F_2(-u,v)}{\partial(v)} - \frac{\partial\theta_{11}}{\partial\beta_2}
\end{aligned}$$

$$\begin{aligned}
\frac{\partial\theta_{13}}{\partial\alpha_1} &= f(-u) - \frac{\partial F_2(-u,v)}{\partial(-u)} \\
\frac{\partial\theta_{13}}{\partial\alpha_2} &= -\frac{\partial F_2(-u,v)}{\partial v} \\
\frac{\partial\theta_{13}}{\partial\beta_1} &= x_1 f(-u) - x_1\frac{\partial F_2(-u,v)}{\partial(-u)} \\
\frac{\partial\theta_{12}}{\partial\beta_2} &= -y_2\frac{\partial F_2(-u,v)}{\partial v}
\end{aligned}$$

$$\begin{aligned}
\frac{\partial\theta_{21}}{\partial\alpha_1} &= \frac{\partial F_2(u,-v)}{\partial u} - \frac{\partial\theta_{11}}{\partial\alpha_1} \\
\frac{\partial\theta_{21}}{\partial\alpha_2} &= \frac{\partial F_2(u,-v)}{\partial(-v)} - \frac{\partial\theta_{11}}{\partial\alpha_2} \\
\frac{\partial\theta_{21}}{\partial\beta_1} &= x_2\frac{\partial F_2(u,-v)}{\partial(u)} - \frac{\partial\theta_{11}}{\partial\beta_1} \\
\frac{\partial\theta_{12}}{\partial\beta_2} &= y_1\frac{\partial F_2(u,-v)}{\partial(-v)} - \frac{\partial\theta_{11}}{\partial\beta_2}
\end{aligned}$$

$$\begin{aligned}
\frac{\partial\theta_{22}}{\partial\alpha_1} &= \frac{\partial F_2(u,v)}{\partial u} - \frac{\partial F_2(-u,v)}{\partial(-u)} - \frac{\partial F_2(u,-v)}{\partial u} + \frac{\partial\theta_{11}}{\partial\alpha_1} \\
\frac{\partial\theta_{22}}{\partial\alpha_2} &= \frac{\partial F_2(u,v)}{\partial v} - \frac{\partial F_2(-u,v)}{\partial v} - \frac{\partial F_2(u,-v)}{\partial(-v)} + \frac{\partial\theta_{11}}{\partial\alpha_2} \\
\frac{\partial\theta_{21}}{\partial\beta_1} &= x_2\frac{\partial F_2(u,v)}{\partial u} - x_1\frac{\partial F_2(-u,v)}{\partial(-u)} - x_2\frac{\partial F_2(u,-v)}{\partial u} + \frac{\partial\theta_{11}}{\partial\beta_1} \\
\frac{\partial\theta_{12}}{\partial\beta_2} &= y_2\frac{\partial F_2(u,v)}{\partial v} - y_2\frac{\partial F_2(-u,v)}{\partial v} - y_1\frac{\partial F_2(u,-v)}{\partial(-v)} + \frac{\partial\theta_{11}}{\partial\beta_2}
\end{aligned}$$

$$\begin{aligned}
\frac{\partial\theta_{23}}{\partial\alpha_1} &= f(u) - f(-u) - \frac{\partial F_2(u,v)}{\partial u} + \frac{\partial F_2(-u,v)}{\partial(-u)} \\
\frac{\partial\theta_{23}}{\partial\alpha_2} &= -\frac{\partial F_2(u,v)}{\partial v} + \frac{\partial F_2(-u,v)}{\partial v} \\
\frac{\partial\theta_{21}}{\partial\beta_1} &= x_2 f(u) - x_1 f(-u) - x_2\frac{\partial F_2(u,v)}{\partial u} + x_1\frac{\partial F_2(-u,v)}{\partial(-u)} \\
\frac{\partial\theta_{12}}{\partial\beta_2} &= -y_2\frac{\partial F_2(u,v)}{\partial v} + y_2\frac{\partial F_2(-u,v)}{\partial v}
\end{aligned}$$

$$\begin{aligned}
\frac{\partial\theta_{31}}{\partial\alpha_1} &= -\frac{\partial F_2(u,-v)}{\partial u} \\
\frac{\partial\theta_{31}}{\partial\alpha_2} &= f(-v) - \frac{\partial F_2(u,-v)}{\partial(-v)} \\
\frac{\partial\theta_{21}}{\partial\beta_1} &= -x_2\frac{\partial F_2(u,-v)}{\partial u} \\
\frac{\partial\theta_{12}}{\partial\beta_2} &= y_1 f(-v) - y_1\frac{\partial F_2(u,-v)}{\partial(-v)}
\end{aligned}$$

