The Annals of Statistics
2001, Vol. 29, No. 4, 1094–1116

AN ALGORITHM FOR CALCULATING Γ-MINIMAX DECISION RULES UNDER GENERALIZED MOMENT CONDITIONS[1]

BY ROGER FANDOM NOUBIAP AND WILFRIED SEIDEL

Universität der Bundeswehr Hamburg

We present an algorithm for calculating a Γ-minimax decision rule, when Γ is given by a finite number of generalized moment conditions. Such a decision rule minimizes the maximum of the integrals of the risk function with respect to all distributions in Γ. The inner maximization problem is approximated by a sequence of linear programs. This approximation is combined with an elimination technique which quickly reduces the domain of the variables of the outer minimization problem. To test for convergence in a final step, the inner maximization problem has to be completely solved once for the candidate of the Γ-minimax rule found by the algorithm. For an infinite, compact parameter space, this is done by semi-infinite programming. The algorithm is applied to calculate robustified Bayesian designs in a logistic regression model and Γ-minimax tests in monotone decision problems.

1. Introduction. Let us consider a class of statistical decision rules for a parameter θ which varies in a σ-compact subset Θ of a Euclidean parameter space. Assume that each decision rule can be represented by a pair (k, y). Here, k is a discrete variable, for example a vector of sample sizes or a vector of numbers of dose levels in experimental design, and y is an additional variable (discrete or continuous) that characterizes the strategy (a vector of critical values, say). Given a loss function, the *risk function* $R(k, y|\theta)$ of the decision rule (k, y), given θ, can be obtained in the usual way; it describes the average loss associated with (k, y), if θ is the true value of the parameter. Let Γ be a class of probability measures on Θ. For each distribution $\pi \in \Gamma$, the *Bayes risk* of (k, y) with respect to π is $r(k, y|\pi) = \int_\Theta R(k, y|\theta)\, d\pi(\theta)$. The Γ-*Bayes risk* of (k, y) is

$$r_{\sup}(k, y|\Gamma) = \sup_{\pi\in\Gamma} r(k, y|\pi)\,, \tag{1}$$

where the supremum over an empty set is defined as $-\infty$. A Γ-*minimax decision rule* (k_Γ, y_Γ) minimizes $r_{\sup}(k, y|\Gamma)$ [Berger (1985)]. For a function $g : \Theta \to \mathbb{R}$,

$$m_g(\pi) = \int_\Theta g(\theta)\, d\pi(\theta) \tag{2}$$

Received July 1998; revised March 2001.

[1]Supported in part by a grant from the Deutsche Forschungsgemeinschaft.

AMS 2000 *subject classifications*. Primary 62C12; secondary 62F35, 90C34.

Key words and phrases. Bayesian robustness, experimental design, Γ-minimax decision rules, minimax problems, monotone decision problems, semi-infinite programming.

is the *generalized moment of g with respect to* π. We shall assume that Γ is given by *r generalized moment conditions*, that is, it is of the form

$$\Gamma = \{\pi\text{: } g_i \text{ is } \pi\text{-integrable and } m_{g_i}(\pi) \in I_i,\ 1 \le i \le r\} \tag{3}$$

for certain functions g_i and closed intervals I_i, $1 \le i \le r$. The aim of the paper is to develop an algorithm for calculating Γ-minimax strategies under generalized moment conditions.

The Γ-Bayes risk of a strategy (k, y) and Γ-minimax strategies are of interest, for example, in robustness studies in Bayesian statistics. Given a prior π, a Bayes strategy minimizes $r(k, y|\pi)$. It may depend strongly on the prior, and often there is not enough knowledge for a precise specification of π. According to Berger (1990), "prior elicitation typically involves the actual specification of only a finite number of features of the prior," and in particular, conditions on a finite number of generalized moments of the prior distribution can be used as a tool to study robustness properties of Bayesian decision rules [Berger (1990); Betrò, Ruggeri and Męczarski (1994)]. A Γ-minimax rule is usually a Bayes rule that is compatible with the specified prior information, and it is usually robust over Γ [Berger (1985)].

A lot of work has been done on the development of theory and algorithms for optimization under generalized moment constraints. A classical text is the paper by Kemperman (1987). In the context of Bayesian robustness and decision analysis, algorithms have been proposed by Dall'Aglio and Salinetti (1994), Smith (1995) and Betrò and Guglielmi (1996). In Betrò, Ruggeri and Męczarski (1994), results of Winkler (1988) have been used to calculate extrema of certain posterior functionals. However, we will employ methods for inner maximization which are more similar to those in Betrò and Guglielmi (1996).

Now calculating (k_Γ, y_Γ) is a minimax problem, where inner maximization of $r(k, y|\pi)$ with respect to $\pi \in \Gamma$ has to be combined with outer minimization of $r_{\sup}(k, y|\Gamma)$ with respect to (k, y). A good algorithm for inner maximization is necessary, but in general not sufficient for an efficient solution of such a minimax problem. The reason is that for minimization with respect to k and y, $r_{\sup}(k, y|\Gamma)$ has to be computed so often that it may be impossible or at least inefficient to search for a good approximation to it in every instance. On the other hand, if one stops inner maximization too early, one might minimize a completely wrong goal function in the outer minimization problem. In this case the solution proposed by the algorithm may be totally different from (k_Γ, y_Γ).

The focus of the paper is on the development of a safeguarded strategy for the minimax problem that keeps control of the interaction between inner maximization and outer minimization. It is based on an increasing sequence of lower bounds for $r_{\sup}(k, y|\Gamma)$ together with upper bounds for $r_{\sup}(k_\Gamma, y_\Gamma|\Gamma)$. For inner maximization, in principle any algorithm can be used that results in such a sequence of functions, converging from below to $r_{\sup}(k, y|\Gamma)$ and with the property that each element can be minimized with respect to (k, y).

We found that in many situations an approach based on linear programming is suitable:

If Θ is finite, $r_{\sup}(k, y|\Gamma)$ is the solution of a linear program with variables $\pi(\{\theta\})$, $\theta \in \Theta$. Even in this case, however, it would be inefficient to calculate the exact value of $r_{\sup}(k, y|\Gamma)$ for each (k, y) needed for the minimization algorithm, as the number of variables in the linear program might be too large. If Θ is large or infinite, we shall approximate it by an increasing sequence $(\Theta_l)_{l\in\mathbb{N}}$ of finite subsets such that $\cup\{\Theta_l: l \in \mathbb{N}\}$ is dense in Θ. Let

$$\Gamma_l = \{\pi \in \Gamma: \pi \text{ is concentrated on } \Theta_l\}. \tag{4}$$

Then $r_{\sup}(k, y|\Gamma_l)$ can be calculated by linear programming in a fast and reliable way. Linear programming also has the advantage that it can be easily implemented and that large sets of moment conditions can be incorporated. For each (k, y), the sequence $(r_{\sup}(k, y|\Gamma_l))_{l\in\mathbb{N}}$ is nondecreasing, and in each approximation step l, $r_{\sup}(k, y|\Gamma_l)$ is based on a discretization of the parameter space which is independent of (k, y). At stage l, let $(k^l, y^l) = (k_{\Gamma_l}, y_{\Gamma_l})$ be a decision rule that minimizes $r_{\sup}(k, y|\Gamma_l)$.

A key feature of our algorithm is an elimination strategy that, in the first approximation steps, drastically reduces the size of the search domain for k in the outer minimization problem. It makes use of the fact that $r_{\sup}(k, y|\Gamma_l)$ is increasing in l. If an exact solution of the inner maximization problem is possible, it may be used for a final test of convergence. For infinite, compact Θ, this maximization problem is solved by semi-infinite programming. For introductions into this rapidly growing area see, for example, Hettich and Kortanek (1993), Goberna and López (1998) or Reemtsen and Rückmann (1998).

We will give sufficient regularity conditions for the following convergence properties, as $l \to \infty$:

C1. $r_{\sup}(k, y|\Gamma_l)$ converges to $r_{\sup}(k, y|\Gamma)$ for each (k, y).

C2. (k^l, y^l) converges (in some sense) to (k_Γ, y_Γ) and $r_{\sup}(k^l, y^l|\Gamma_l)$ converges from below to $r_{\sup}(k_\Gamma, y_\Gamma|\Gamma)$.

Now consider the following regularity conditions:

R1. k varies in a finite set K. For each $k \in K$, y is an element of a compact subset $\mathcal{Y}_k$ of $(\mathbb{R} \cup \{\pm\infty\})^\nu$ for some ν and $R(k, y|\theta)$ is continuous on $\mathcal{Y}_k \times \Theta$. Moreover, g_i is continuous for $1 \le i \le r$.

Often $\mathcal{Y}_k$ is a real interval or rectangle. In the simplest case, k is an integer and y a real number. However, in "sequentially planned decision procedures," for example, k and y have a more complicated structure [see Schmitz (1992)]. If k is related to a sample size, it is in general possible to restrict its domain to a finite set: for example, there may be a threshold where sampling costs exceed the losses caused by wrong decisions (see Section 3).

If R1 is satisfied, C2 follows from C1. Property C1 is trivial, if Θ is finite. In general, it seems at first glance that C1 follows from R1, because each measure $\pi \in \Gamma$ can be approximated arbitrarily well with respect to the weak topology by measures π_l on Θ_l, and since $R(k, y|\theta)$ is continuous on Θ, $r(k, y|\pi_l)$ converges to $r(k, y|\pi)$. However, π_l need not satisfy the moment conditions. The following simple example shows that for $\Gamma \neq \varnothing$ it is even possible that all sets Γ_l are empty.

EXAMPLE 1.1. Let $\Theta = [0, 1]$, $\theta_0 \in \Theta$, $r = 1$, $g_1(\theta) = |\theta - \theta_0|$ and $I_1 = \{0\}$. If θ_0 is not an element of $\cup\Theta_l$, then $\Gamma_l = \varnothing$ for each l, whereas Γ contains exactly one element, namely the unit measure that puts mass 1 to the point θ_0. Consequently, any attempt to approximate $r_{\sup}(k, y|\Gamma)$ for some (k, y) by $r_{\sup}(k, y|\Gamma_l)$ must fail.

Sufficient conditions for C1 are given in Sections 5 and 6. It may be hard to verify these in practical applications, but at least for compact Θ, this is not really necessary: the algorithm provides a lower and an upper bound for the function value in the minimax point, and if both coincide, we know that the desired solution has been found. For compact Θ and $\Gamma \neq \varnothing$, convergence has been observed in all examples considered so far.

