KYBERNETIKA — VOLUME 27 (1991), NUMBER 1

ON NON-NORMAL ASYMPTOTIC BEHAVIOR OF OPTIMAL SOLUTIONS FOR STOCHASTIC PROGRAMMING PROBLEMS AND ON RELATED PROBLEMS OF MATHEMATICAL STATISTICS*

JITKA DUPAČOVÁ

Non-normal asymptotic behavior of estimated optimal solutions for stochastic programming problems can appear if the true optimal solution belongs to the boundary of the set of feasible solutions. The type of the non-normal asymptotic distribution is described and the existing conditions under which asymptotic normality is retained are discussed and compared. The results are related to asymptotic properties of restricted estimates and illustrated for L_1 and L_2 estimates of parameters in restricted linear regression models.

1. INTRODUCTION

We shall consider the stochastic programming problem:
find $x(P_0)$ that minimizes $f(x, P_0)$ on a given nonempty closed set $M \subset \mathbb{R}^n$,
where

$$f(x, P_0) = \int_\Omega h(x, \omega)\, P_0(\mathrm{d}\omega) \tag{1}$$

where

P_0 is a given probability measure on $(\Omega, \mathscr{B})$,
h: $\mathbb{R}^n \times \Omega \to \mathbb{R}^1$ is a given function such that $h(x, \cdot)$ is continuous on Ω and $h(\cdot, \omega)$ is locally Lipschitz for all $\omega \in \Omega$ with a P_0-integrable Lipschitz modulus.

In many practical situations, the probability measure P_0 is not known completely and we use its estimate, say P_N, based on the observed data. Accordingly, we solve the problem

$$\text{minimize} \quad f(x, P_N) \quad \text{on the set } M \tag{2}$$

instead of the original one and we use its optimal solution $x(P_N)$ at the place of the true optimal solution $x(P_0)$.

* Presented at the "Kolloquium über Mathematische Statistik im Rahmen der Wissenschaftlichen Kolloquien der Universität Hamburg und der Karls-Universität Prag", Hamburg, June 1989.

Except for the constraints, the approximate problem (2) has the form familiar from M-estimation: with P_N – the empirical probability measure concentrated at the points $\omega_1, \ldots, \omega_N$, the problem (2) reads

$$\text{minimize } \frac{1}{N} \sum_{i=1}^{N} h(x, \omega_i) \quad \text{on the set } M\,. \tag{3}$$

Asymptotic behavior of the optimal value of (3) was studied in connection with hypotheses testing (see e.g. [3], [14], [28]), asymptotic behavior of optimal solutions of (3) is connected with properties of restricted estimators; for examples see [10].

To get statistical properties of the optimal solution $x(P_N)$ of (2) and of the optimal value

$$\varphi(P_N) = f(x(P_N), P_N) = \min_{x \in M} f(x, P_N) \tag{4}$$

of the estimated objective function we shall imbed (2) into a family of parametric optimization problems

$$\text{minimize} \quad f(x, P) \quad \text{on the set } M. \tag{5}$$

The parameter P in (5) is supposed to belong to a subset $\mathscr{P}$ of probability measures on $(\Omega, \mathscr{B})$ such that the true probability measure $P_0 \in \mathscr{P}$ and $\mathscr{P}$ is metrized in a suitable way. We shall assume that there are some persistence and stability results available (see e.g. Theorem 4.2 of [25] or Theorem 3.9 of [10]) and we shall concentrate on local sensitivity analysis of program (5) with respect to the parameter – the underlying probability measure. Sensitivity analysis of program (5) and asymptotic behavior of the estimated optimal solutions and of the estimated optimal value are closely connected with differentiability properties of the optimal solutions and of the optimal value function φ at the true parameter value $P = P_0$, respectively. The weakest concept of differentiability uses Gâteaux derivates or directional derivatives (in the case that $\mathscr{P}$ is a parametric family of probability measures indexed by a real vector parameter y); for results connected with other types of directional derivatives and their application see [30].

For to obtain results on differentiability of the optimal solution $x(P)$ of (5) at $P = P_0$ we shall assume that $x(P_0)$ is locally unique minimizer of (1) on the set M. We shall also assume that the optimal solutions $x(P)$ of (5) are P_0-almost surely unique for all P belonging to a neighborhood of P_0.

We shall mostly consider the set $M \subset \mathbb{R}^n$ of feasible solutions defined by explicit constraints, say

$$M = \{x \in \mathbb{R}^n\colon g_i(x) = 0,\ i = 1, \ldots, m,\ g_i(x) \leqq 0,\ i = m + 1, \ldots, m + p\} \tag{6}$$

with C^1 functions g_i, $i = 1, \ldots, m + p$, so that the Lagrangian approach can be applied. There are different assumptions that guarantee the existence of unique local minimizers of (1) on the set (6); see e.g. [11], some of them will be formulated later. If the corresponding Lagrange multipliers are unique, too, one can get their derivatives as a byproduct. Moreover, under quite general assumptions, the Gâteaux or directional derivative of the optimal solution $x(P)$ at $P = P_0$ equals the optimal solution of a convex linear program being linear if and only if the strict complementarity conditions (see Assumption A3 in Section 2) are fulfilled. The form of the quadratic program can be obtained from [16]. In spite of the fact, that [16] considers (nonstochastic) nonlinear programs dependent on a real vector parameters the results can be applied in the context of stochastic programming to obtain Gâteaux derivatives of optimal solutions (cf. [5], [6], [31]).