$$\begin{aligned}
\frac{\partial\theta_{32}}{\partial\alpha_1} &= -\frac{\partial F_2(u,v)}{\partial u} + \frac{\partial F_2(u,-v)}{\partial(u)} \\
\frac{\partial\theta_{23}}{\partial\alpha_2} &= f(v) - f(-v) - \frac{\partial F_2(u,v)}{\partial v} + \frac{\partial F_2(u,-v)}{\partial(-v)} \\
\frac{\partial\theta_{21}}{\partial\beta_1} &= -x_2\frac{\partial F_2(u,v)}{\partial u} + x_2\frac{\partial F_2(u,-v)}{\partial(u)} \\
\frac{\partial\theta_{12}}{\partial\beta_2} &= y_2 f(v) - y_1 f(-v) - y_2\frac{\partial F_2(u,v)}{\partial v} + y_1\frac{\partial F_2(u,-v)}{\partial(-v)}
\end{aligned}$$

$$\begin{aligned}
\frac{\partial\theta_{33}}{\partial\alpha_1} &= -f(u) - \frac{\partial F_2(u,v)}{\partial u} \\
\frac{\partial\theta_{33}}{\partial\alpha_2} &= -f(v) - \frac{\partial F_2(u,v)}{\partial v} \\
\frac{\partial\theta_{21}}{\partial\beta_1} &= -x_2 f(u) - x_2\frac{\partial F_2(u,v)}{\partial u} \\
\frac{\partial\theta_{12}}{\partial\beta_2} &= -y_2 f(v) - y_2\frac{\partial F_2(u,v)}{\partial v}
\end{aligned}$$

6.4.2 Some results

In tables 6.4 to 6.6 below, we report the D-optimal solutions in respect of the logistic distribution for the four cases $\{(-u,0,u);(-v,0,v)\}$, $\{(-u,u);(-u,u)\}$, $\{(-u,u);(-v,0,v)\}$ and $\{(-u,u);(-u,0,u)\}$ and the skewed logistic ($m = 2/3$) and complimentary log-log distribution for the case $\{(-u,v);(u,v)\}$. The results have the same patterns and characteristics as we found in two parameter case.

Table 6.4: D-optimal cutpoints for: logistic distribution; bivariate approach; four parameters.

ψ	Form$\{(-u,0,u);(-v,0,v)\}$		Form $\{(-u,u);(-u,u)\}$	
	(u^*,v^*)	$\Phi^*(u^*,v^*)$	u^*	$\Phi^*(u^*)$
0.001	(1.6200, 1.6200)	1.6509	1.3200	0.6860
0.010	(1.6200, 1.6200)	-0.2745	1.3200	-1.2635
0.100	(1.7800, 1.7800)	-1.8097	1.37000	-2.6450
0.200	(1.8700, 1.8700)	-2.1476	1.4000	-2.8887
0.300	(1.9200, 1.9200)	-2.2988	1.4300	-2.9895
0.400	(1.9400, 1.9400)	-2.3815	1.4400	-3.0423
0.500	(1.9600, 1.9600)	-2.4306	1.4500	-3.0730
0.600	(1.9700, 1.9700)	-2.4607	1.4600	-3.0915
0.700	(1.9700, 1.9700)	-2.4791	1.4600	-3.1027
0.800	(1.9800, 1.9800)	-2.4898	1.4700	-3.1092
0.900	(1.9800, 1.9800)	-2.4951	1.4700	-3.1124
0.990	(1.9800, 1.9800)	-2.4966	1.4700	-3.1133
0.999	(1.9800, 1.9800)	-2.4966	1.4700	-3.1133
2.000	(1.9600, 1.9600)	-2.4306	1.5100	-3.0730
3.000	(1.9300, 1.9300)	-2.3316	1.4300	-3.0106
4.000	(1.9000, 1.9000)	-2.2357	1.4200	-2.9480
5.000	(1.8700, 1.8700)	-2.1476	1.4000	-2.8887
10.00	(1.7800, 1.7800)	-1.8097	1.3700	-2.6450
20.00	(1.7100, 1.7100)	-1.3991	1.3400	-2.3145
50.00	(1.6400, 1.6400)	-0.7838	1.3300	-1.7588
100.0	(1.6200, 1.6200)	-0.2745	1.3200	-1.2635

Table 6.5: D-optimal cutpoints for: logistic distribution; bivariate approach; four parameters.