As an application of our algorithm, we calculate robustified Bayesian designs in a nonlinear regression model, where the space of designs considered is restricted in a suitable way. The model is introduced in Section 2; computational details and results are presented in Section 8. A general class of decision problems for which the regularity conditions are usually satisfied and where our approach works very efficiently is the class of so-called monotone decision problems. These are outlined together with applications to quality control in Section 3; numerical examples are given in Section 9.

2. Robustified Bayesian designs for binary data. Consider a binary response experiment, where for $i = 1, \ldots, s$, n_i subjects are administered a stimulus at dose level x_i. Let $n = \sum_{i=1}^{s} n_i$ be the total sample size. We assume that each experiment results in a binary outcome with success probability $p_i = F(z_i) = 1/(1 + e^{-z_i})$, where $z_i = \beta(x_i - \mu)$ is the standardized dose.

Within this parameterization β represents the slope of the dose–response curve and μ is the median effective dose, defined as the concentration that is expected to produce a response in 50% of the experimental units. The parameter space is $\Theta = \{\theta = (\mu, \beta) |\ \beta > 0\}$.

Any exact design may be represented as a probability measure ξ, where $\xi(x_i) = \lambda_i = n_i/n$ denotes the relative proportion of total observations at x_i. Let Ξ denote the space of designs considered. A design is chosen to minimize some real-valued function R of the asymptotic variance–covariance matrix of the maximum likelihood estimates of the model parameters, given as the inverse of the information matrix

$$I(\xi, \theta) = n \begin{pmatrix} \beta^2 S_0 & -S_1 \\ -S_1 & \dfrac{1}{\beta^2} S_2 \end{pmatrix},$$

with $S_0 = \sum_{i=1}^{s} \lambda_i w(z_i)$, $S_1 = \sum_{i=1}^{s} \lambda_i w(z_i) z_i$, $S_2 = \sum_{i=1}^{s} \lambda_i w(z_i) z_i^2$ and $w(z_i) = (F'(z_i))^2 / F(z_i)(1-F(z_i)) = e^{-z_i}/(1+e^{-z_i})^2$, where $F'(z)$ denotes the derivative of F at the point z.

Design criteria R of main interest are the D-criterion $R(\xi|\theta) = -\det I(\xi, \theta) = -n^2[S_2 S_0 - S_1^2]$, minimizing the generalized variance of the parameter estimates, and the C-criterion $R(\xi|\theta) = c' I(\xi, \theta)^{-1} c$ with $c \in \mathbb{R}^2$, minimizing the variance of a linear combination of the parameter estimates. Unfortunately, in nonlinear models the information matrix usually depends on the unknown parameters. Therefore many concepts for the construction of optimal designs use prior information about these parameters, for example, by considering a prior distribution (Bayesian designs) or a "good guess" θ_0 (locally optimal designs). However, analyses into the robustness of such designs show that their efficiency can be quite poor, if the prior information is misspecified.

In the model considered here, Sitter (1992) starts from an initial guess $\theta_0 = (\mu_0, \beta_0)$ for the parameter $\theta = (\mu, \beta)$ and constructs robustified versions of locally optimal designs by introducing a rectangular region around θ_0 that reflects the experimenter's uncertainty in the prior information about θ and by applying a minimax criterion. He observes, however, that the corresponding optimization problem is mathematically intractable and numerically too difficult, if the set Ξ of designs considered is too rich (e.g., the set of all designs symmetrical about μ_0). Therefore he restricts Ξ to the set of balanced designs that allocate an equal number of observations to each of s points symmetrically placed about μ_0, with constant distance between adjacent points. He calculates the designs by nesting Nelder–Mead simplex algorithms [see Nelder and Mead (1965)] and it is not clear if there are mechanisms to protect against problems resulting from multimodality.

We will restrict Ξ in several steps. The most general set is that of all designs symmetrical about μ_0, with fixed total sample size n. We will start from points $u_1 \le u_2 \le \cdots \le u_s$, placed symmetrically about zero, and put $x_i = \mu_0 + u_i/\beta_0$. Then such a design can be represented by two vectors k and y: $k = (n_1, \ldots, n_s)$ is the vector of sample sizes such that $n_i = n_{s+1-i}$, $i = 1 \ldots s$ and $n_1 + \cdots + n_s = n$. Let $d_i = x_{i+1} - x_i$. Then $d_i = d_{s-i}$, and there are $r = s/2$ different values if s is even, and $r = (s-1)/2$, if s is odd. Therefore the design points are uniquely represented by the vector $y = (d_1, \ldots, d_r)$ with $d_i > 0$.

In a second step, we will assume constant dose spacings $d_1 = \cdots = d_r = d$. Then y simplifies to a one-dimensional variable $y = d$. Finally, we will assume that in addition, $n_1 = \cdots = n_s$. In this balanced case, we have $\lambda_i = 1/s$ for all i and k reduces to the one-dimensional integer $k = s$. Here we may set $n = 1$ w.l.o.g. The latter are the designs considered by Sitter (1992).

Using the above representation, the standardized doses can be simplified to yield $z_i = \beta(x_i - \mu) = b(u_i - m)$ with $m = \beta_0(\mu - \mu_0)$, $b = \beta/\beta_0$. Since the transformed parameter space is only dependent on m and b, the design problem can be solved for the special case $\mu_0 = 0$ and $\beta_0 = 1$ without loss of generality. From these results the minimax design for any other value of θ_0 can be obtained.

Sitter (1992) models the experimenter's uncertainty in the prior information about θ by a rectangular region around θ_0. As a common generalization of robustified locally optimal and of Bayesian designs, we will model uncertainty by a set Γ of distributions π on Θ with expectation θ_0 and which meet certain constraints on their variance and covariance. It will be described by moment conditions in the following way: for $i = \mu, \beta$, let $g_{i,1}(\mu, \beta) = i$, $m_1(i) = m_{g_{i,1}}(\pi)$, $g_{i,2}(\mu, \beta) = i^2$, $m_2(i) = m_{g_{i,2}}(\pi)$, $g_{\mu\beta}(\mu, \beta) = \mu\beta$ and $m_2(\mu, \beta) = m_{g_{\mu\beta}}(\pi)$. Then $m_1(i)$ is the expectation and $m_2(i) - (m_1(i))^2$ the variance of i; $m_2(\mu, \beta) - m_1(\mu)m_1(\beta)$ the covariance of μ and β. Let $I_1(i)$, $I_2(i)$ and $I_2(\mu, \beta)$ denote the closed intervals representing information about the moments $m_1(i)$, $m_2(i)$ and $m_2(\mu, \beta)$. We shall assume that the expectations are equal to the initial guesses, that is, $I_1(\mu) = \{0\}$ and $I_1(\beta) = \{1\}$. Then $m_2(\mu)$ and $m_2(\beta) - 1$ are the variances of μ and β, respectively, and we will consider different bounds on these variances. Moreover, $m_2(\mu, \beta)$ is here the covariance of μ and β. Usually there will be no prior information about covariances; however, for fixed variances of μ and β we may study the effects of low or high correlation by different choices of $I_2(\mu, \beta)$.

3. Monotone testing problems and sampling inspection. Let $\Theta \subset \mathbb{R}$ and let $l_a(\theta)$ and $l_r(\theta)$ be bounded loss functions associated with the two possible final decisions of a test, labelled **a** ("accept") and **r** ("reject" a certain hypothesis). For fixed sample size k, the decision is based on a real-valued test statistic X, usually a sufficient statistic. In a *monotone testing problem* the distribution of X is assumed to have a monotone likelihood ratio for each k, and the loss functions reflect a one-sided test for θ in the sense of (8.9) in Berger [(1985), page 529]. We will in addition assume that *sampling costs* $s(k)$ are given, considered as nondecreasing and unbounded in k.

For example, in sampling inspection, **a** (**r**) means accepting (rejecting) a certain batch. Here the parameter θ is closely related to quality; in fact, quality is measured by a "fraction nonconforming" $p(\theta) \in [0, 1]$ in the batch or in the production process, it is assumed to be a function of θ. A variety of loss functions can be transformed into the so-called Stange cost model: assume that (relative) sampling costs per item $q > 0$ and a break even quality level p_0 are given. Then the sampling costs are $s(k) = k \times q$, and the loss functions are $l_a(\theta) = \max\{0, p(\theta) - p_0\}$ and $l_r(\theta) = \max\{0, p_0 - p(\theta)\}$. Consequently, a fraction nonconforming less than p_0 should be accepted, otherwise rejected. Together with a Stange cost model, the distributions considered below constitute monotone testing problems; Example 3.2 shows that sometimes models with more than one parameter can be reparametrized in terms of a monotone testing problem.

Theorem 6 in Chapter 8 of Berger (1985) implies that in a monotone testing problem, the class of *monotone decision rules* is essentially complete. These are of the form "accept (reject) if $X < y$ ($X > y$), with possible randomization if $X = y$" for some $y \in \mathbb{R}$.

Randomization is only necessary in discrete models. Here y is discrete; however, interpolating its range using the randomization probability results in a continuous variable. Therefore in a monotone testing problem, optimal strategies can be characterized by a pair $d = (k, y)$, where y is one-dimensional. In quality control, (k, y) is also called a *sampling plan*. An important property for optimization is that $r_{\sup}(k, y|\Gamma)$ is unimodal in y: it can be proved exactly in the same way as in Thyregod [(1974), Theorem 2] that for nonrandomized decision rules, the Bayes risk $r(k, y|\pi)$ is a quasiconvex function of y for each distribution π and each sample size k [a real-valued function f is called *quasiconvex*, if $f(y) \leq \max\{f(x), f(z)\}$ holds for all $x < y < z$ in the domain of f]. It is then easy to see that this property also holds for a suitable representation of randomization, resulting in piecewise linearization. As the supremum of each family of quasiconvex functions is quasiconvex, it also holds that $r_{\sup}(k, y|\Gamma)$ is a quasiconvex function of y for each sample size k and each set Γ of probability measures on Θ.

The Γ-minimax principle has been proposed in sampling inspection by Krumbholz (1982); he considers prior information essentially of type $P(p(\theta) \leq a) \geq \gamma$. With our approach, Γ-minimax sampling plans (k_Γ, y_Γ) can be calculated for arbitrary generalized moments of the prior distribution. In particular, we considered the following types of models.