If $M \subset \mathbb{R}^n$ is a general closed nonempty set, an alternative approach can be used. The necessary condition for $x(P)$ to be a local minimizer of (5) reads

$$0 \in \partial_x f(x(P), P) + N_M(x(P)) \tag{7}$$

where $N_M(x)$ denotes the normal cone to M at the point $x \in M$, e.g., for M convex polyhedral

$$N_M(x) = \{u \in \mathbb{R}^n : (x' - x)^{\mathrm{T}} u \leqq 0 \quad \forall x' \in M\} ,$$

and $\partial_x f(x, P)$ is subdifferential of $f(\cdot, P)$ at x as defined by Clarke [4] for locally Lipschitz functions. It means that there should exist

$$v_0(P) \in \partial_x f(x(P), P) \quad \text{and} \quad v_M(P) \in N_M(x(P))$$

such that

$$0 = v_0(P) + v_M(P).$$

If $f(\cdot, P)$ is differentiable at $x(P)$ we get a simple necessary condition

$$- \nabla_x(f(x(P), P) \in N_M(x(P)) .$$

In this case, the necessary condition for differentiability of $x(P)$ at $P = P_0$ (in the sense of an *affine* approximation of its graph at $P = P_0$) reads

$$- \nabla_x f(x(P_0), P_0) \in \operatorname{int} N_M(x(P_0)) \tag{8}$$

(cf. [24]). Moreover, for the set M explicitly defined as (6) with C^1 functions g_i, $i = 1, \ldots, m + p$ such that gradients of constraints active at $x(P_0)$ are linearly independent, condition (8) is equivalent to the mentioned strict complementarity conditions. If (8) is not fulfilled, one can still obtain Gâteaux or directional derivatives of the optimal solution and arrives again at a quadratic program [24].

Also the assumed form of the set $\mathscr{P}$ of the considered probability measures influences the solution technique that has to be adapted to the properties of the corresponding parameter space.

If $\mathscr{P}$ is a parametric family of probability measures indexed by a real vector $y \in Y \subset \mathbb{R}^q$,

$$\mathscr{P} = \{P_y, y \in Y\} , \tag{9}$$

the deterministic sensitivity analysis of (5) reduces evidently to the mentioned analysis for nonlinear parametric programs with respect to a real parameter vector. As the true parameter vector, say η, need not be known its estimate is used to obtain an estimate of the probability measure $P_0 := P_\eta$ and an approximate stochastic program is solved. Statistical properties of the estimate of the optimal value and of the optimal solution depend on the continuity and differentiability properties of the optimal value and of the optimal solution of the parametric program at the true parameter value η.

Quite different methods have been used in the statistical approach in which a sequence of problems (2) is considered for an increasing sample size N. Consistency of the optimal solutions $x(P_N)$ can be obtained under relatively mild assumptions (cf. Theorem 3.9 of [10]). In comparison with the unconstrained case as done in [15] there is a limiting assumption for asymptotic normality of $x(P_N)$, namely

$$\sqrt{(N)}\,(v_M(P_0) - v_M(P_N)) \to 0 \quad \text{in probability} \tag{10}$$

(cf. [10], assumption 4.7 (iii)). However, this assumption is not necessary and we shall see that it can be omitted if the strict complementarity conditions or condition (8) hold true. The reason is simple: Under strict complementarity or under (8), the problem

$$\text{minimize } f(x, P_0) \quad \text{on the set } M \tag{11}$$

reduces locally to minimization on an affine subspace.

If none of the mentioned sufficient conditions is fulfilled one can try to approximate the difference $f(x, P_N) - f(x(P_0), P_N)$ by a quadratic function whose coefficients of the linear term are asymptotically normal and to minimize it on a conical set that approximates locally the set M on a neighborhood of $x(P_0)$. The properties of the optimal solution of the resulting quadratic program give again a key to results on asymptotic distribution of $x(P_N)$, cf. [29].

If $h(\cdot, \omega)$ for all $\omega \in \Omega$ and g_i, $i = 1, \ldots, m + p$, are continuously differentiable, stability and sensitivity analysis for stochastic program (1), (6) can be studied in the Banach space of C^0 or C^1 functions defined on a compact subset U of $\mathbb{R}^n$ (cf. [18], [19], [30]). This approach uses central limit theorem for random variables, such as $N^{-1} \sum_{i=1}^{N} h(x, \omega_i)$ or $N^{-1} \sum_{i=1}^{N} \nabla_x h(x, \omega_i)$, from this Banach space.

Similar problems have been recognized and partly treated in statistics in connection with hypotheses testing and later with restricted M-estimation, isotonic regression, etc. One of the first papers in this direction [3] studies asymptotic properties of the maximum likelihood test statistics in case that the true parameter value belongs to the boundary of the set corresponding to the considered hypothesis and/or to the alternative. In this paper, the idea of quadratic approximation of the fitting function and of conical approximation of the set was initiated. One recognizes immediately that the strict complementarity conditions are not fulfilled if the true (unconstrained) parameter lies on the boundary of the set of the considered parameter values. The asymptotic distribution of the test statistics is connected with that of the optimal value of the quadratic program and with the chi-squared-bar statistics (cf. e.g., [1], [14], [23], [28]). Similar quadratic program appears also in nonparametric statistics, cf. [2].

In this paper we shall concentrate on the sensitivity analysis for optimal solutions with implications about asymptotic properties of restricted estimators. We shall mostly consider the case of the set M defined by explicit constraints as in (6); the parallel results for an abstract set M can be found in [8], [9].

2. SENSITIVITY ANALYSIS AND ASYMPTOTIC BEHAVIOR OF OPTIMAL SOLUTIONS

As our starting point, we shall review results on stability for the parametric program

$$\text{minimize} \quad f(x, y) \quad \text{on the set } M \tag{12}$$

where M is given by (6) and y is a parameter that belongs to a given set $Y \subset \mathbb{R}^q$. We shall thus obtain a basis for statistical sensitivity analysis and for comparison of strict complementarity conditions with condition (10). We shall denote by η the true (or reference) parameter value.