ψ	Form $\{(-u,u);(-v,0,v)\}$		Form $\{(-u,u);(-u,0,u)\}$	
	(u^*,v^*)	$\Phi^*(u^*,v^*)$	u^*	$\Phi^*(u^*)$
0.001	(1.4100, 1.4200)	0.9158	1.4100	0.9149
0.010	(1.3800, 1.4500)	-0.9557	1.4200	-0.9705
0.100	(1.3400, 1.7800)	-2.2614	1.5100	-2.3273
0.200	(1.3800, 1.8800)	-2.5296	1.6500	-2.5942
0.300	(1.4100, 1.9300)	-2.6495	1.6000	-2.7104
0.400	(1.4300, 1.9500)	-2.7148	1.6200	-2.7731
0.500	(1.4500, 1.9600)	-2.7534	1.6000	-2.8100
0.600	(1.4600, 1.9700)	-2.7770	1.6300	-2.8326
0.700	(1.4600, 1.9800)	-2.7913	1.6400	-2.8464
0.800	(1.4700, 1.9800)	-2.7996	1.6400	-2.8543
0.900	(1.4700, 1.9800)	-2.8038	1.6500	-2.8582
0.990	(1.4700, 1.9800)	-2.8050	1.6500	-2.8594
0.999	(1.4700, 1.9800)	-2.8050	1.6500	-2.8594
2.000	(1.4500, 1.9600)	-2.7534	1.6300	-2.8100
3.000	(1.4200, 1.9400)	-2.6754	1.6000	-2.7353
4.000	(1.4000, 1.9100)	-2.5995	1.5800	-2.6620
5.000	(1.3800, 1.8800)	-2.5296	1.5600	-2.5942
10.00	(1.3400, 1.7800)	-2.2614	1.5100	-2.3273
20.00	(1.3400, 1.7000)	-1.9357	1.4600	-1.9878
50.00	(1.3600, 1.5000)	-1.4168	1.4300	-1.4452
100.0	(1.3800, 1.4500)	-0.9557	1.4200	-0.9705

Table 6.6: D-optimal cutpoints for: bivariate approach, skewed logistic (m=2/3) and Complementary log-log distributions; form $\{(u,v);(u,v)\}$

ψ	Skewed logistic		Complementary log-log	
	(u^*,v^*)	$\Phi^*(u^*,v^*)$	(u^*,v^*)	$\Phi^*(u^*,v^*)$
0.001	(-2.1900, 0.8200)	0.0049	(-1.6700, 0.5700)	2.6020
0.010	(-2.1900, 0.8300)	-1.9444	(-1.6500, 0.5900)	0.6749
0.100	(-2.1300, 0.9500)	-3.3225	(-1.3500, 0.8400)	-0.5050
0.200	(-2.1100, 1.0300)	-3.5661	(-1.2500, 0.9200)	-0.6493
0.300	(-2.1200, 1.0700)	-3.6684	(-1.2600, 0.9400)	-0.7131
0.400	(-2.1400, 1.0900)	-3.7228	(-1.2700, 0.9500)	-0.7484
0.500	(-2.1500, 1.1000)	-3.7546	(-1.2800, 0.9500)	-0.7695
0.600	(-2.1600, 1.1000)	-3.7740	(-1.2800, 0.9600)	-0.7825
0.700	(-2.1700, 1.1100)	-3.7857	(-1.2900, 0.9600)	-0.7903
0.800	(-2.1700, 1.1100)	-3.7925	(-1.2900, 0.9600)	-0.7949
0.900	(-2.1700, 1.1100)	-3.7959	(-1.3000, 0.9600)	-0.7972
0.990	(-2.1700, 1.1100)	-3.7968	(-1.3000, 0.9600)	-0.7978
0.999	(-2.1700, 1.1100)	-3.7968	(-1.3000, 0.9600)	-0.7978
2.000	(-2.1500, 1.1000)	-3.7546	(-1.2700, 0.9500)	-0.7686
3.000	(-2.1300, 1.0800)	-3.6898	(-1.2600, 0.9500)	-0.7219
4.000	(-2.1100, 1.0600)	-3.6253	(-1.2400, 0.9400)	-0.6737
5.000	(-2.0900, 1.0500)	-3.5646	(-1.2300, 0.9300)	-0.6270
10.00	(-2.0500, 1.0100)	-3.3170	(-1.2000, 0.9100)	-0.4242
20.00	(-2.0200, 1.0000)	-2.9842	(-1.1900, 0.8800)	-0.1294
50.00	(-2.0200, 0.9600)	-2.4271	(-1.1900, 0.8400)	0.3930
100.0	(-2.0200, 0.9500)	-1.9316	(-1.1900, 0.8300)	0.8724

6.4.3 Special case: $\psi = 1$.

We again consider the special case when $\psi = 1$, i.e. the two variables X and Y are independent (symmetric distribution). We have $\theta_{ij} = \theta_{i.} \times \theta_{.j}$. $\theta_{i.}$ depends only on α_1 and β_1, $\theta_{.j}$ depends only on α_2 and β_2. Denote the information matrix 6.22 by:

$$I(\underline{\gamma}|u,v) = \begin{bmatrix} D_{11} & D_{12} \\ D_{21} & D_{22} \end{bmatrix} \tag{6.23}$$