EXAMPLE 3.1 (Sampling by attributes). This means that the decision is based on the number X of nonconforming items in a sample of size k. In the *hypergeometric model*, θ is the number and $p(\theta)$ the fraction of nonconforming items in a batch of size N. Then $\Theta = \{0, 1, \ldots, N\}$ is discrete and conditionally on θ, the observation has a hypergeometric distribution. In the *binomial model*, $\Theta = [0, 1]$ and $p(\theta) = \theta$, usually considered as the fraction of nonconforming items in the production process. Conditionally on θ, X has a binomial distribution with parameters k and θ.

EXAMPLE 3.2 (Sampling by variables, upper specification limit). Here, an item is considered as nonconforming if a continuous quality characteristic ξ exceeds an upper specification limit U. It is assumed that ξ is normally distributed with expectation μ and variance σ^2, which are parameters of the production process. Then the fraction nonconforming is given as $p(\mu, \sigma^2) = P(\xi > U) = \Phi((\mu - U)/\sigma)$, where Φ denotes the cumulative distribution function of the standard normal distribution. In a sample of size k, let $\overline{\xi_k}$ be the sample mean and S_k the sample standard deviation. If σ^2 is unknown, the usual test statistic is $X = \sqrt{k}(\overline{\xi_k} - U)/S_k$; it has a noncentral t distribution with $k - 1$ degrees of freedom and noncentrality parameter $\lambda = \sqrt{k}(\mu - U)/\sigma$. Thus the fraction nonconforming and the distribution of X depend on μ and σ only in terms of $\theta = (\mu - U)/\sigma$. Here it is not advantageous to restrict to a compact parameter space from the beginning, as small values of σ (which represent a favorable production process) tend to inflate θ.

In monotone testing problems with continuous loss functions, it is often possible to assume that for each k, y varies in a compact set $\mathscr{Y}_k$ [in Example 3.2 $\mathscr{Y}_k = [-\infty, \infty]$, with the obvious interpretation of $(k, -\infty)$ and (k, ∞), namely rejecting and accepting regardless of the outcome of the sample] such that $R(k, y|\theta)$ is continuous on $\mathscr{Y}_k \times \Theta$.

Moreover, k can be restricted to a finite set K: let $l_{\min}(\theta) = \min\{l_a(\theta), l_r(\theta)\}$ and $r(l_{\min}|\Gamma) = \sup_{\pi\in\Gamma} \int_\Theta l_{\min}(\theta)\, d\pi(\theta)$. For an arbitrary decision rule (k, y), let $k_{\sup}(k, y) = \min\{\tilde{k}\colon s(\tilde{k}) + r(l_{\min}|\Gamma) \geq r_{\sup}(k, y|\Gamma)\}$. Then it is easy to see that $k_\Gamma \leq k_{\sup}(k, y)$. In discretization step l, $k^l_{\sup}(k, y)$ may be defined in the same way; it is an upper bound for k^l.

4. Outer minimization for compact Θ. The next sections deal with the solution of the general minimax problem stated in the introduction under the assumption that the regularity conditions R1 in the introduction are satisfied. We will first consider only compact parameter spaces (Sections 4 and 5).

THEOREM 4.1. *For each k, the function $r_{\sup}(k, y|\Gamma)$ and the functions $r_{\sup}(k, y|\Gamma_l)$ are continuous in y.*

PROOF. The assertion follows because for each k, the function $R(k, y|\theta)$ is uniformly continuous on the (compact) set $\mathscr{Y}_k \times \Theta$.

The results of this section will be proved under the assumption that for each (k, y), the sequence $(r_{\sup}(k, y|\Gamma_l))_{l\in\mathbb{N}}$ is nondecreasing and converges from below to $r_{\sup}(k, y|\Gamma)$ (sufficient conditions for this will be given in Section 5).

Since K is finite and for each k the function $r_{\sup}(k, \cdot|\Gamma_l)$ is continuous on the compact set $\mathscr{Y}_k$, existence of a (global) minimizer (k^l, y^l) of $r_{\sup}(k, y|\Gamma_l)$ is guaranteed, although it need not be uniquely determined (of course, the corresponding function value is uniquely determined). Moreover, each sequence $\left(k^l, y^l\right)_{l=1,2,\ldots}$ has an accumulation point (k^*, y^*), which is a solution of the minimax problem, as the following theorem shows. If (k^*, y^*) is not unique, it is irrelevant which value is chosen.

THEOREM 4.2. *Each accumulation point (k^*, y^*) is a minimal point of $r_{\sup}(k, y|\Gamma)$, and $r_{\sup}(k^l, y^l|\Gamma_l)$ converges from below to $\min_{k\in K, y\in\mathscr{Y}_k} r_{\sup}(k, y|\Gamma)$.*

PROOF. Let $R^{\text{opt}} = \min_{k\in K, y\in\mathscr{Y}_k} r_{\sup}(k, y|\Gamma)$ and $R^{\text{opt}}_l = r_{\sup}(k^l, y^l|\Gamma_l)$. As K is finite and $\mathscr{Y}_k$ is compact for each k, the sequence $\left(r_{\sup}(k, y|\Gamma_l)\right)_l$ converges uniformly from below to $r_{\sup}(k, y|\Gamma)$. From this, monotone convergence of R^{opt}_l to R^{opt} may be easily deduced. To show that (k^*, y^*) is a minimizer of $r_{\sup}(k, y|\Gamma)$, we may assume w.l.o.g. that the sequence $\left(k^l, y^l\right)_l$ converges to (k^*, y^*). Obviously, $r_{\sup}(k^l, y^l|\Gamma)$ converges to $r_{\sup}(k^*, y^*|\Gamma)$. Let $\varepsilon_l = \|r_{\sup}(\cdot, \cdot|\Gamma) - r_{\sup}(\cdot, \cdot|\Gamma_l)\|_\infty$. Then $\varepsilon_l \to 0$ and $r_{\sup}(k^l, y^l|\Gamma) \leq R^{\text{opt}}_l + \varepsilon_l \leq R^{\text{opt}} + \varepsilon_l$; therefore $r_{\sup}(k^l, y^l|\Gamma)$ converges to R^{opt}. This proves $r_{\sup}(k^*, y^*|\Gamma) = R^{\text{opt}}$.

4.1. *Successive elimination.* The approximation technique presented above yields an efficient minimax algorithm, if it is combined with a strategy for excluding values of k during the computation of (k^l, y^l), using information on k^{l-1}. Let

$$M_l(k) = \min_{y \in \mathscr{Y}_k} r_{\sup}(k, y|\Gamma_l).$$

Existence of $M_l(k)$ follows because $r_{\sup}(k, \cdot|\Gamma_l)$ is continuous on the compact set $\mathscr{Y}_k$. If $r_{\sup}(k, \cdot|\Gamma_l)$ is quasiconvex, as for example in monotone decision problems, each strict local minimum is also a global one. In all other cases, one has to safeguard against multimodality, for example by using some version of a multistart strategy. However, in this case it is even more important to be able to reduce the inner maximization problem to a drastically simplified version for most of the values of k, because it has to be solved much more often.

We construct a sequence $(m_l(k))_{l=1,2,\ldots}$ of lower bounds for $M_l(k)$ as follows:

$l = 1$. Compute $M_1(k)$ and set $m_1(k) = M_1(k)$ for each $k \in K$. Calculate (k^1, y^1).

$l \geq 2$. Assume that for each $k \in K$ a lower bound $m_{l-1}(k)$ for $M_{l-1}(k)$ is known and that (k^{l-1}, y^{l-1}) has been computed in the previous step. Calculate $M_l(k^{l-1})$. For all $k \in K_l = \{k \in K: m_{l-1}(k) < M_l(k^{l-1})\}$ compute $M_l(k)$. Put

$$m_l(k) = \begin{cases} M_l(k), & \text{if } k \in K_l \cup \{k^{l-1}\}, \\ m_{l-1}(k), & \text{otherwise.} \end{cases}$$

For calculating (k^l, y^l) all points k which are different from k^{l-1} and do not belong to K_l can be excluded; since $r_{\sup}(k, y|\Gamma_l) \geq r_{\sup}(k, y|\Gamma_{l-1})$ for all (k, y), it holds for each $k \in K \backslash K_l$ that

$$M_l(k) \geq M_{l-1}(k) \geq m_{l-1}(k) \geq M_l(k^{l-1}) \geq M_l(k^l).$$

Therefore $M_l(k)$ is calculated only for k^{l-1} and those k which belong to K_l, and k^l is chosen among these values. Observe that a value of k that has been excluded in a previous step may well be included. Using information about the particular structure of the problem, a considerable further reduction of the set of possible values of k is often possible; see Section 8.1.

4.2. *Stopping criterion.* A possible strategy might be to stop the algorithm if

$$r_{\sup}(k^l, y^l|\Gamma_{l+1}) - r_{\sup}(k^l, y^l|\Gamma_l) \leq \varepsilon \tag{5}$$

for some $\varepsilon > 0$. This, however, does not necessarily guarantee convergence. On the other hand, if one is able to compute for pregiven (k, y) the exact value of $r_{\sup}(k, y|\Gamma)$, the special structure of minimax problems can be exploited to derive a more reliable criterion. This is based on the following lemma.

LEMMA 4.1 (Stopping criterion).

(a) *If* $r_{\sup}(k^l, y^l|\Gamma) = r_{\sup}(k^l, y^l|\Gamma_l)$ *then*:

(i) (k^l, y^l) *is a minimal point of* $r_{\sup}(\cdot, \cdot|\Gamma)$ *and*

(ii) $r_{\sup}(k^l, y^l|\Gamma_l) = \min_{k\in K, y\in\mathscr{Y}_k} r_{\sup}(k, y|\Gamma)$.