As we already know, program (12) can be directly related to sensitivity analysis for a parametric family $\mathscr{P}$ of probability measures (9). Moreover, it appears in connection with computing Gâteaux derivatives of optimal solution $x(P)$ at $P = P_0$: The Gâteaux derivative $x'(P_0; P - P_0)$ of the optimal solution $x(P)$ at P_0 in the direction of $P - P_0$ is defined as

$$x'(P_0; P - P_0) = \lim_{t \to 0_+} \frac{x(P_t) - x(P_0)}{t}$$

where $P_t = (1 - t) P_0 + tP$, $0 \leqq t < 1$ is the probability measure P_0 contaminated by P. If a suitable result on persistence holds true then

$$\{x(P_t)\} = \arg\min_{x \in M} [(1 - t) f(x, P_0) + t f(x, P)] .$$

The corresponding parametric program (12) depends on the scalar parameter t and its objective function

$$f(x, t) := (1 - t) f(x, P_0) + t f(x, P)$$

is linear in t. In this case, the reference parameter value is $t = 0$.

It would be possible to present the results for the set M depending on the parameter y, too; see e.g. [6]. However, we want to concentrate on the basic ideas and to keep the presentation as simple as possible.

Denote

$$L(x, w, y) = f(x, y) + \sum_{i=1}^{m+p} w_i g_i(x) \tag{13}$$

the Lagrange function corresponding to program (12). Accordingly, $L(\cdot, \cdot, y)$ is defined on $\mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^p_+$. For a given $y \in Y$ we shall denote by $x(y)$ the isolated local minimizer of (12) and we shall consider the whole set $W(y)$ of Lagrange multipliers that together with $x(y)$ fulfil the Kuhn-Tucker conditions. For C^1 functions $f(\cdot, y)$, g_i, $i = 1, \ldots, m + p$ it means that $w(y) \in W(y)$ if and only if the Kuhn-Tucker conditions

$$\nabla_x L(x(y), w(y), y) = 0 \tag{14}$$

$$x(y) \in M\,, \quad w_i(y) \geqq 0\,, \quad i = m + 1, \ldots, m + p \tag{15}$$

$$w_i(y) \,.\, g_i(x(y)) = 0\,, \quad i = m + 1, \ldots, m + p\,, \tag{16}$$

are fulfilled; the conditions (16) are so called complementarity conditions. We shall denote

$$I(y) = \{i\colon g_i(x(y)) = 0\} \cap \{m + 1, \ldots, m + p\}$$

the set of indices of the inequality constraints that are active at $x(y)$.

Let us list some of frequently used assumptions, they are related to the program

$$\text{minimize} \quad f(x, \eta) \quad \text{on the set } M$$

that corresponds to the true parameter vector η.

A1. *Differentiability.* For all y belonging to a neighborhood of η, the functions $f(\cdot, y)$, g_i, $i = 1, \ldots, m + p$ are C^2-functions on a neighborhood of $x(y)$. Furthermore, continuous derivatives $\nabla_{xy} f$ exist on a neighborhood of $[x(\eta), \eta]$.

A2. *Linear independence condition.* The gradients

$$\nabla_x g_i(x(\eta))\,, \quad i \in I(\eta) \cup \{1, \ldots, m\}$$

are linearly independent.

A3. *Strict complementarity conditions.* For $i \in \{m + 1, \ldots, m + p\}$,

$$w_i(\eta) = 0 \Leftrightarrow i \notin I(\eta)\,.$$

A4. *The second order sufficient condition.* The inequality

$$z^{\mathrm{T}}\, \nabla_{xx} L(x(\eta), w(\eta), \eta)\, z > 0 \tag{18}$$

holds true for each $z \neq 0$ such that

$$z^{\mathrm{T}}\, \nabla_x g_i(x(\eta)) = 0\,, \quad i = 1, \ldots, m\,, \tag{19}$$

$$z^{\mathrm{T}}\, \nabla_x g_i(x(\eta)) = 0\,, \quad i \in I(\eta) \quad \text{such that} \quad w_i(\eta) > 0 \tag{20}$$

$$z^{\mathrm{T}}\, \nabla_x g_i(x(\eta)) \leqq 0\,, \quad i \in I(\eta) \quad \text{such that} \quad w_i(\eta) = 0\,. \tag{21}$$

A5. *The strong second order sufficient condition.* The inequality (18) holds true for each $z \neq 0$ such that

$$z^{\mathrm{T}}\, \nabla_x g_i(x(\eta)) = 0\,, \quad i = 1, \ldots, m\,,$$

$$z^{\mathrm{T}}\, \nabla_x g_i(x(\eta)) = 0\,, \quad i \in I(\eta) \quad \text{such that} \quad w_i(\eta) > 0\,.$$

The assumptions A1 – A4 imply uniqueness and continuous differentiability of $x(y)$ and $w(y)$ for $y \in O(\eta)$ with the optimal value function

$$\varphi(y) = f(x(y), y)$$

twice continuously differentiable on $O(\eta)$. This situation changes essentially if the strict complementarity conditions A3 are not fulfilled.