We can transform the elements of $I(\underline{\gamma}|u,v)$ as follows:

$$\begin{aligned}
D_{11} &= \begin{bmatrix} \sum_{i=1}^{I}\sum_{j=1}^{J}\frac{1}{\theta_{ij}}\left(\frac{\partial\theta_{ij}}{\partial\alpha_1}\right)^2 & \sum_{i=1}^{I}\sum_{j=1}^{J}\frac{1}{\theta_{ij}}\frac{\partial\theta_{ij}}{\partial\alpha_1}\frac{\partial\theta_{ij}}{\partial\beta_1} \\ \sum_{i=1}^{I}\sum_{j=1}^{J}\frac{1}{\theta_{ij}}\frac{\partial\theta_{ij}}{\partial\alpha_1}\frac{\partial\theta_{ij}}{\partial\beta_1} & \sum_{i=1}^{I}\sum_{j=1}^{J}\frac{1}{\theta_{ij}}\left(\frac{\partial\theta_{ij}}{\partial\beta_1}\right)^2 \end{bmatrix} \\
&= \begin{bmatrix} \sum_{i=1}^{I}\sum_{j=1}^{J}\frac{1}{\theta_{i.}\theta_{.j}}\left(\frac{\partial\theta_{i.}}{\partial\alpha_1}\theta_{.j}\right)^2 & \sum_{i=1}^{I}\sum_{j=1}^{J}\frac{1}{\theta_{i.}\theta_{.j}}\left(\frac{\partial\theta_{i.}}{\partial\alpha_1}\theta_{.j}\right)\left(\frac{\partial\theta_{i.}}{\partial\beta_1}\theta_{.j}\right) \\ \sum_{i=1}^{I}\sum_{j=1}^{J}\frac{1}{\theta_{i.}\theta_{.j}}\left(\frac{\partial\theta_{i.}}{\partial\alpha_1}\theta_{.j}\right)\left(\frac{\partial\theta_{i.}}{\partial\beta_1}\theta_{.j}\right) & \sum_{i=1}^{I}\sum_{j=1}^{J}\frac{1}{\theta_{i.}\theta_{.j}}\left(\frac{\partial\theta_{i.}}{\partial\beta_1}\theta_{.j}\right)^2 \end{bmatrix} \\
&= \begin{bmatrix} \left(\sum_{i=1}^{I}\frac{1}{\theta_{i.}}\left(\frac{\partial\theta_{i.}}{\partial\alpha_1}\right)^2\right)\left(\sum_{j=1}^{J}\theta_{.j}\right) & \left(\sum_{i=1}^{I}\frac{1}{\theta_{i.}}\left(\frac{\partial\theta_{i.}}{\partial\alpha_1}\right)\left(\frac{\partial\theta_{i.}}{\partial\beta_1}\right)\right)\left(\sum_{j=1}^{J}\theta_{.j}\right) \\ \left(\sum_{i=1}^{I}\frac{1}{\theta_{i.}}\left(\frac{\partial\theta_{i.}}{\partial\alpha_1}\right)\left(\frac{\partial\theta_{i.}}{\partial\beta_1}\right)\right)\left(\sum_{j=1}^{J}\theta_{.j}\right) & \left(\sum_{i=1}^{I}\frac{1}{\theta_{i.}}\left(\frac{\partial\theta_{i.}}{\partial\beta_1}\right)^2\right)\left(\sum_{j=1}^{J}\theta_{.j}\right) \end{bmatrix}
\end{aligned}$$

Finally, since $\sum_{j=1}^{J} \theta_{.j} = 1$, we have:

$$D_{11} = \begin{bmatrix} \sum_{i=1}^{I} \frac{1}{\theta_{i.}} \left(\frac{\partial \theta_{i.}}{\partial \alpha_1}\right)^2 & \sum_{i=1}^{I} \frac{1}{\theta_{i.}} \left(\frac{\partial \theta_{i.}}{\partial \alpha_1}\right) \left(\frac{\partial \theta_{i.}}{\partial \beta_1}\right) \\ \sum_{i=1}^{I} \frac{1}{\theta_{i.}} \left(\frac{\partial \theta_{i.}}{\partial \alpha_1}\right) \left(\frac{\partial \theta_{i.}}{\partial \beta_1}\right) & \sum_{i=1}^{I} \frac{1}{\theta_{i.}} \left(\frac{\partial \theta_{i.}}{\partial \beta_1}\right)^2 \end{bmatrix}$$

$$= I_U$$

Similarly, we have:

$$D_{22} = \begin{bmatrix} \sum_{j=1}^{J} \frac{1}{\theta_{.j}} \left(\frac{\partial \theta_{.j}}{\partial \alpha_2}\right)^2 & \sum_{j=1}^{J} \frac{1}{\theta_{.j}} \left(\frac{\partial \theta_{.j}}{\partial \alpha_2}\right) \left(\frac{\partial \theta_{.j}}{\partial \beta_2}\right) \\ \sum_{j=1}^{J} \frac{1}{\theta_{.j}} \left(\frac{\partial \theta_{.j}}{\partial \alpha_2}\right) \left(\frac{\partial \theta_{.j}}{\partial \beta_2}\right) & \sum_{j=1}^{J} \frac{1}{\theta_{.j}} \left(\frac{\partial \theta_{.j}}{\partial \beta_2}\right)^2 \end{bmatrix}$$

$$= I_V$$

and:

$$D_{12} = D_{21} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$$

where I_U and I_V are the information matrices in the univariate case for U and V separately. So, in the four parameter model, where X and Y are independent, the Fisher information matrix turns out to be:

$$I(\underline{\gamma}|U, V) = \begin{bmatrix} I_U & 0 \\ 0 & I_V \end{bmatrix}$$

This is to be expected. So in the independent case, the optimal cutpoints will be similar to the ones we obtain in corresponding univariate cases when ψ close to 1.