(b) *If* $r_{\sup}(k^l, y^l|\Gamma) \le r_{\sup}(k^l, y^l|\Gamma_l) + \varepsilon$ *for some* $\varepsilon > 0$, *then*

$$\min_{k\in K, y\in\mathscr{Y}_k} r_{\sup}(k, y|\Gamma) \le r_{\sup}(k^l, y^l|\Gamma_l) + \varepsilon.$$

PROOF. The assertion follows because

$$r_{\sup}(k^l, y^l|\Gamma_l) \le \min_{k\in K,\, y\in\mathscr{Y}_k} r_{\sup}(k, y|\Gamma) \le r_{\sup}(k^l, y^l|\Gamma).$$

In the algorithm, the stopping criterion is applied as follows: as we do not want to perform exact inner maximization too often, we proceed with the iteration until (5) is satisfied for some $\varepsilon = \varepsilon_1$. Then we compute either the exact value of $r_{\sup}(k^l, y^l|\Gamma)$ by solving the corresponding linear program, if Θ is finite, or a "good approximation" $\tilde{r}_{\sup}(k^l, y^l|\Gamma)$ of $r_{\sup}(k^l, y^l|\Gamma)$ using Algorithm 1 in Section 5.2, if Θ is not finite. If $r_{\sup}(k^l, y^l|\Gamma) - r_{\sup}(k^l, y^l|\Gamma_l) \le \varepsilon_2$ or $\tilde{r}_{\sup}(k^l, y^l|\Gamma) - r_{\sup}(k^l, y^l|\Gamma_l) \le \varepsilon_2$ for some ε_2, then we set $(k_\Gamma, y_\Gamma) = (k^l, y^l)$ and stop. □

5. Inner maximization for compact Θ. In this section, we deal with calculation of $r_{\sup}(k, y|\Gamma)$ for fixed (k, y), the primal optimization problem, if Θ is an infinite, compact set. The regularity conditions R1 in the introduction are assumed to be satisfied.

On several occasions, we have to state theorems simultaneously for Θ and for Θ_l. Recall that Θ_l is finite, thus compact; we shall introduce the necessary concepts for an arbitrary compact subset Ω of Θ. In this case, g_i and $R(k, y|\theta)$ denote the restrictions of the corresponding functions to Ω. For a probability measure π on Ω and for real numbers c_i, $1 \le i \le r$, the vectors $\mathbf{m}(\pi)$ and $\mathbf{c}$ are defined by $\mathbf{m}(\pi) = (m_{g_1}(\pi), \ldots, m_{g_r}(\pi))$ and $\mathbf{c} = (c_1, \ldots, c_r)$. Let Γ be the set of probability measures on Ω restricted by $\mathbf{m}(\pi) = \mathbf{c}$ (equality constraints) or by $\mathbf{m}(\pi) \le \mathbf{c}$ (inequality constraints). We shall give a sufficient condition for convergence of $r_{\sup}(k, y|\Gamma_l)$ to $r_{\sup}(k, y|\Gamma)$ and present an algorithm for calculating $r_{\sup}(k, y|\Gamma)$, both for $\Omega = \Theta$ and for Γ_l being defined by (4).

5.1. *The dual problem.* We define $c_0 = 1$ and $g_0(\theta) = 1$ for all $\theta \in \Omega$. Then each probability measure $\pi \in \Gamma$ satisfies the additional condition $\int_\Omega g_0(\theta)\ d\pi(\theta) = c_0$. Depending on the structure of Γ, let S be the set

$$S = \left\{\mathbf{s} = (s_0, s_1, \ldots, s_r) \in \mathbb{R}^{r+1}\colon \sum_{i=0}^{r} s_i g_i(\theta) \ge R(k, y|\theta) \quad \forall\, \theta \in \Omega\right\},$$

if Γ is defined by equality constraints, or the restriction of this set to all $\mathbf{s}$ with $s_0 \in \mathbb{R}$ and $s_j \ge 0$, $j = 1, \ldots, r$, if Γ is defined by inequality constraints.

These sets are not empty, since they contain the point $(\bar{s}_0, \bar{s}_1, \ldots, \bar{s}_r)$ defined by $\bar{s}_0 = 1 + \max_{\theta\in\Omega} R(k, y|\theta)$ and $\bar{s}_i = 0$, $i = 1, 2, \ldots, r$. Consider the following linear semi-infinite problem:

$$\text{find } I(S) = \inf\left\{\sum_{i=0}^{r} c_i s_i : \mathbf{s} \in S\right\}. \tag{6}$$

It is the dual problem of the calculation of $r_{\sup}(k, y|\Gamma)$: obviously, $r_{\sup}(k, y|\Gamma) \leq I(S)$ holds. If $\Gamma \neq \varnothing$, then $r_{\sup}(k, y|\Gamma) = I(S)$. For equality constraints, this result is well known; see, for example, Chapter 8 in Goberna and López (1998) or Theorem 1 in Gustafson (1973). For inequality constraints, see, for example, Theorem, 2.1 in Lai and Wu (1992).

5.2. *Main results and algorithms.* For $\mathbf{g} = (g_0, g_1, \ldots, g_r)$, let

$$C(\mathbf{g}(\Omega)) = \left\{\mathbf{z} \in \mathbb{R}^{r+1} : \mathbf{z} = \sum_{j=1}^{q} \alpha_j \mathbf{g}(\theta_j),\ \theta_j \in \Omega, \alpha_j \geq 0, q < \infty\right\}$$

denote the *convex cone generated by* $\mathbf{g}(\Omega)$ and

$$M_{\mathbf{g}}(\Omega) = \Big\{\mathbf{z} \in \mathbb{R}^{r+1} : \int_{\Omega} g_i(\theta)\, d\pi(\theta) = z_i, \quad i = 0, 1, \ldots, r \quad \text{for at least one nonnegative finite measure } \pi \text{ on } \Omega\Big\}$$

the *moment cone* of the moments of measures on Ω with respect to $\mathbf{g}$. It is well known that $M_{\mathbf{g}}(\Omega) = C(\mathbf{g}(\Omega))$ [see, e.g., Gustafson (1973), Lemma 3].

ASSUMPTION 5.1. For some $\mathbf{s}^0 \in S$, the lower level set $N(\mathbf{s}^0, \mathbf{c}, S) = \{\mathbf{s} \in S : \sum_{i=0}^{r} c_i s_i \leq \sum_{i=0}^{r} c_i s_i^0\}$ is compact.

REMARK 5.1. Assumption 5.1 will be needed because it guarantees the existence of a point $\mathbf{s}^* \in S$ with $\sum_{i=0}^{r} c_i s_i^* = I(S)$ $(= r_{\sup}(k, y|\Gamma))$. In general, such a point need not exist, in contrast to the primal problem, where there is always a (discrete) measure which maximizes the goal function, if the latter is bounded on the feasible region.

THEOREM 5.1. *A sufficient condition for Assumption* 5.1 *is that* $(1, c_1, \ldots, c_r) \in \operatorname{int} C(\mathbf{g}(\Omega))$, *where* int *A denotes the interior of a set A. If the moment conditions are defined by equality constraints, the condition is also necessary.*

Theorem 5.1 is proved in Section 5.3.

REMARK 5.2. Suppose that $\mathbf{c} = (c_0, c_1, \ldots, c_r)$ has a *maximal representation*, that is, that $\mathbf{c} = \sum_{j=0}^{r} t_j \mathbf{g}(\theta_j)$ for certain points $\theta_0, \theta_1, \ldots, \theta_r$ in Θ with $t_j > 0$ and such that the vectors $\mathbf{g}(\theta_0), \ldots, \mathbf{g}(\theta_r)$ are linearly independent. Then $\mathbf{c} \in \operatorname{int} C(\mathbf{g}(\Omega))$ for each subset Ω of Θ that contains the points $\theta_0, \theta_1, \ldots, \theta_r$.

DEFINITION 5.1. For $M > 0$, let

$$I(S, M) = \inf \left\{ \sum_{i=0}^{r} c_i s_i : \mathbf{s} \in S,\ |s_i| \le M,\ i = 0, 1, \ldots, r \right\}.$$

Let $\Omega = \Theta$. Calculation of $I(S, M)$ is a semi-infinite optimization problem. Below, we present a standard algorithm based on a given finite discretization Θ_{l_0} of Θ. This algorithm is examined in Reemtsen and Görner (1998) under the name of implicit one-point exchange algorithm and has been employed in a Bayesian context in Betrò and Guglielmi (1996). For $\mathbf{s} \in S$, let $G(\mathbf{s}) = \sum_{i=0}^{r} c_i s_i$.

ALGORITHM 1. *Let $\Theta^0 = \Theta_{l_0}$. A sequence $(\mathbf{s}^n)_{n\in\mathbb{N}}$ in S and a sequence $(\Theta^n)_{n\in\mathbb{N}}$ of finite subsets of Θ is calculated as follows*:

(a) $G(\mathbf{s}^n) = \min\{G(\mathbf{s}): \mathbf{s} \in S^{n-1}(M)\}$, *where* $S^{n-1}(M) = S^{\mathrm{eq}} = \{\mathbf{s}: \sum_{i=0}^{r} s_i \times g_i(\theta) \ge R(k, y|\theta), \theta \in \Theta^{n-1}, |s_i| \le M,\ i = 0, 1, \ldots, r\}$, *if* Γ *is defined by equality constraints, and* $S^{n-1}(M) = S^{\mathrm{eq}} \cap \{\mathbf{s}: s_i \ge 0,\ i = 1, \ldots, r\}$, *if* Γ *is defined by inequality constraints*.

(b) $\Theta^n = \Theta^{n-1} \cup \{\theta^n\}$, *where* θ^n *is defined by*

(c) $\sum_{i=0}^{r} s_i^n g_i(\theta^n) - R(k, y|\theta^n) = \min\{\sum_{i=0}^{r} s_i^n g_i(\theta) - R(k, y|\theta), \theta \in \Theta\}$.

THEOREM 5.2.

(a) *The sequence $(G(\mathbf{s}^n))_{n\in\mathbb{N}}$ converges from below to $I(S, M)$.*

(b) *Each accumulation point $\mathbf{s}^*$ of the sequence $(\mathbf{s}^n)_{n\in\mathbb{N}}$ is a minimal point of $G(\mathbf{s})$ subject to $\mathbf{s} \in S$, $|s_i| \le M,\ i = 0, 1, \ldots, r$.*

Theorem 5.2 can be proved in the same way as Theorem 2.8 in Reemtsen and Görner (1998).