Theorem 1 [16]. Let assumptions A1, A2, A5 hold true for the Kuhn-Tucker point $[x(\eta), w(\eta)]$ of (17). Then for all $y \in O(\eta)$, there are Lipschitz continuous vector functions $x\colon O(\eta) \to \mathbb{R}^n$ $w\colon O(\eta) \to \mathbb{R}^m \times \mathbb{R}^p_+$ such that $[x(y), w(y)]$ is Kuhn-Tucker point for (12) for which assumptions A2 and A5 are fulfilled. Moreover, $x(y)$ and $w(y)$ are directionally differentiable in any direction $s \neq 0$. The directional derivative $x'(\eta; s)$ is the unique optimal solution of the quadratic program

$$\text{minimize} \quad \left[\tfrac{1}{2} z^{\mathrm{T}} \nabla_{xx} L(x(\eta), w(\eta), \eta)\, z + z^{\mathrm{T}} \nabla_{xy} L(x(\eta), w(\eta), \eta)\, s\right] \tag{22}$$

subject to constraints (19) – (21).

The components $w_i'(\eta; s)$ of the directional derivative $w'(\eta; s)$ are equal for $i \in \in I(\eta) \cup \{1, \ldots, m\}$ to the corresponding Lagrange multipliers of the quadratic program (19) – (22) and $w_i'(\eta; s) = 0$ otherwise.

This result has been further generalized in different ways, e.g., the assumption A2 has been replaced by the Mangasarian-Fromowitz constraint qualification and by the constant rank condition:

A6. *Mangasarian-Fromowitz constraint qualification* holds at $x(\eta)$, i.e.

a) $\nabla_x g_i(x(\eta))$, $i = 1. \ldots, m$ are linearly independent;

b) there is z such that

$$\nabla_x g_i(x(\eta))^{\mathrm{T}} z = 0\,, \quad i = 1, \ldots, m\,,$$
$$\nabla_x g_i(x(\eta))^{\mathrm{T}} z < 0\,, \quad i \in I(\eta)\,.$$

A7. *The constant rank condition* holds at $x(\eta)$, i.e., for any subset $K \subset I(\eta)$ the family of gradients

$$\nabla_x g_i(x(y))\,, \quad i \in K \cup \{1, \ldots, m\}$$

remains of constant rank for all y belonging to a neighborhood of η.

Under assumption A6, the Lagrange multipliers $w(\eta)$ corresponding to $x(\eta)$ need not be unique but the set $W(\eta)$ is a nonempty bounded polyhedral set [12]. It can be proved (cf. [20]) that assumptions A1, A6 together with existence of Kuhn-Tucker point and with assumption A5 fulfilled at $x(\eta)$ for all $w \in W(\eta)$ are the weakest ones under which the optimal solution $x(y)$ of (12) is locally unique. There also exist results on directional differentiability of $x(y)$ at $y = \eta$, see e.g. [21], [27], the first of them will be formulated below.

Theorem 2 [21] Let assumptions A1, A6, A7 hold true at $x(\eta)$ and let assumption A5 be fulfilled at $x(\eta)$ for each $w \in W(\eta)$. Then for all y in a neighborhood of η, there exists a locally unique continuous local minimizer $x(y)$ of (12). Moreover $x(y)$ is directionally differentiable at η in any direction $s \neq 0$ and its directional derivative $x'(\eta; s)$ uniquely solves the following quadratic program for an extremal point $w \in W(\eta)$:

$$\text{minimize} \quad [\tfrac{1}{2}z^{\mathrm{T}}\, \nabla_{xx}L(x(\eta), w, \eta)\, z + z^{\mathrm{T}}\, \nabla_{xy}L(x(\eta), w, \eta)\, s]$$

subject to

$$\nabla_x g_i(x(\eta))^{\mathrm{T}}\, z = 0\,, \quad i = 1, \ldots, m\,,$$

$$\nabla_x g_i(x(\eta))^{\mathrm{T}}\, z = 0\,, \quad i \in K(w; \eta)\,,$$

$$\nabla_x g_i(x(\eta))^{\mathrm{T}}\, z \leqq 0\,, \quad i \in I(\eta) - K(w; \eta)\,,$$

where $K(w; \eta) = \{k \in I(\eta): w_k > 0\}$.

If we solve the parametric program (12) for an estimate y_N of the true parameter η that is asymptotically normal,

$$\sqrt{(N)}\,(y_N - \eta) \sim \mathcal{N}(0, \Sigma) \tag{23}$$

then under assumptions of Theorem 1, the distribution of $\sqrt{(N)}\,(x(y_N) - x(\eta))$ is asymptotically equivalent to that of $x'(\eta; \sqrt{(N)}\,(y_N - \eta))$, i.e. to the unique solution of the quadratic program (19)–(22) with asymptotically normal parameter vector $s = \sqrt{(N)}\,(y_N - \eta)$. Optimal solutions of strictly convex parametric quadratic programs such as (19)–(22) are known to be piece-wise linear, polyhedral, Lipschitz continuous vector function on the whole parameter space $\mathbb{R}^q$. It means that the parameter space can be decomposed into finitely many convex polyhedral sets, say, S_h, $h = 1, \ldots, H$, the stability sets of the quadratic program (19)–(22) with respect to the parameter s. For each h, $x'(\eta; s)$ is linear on cl S_h, differentiable on int S_h; it can be represented in more than one way on the boundary of S_h.

This decomposition implies the decomposition of the probability space in our case and we can compute, e.g., the distribution function

$$\mathsf{P}\{x'(\eta; \sqrt{(N)}\,(y_N - \eta)) \leqq a\} = \sum_{h=1}^{H} \mathsf{P}\{x'(\eta; \sqrt{(N)}\,(y_N - \eta)) \leqq$$

$$\leqq a \mid \sqrt{(N)}\,(y_N - \eta) \in S_h\} \,.\, \mathsf{P}\{\sqrt{(N)}\,(y_N - \eta) \in S_h\}\,. \tag{24}$$

We shall briefly call asymptotic distributions of the type (24) with $x'(\eta; \sqrt{(N)}\,(y_N - \eta))$ linear in $\sqrt{(N)}\,(y_N - \eta)$ on S_h (a convex polyhedral set) and with $\sqrt{(N)}\,(y_N - \eta)$ asymptotically normal a "mixture of asymptotically normal distributions conditioned by convex polyhedral sets". The results can be summarized now.