Chapter 7

Conclusions and Future Work

7.1 Conclusions

This thesis has considered the determination of a set (or sets) of cutpoints (in a categorized survey question) in a number of cases (one point designs, multiple point designs, a bivariate approach) by using various methods of finding optimal solutions (search methods, graphical approach, multiplicative algorithms). We now summarize the whole work and draw conclusions.

We first introduced the design problem for linear and non-linear models including some basic concepts and their properties such as information matrices or design criteria.

Then, in chapter 2 we started from some applications to construct our formal problem. A generalized linear model for categorical responses was assumed with two parameters of interest μ and σ (or α and β). Depending on which parameter will be estimated (or both of them), the design objectives (design criteria) were defined.

In chapter 3 we considered the one design point problem with k categories. The formula of the information matrix in our particular case was constructed. We used search methods to find the optimal solutions (optimal cutpoints, corresponding c.d.f function and optimal criterion values) for $k = 3, 4, 5, 6$. Five standard optimal criteria (D-, A-, e_1, e_2 and E-optimality) and four symmetric distributions (logistic, normal, double exponential and double reciprocal) were considered. We saw that when the number of categories (or cutpoints) increases, the criterion values also increase but level off. In general, we concluded that a reasonable number of categories used in each case is 4 or 5. We also checked analytically the increase of criterion values when we insert one cutpoint between two consecutive cutpoints. We found that the difference between the new information matrix (after inserting cutpoint) and the old information matrix (before inserting cutpoint) is a non-negative definite matrix, which means that the values of standard criteria always increase when the number of categories increase in this way. At the end of the chapter, results for asymmetric distributions in the three category case were reported and contour plots were produced to illustrate results found by search methods.

To overcome the limitation of search methods (limited number of cutpoints) especially for asymmetric distributions, we introduced multiplicative algorithm in chapter 4 to find optimal cutpoints and optimal cell probabilities. We used this algorithm to verify the results found by search methods and extended it to find optimal solutions in the asymmetric distribution cases when the number of categories is bigger than two.

In chapter 5, we extended our considerations to the case of multiple design points. In this case, several sets of cutpoints available. Respondents will be divided into several groups and each group will be offered a common set of cutpoints. The optimal design problem turns out to be determining the sets of cutpoints and the respective weights optimally. We constructed this multiple point design problem and introduced the expected information matrix. We focused on considering two main cases. In the first case, we assumed equal weights and constraints on the cell probabilities and hence cutpoints. We used the concept of latin-squares to construct the design points when there were three and four cell probabilities. The constraints limit our consideration to one or two variable optimizing problems. So we can use search methods or a graphical approach to find optimal solutions.
In the second case, there was no constraints and arbitrary weights. We used the multiplicative algorithm to find optimal solutions in this case. The algorithm deals with both cell probabilities and design weights.
We also considered the case of unequal numbers of cutpoints across design points.
The general result in terms of criterion values was that it would be usually better to use one or two point designs since criterion values are higher at these designs than at designs with more support points.

Chapter 6 was devoted to the problem of the bivariate approach. This approach is motivated from the two stage process or the double bound model in contingent valuation studies. This generates an extension of our problem in which we wish to find a set or sets of cutpoints in each of two dimensions. The bivariate problem was set up with the use of the Plackett copula and the joint distribution function. The coefficient of association ψ represents the relationship between two variables. We focused on two main cases. In the first case, the two marginal distributions are identical in their parameters. We have two parameter problem. In the other case, a four parameter problem was considered when the two marginal distributions are different in their parameters. In both cases, we used search methods to find optimal solutions for a variety of situations. We considered the changes of optimal cutpoints and criterion values corresponding to the change in parameter ψ. We also investigated the information matrix in the independent case ($\psi = 1$) and compared findings with those of the univariate case. Results found are to be expected.

7.2 Future work

We now list some topics we will pursue in the future

7.2.1 Conditional approach.

Whether or not we assume a change in distribution between bids, one approach to design construction when bids are offered iteratively, particularly when there is a time gap between offers, is to consider designing for the second or next stage by changing the *cdf* $F(z)$ to that corresponding to the conditional distribution of X (or X_2) given the response to the first or previous bid. See Gunduz and Torsney (2002b). Design points could still be sets of cut-points.