If Assumption 5.1 is satisfied for $\Omega = \Theta$, then for each (k, y), $r_{\sup}(k, y|\Gamma)$ can be calculated by Algorithm 1: let $\mathbf{s}^*$ be given as in Remark 5.1 and let $M^* = \max\{|s_i^*|: i = 0, 1, \ldots, r\}$. Then for each $M \ge M^*$, Algorithm 1 converges to $r_{\sup}(k, y|\Gamma)$. In our implementation, we stop the iteration if $\sum_{i=0}^{r} s_i^n g_i(\theta^n) - R(k, y|\theta^n) \ge -\tilde{\varepsilon}$, where $\tilde{\varepsilon}$ is a pregiven positive constant. Observe that if $\sum_{i=0}^{r} s_i^n g_i(\theta^n) - R(k, y|\theta^n) \ge 0$, then $\mathbf{s}^n \in S$ and $G(\mathbf{s}^n) = I(S, M)$ holds. This stopping criterion can be found, for example, in Goberna and López [(1998), Algorithm 11.4.1]. In part (c) of Algorithm 1, a global optimization problem has to be solved. If Θ is a low-dimensional set and if the functions g_i and $R(k, y|\cdot)$ are sufficiently smooth (which is usually the case in statistical applications), this can be done by considering only the values at a sufficiently fine grid $\widehat{\Theta}$ on Θ. Note, however, that this yields the dual solution of the linear program for $\widehat{\Theta}$, so it is equivalent to calculating $r_{\sup}(k, y|\widehat{\Gamma})$ by the primal LP (linear program).

Usually Algorithm 1 is embedded into the main algorithm; $r_{\sup}(k^*, y^*|\Gamma_{l_0})$ has been calculated for some (k^*, y^*) and some discretization level l_0. Then

one may start Algorithm 1 at Θ_{l_0}, and if the vector $\tilde{\mathbf{s}}$ of the dual variables of the optimal solution of the primal problem is provided by the LP-algorithm, $\mathbf{s}^1 = \tilde{\mathbf{s}}$ may be used in part (a).

REMARK 5.3. In Reemtsen and Görner (1998) it is mentioned that Algorithm 1 suffers in practice from the monotonic growth of the constraint set and the large number of costly computations of global minimizers in step (c). Moreover, instability may occur especially when step (c) picks a point θ^n very close to some point already included in the finite set Θ^{n-1}. In such a case the finite linear program may encounter big problems because the set of constraints is "almost" linearly dependent. Alternative methods for which convergence has been proved can be found in Goberna and López [(1998), Algorithm 11.4.2] and in Hettich and Zencke [(1982), Algorithm 5.2.10]. Roleff (1979) proposed an algorithm (without convergence proof) which intends to overcome the monotonic growth of the constraint set by eliminating some old constraints in each step. Convergence of a special case of Roleff's algorithm is shown in Lai and Wu (1992). However, the alternative methods mentioned in this remark also have certain disadvantages. Up to now, there is no algorithm that overcomes all drawbacks of Algorithm 1 simultaneously.

The next theorem, which can be proved using duality and Corollary 2.9 in Reemtsen and Görner (1998), gives a sufficient condition for convergence of $r_{\sup}(k, y|\Gamma_l)$ to $r_{\sup}(k, y|\Gamma)$.

THEOREM 5.3. *Let $(\Theta_l)_{l\in\mathbb{N}}$ be an increasing sequence of subsets of Θ such that $\cup\{\Theta_l: l \in \mathbb{N}\}$ is dense in Θ. Assume that for some index l_0, Assumption* 5.1 *holds for $\Omega = \Theta_{l_0}$. Then $r_{\sup}(k, y|\Gamma_l)$ converges from below to $r_{\sup}(k, y|\Gamma)$.*

Clearly it is hard to verify that Assumption 5.1 holds for some $\Omega = \Theta_{l_0}$, and l_0 may depend on (k, y). However, according to Theorem 5.1, a sufficient condition is that $(1, c_1, \ldots, c_r) \in \operatorname{int} C(\mathbf{g}(\Theta_{l_0}))$ [in this case, l_0 is independent of (k, y)]. In fact, more can be said.

THEOREM 5.4. *Let $(\Theta_l)_{l\in\mathbb{N}}$ be an increasing sequence of subsets of Θ such that $\cup\{\Theta_l: l \in \mathbb{N}\}$ is dense in Θ. If $(1, c_1, \ldots, c_r) \in \operatorname{int} C(\mathbf{g}(\Theta))$, then there is an index l_0 such that $(1, c_1, \ldots, c_r) \in \operatorname{int} C(\mathbf{g}(\Theta_{l_0}))$.*

Theorem 5.4 can be proved using Theorem 5.1 in this paper and part (b) of Corollary 2.5 in Reemtsen and Görner (1998).

In our main algorithm, $r_{\sup}(k, y|\Gamma)$ is approximated by $r_{\sup}(k, y|\Gamma_l)$ for an increasing sequence $(\Theta_l)_{l\in\mathbb{N}}$ of subsets of Θ. An actual candidate (k^l, y^l) for (k_Γ, y_Γ) and a lower bound $L = r_{\sup}(k^l, y^l|\Gamma_{l+1})$ for $r_{\sup}(k^l, y^l|\Gamma)$ is calculated in each step. To test for convergence (see Section 4.2), $I(S, M)$ is calculated for sufficiently large M, starting with the discretization Θ_{l+1}. Now suppose that M has been chosen too small or that Assumption 5.1 is not satisfied for

Θ, meaning that Algorithm 1 need not converge to $r_{\sup}(k^l, y^l|\Gamma)$. However, in this case it converges to $I(S, M) > r_{\sup}(k^l, y^l|\Gamma)$, hence $L < I(S, M)$, and the main algorithm does not stop. Moreover, it cannot stop, if the sequence $(r_{\sup}(k^l, y^l|\Gamma) - r_{\sup}(k^l, y^l|\Gamma_l))_{l\in\mathbb{N}}$ does not converge to zero. Conversely, if it stops, then (k_Γ, y_Γ) or at least a good approximation to it has been found.

5.3. *Proof of Theorem* 5.1. The proof of Theorem 5.1 is based on Lemmas 5.1 and 5.2 and Remark 5.4 below.

LEMMA 5.1 [cf., e.g., Hettich and Zencke (1982), Theorem 3.1.11]. *Let $V \subset \mathbb{R}^n$ be such that the convex cone $C(V)$ generated by V is closed. Then for all $\mathbf{c} \in \mathbb{R}^n$ exactly one of the following two statements holds true*:

(a) $\mathbf{c} \in \operatorname{int} C(V)$; *that is*, $\mathbf{c}$ *belongs to the interior of* $C(V)$.
(b) *The system* $\sum_{i=1}^n c_i x_i \geq 0$; $\sum_{i=1}^n v_i x_i \leq 0\ \forall \mathbf{v} \in V$ *has a solution* $\mathbf{x} \neq \mathbf{0}$.

LEMMA 5.2 [cf. Hettich and Zencke (1982), Theorem 3.2.7]. *Let $X \subseteq \mathbb{R}^n$ be nonempty, convex and closed and let $\mathbf{c} \in \mathbb{R}^n$. Then the following properties are equivalent*:

(i) *The system* $\sum_{i=1}^n c_i \xi_i \leq 0$, $\xi + X \subset X$, *where* $\xi + X \subset X$ *means* $\xi + \mathbf{y} \in X$ *for all* $\mathbf{y} \in X$, *has no solution* $\xi \neq \mathbf{0}$.
(ii) *There exists* $\mathbf{x}^0 \in X$ *such that the lower level set* $N(\mathbf{x}^0, \mathbf{c}, X) = \{\mathbf{x} \in X: \sum_{i=1}^n c_i x_i \leq \sum_{i=1}^n c_i x_i^0\}$ *is compact*.
(iii) *For all* $\mathbf{x}^0 \in X$ *the lower level set* $N(\mathbf{x}^0, \mathbf{c}, X)$ *is compact and the solution set* $X_{\text{opt}} = \{\bar{\mathbf{x}} \in X: \sum_{i=1}^n c_i \bar{x}_i = \inf_{x\in X} \sum_{i=1}^n c_i x_i\}$ *is nonempty, convex and compact*.

REMARK 5.4. Let (Γ, S) be a dual pair as defined in Section 5.1 and let $\mathbf{s} = (s_0, s_1, \ldots, s_r)$. If S corresponds to equality constraints, it holds that $\mathbf{s} + S \subset S \leftrightarrow \sum_{i=0}^r s_i g_i(\theta) \geq 0\ \forall \theta \in \Omega$. If S corresponds to inequality constraints, it holds that $\mathbf{s} + S \subset S \leftrightarrow s_0 \in \mathbb{R}$, $s_i \geq 0$, $i = 1, \ldots, r$, $\sum_{i=0}^r s_i g_i(\theta) \geq 0\ \forall \theta \in \Omega$.

PROOF. "$\longleftarrow$" is easy to see, so let us show "$\longrightarrow$." Let $\mathbf{s} + S \subset S$, then for all $\mathbf{x} \in S$ and for all $n \in \mathbb{N}$ it holds that $n\mathbf{s} + \mathbf{x} \in S$, that is,

$$n \sum_{i=0}^r s_i g_i(\theta) + \sum_{i=0}^r x_i g_i(\theta) - R(k, y|\theta) \geq 0 \qquad \forall \theta \in \Omega. \tag{7}$$

Suppose now that $\sum_{i=0}^r s_i g_i(\tilde{\theta}) < 0$ for some $\tilde{\theta} \in \Omega$. Then we choose n so big that $n \sum_{i=0}^r s_i g_i(\tilde{\theta}) + \sum_{i=0}^r x_i g_i(\tilde{\theta}) - R(k, y|\tilde{\theta}) < 0$ and obtain a contradiction to (7). Furthermore, $s_i \geq 0$, $i = 1, \ldots, r$, if S corresponds to inequality constraints, because for all n, $ns_i + x_i \geq 0$, $i = 1, \ldots, r$. □

PROOF OF THEOREM 5.1. The set $C(\mathbf{g}(\Omega))$ is closed, as it coincides with the closed set $M_{\mathbf{g}}(\Omega)$. By Lemma 5.1, $(1, c_1, \ldots, c_r) \in \operatorname{int} C(\mathbf{g}(\Omega))$ if and only if the system $\sum_{i=0}^r c_i s_i \leq 0$, $\sum_{i=0}^r s_i g_i(\theta) \geq 0\ \forall \theta \in \Omega$ has no solution. By Remark 5.4 this implies that the system $\sum_{i=0}^r c_i s_i \leq 0$, $\mathbf{s} + S \subset S$ has no solution;

for S corresponding to equality constraints these conditions are equivalent. The set S is nonempty, closed and convex; therefore the assertion follows from Lemma 5.2. □

6. σ-compact parameter space. Let Θ be a σ-compact subset of an Euclidean parameter space and assume that the regularity conditions R1 in the Introduction are satisfied. Moreover, we shall assume that the risk function $R(k, y|\theta)$ is bounded on Θ for each (k, y), then $r_{\sup}(k, y|\Gamma)$ is finite for each set Γ. Note that $r_{\sup}(k, y|\Gamma)$ is continuous in y, if Γ is concentrated on a (finite or) compact subset of Θ, whereas it need not be continuous for arbitrary sets Γ.