Theorem 3 [8]. Under assumptions of Theorem 1 and for asymptotically normal $\sqrt{(N)}\,(y_N - \eta)$, $\sqrt{(N)}\,(x(y_N) - x(\eta))$ is asymptotically distributed as a mixture of asymptotically normal distributions conditioned by convex polyhedral sets.

Moreover, thanks to similar continuity properties of the Lagrange multipliers of the parametric quadratic program (19)–(22), convergence in distribution holds also for the Lagrange multipliers corresponding to $s = \sqrt{(N)}\,(y_N - \eta)$ and, consequently, for $\sqrt{(N)}\,(w(y_N) - w(\eta))$. It means, inter alia, that under assumptions of Theorem 1 and for asymptotically normal $s = \sqrt{(N)}\,(y_N - \eta)$, we have $\sqrt{(N)}\,(w(y_N) - w(\eta))$ bounded in probability (cf. [26]),

$$w(y_N) - w(\eta) = O_p(N^{-1/2})\,. \tag{25}$$

Under assumptions of Theorem 1, the necessary and sufficient conditions for differentiability of the Kuhn-Tucker point $[x(y), w(y)]$ of (12) at $y = \eta$ and thus for asymptotic normality of both $x(y_N)$ and $w(y_N)$ are the strict complementarity conditions A3 (cf. [24]). Let's compare this result with the condition (10) specified to the case of parametric family $\mathscr{P} = \{P_y, y \in Y\}$ in which the true probability measure $P_0 := P_\eta$ is estimated by $P_N := P_{y_N}$, the set M is given by (6) and assumptions of Theorem 1 are fulfilled.

Notice that condition (10) was used in [10] under more general circumstances, with essentially weakened differentiability assumptions and that for an abstract set M, no constraint qualification such as A2, was needed. Accordingly, existence and uniqueness of Lagrange multipliers were not used. We shall thus relate condition (10) and strict complementarity conditions, both of which have been found connected with results on asymptotic normality of optimal solutions $x(y_N)$, only in a special case.

First of all notice that under A1 and A2, there is a one-to-one relationship between the vectors v_M and the Lagrange multipliers:

$$v_M(y) = \sum_{i=1}^{m+p} \nabla g_i(x(y))\, w_i(y)\,. \tag{26}$$

Together with

$$v_0(y) = \nabla_x(f(x(y), y) \tag{27}$$

it means that

$$v_M(y) = -\nabla_x f(x(y), y) \tag{28}$$

is uniquely defined. Furthermore, under A1, A2, A4, necessary conditions for $x(y)$ to be an isolated local minimizer of (12) are the Kuhn-Tucker conditions (14)–(16). With (23) and A1, A2, A4, the additional assumption A3 – the strict complementarity conditions – implies asymptotic normality of Lagrange multipliers $w(y_N)$ and, according to (26), of the vectors $v_M(y_N)$ as well. We obtain (cf. [26])

$$v_M(y_N) - v_M(\eta) = O_p(N^{-1/2})$$

in distinction to (10).

Let's have a look at the crucial part of Theorem 4.8 of [10] about asymptotic normality of optimal solutions. In our present notation, asymptotic equivalence of

$$\sqrt{(N)}\,(\nabla_x f(x(y_N), \eta) - \nabla_x f(x(\eta), \eta))$$

and

$$\sqrt{(N)}\,(v_M(\eta) - v_M(y_N)) - \sqrt{(N)}\,(\nabla_x f(x(\eta), y_N) + v_M(\eta)) \tag{29}$$

is used and asymptotic normality of (29) obtained by means of (10). Now, under A1, A2 and strict complementarity conditions A3, both terms in (29) are not only asymptotically normal, but they are asymptotically distributed as linear transformations of $\sqrt{(N)}\,(y_N - \eta)$. In this case, the result on asymptotic normality of $\sqrt{(N)}\,(x(y_N) - x(\eta))$ can be obtained without assuming (10).

3. ASYMPTOTIC PROPERTIES OF ESTIMATES IN RESTRICTED LINEAR MODELS

We shall discuss asymptotic properties of the least squares and L_1 estimates of the p-dimensional parameter vector β in the linear model

$$y = X\beta + e \tag{30}$$

in the case that the true parameter vector β is supposed to fulfil linear constraints, say,

$$\beta \in M := \{b\colon Ab \leqq d\} \tag{31}$$

with a given matrix $A(m, p)$ and a vector $d \in \mathbb{R}^m$. These restrictions correspond to the prior knowledge, they may arrise in connection with formulation of hypothesis or alternative, etc.

The same ideas can be used to obtain asymptotic properties of other types of restricted estimates even in the case that the estimated parameter vector belongs to the boundary of the considered set M. They apply, e.g., to the restricted maximum likelihood estimates used for misspecified models [32] to estimate the minimizer of the Kullback-Leibler Information Criterion.

The notation used in this section conforms to that established in statistics.