7.2.2 Extensions to higher dimensions.

In the two-category case we often view one of the two categories as a "response" of interest. If this equates to "$X \leq x$" then:

$$P(\,response\,) = F(\alpha + \beta x)$$

Sitter and Torsney (1995a, 1995b) consider the extension of this model to two and to more than two design variables respectively, so that:

$$P(\,response\,) = F(\alpha + \underline{\beta}^T \underline{u}), \;\; \underline{u} \in \underline{\mathcal{U}} \subseteq \Re^m$$

Following Ford. Torsney and Wu (1992) they consider a linear transformation from $\underline{u}$ to $\underline{z}$ such that $z_1 = (\alpha + \underline{\beta}^T \underline{u})$ with the remaining z_j to be chosen by the experimenter. With the possible exception of z_1, $\underline{z}$ must be bounded. This will be the case if $\underline{U}$ is bounded. They argue that any design space $\underline{Z}$ must be equivalent to a subset of:

$$\underline{Z}_w = \{z_1 \in [A, B], \; -1 \leq z_j \leq 1, \; j = 2, 3, \ldots, m\},$$

where $[A, B]$ is the sample space of $F(z)$. This is a widest or 'largest' possible design space.

On $\underline{Z}_w$ observations need only be taken at $z_j = \pm 1$, $j = 2, 3, \ldots, m$, while z_1 plays the role of z in section 2.6 in chapter 2. In fact the total weight at a value of z_1 can be split uniformly across the 2^{m-1} combinations of $z_j = \pm 1$ or over subsets of these forming Hadamard matrices. Hence we can focus on the marginal design on z_1. Solutions have a similar structure to the one-design variable case. This includes the case of z_1 restricted to a subset $[a, b]$ of $[A, B]$. See Torsney and Gunduz (2000, 2002).

We could now consider the possibility of two or more cut-points for z_1. In this case the matrix Z in the definition of the information matrix I_Z changes. Suppose the cut-points are $\underline{z}_1 = (z_{11}, z_{12}, \ldots, z_{1(k-1)})^T$. Then in the case $m = 3$ there are four possible forms for Z, namely:

$$(\underline{1}_{k-1}|\underline{z}_1| - \underline{1}_{k-1}| - \underline{1}_{k-1})^T, \ (\underline{1}_{k-1}|\underline{z}_1| - \underline{1}_{k-1}| + \underline{1}_{k-1})^T,$$
$$(\underline{1}_{k-1}|\underline{z}_1| + \underline{1}_{k-1}| - \underline{1}_{k-1})^T, \ (\underline{1}_{k-1}|\underline{z}_1| + \underline{1}_{k-1}| + \underline{1}_{k-1})^T.$$

These correspond to the four possible combinations $(z_2, \ z_3) = (\pm 1, \pm 1)$. Suppose we use this set of cut-points for all design points (i.e. for all respondents regardless of their z_2 and z_3 status). Then this implies a one-point marginal design on $\underline{z}_1$. If we wish to choose $\underline{z}_1$ to optimize any of the standard criteria, then the information matrix I_Z must be replaced by it's sum over the above four Z-matrices.

For the logistic and normal/probit choices of $F(z)$ this optimization problem again reduces to an univariate optimization in the case $k = 3, 4$ as $\underline{z}_1$ must be of the form $(-z, z)^T$ or $(-z, 0, z)^T$ respectively.

7.2.3 Using bivariate normal distribution in the bivariate approach

In the bivariate approach considered in chapter 6, we used the Plackett copula as a joint distribution function between two variables U and V. We see that the optimal criterion values and optimal cutpoints are not very sensitive to the change in the coefficient of association ψ. Maybe this will not be the case with the bivariate normal. After standardization, the bivariate normal distribution between U and V has the following form:

$$p(u, v; \rho) = \frac{1}{2\pi\sqrt{1-\rho^2}} exp\left[- \frac{z}{2(1-\rho^2)} \right]$$

where

$$z = u^2 - 2\rho uv + v^2$$

ρ is coefficient of correlation, $-1 \le \rho \le 1$.

$$F_2(u, v; \rho) = \int_{-\infty}^{u} \int_{-\infty}^{v} p(x, y, \rho) d_x d_y$$

However, in order to calculate $F_2(u, v; \rho)$ and other derivatives, we have to use simulation techniques.

7.2.4 The multivariate approach

We can extend the bivariate approach to the multivariate approach in which we wish to find a set or sets of cutpoints in each of multiple dimensions. In this case, we have to use a multivariate distribution. Note that this problem will be more complicated in computation if we use the multivariate normal distribution as we need to calculate multiple integrals.

7.2.5 The use of multiplicative algorithms for bivariate and multivariate approach

We can extend the use of the multiplicative algorithm to find the optimal solutions in the case of bivariate approach. The problem arising as the relationship between θ_{ij}'s and (x_i, y_j) is not one to one. We should consider to assume some constraints between θ_{ij}'s. With this idea, we can apply the algorithm for multivariate approach.

Bibliography

[1] Abdelbasit, K. and Plackett, R. (1983). Experiment design for binary data. *Journal of the American Statistical Association*, 78(381):90-98.