Often it is argued that in practice, it is always possible to assume that the parameter space is compact. In presence of moment conditions, however, this is not obvious.

EXAMPLE 6.1. Let $\Theta = \mathbb{R}$ and assume that for $g(\theta) = \theta$, Γ is given by the condition $m_g(\pi) = 0$. Consider a sequence (x_n) in $\mathbb{R}$ such that $R(k, y|x_n)$ converges to $R_{\sup} = \sup\{R(k, y|\theta)\colon \theta \in \Theta\}$. Let $y_n = (1 - n)x_n$ and let π_n be the measure that puts mass $1/n$ to the point y_n and mass $1 - (1/n)$ to the point x_n. Then $m_g(\pi_n) = 0$ and $r(k, y|\pi_n)$ converges to $R_{\sup}$. Consequently, $r_{\sup}(k, y|\Gamma) = R_{\sup}$, and the Γ-minimax strategy is the same as the minimax strategy without prior information. For a compact subset of Θ, however, this assertion does not necessarily hold.

Let $(\Omega_l)_{l\in\mathbb{N}}$ be an increasing sequence of compact subsets of Θ such that $\cup\{\Omega_l\colon l \in \mathbb{N}\} = \Theta$, and let $\widetilde{\Gamma}_l$ be the restriction of Γ to measures concentrated on Ω_l.

LEMMA 6.1. *For each* (k, y), $r_{\sup}(k, y|\widetilde{\Gamma}_l)$ *converges from below to* $r_{\sup}(k, y|\Gamma)$.

PROOF. In Winkler (1988) it is shown that $r_{\sup}(k, y|\Gamma)$ can be approximated arbitrarily well by $r(k, y|\nu)$ for measures ν with finite support. For each of these measures, however, there exists an index l such that ν is concentrated on Ω_l. □

It can be shown that the sequence of Γ-minimax strategies for $(\widetilde{\Gamma}_l)$ converges in some sense to a Γ-minimax strategy for the original set Γ. Therefore, one could start with a large compact set Ω^*, calculate a Γ-minimax strategy, calculate another strategy for a set $\Omega^+ \supset \Omega^*$, and iterate this procedure, until convergence seems to be achieved. Each strategy is approximated by a sequence of linear programs, and we will show now that there is some kind of a "diagonal sequence" that approximates the desired strategy.

Let Θ_l be an increasing sequence of finite subsets of Θ such that $\cup\{\Theta_l\colon l \in \mathbb{N}\}$ is dense in Θ, and let Γ_l be the restriction of Γ to measures concentrated on

Θ_l. For each l, let (k^l, y^l) be a Γ_l-minimax strategy. There exists for each l a compact set Ω_l such that $\Theta_l \subset \Omega_l$ and $\Omega_l \uparrow \Theta$. For each i, l, let $\Gamma_{i|l}$ be the set of all measures in Γ concentrated on $\Theta_i \cap \Omega_l$ and Γ_l^* the restriction of Γ to measures concentrated on Ω_l.

THEOREM 6.1. *Assume that for each* l,

$$\lim_{i\to\infty} r_{\sup}(k, y|\Gamma_{i|l}) = r_{\sup}(k, y|\Gamma_l^*)$$

holds for all (k, y). *Then the sequence* $(k^l, y^l)_{l\in\mathbb{N}}$ *has an accumulation point* $(\tilde{k}; \tilde{y})$; *each accumulation point* $(\tilde{k}, \tilde{y})$ *is a minimal point of* $r_{\sup}(k, y|\Gamma)$ *and* $r_{\sup}(k^l, y^l|\Gamma_l)$ *converges from below to* $\min_{k\in K,\, y\in\mathscr{Y}_k} r_{\sup}(k, y|\Gamma)$.

PROOF. For each l, $r_{\sup}(k, y|\Gamma_l) \le r_{\sup}(k, y|\Gamma)$ and therefore $r_{\sup}(k^l, y^l|\Gamma_l) \le \inf_{k\in K,\, y\in\mathscr{Y}_k} r_{\sup}(k, y|\Gamma)$. The sequence $r_{\sup}(k^l, y^l|\Gamma_l)$ is nondecreasing and bounded from above by $\inf_{k\in K, y\in\mathscr{Y}_k} r_{\sup}(k, y|\Gamma)$. Thus it converges to some limit G with

$$G \le \inf_{k\in K,\, y\in\mathscr{Y}_k} r_{\sup}(k, y|\Gamma) \le r_{\sup}(\tilde{k}, \tilde{y}|\Gamma). \tag{8}$$

Now we show that $r_{\sup}(\tilde{k}, \tilde{y}|\Gamma) \le G$ [then it follows that $r_{\sup}(\tilde{k}, \tilde{y}|\Gamma) = G = \inf_{k\in K,\, y\in\mathscr{Y}_k} r_{\sup}(k, y|\Gamma)$ and the theorem is proved]. Suppose that $r_{\sup}(\tilde{k}, \tilde{y}|\Gamma) > G$. Let $\varepsilon = r_{\sup}(\tilde{k}, \tilde{y}|\Gamma) - G$ (> 0). By Lemma 6.1 there exists $\nu \in \mathbb{N}$ such that $r_{\sup}(\tilde{k}, \tilde{y}|\Gamma_\nu^*) \ge r_{\sup}(\tilde{k}, \tilde{y}|\Gamma) - \varepsilon/4$. Moreover, by the assumption of the theorem there exists $\rho \in \mathbb{N}$ such that $r_{\sup}(\tilde{k}, \tilde{y}|\Gamma_{\rho|\nu}) \ge r_{\sup}(\tilde{k}, \tilde{y}|\Gamma_\nu^*) - \varepsilon/4$. Thus

$$r_{\sup}(\tilde{k}, \tilde{y}|\Gamma_{\rho|\nu}) \ge r_{\sup}(\tilde{k}, \tilde{y}|\Gamma) - \varepsilon/2. \tag{9}$$

We assume w.l.o.g. that $(k^l, y^l)_l$ converges to $(\tilde{k}, \tilde{y})$, and since $r_{\sup}(k, y|\Gamma_{\rho|\nu})$ is continuous in y and there are only finitely many k, there exists $\tau \in \mathbb{N}$, $\tau > \rho$ such that $r_{\sup}(k^\tau, y^\tau|\Gamma_{\rho|\nu}) \ge r_{\sup}(\tilde{k}, \tilde{y}|\Gamma_{\rho|\nu}) - \varepsilon/4$. Taking relation (9) into account we obtain $r_{\sup}(k^\tau, y^\tau|\Gamma_{\rho|\nu}) \ge r_{\sup}(\tilde{k}, \tilde{y}|\Gamma) - 3\varepsilon/4 = r_{\sup}(\tilde{k}, \tilde{y}|\Gamma)/4 + 3G/4 > G$. On the other hand, from $\tau > \rho$ we have $r_{\sup}(k^\tau, y^\tau|\Gamma_{\rho|\nu}) \le r_{\sup}(k^\tau, y^\tau|\Gamma_\tau) \le G$. This is a contradiction. □

As a consequence of Theorem 6.1, we can approximate (k_Γ, y_Γ) by our algorithm, where the elimination technique described in Section 4.1 can be applied. There is, however, no upper bound for $r_{\sup}(k, y|\Gamma)$ which could be used as a test for convergence. Consequently we stop the algorithm, if the "weak" criterion (5) in Section 4.2 is satisfied for some $\varepsilon > 0$. In addition, one can calculate the exact value of $r_{\sup}(k^l, y^l|\widetilde{\Gamma})$ for the restriction $\widetilde{\Gamma}$ of Γ to a "large" compact subset of Θ.

7. Comments on the regularity conditions. Convergence of the algorithms has been proved under certain regularity conditions. If possible, these

should be checked before running the algorithm. This is not always possible, and then some care is needed to interpret the results.

As an example, let us discuss compactness of the space $\mathscr{Y}_k$. In many applications, it is not compact from the beginning. Consider Example 3.2 (sampling by variables, upper specification limit). Here it can be shown that with $\mathscr{Y}_k = [-\infty, \infty]$ and the obvious interpretation of $(k, -\infty)$ and (k, ∞), the risk function $R(k, y|\theta)$ can be extended to a continuous function on $\mathscr{Y}_k \times \Theta$. Quasi-convexity of $r_{\sup}(k, y|\Gamma_l)$ and $r_{\sup}(k, y|\Gamma)$ in y guarantees that a minimum in the interior of $\mathscr{Y}_k$ is found or convergence to $\pm\infty$ can be established, at least within statistically meaningful bounds.

If K is finite and $\mathscr{Y}_k$ is not compact, but the regularity conditions for convergence hold for each compact subset of each $\mathscr{Y}_k$, then the following is easy to prove: if there exists a sequence $(k^l, y^l)_{l=1,2,\ldots}$ of global minimizers of $r_{\sup}(k, y|\Gamma_l)$ that has an accumulation point (k^*, y^*), then (k^*, y^*) minimizes $r_{\sup}(k, y|\Gamma)$. Of course neither the existence of global minimizers nor that of an accumulation point is guaranteed in this case. However, if one finds a global minimizer in each iteration step, the above stopping criteria are satisfied at step l and the difference between (k^{l-1}, y^{l-1}) and (k^l, y^l) is small, then a solution of the minimax problem has been found.

Another crucial point is that the stopping criterion may never be satisfied, for example, due to a bad choice of the constant M in the inner maximization algorithm or because Assumption 5.1 is not satisfied. Then the lower and the upper bound for the function value in the candidate for the minimax point may not coincide and the algorithm may not stop or even the "weak" stopping criterion may not be satisfied. We are grateful to a referee who showed a way out of this difficulty: one should try to find a "good" starting set Θ_0. This might be chosen in the case that Γ is defined by equality constraints such that it has $r+1$ points and, if $\mathbf{g}(\Theta_0)$ denotes the matrix $\{g_i(\theta_j^0)\}$, the linear system $\mathbf{g}(\Theta_0)\mathbf{x} = \mathbf{c}$ has a unique solution, that is, $\det \mathbf{g}(\Theta_0) \neq 0$ and $\mathbf{x} = \mathbf{g}(\Theta_0)^{-1}\mathbf{c}$. In the case of inequalities, one may choose some $\mathbf{c}' \leq \mathbf{c}$. In this case, $\mathbf{c}$ (resp. $\mathbf{c}'$) has a maximal representation (see Remark 5.2) and Assumption 5.1 is satisfied.