3.1. Restricted ordinary least squares regression

Assume that $X = (x_{ij})$ is a given (N, p) matrix of full column rank, e is an N-dimensional random vector of errors with

$$\mathsf{E}e = 0\,, \quad \operatorname{var} e = \sigma^2 I\,, \quad \sigma^2 > 0 \quad \text{an unknown parameter}\,, \tag{32}$$
$$y = X\beta + e$$

is an N-dimensional random vector of observations and β is a p-dimensional unknown vector of parameters to be estimated. The restricted ordinary least squares estimate $b_{2,N}$ of β is the optimal solution of the quadratic program

$$\text{minimize} \quad \|y - Xb\|_{2,N} := \sum_{k=1}^{N} (y_k - \sum_{j=1}^{p} x_{ij} b_j)^2 \tag{33}$$

on the set

$$M = \{b \in \mathbb{R}^p: Ab \leqq d\}, \tag{34}$$

see e.g. [17], [22]. This problem can be also understood as projection of y on the convex polyhedral set $M(X) = \{u: u = Xb,\ b \in M\}$.

If N is fixed and e is multinormal $\mathcal{N}(0, \sigma^2 I)$, the exact distribution of $b_{2,N}$ is that of the (unique) optimal solution of the quadratic program (33), (34) or, equivalently, of

$$\text{minimize} \quad [b^T X^T X b - 2y^T X b] \quad \text{on the set } M$$

with normally distributed vector of coefficients $(2X^T y)$ in the linear term of the objective function. The same quadratic program or projection problem appears in maximum likelihood estimation and testing, cf. [14]. Notice that the objective function (33) can be written as

$$(b - \beta)^T X^T X(b - \beta) - 2e^T X(b - \beta) + e^T e \tag{35}$$

and its minimal value appears in the test statistics.

Asymptotic properties of $b_{2,N}$ for $N \to \infty$ can be studied under suitable assumption about errors e_k, such that e_k are i.i.d. with $\mathsf{E} e_k = 0$, $\operatorname{var} e_k = \sigma^2 > 0$ and finite and about behavior of the matrix $X^T X$ for $N \to \infty$, e.g.,

$$N^{-1} X^T X \to C \quad \text{with} \quad C \quad \text{a positive definite matrix},$$

$$x^k = (x_{k1}, \ldots, x_{kp}), \quad k = 1, 2, \ldots, \quad \text{are of uniformly bounded norm}.$$

Under these assumptions, the true parameter vector β solves the (unconstrained) minimization problem

$$\text{minimize} \quad (b - \beta)^T C(b - \beta)$$

(compare with (35)). If $\beta \in \operatorname{int} M$, $\sqrt{(N)}\,(b_{2,N} - \beta)$ is asymptotically normal $\mathcal{N}(0, \sigma^2 C^{-1})$. If $\beta \in \operatorname{bnd} M$, strict complementarity conditions are not fulfilled for

$$\text{minimize} \quad (b - \beta)^T C(b - \beta) \quad \text{on the set } M$$

and the asymptotic distribution of $b_{2,N}$ is a mixture of asymptotically normal distributions conditioned by convex polyhedral sets, the stability sets of the quadratic program

$$\text{minimize} \quad (b - \beta)^T C(b - \beta) - p^T (b - \beta) \quad \text{on the set } M \tag{36}$$

with respect to the parameter p. Out of all stability sets, only those containing the parameter value $p = 0$ enter the formulas for asymptotic distribution; see [7], [9] and the discussion in Section 2.

The result can be used for hypothesis testing: Assume that σ^2 is known, $\{b: Ab = d\} \neq \emptyset$ and consider the test of hypothesis

$$\mathrm{H}_0: \beta \in M_0 := \{b: Ab = d\}$$

against alternative

$$\mathrm{H}_1: \beta \in M_1 := \{b: A_1 b \leqq d_1\}$$

where A_1 is a (q, p) submatrix of A and d_1 is the corresponding subvector of d. The likelihood ratio test rejects H_0 for large values of

$$\min_{b \in M_0} \|y - Xb\|_{2,N} - \min_{b \in M_1} \|y - Xb\|_{2,N}$$

distributed as $\bar{\chi}^2$. As the $\bar{\chi}^2$ statistics is defined as

$$\bar{\chi}^2 = y^T V^{-1} y - \min_{u \in K} (y - u)^T V^{-1}(y - u)$$

with y normal, $\mathcal{N}(0, V)$, V positive definite and K – a convex cone it is essential that M_0 is an affine linear space, $M_0 \subset M_1$ and M_1 is convex polyhedral; for details consult [27].

As the last possibility, consider the case of $\beta \notin M$, i.e., of misspecified model, and put

$$\{\tilde{\beta}\} = \arg \min_{b \in M} (b - \beta)^T C(b - \beta) . \tag{37}$$

It means that $\tilde{\beta} \neq \beta$ and, necessarily, $\tilde{\beta} \in \operatorname{bnd} M$. According to [10], $b_{2,N} \to \tilde{\beta}$ a.s. It is interesting that is this case, asymptotic normality of $\sqrt{(N)}\,(b_{2,N} - \tilde{\beta})$ can be obtained provided that for quadratic program (36) strict complementarity holds true or, equivalently, if parameter $p = 0$ belongs to the interior of a stability set for the perturbed quadratic program (36). In the opposite case, we again arrive at the mixture of asymptotically normal distributions conditioned by convex polyhedral sets.

3.2. Restricted L_1-regression

The restricted L_1 estimate $b_{1,N}$ of coefficients in linear regression model solves the program

$$\text{minimize} \quad \|y - Xb\|_{1,N} := \sum_{k=1}^{N} |y_k - \sum_{j=1}^{p} x_{kj} b_j| \quad \text{on the set } M . \tag{38}$$

We shall treat the case of random regressors as done, e.g., in [7] or in [13] for the unconstrained case. To this purpose assume that y_k, x^k are i.i.d. observations of random variable η and random vector ξ, respectively. For the true parameter vector β, the random vector ξ and $\varepsilon = \eta - \beta^T \xi$ are supposed to be independent with densities h_ξ and h_ε, respectively. Moreover, the following assumptions (or similar assumptions of [7]) have to be satisfied:

(i) The density h_ε is positive and bounded on a neighborhood of 0.