[2] Agresti, A. (1996). *An Introduction to Categorical Data Analysis.* John Wiley & Sons, Inc.

[3] Alahmadi, A.M. (1993). *Algorithms for the Construction of Constrained and Unconstrained Optimal Designs.* PhD thesis, Department of Statistics, University of Glasgow.

[4] Alberini, A. (1995). Optimal designs for discrete choice contingent valuation surveys: single bound, double bound, and bivariate model. *Journal of Enviroimental Economics and Management*, 28:287-306.

[5] Alberini, A. (1995). Efficiency vs bias of willingness to pay estimates: Bivariate and interval-data models. *Journal of Enviroimental Economics and Management*, 29:169-180.

[6] Alberini, A., Carson, R.T. and Kanninen, B.J. (1997). Modelling response incentive effects in dichotomous choice contingent valuation data. *Land Economics and Management*, 73:309-324.

[7] Atkinson, A. and Donev, A. (1992). *Optimum experimemtal designs.* Oxford Statistical Science Series .8. Oxford Science Publications.

[8] Atkinson, A. C. (1996). The usefulness of optimum experimental designs. *J. Roy. Stat. Soc. Ser. B*, 58(1):59-76.

[9] Bishop, R. and Heberlein, T. (1979). Measuring values of extramarket goods: are indirect measures biased? *Amer.J.Agri.Econ.*, 61:927-930.

[10] Box, G. and Lucas, H. (1959). Design of experiments in non-linear situation *Biometrika*, 46:77-90.

[11] Chernoff, H. (1979). Sequential analysis and optimal design. Revised ed. Philadelphia: SIAM.

[12] Cameron, T. A. (1989). Sample design for Estimation Efficiency in Probit-bases Discrete Choice Contingent Valuation Models. Working Paper, Department of Economics, University of California, Los Angeles.

[13] Cameron, T.A. and Quiggin, J. (1994). Estimation using contingent valuation data from a 'dichotomous choice with follow-up' questionaire. *Journal of Enviroimental Economics and Management*, 27:218-234.

[14] Cooper C. J. (1993). Optimal bid selection for dichotomous choive contingent valuation surveys. *Journal of Environmental Economics and Management*, 24:25-40.

[15] Cooper, J. and Loomis, J. (1992). Sensitivity of willingness to pay estimates to bid design in dichotomous choice contigent valuation. *Land Economics*, 68:211-224.

[16] Donaldson, C. (1993). Theory and practice of willingness to pay for health care. HERU Discussion Paper 01/93, Dept. Pub. Health, Univ of Aberdeen.

[17] Elving, G. (1952). Optimal allocation in linear regression theory. *Ann. Math. Stat.*, 23:255-262.

[18] Fedorov, V. (1972). *Theory of Optimal Experiments - Originally Published in Russia under the title THEORIYA OPTIMAL'NOGO EKSPERIMENTA by izdatel'stvo Moskovskogo Universiteta, 1969.* Academic Press, first edition.

[19] Ford, I., Torsney, B., and Wu, C. (1992). The use of a canonical form in the construction of locally optimal designs for non-linear problems. *J. Roy. Stat. Soc. Ser. B*, 54(2):569-583.

[20] Genius, M. and Strazzera, E. (2005). Modelling Elicitation Effects in Contingent Valuation Studies: A Monter Carlo Analysis of the Bivariate Approach. Working Paper, Center of Research Economics North South, University of Cagliari, Uiversity of Sassari.

[21] Hanemann, W., J. Loomis, and B. Kanninen (1991). Statistical Efficiency of Double-Bounded Dichotomous Choice Contingent Valuation. *American Journal of Agricultural Economics*, 73:1255-1263.

[22] John, R. and Draper, N (1975). *D*-Optimality for regression designs: a review. *Technometrics*, 17:15-23.

[23] Kanninen, B. (1993). Optimal experimental design for double-bounded dichotomous choice contingent valuation. *Land Economics*, 69(2):138-146.

[24] Karlin, S. and Studden, W. (1966). Optimal experimental designs. *Ann. Math. Stat.*, 37:783-815.

[25] Kiefer, J. (1959). Optimum experimental designs. *J. Roy. Stat. Soc. Ser. B*, 21:272-319.

[26] Kiefer, J. and Wolfowitz. J (1959). Optimum designs in regression problems. *Ann. Math. Stat.*, 30(2):271-294.

[27] Kiefer, J. (1961). Optimum designs in regression problems, ii. *Ann. Math. Stat.*, 32(1):298-325.

[28] Kiefer, J. (1952). Two more criteria equivalent to D-optimality of designs *Ann. Math. Stat.*, 33:793-796.

[29] Kling, C. L. (1997). The gains from combining travel cost and contingent valuation data to value non-market goods. *Land Economics*, 73:428-4439.

[30] Mandal, S. (2000). Construction of optimising dsitributions with applications in estimation and optimal design. *PhD thesis, Department of Statistics, University of Glasgow.*

[31] Maldal, S. and Torsney, B. (2000). Algorithm for construction of optimazing distributions. *Communications in Statistics-Theory and Methods*, 29:1219-1231.