A last remark refers to the stopping criterion motivated by Lemma 4.1. If Θ is not finite, an approximation $\tilde{r}_{\sup}(k^l, y^l|\Gamma)$ of $r_{\sup}(k^l, y^l|\Gamma)$ is calculated using Algorithm 1. If $\mathbf{s}^n$ is not a feasible point (see Algorithm 1 and the discussion of its stopping criterion), there will be an approximation error. Although the distance between $\tilde{r}_{\sup}(k^l, y^l|\Gamma)$ and $r_{\sup}(k^l, y^l|\Gamma)$ will usually be small, the inequality for $\min_{k\in K,\, y\in\mathscr{Y}_k} r_{\sup}(k, y|\Gamma)$ in part (b) of Lemma 4.1 is satisfied in a strict sense only if $r_{\sup}(k^l, y^l|\Gamma)$ is equal to $\tilde{r}_{\sup}(k^l, y^l|\Gamma)$.

8. Robust Bayesian designs: computation and examples.

8.1. *Computational issues.*

8.1.1. *Outer minimization.* In general, $r_{\sup}(k, y|\Gamma)$ is not unimodal in y. For example, if y is one-dimensional (equidistant design points), then in many cases $r_{\sup}(k, y|\Gamma_l)$ has two or three local minima. Sometimes it depends on the

discretization Θ_l which minimum is the global one. Moreover, $r_{\sup}(k, y|\Gamma)$ is not necessarily differentiable in y (it is a general property of minimax problems of type $\min_y \sup_z f(y, z)$ that even for differentiable f, the maximum function $\sup_z f(y, z)$ is often not differentiable, especially in the local minimizers $\tilde{y}$). Taking into account these properties, we started a local minimizer of Nelder–Mead simplex type, which is especially suitable for nondifferentiable functions [see Nelder and Mead (1965); Parkinson and Hutchinson (1972)], from a grid of initial points. For one-dimensional y, we used $\{0.1, 1, 2, \ldots, 10\}$; for $y \in [0, \infty]^r$, we used $\{0.1, 2, 4, 6, 8\}^r$.

Let $\mathscr{Y}$ denote the space of all possible values of y; it is not compact. Moreover, K is not finite, if k is one-dimensional; in our algorithm, we restricted it to a finite set (depending on the problem) in this case (for unbalanced designs with fixed total sample size, K is finite from the beginning). From our choice of starting values, our numerical observations (only a few local minima of $r_{\sup}(k, y|\Gamma_l)$ with respect to y and convergence to the same minima from different starting points, components of the local minimizers $\tilde{y}$ bounded away from zero and not too large), and in the spirit of the discussion in Section 7, one may conclude that the calculated designs are minimax at least in the space of designs with k restricted to the considered set and components of y not much larger than the largest component of the starting values. The assumption that these designs are minimax in the class of all designs of the considered type is of heuristic nature only, although it is supported by additional numerical experiments (also with respect to the variable k).

For stopping the algorithm, criterion (5) in Section 4.2 is used with $\varepsilon = 10^{-6}$. As an additional condition (see Section 7) we checked if $k^{l-1} = k^l$ and $y^{l-1} \approx y^l$ by tracing (k^l, y^l) without using a formal criterion. To make balanced and unbalanced designs comparable, the criterion function is always calculated with $n = 1$.

8.1.2. *Inner maximization.* The most important feature here is the choice of the discretization of Θ. Consider positive integers n_β, n_μ and equidistant discretizations p_i, $i = 1, \ldots, n_\beta$ and q_j, $j = 1, \ldots, n_\mu$ of $]0, 1[$. These were transformed into a discretization $\{(\mu_{i,j}, \beta_i)), i = 1, \ldots, n_\beta, j = 1, \ldots, n_\mu\}$ of Θ as follows:

1. Let G denote the cumulative distribution function of some (prior) distribution of β. Then $\beta_i = G^{-1}(p_i)$.
2. $\mu_{i,j}$ is defined by the relations $q_j = 1/(1+\exp(-z_{i,j}))$ and $z_{i,j} = \beta_i(x-\mu_{i,j})$, where $x = 0$.

For G, we used the cumulative distribution function of the lognormal distribution with expectation 1 and variance v_β, where v_β is chosen as the upper bound for the variance of β in the moment conditions defining Γ.

To define the discretization Θ^l in step l, we chose this discretization with parameters n_β^l and n_μ^l, depending on l. Usually we started with $n_\mu^1 = n_\beta^1 = 70$ or $= 140$ and updated it according to $n_\mu^{l+1} = 2 \times n_\mu^l$ and $n_\beta^{l+1} = 2 \times n_\beta^l$.

8.2. *Results for the D-criterion.* Let us first start with some symmetrical designs with possibly different dose spacings. We considered a maximum number of six design points; that is, we calculated the designs for a total sample size of $n = 6$. The Γ-minimax risk is divided by n^2 to make the results better comparable with the balanced case, where we assumed $n = 1$. For $I_2(\mu) = [0, 0.1]$ and $I_2(\beta) = [1, 1.1]$ we obtained a two-point design with $k_\Gamma = (3, 3)$, $y_\Gamma = d_1 = 2.926$ and $r_{\sup}(k_\Gamma, y_\Gamma|\Gamma)/n^2 = -0.040275$. For $I_2(\mu) = [0, 0.5]$ and $I_2(\beta) = [1, 1.5]$, the design is concentrated on six points with $k_\Gamma = (1, 1, 1, 1, 1, 1)$, $y_\Gamma = (d_1, d_2, d_3) = (2.461, 1.711, 2.551)$ and $r_{\sup}(k_\Gamma, y_\Gamma|\Gamma)/n^2 = -0.018190$.

As the number of starting points for minimization in y gets very large with increasing dimension r of this problem, the number s of different design points that can be considered by this version of the algorithm is restricted to values not much larger than ten (these designs may also be applied for a larger total sample size; however, then they need not be optimal. A more sophisticated version of the algorithm might apply the elimination strategy also to the range of y). Usually there are local minima y^* of $r_{\sup}(k_\Gamma, y|\Gamma)$ different from y_Γ but with almost the same function value; moreover, there are often also equidistant designs with a not much larger value of the design criterion. In the next step we will consider only equidistant designs.

For a total sample size of $n = 14$, the set K of the vectors of admissible sample sizes consists of 127 elements. Table 1 shows some selected equidistant D-optimal designs for different bounds on the variances. The resulting designs are often nearly balanced, which might be an additional argument for considering balanced designs. Moreover, these are good examples where the elimination method is efficient: starting from a discretization of 140×140 points, we obtain 280×280 points for $l = 2$. In this second step, the size of the set of values of k that has to be considered is usually drastically reduced (one or three nearly balanced designs with different values of s, say). Sometimes the algorithm stops in the next step, sometimes it needs one additional iteration. Therefore, a large number of vectors k which result in a high risk are ruled out already at an early approximation step without wasting too much computing time.

TABLE 1
Equidistant designs for total sample size $n = 14$

$I_2(\mu)$	$I_2(\beta)$	k_Γ	y_Γ	$r_{\sup}(k_\Gamma, y_\Gamma\|\Gamma)/n^2$
[0, 0.1]	[1, 1.1]	(7, 7)	2.926	−0.040275
[0, 0.1]	[1, 2]	(1, 2, 1, 2, 1, 1, 2, 1, 2, 1)	0.374	−0.013219
[0, 0.5]	[1, 1.1]	(6, 2, 6)	1.707	−0.034700
[0, 0.5]	[1, 1.5]	(4, 3, 3, 4)	1.172	−0.018215
[0, 1]	[1, 1.1]	(4, 1, 1, 2, 1, 1, 4)	0.698	−0.030473
[0, 1]	[1, 1.5]	(2, 1, 1, 2, 2, 2, 1, 1, 2)	0.599	−0.015423
[0, 1]	[1, 2]	(2, 2, 3, 3, 2, 2)	0.946	−0.007771
[0.5, 0.5]	[1.5, 1.5]	(4, 3, 3, 4)	1.174	−0.018216

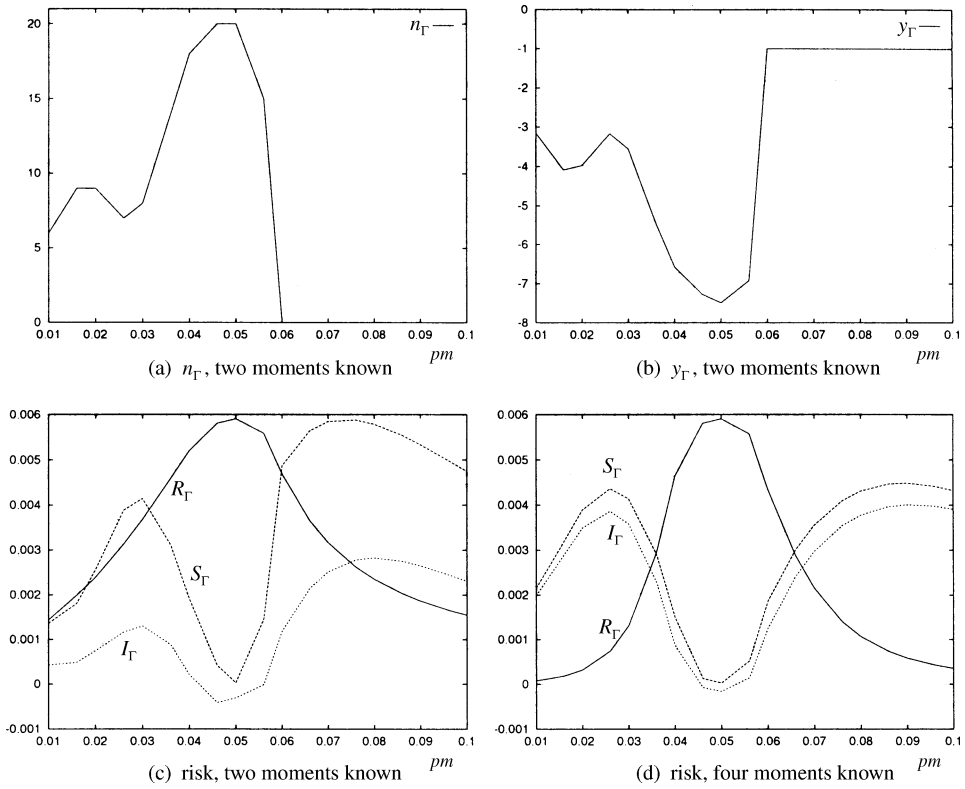
TABLE 2
Equidistant, balanced designs

$I_2(\mu)$	$I_2(\beta)$	k_Γ	y_Γ	$r_{\sup}(k_\Gamma, y_\Gamma \mid \Gamma)$
[0, 0.1]	[1, 1.1]	2	2.926	−0.040275
[0, 0.1]	[1, 2]	12	0.305	−0.013211
[0, 0.5]	[1, 1.1]	2	3.104	−0.034402
[0, 0.5]	[1, 1.5]	4	1.236	−0.01823
[0, 1]	[1, 1.1]	3	1.982	−0.030131
[0, 1]	[1, 1.5]	4	1.372	−0.015342
[0, 1]	[1, 2]	10	0.519	−0.007517
[0.5, 0.5]	[1.5, 1.5]	4	1.241	−0.018232