(ii) $\operatorname{med} \varepsilon = 0$

(iii) $\max_{1 \leq k \leq N} |x^k| = o_p(N^{1/2})$, $\mathsf{E}_{P_0} \xi \xi^T = \tilde{\Sigma}$, a positive definite matrix.

Notice that the assumption $\operatorname{med} \varepsilon = 0$ can be replaced by

$$\{\beta\} = \arg \min \mathsf{E}_{P_0} |\eta - b^T \xi| . \tag{39}$$

We shall again relate β to the solution $\tilde{\beta}$ of the constrained optimization problem

$$\text{minimize} \quad \mathsf{E}_{P_0}|\eta - b^{\mathrm{T}}\xi| \quad \text{on the set } M . \tag{40}$$

There are three possibilities:

a) $\beta \in \operatorname{int} M$; in this case $\beta = \tilde{\beta}$ and asymptotic normality of $\sqrt{(N)}\,(b_{1,N} - \beta)$ follows (see e.g. [7], [13]).

b) $\beta \in \operatorname{bnd} M$; in this case $\beta = \tilde{\beta}$ and $b_{1,N} \to \beta$ a.s. (cf. [7]).

As the strict complementarity conditions are not fulfilled for (40) at $b = \tilde{\beta}$, $\sqrt{(N)}\,(b_{1,N} - \beta)$ is not asymptotically normal. To obtain a result on asymptotic behavior of $b_{1,N}$ we shall use an approximating quadratic program.

c) $\beta \notin M$, i.e., the case of misspecified model with $\beta \neq \tilde{\beta}$ in which $b_{1,N} \to \tilde{\beta}$ a.s. (cf. [7]).

If strict complementarity conditions are fulfilled for (40) at $b = \tilde{\beta}$ asymptotic normality of $\sqrt{(N)}\,(b_{1,N} - \tilde{\beta})$ can be proved, cf. [7]. For the opposite case, the asymptotic distribution of $b_{1,N}$ has not yet been specified; the quadratic approximation of the objective function (38) obtained in [13] holds true only under assumption med $\varepsilon = 0$ that is not fulfilled in this case.

We shall concentrate on the case b). In [13], we can find the desired quadratic approximation

$$\sum_{k=1}^{N} |y_k - (\beta + N^{-1/2}z)^{\mathrm{T}} x^k| - \sum_{k=1}^{N} |y_k - \beta^{\mathrm{T}} x^k| =$$
$$= 2z^{\mathrm{T}} s_N + z^{\mathrm{T}} \tilde{\Sigma} z \, h_\varepsilon(0) + o_{P_0}(1) \tag{41}$$

that is valid under assumptions (i)–(iii) uniformly for all $\|z\| \leqq K$, with s_N asymptotically normal $\mathcal{N}(0, \frac{1}{4}\tilde{\Sigma})$. We want to use consistent estimates $b_{1,N}$ that belong to the convex polyhedral set M and we assume that $\operatorname{int} M \neq \emptyset$. The optimal choice of $b_{1,N}$ based on the approximation (41) corresponds to an optimal choice of z, i.e., to the solution of quadratic program

$$\text{minimize} \quad z^{\mathrm{T}} s_N + \tfrac{1}{2} z^{\mathrm{T}} \tilde{\Sigma} z \, h_\varepsilon(0)$$
$$\text{subject to} \quad \|z\| \leqq K \quad \text{and} \quad \beta + N^{-1/2} z \in M .$$

On a neighborhood of β, M can be approximated by its tangent cone $T_M(\beta)$ and we obtain quadratic program

$$\text{minimize} \quad [z^{\mathrm{T}} s_N + \tfrac{1}{2} z^{\mathrm{T}} \tilde{\Sigma} z \, h_\varepsilon(0)] \quad \text{subject to} \quad z \in T_M(\beta) \tag{42}$$

in which s_N is asymptotically normal and does not depend on z and whose optimal solutions z_N provide a local approximation of $\sqrt{(N)}\,(b_{1,N} - \beta)$. The optimal solutions z_N can be asymptotically normal if and only if $T_M(\beta)$ is a subspace what was explicitly excluded in the considered case b. The asymptotic behavior of z_N and thus of $\sqrt{(N)}\,(b_{1,N} - \beta)$, can be again obtained via the mentioned analysis of the perturbed quadratic program (42) as summarized in the following theorem:

Theorem 4. Let for the true parameter vector β defined by (39), the random vectors ξ and $\varepsilon = \eta - \beta^{\mathrm{T}}\xi$ be independent with densities h_ξ and h_ε, respectively. Let assumptions (i)–(iii) be fulfilled and let $\beta \in \operatorname{bnd} M$. Then there is a consistent sequence of estimates $b_{1,N}$ obtained as

$$b_{1,N} \in \arg\min_{b \in M} \sum_{k=1}^{N} |y_k - b^{\mathrm{T}} x^k|$$

and $\sqrt{(N)}\,(b_{1,N} - \beta)$ is asymptotically equivalent to z_N, the unique solution of the quadratic program (42). Accordingly, $\sqrt{(N)}\,(b_{1,N} - \beta)$ is asymptotically distributed as a mixture of asymptotically normal distributions conditioned by convex polyhedral sets.

A similar result can be found in [29] under different assumptions needed for uniqueness of Lagrange multipliers and the statistical assumptions strengthened to continuity of the density h_ε on a neighborhood of 0.