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

[33] Minkin, S. (1987). Optimal designs for binary data. *J. Amer. Statist. Assoc.*, 82:1098-1103.

[34] Musrati, A. (1992). Locally optimal designs for binary and weighted regression models. *Master's thesis, Department of Statistics, University of Glasgow.*

[35] Nelsen, R. B. (1999). *An Introdution to copulas.* Lecture Notes in Statistics. New York: Springer-Verlag.

[36] Nyquist, H. (1992). Optimal designs of discrete response experiments in contingent valuation studies. *The Review of Economics and Statistics*, 74:559-563.

[37] Pukelsheim, F. (1993). *Optimal Design of Experiment.* John Wiley, New York.

[38] Pukelsheim, F. and Torsney, B. (1991). Optimal weights for experimental designs on linearly independent support points *The Annals of Statistics.*, 19(3):1614-1625.

[39] Ryan, M. and Ratcliffe, J (2000). Some issues in the application of closed-ended willingness to pay studies to valuing health googs: an application to antenatal care in Scotland. *Applied Economics*, 32(5):643-651.

[40] Sebastiani, P. and Settimi, R. (1997). A note on D-optimal designs for a logistic regression model. *Journal of Statistical Planning and Inference*, 59:359-368.

[41] Silvey, S. (1980). *Optimal Design. An Introduction to the Theory for parameter Estimation.* Chapman and Hall.

[42] Silvey, S. D., Titterington, D. M., and Torsney, B. (1978). An algorithm for optimal designs on a finite design space. *Communications in Statistics A*, 7:1379-1389.

[43] Sitter, R. and Fainaru, I. (1997). Optimal designs for the logit and probit models for binary data. *Canadian Journal of Statistics*, 25:175-189.

[44] Sitter, R. and Torsney, B. (1995a). *D*-optimal designs for generalized linear models. In Kitsos, C. and Muller, W., editors, *MODA4-Advances in Model Oriented Data Analysis*, page 87-102. Phisica-Verlag. Refereed Proceedings of the 4th international workshop in Spetses, Greece, June 5-9, 1995.

[45] Sitter, R. and Torsney, B. (1995b). Optimal designs for binary response experiments with two design variables. *Statistica Sinica*, 5:405-419.

[46] Smith, R. Olsen, J. A and Harris, A. (1999). A Review of Methodological Issues in the Conduct of WTP Studies in Health Care II: Administration of a CV Survey. Working Paper, Health Economics Unit, Center for Health Program Evaluation, Monash University.

[47] Torsney, B. (1981). Algorithms for a constrained optimization problem wit application in statistics and optimum design. *PhD thesis, Department of Statistics, Univesity of Glasgow.*

[48] Torsney, B. and Alahmadi, A. M. (1992). Further developmentof algorithms for constructing optimizing distributions. In Fedorov, V., Muller,

W., and Vuchkov, I., editors, *Proc. 2nd IIASA Model Oriented Data Workshop in St. Kyrik, Bulgaria*, Pages 121-129, Physica-Verlag GmbH Co. P.O. Box 10 52 80 69042 Heildelberg, FRG. Physica-Verlag.

[49] Torsney, B. and Alahmadi, A. M. (1995). Designing for minimally dependent observations. *Statica Sinica*, 5: 499-514.

[50] Torsney, B. and Musrati, A. (1993). On the construction of optimal designs with applications to binary response and to weighted regression models. In Muller, W., Wynn, H., and Zhigljavsky, A., editors, *Proc. 3rd Model Oriented Data Workshop in Petrodvorets, Russia*, Volume 3, Pages 37-52, Physica-Verlag GmbH Co. P.O. Box 10 52 80 69042 Heildelberg, FRG. Physica-Verlag.

[51] Torsney, B. and Gunduz, N. (1999). A brief review of optimal designs in contingent valuation studies. *Tatra Mountains Mathematics Publications*, 17:185-195

[52] White, L. V (1973). An extension of the general equivalence theorem to non-linear models. *Biometrika.*, 60(2):345-348.

[53] Whittle, P. (1973). Some general points in the theory of optimal experimental design. *J. Roy. Stat. Soc. Ser. B*, 35:123-130.

[54] Wu. C. F. J (1988). Optimal design for percentile estimation of a quantal response curve. In Y. Dodge, V. Fedorov and H. Wynn editor, *Optimal Design and Analysis of Experiments*, pages 213-222. Elsevier Science Publishers B.V. (North Holland).

[55] Wu, C. F. J. and Hamada, M. (2000). *Experiment: Planning, Analysis, and Parameter Design Optimization.* John Wiley & Sons, New York.

[56] Wynn, H. (1972). Results in the theory and construction of D-optimum experimental designs. *J. Roy. Stat. Soc. Ser. B*, 34:133-147,170-186. (with discussion).