If unbalanced designs are admitted, the number of elements in K grows very fast with n. Therefore even equidistant designs can be calculated only for moderate values of n, although the results are applicable as an approximation for arbitrarily large total sample sizes. If the set of designs considered is further reduced, arbitrary values of n can be considered. Table 2 shows optimal balanced designs for the same types of prior information as in Table 1. Obviously the results are comparable. For the interpretation from a statistical point of view, observe that $I_2(\mu) = [0, 0.1]$ and $I_2(\beta) = [1, 1.1]$ represents strong belief in the initial guess. According to Sitter (1992), the locally optimal design is concentrated on two points, in the same way as the designs in our examples. When less is known about the parameters a priori, the recommended design is more spread out and has more design points. This observation has already been made by Sitter in his model.

To give an example of the effect of correlation, let $I_2(\mu) = \{0.5\}$ and $I_2(\beta) = \{1.5\}$, so the variance of μ and β is assumed to be 0.5. The designs and the Γ-minimax risk are nearly identical to those for $I_2(\mu) = [0, 0.5]$ and $I_2(\beta) = [1, 1, 5]$. Now let $I_2(\mu, \beta) = [0, 0.05]$, which means that the correlation between μ and β does not exceed 0.1. It turns out that this condition has no influence at all. On the other hand, for $I_2(\mu, \beta) = [0.45, \infty]$ (i.e., the correlation between μ and β is at least 0.9) we obtain $k_\Gamma = (4, 1, 2, 2, 1, 4)$, $y_\Gamma = 0.674$ and $r_{\sup}(k_\Gamma, y_\Gamma|\Gamma)/n^2 = -0.019790$ in the unbalanced and $(k_\Gamma, y_\Gamma, r_{\sup}(k_\Gamma, y_\Gamma|\Gamma)) = (3, 1.618, -0.019604)$ in the balanced case.

9. Monotone testing problems in sampling inspection: examples. In *attributes sampling* (Example 3.1), *hypergeometric model*, Θ is discrete and the inner maximization problem is a linear problem which can in principle be solved exactly. However, considerable gains in computation time can be achieved by our elimination technique compared to a naive approach where the inner maximization problem is solved exactly for all (k, y) needed for the solution of the outer minimization problem (for certain types of prior information, it was possible to cut down the computing time from several hours to a few minutes). In the *binomial model*, Θ is a compact interval and $r_{\sup}(k^l, y^l|\Gamma)$ is calculated using Algorithm 1 in Section 5.2.



(a) n_Γ, two moments known (b) y_Γ, two moments known

(c) risk, two moments known (d) risk, four moments known

FIG. 1.

Let us now show some typical Γ-minimax sampling plans for *sampling by variables, σ^2 unknown* (Example 3.2). Let Γ represent exact knowledge of the first ordinary moments of the prior and let a Stange cost model be given with parameters $p_0 = 0.05$ and $q = 10^{-4}$ (note that the value of q in this version of the Stange cost model is the normalized value of "real" sampling costs).

For illustration purposes only, we assume for simplicity that the true prior π is $N(a, \tau^2)$, the normal distribution with expectation a and variance τ^2. We choose $\tau^2 = 0.03$, which means that the dispersion of the prior is about 17% of the dispersion of the quality characteristic. To illustrate a in terms of the fraction nonconforming, observe that $p(a)$ is the fraction nonconforming if θ coincides with its expectation a [however, $p(a)$ is *not* the expectation of the fraction nonconforming]. Therefore $p(a)$ may be considered as an interpretation of the location parameter a in terms of the fraction nonconforming. In our examples, we show the sampling plans corresponding to the first moments of $\pi = N(p^{-1}(pm), 0.03)$ for pm varying in the interval $[0, 0.1]$.

To measure the quality of prior information, let us start from the minimax sampling plan. This corresponds to the situation when no prior information is available, it minimizes the maximum value of the risk function. Here it is

given by $(k_0, y_0) = (17, -6.9)$ with a maximum risk of 0.005964. The range of the gains that may be achieved by using (k_Γ, y_Γ) instead of (k_0, y_0) is delimited by $S_\Gamma = \sup\{r(k_0, y_0|\pi) - r(k_\Gamma, y_\Gamma|\pi)\colon \pi \in \Gamma\}$ (maximum gain) and by $I_\Gamma = \inf\{r(k_0, y_0|\pi) - r(k_\Gamma, y_\Gamma|\pi)\colon \pi \in \Gamma\}$ (minimum gain). If Γ is a singleton, S_Γ and I_Γ coincide. These bounds are shown together with $R_\Gamma = r_{\sup}(k_\Gamma, y_\Gamma|\Gamma)$ in Figure 1. Parts (a)–(c) correspond to the situation where the first two moments are known, whereas (d) shows the risk functions, when the first four moments are known. Here $(k_\Gamma, y_\Gamma) = (0, 0)$ for pm (roughly) ≤ 0.04 and $(k_\Gamma, y_\Gamma) = (0, -1)$ for $pm \geq 0.06$. For the interpretation, observe that $(k, y) = (0, -1)$ $(= (0, 0))$ means rejecting (accepting) without sampling. The smallest possible sample size not equal to 0 is $k = 2$.

Acknowledgments. We thank the Co-Editor, James Berger, an Associate Editor and the referees for valuable suggestions that considerably improved the article.

REFERENCES

BERGER, J. O. (1985). *Statistical Decision Theory and Bayesian Analysis*. Springer, New York.

BERGER, J. O. (1990). Robust Bayesian analysis: sensitivity to the prior. *J. Statist. Plann. Inference* **25** 303–328.

BETRÒ, B., RUGGERI, F. and MĘCZARSKI, M. (1994). Robust Bayesian analysis under generalized moment conditions. *J. Statist. Plann. Inference* **41** 257–266.

BETRÒ, B. and GUGLIELMI, A. (1996). Numerical robust bayesian analysis under generalized moment conditions (with discussion) In *Bayesian Robustness* (J. O. Berger, B. Betrò, E. Moreno, L. R. Pericchi, F. Ruggeri, G. Salinetti and L. Wasserman, eds.) 3–19. IMS, Hayward, CA.

DALL'AGLIO, M. and SALINETTI, G. (1994). Bayesian robustness, moment problem and semi-infinite linear programming, Poster Presented at the Fifth Valencia Meeting, Alicante (Spain), 5–9 June 1994.

GOBERNA, M. A. and LÓPEZ, M. A. (1998). *Linear Semi-Infinite Optimization*. Wiley, Chichester.

GUSTAFSON, S. A. (1973). Nonlinear systems in semi-infinite programming. In *Numerical Solution of Systems of Nonlinear Algebraic Equations* (G. D. Byrne and C. A. Hall, eds.) 63–99. Academic Press, New York.

HETTICH, R. and KORTANEK, K. O. (1993). Semi-infinite programming: theory, methods and applications. *SIAM Rev*. **35** 380–429.

HETTICH, R. and ZENCKE, P. (1982). *Numerische Methoden der Approximation und semi-infiniten Optimierung*. Teubner, Stuttgart.

KEMPERMAN, J. H. B. (1987). Geometry of the moment problem. In *Moments in Mathematics* (H. J. Landau, ed.) **37** 1–15. Amer. Math. Soc., Providence, RI.

KRUMBHOLZ, W. (1982). Die Bestimmung einfacher Attributprüfpläne unter Berücksichtigung von unvollständiger Vorinformation. *Allgemeines Stat. Archiv* **66** 240–253.

LAI, H. C. and WU, S. Y. (1992). On linear semi-infinite programming problems, an algorithm. *Numer. Funct. Anal. Optim.* **13** 287–304.

NELDER, J. A. and MEAD, R. (1965). A Simplex method for function minimization. *Comput. J*. **7** 308–313.

PARKINSON, J. M. and HUTCHINSON, D. (1972). An investigation into the efficiency of variants of the simplex method. In *Numerical Methods for Nonlinear Optimization* (F. A. Lootsma, ed.) 115–135. Academic Press, New York.

REEMTSEN, R. and GÖRNER, S. (1998). Numerical methods for semi-infinite programming: a survey. In *Semi-Infinite Programming* (R. Reemtsen, and J. J. Rückmann, eds.) 195–275. Kluwer, Boston.

REEMTSEN, R. and RÜCKMANN, J. J. (eds.) (1998). *Semi-Infinite Programming*. Kluwer, Boston.
ROLEFF, K. (1979). A stable multiple exchange algorithm for linear SIP, *Lecture Notes in Control and Inform. Sci*. **15** 83–96. Springer, Berlin.
SCHMITZ, N. (1992). *Optimal Sequentially Planned Decision Procedures. Lecture Notes in Statist.* **79**. Springer, Berlin.
SITTER, R. R. (1992). Robust designs for binary data. *Biometrics* **48** 1145–1155.
SMITH, J. E. (1995). Generalized Chebychev inequalities: theory and application. In *Decision Analysis, Operations Research* **43** 807–825.
THYREGOD, P. (1974). Bayesian single sampling acceptance plans for finite lot sizes. *J. Roy. Statist. Soc. Ser. B* **36** 305–319.
WINKLER, G. (1988). Extreme points of moment sets. *Math. Oper. Res*. **13** 581–587.

UNIVERSITÄT DER BUNDESWEHR HAMBURG
D-22039 HAMBURG
GERMANY
E-MAIL: roger.noubiap@unibw-hamburg.de
wilfried.seidel@unibw-hamburg.de