(Received December 8, 1989.)

REFERENCES

[1] R. E. Barlow, J. Bartholomew, J. N. Bremner and H. D. Brunk: Statistical Inference under Order Restrictions: The Theory and Application of Isotonic Regression. J. Wiley, New York 1972.

[2] K. Behnen and G. Neuhaus: Rank Tests with Estimated Scores and Their Application. (Teubner Skripten zur Mathematischen Stochastik.) Teubner, Stuttgart 1989.

[3] H. Chernoff: On the distribution of the likelihood ratio. Ann. Math. Statist. *25* (1954), 573–578.

[4] R. Clarke: Optimization and Nonsmooth Analysis. Wiley Interscience, New York 1983.

[5] J. Dupačová: Stability in stochastic programming with recourse: Contaminated distributions. Math. Programming Study *27* (1986), 133–144.

[6] J. Dupačová: Stochastic programming with incomplete information: A survey of results on postoptimization and sensitivity analysis. Optimization *18* (1987), 507–532.

[7] J. Dupačová: Asymptotic properties of restricted L_1-estimates of regression. In: Statistical Data Analysis Based on L_1-norm and Related Methods (Y. Dodge, ed.), North-Holland, Amsterdam 1987, pp. 263–274.

[8] J. Dupačová: On nonnormal asymptotic behavior of optimal solutions of stochastic programming problems: The parametric case. In: Proc. 4th Prague Symp. Asymptotic Statistics (P. Mandl, M. Hušková, eds.), Charles University, Prague 1989, pp. 205–214.

[9] J. Dupačová: On statistical sensitivity analysis in stochastic programming. Ann. Oper. Res. *29* (1990).

[10] J. Dupačová and R. J.-B. Wets: Asymptotic behaviour of statistical estimators and of optimal solutions of stochastic optimization problems. Ann. Statist. *16* (1988), 1517–1549.

[11] V. Fiacco and J. Kyparisis: Sensitivity analysis in nonlinear programming under second order assumptions. In: Systems and Optimization (A. Bagich, H. Th. Jongen, eds.), Springer-Verlag, Berlin–Heidelberg–New York 1985, pp. 74–97.

[12] J. Gauvin: A necessary and sufficient regularity condition to have bounded multipliers in nonconvex programming. Math. Programming *12* (1977), 136–138.

[13] S. van de Geer: Asymptotic Normality of Minimum L_1-norm Estimators in Linear Regression. Report MS-R8806 Centre for Mathematics and Computer Science, Amsterdam 1988.

[14] C. Gouriéroux, A. Holly and A. Monfort: Likelihood ratio test, Wald test and Kuhn-Tucker test in linear models with inequality constraints on the regression parameters. Econometrica *50* (1982), 63—80.

[15] P. J. Huber: Robust Statistics. J. Wiley, New York 1981.

[16] K. Jittorntrum: Solution point differentability without strict complementarity in nonlinear programming. Math. Programming Study *21* (1984), 127—138.

[17] G. G. Judge and T. Takayama: Inequality restriction in regression analysis. J. Amer. Statist. Assoc. *61* (1966), 166—181.

[18] A. J. King: Asymptotic Behaviour of Solutions in Stochastic Optimization: Nonsmooth Analysis and the Derivation of Non-normal Limit Distribution. Ph. D. Dissertation, Univ. of Washington 1986.

[19] A. J. King and R. T. Rockafellar: Asymptotic theory for solutions in generalized M-estimation and stochastic programming. Preprint, 1989.

[20] M. Kojima: Strongly stable stationary solutions in nonlinear program. In: Analysis and Computation of Fixed Points (S. M. Robinson, ed.), Academic Press, New York 1980.

[21] J. Kyparisis: Sensitivity analysis for nonlinear programs and variational inequalities with nonunique multipliers. Math. Oper. Res. (to appear).

[22] Ch. K. Liew: Inequality constrained Least-squares estimation. J. Amer. Statist. Assoc. *71* (1976), 746—751.

[23] T. Robertson, F. T. Wright and R. L. Dykstra: Order Restricted Statistical Inference (Wiley Series in Probability and Statistics). J. Wiley, New York 1988.

[24] S. M. Robinson: Local structure of feasible sets in nonlinear programming. Part III: Stability and sensitivity. Math. Programming Study *30* (1986), 45—66.

[25] S. M. Robinson: Local epi-continuity and local optimization. Math. Programming *37* (1987), 208—222.

[26] R. J. Serfling: Approximation Theorems in Mathematical Statistics. J. Wiley, New York 1980.

[27] A. Shapiro: Sensitivity analysis of nonlinear programs and differentiability properties of metric projections. SIAM J. Control Optim. *26* (1988), 628—645.

[28] A. Shapiro: Towards a unified theory of inequality constrained testing in multivariate analysis. Internat. Statist. Rev. *56* (1988), 49—62.

[29] A. Shapiro: Asymptotic properties of statistical estimators in stochastic programming. Ann. Statist. *17* (1989), 841—858.

[30] A. Shapiro: Gâteaux, Fréchet and Hadamard Directional Differentiability of Functional Solution in Stochastic Programming. Oper. Res. and Statistics Series No. 395, Technion, Haifa, Israel 1989.

[31] A. Shapiro: On differental stability in stochastic programming. Math. Programming *A47* (1990), 107—115.

[32] H. White: Maximum likelihood estimation of misspecified models. Econometrica *50* (1982), 1—25.

Prof. RNDr. Jitka Dupačová, DrSc., Universita Karlova, matematicko-fyzikální fakulta (Charles University — Faculty of Mathematics and Physics), Sokolovská 83, 186 00 Prague 8. Czechoslovakia.