Yokohama Publishers
ISSN 1349-8169 ONLINE JOURNAL
Pacific Journal of Optimization
Vol. 1, No. 1, pp. 179–190, January 2005

STATISTICAL INFERENCE FOR STOCHASTIC PROGRAMMING

Jinde Wang

*This paper is dedicated to Professor R.T. Rockafellar on the occasion of his 70th birthday.**

Abstract: In many practical problems available are only some sample points, not the full distribution information, of the random variables involved in the given stochastic programming problems. Then a sample problem can be formed. The optimal solution x_N of this sample program is a point estimation of the unknown "true" optimal solution x^* of the original stochastic program. Then we have to make a statistical inference for the true solution. The methods of making statistical inference in classical statistics do not apply for stochastic programs. In this paper we study how this kind of statistical inference can be made. We will construct the confidence regions (including confidence intervals) for x^* and the linear form $v'x^*$.

Key words: *statistical inference, stochastic programs, confidence region*

Mathematics Subject Classification: *90C15*

---

1 Introduction

Suppose we have the following stochastic program

$$\begin{array}{ll} \min & Ef(x,\xi) \\ \texttt{s.t.} & x \in S \subset R^n, \end{array} \tag{1}$$

where ξ is a random vector and

$$S = \{x : g_i(x) \leq 0,\ i = 1, \dots, p;\ h_j(x) = 0,\ j = p+1, \dots, q\}.$$

Denote by x^* the unique optimal solution and by (λ^*, μ^*) the corresponding Lagrangian multipliers of program (1).

In practice it is quite often that we do not have the full information on the probability distribution of ξ, but only some sample points of ξ are available. Let $t_1, \dots, t_N$ be the sample points of ξ. Then we can form a sample program of (1) as

$$\begin{array}{ll} \min & N^{-1}\sum_{k=1}^N f(x,t_k) \\ \texttt{s.t.} & x \in S. \end{array} \tag{2}$$

*Prof. R.T. Rockafellar (coauthored with Alan King) investigated the asymptotic distribution of the optimal solution of stochastic programming. Their result is a foundation of statistical inference for stochastic programming.

Copyright © 2005 Yokohama Publishers http://www.ybook.co.jp

Denote the optimal solution of (2) by x_N and the associated Lagrangian multiplier vector by (λ_N, μ_N). The solution (x_N, λ_N, μ_N) is a point estimator of the unknown true solution (x^*, λ^*, μ^*). The next step is to infer, based on (x_N, λ_N, μ_N), where the "true" solution could be.

Several authors have done some work on the statistical problems of stochastic programs. For example, Dupacova and Wets (1988) and Wets (1991) showed that x_N is a consistent estimator of x^* and $N^{\frac{1}{2}}(x_N - x^*)$ is approximately normal under some regular conditions; King (1986) and King and Rockafellar (1993) pointed out that usually $N^{\frac{1}{2}}(x_N - x^*)$ is not approximately normal, but is approximately piecewise normally distributed. They proved that $N^{\frac{1}{2}}(x_N - x^*)$ converges in distribution to the optimal solution of a stochastic quadratic program. Similar results are given also by Shapiro (1988, 1993). Zhao (1997) obtained a similar but more direct result that $N^{\frac{1}{2}}(x_N - x^*)$ converges in distribution to the optimal solution z^* of the following stochastic quadratic program:

$$\begin{array}{ll} \min & \frac{1}{2}z'Hz + z'\eta \\ \texttt{s.t.} & z \in D \subset R^n, \end{array}$$

where η has normal distribution $N(0, K)$ with $K = \operatorname{cov} L(x^*, \xi)$, where

$$\begin{aligned} L(x,\xi) &= f(x,\xi) + \lambda' g(x) + \mu' h(x), \\ H &= E\nabla^2_{xx} L(x^*, \xi) = E[\nabla^2_{xx} f(x^*, \xi) + \lambda^{*\prime}\nabla^2_{xx} g(x^*) + \mu^{*\prime}\nabla^2_{xx} h(x^*)], \\ D &= \{z : \nabla_x g_i(x^*)'z \le 0,\ i \in I^=; \nabla_x g_i(x^*)'z = 0,\ i \in I^+; \\ &\qquad \nabla_x h_j(x^*)'z = 0,\ j = p+1, \ldots, q\}, \\ I^= &= \{i : g_i(x^*) = 0,\ \lambda_i^* = 0, i = 1, \ldots, p\}, \\ I^+ &= \{i : g_i(x^*) = 0,\ \lambda_i^* > 0, i = 1, \ldots, p\}, \\ \lambda^* &= \{\lambda_1^*, \ldots, \lambda_p^*\}, \\ \lambda_N^* &= \{\lambda_{N1}^*, \ldots, \lambda_{Np}^*\}. \end{aligned}$$

As η has a symmetric distribution, the distribution of z is the same as that of the optimal solution of the program

$$\begin{array}{ll} \min & \frac{1}{2}z'Hz - z'\eta \\ \texttt{s.t} & z \in D. \end{array} \tag{3}$$

In classical statistical problems, having the estimator and the associated (approximate) distribution is enough for making statistical inference. However it is not the case in making statistical inference for stochastic programs. The main causes for that are as follows: Firstly, since there is no explicit formula of optimal solutions of quadratic programs and the index subsets $I^=$ and I^+ are unknown (they are related to the unknown solution x^*), we do not know the distribution of z^*; secondly, as $N^{\frac{1}{2}}(x_N - x^*)$ is piecewise normally distributed, we do not know which piece of the distribution of $N^{\frac{1}{2}}(x_N - x^*)$ should be used for the given data; thirdly it is not known how to treat constraints in the problems. These points are not considered in the classical statistics. Thus we have to find ways to solve these problems. That is the purpose of this article.

This paper is organized as follows. In Section 2 we give the explicit expression of the asymptotic distribution of $N^{\frac{1}{2}}(x_N - x^*)$ in each possible case. Section 3 is devoted to constructing confidence intervals and for linear forms $v'x^*$ under the assumption that the constraints in program (1) are linear. The confidence region for x^* is constructed in Section 4.

$\boxed{2}$ The Representation of Asymptotic Distribution

To obtain the (approximate) distribution of $N^{\frac{1}{2}}(x_N - x^*)$ is the most important step for making statistical inference on x^*. As mentioned in section 1, it has been proved that $N^{\frac{1}{2}}(x_N - x^*)$ converges in distribution to the optimal solution z^* of the program (3). However the theory of quadratic programming does not provide an explicit formula for the optimal solutions. Thus the distribution of z^* is still not ready. Here we go another way to solve this problem.

For simplicity of notation, denote the matrix $(\nabla_x g_i(x^*), i \in I^=)$ by $\nabla_x g^=$, $(\nabla_x g_i(x^*), i \in I^+)$ by $\nabla_x g^+$ and $\nabla_x h_j(x^*),\ j = p+1, \ldots, q$ by $\nabla_x h$. Let

$$H_0 = \begin{pmatrix} H & \nabla_x g^+ & \nabla_x h \\ (\nabla_x g^+)' & 0 & 0 \\ (\nabla_x h)' & 0 & 0 \end{pmatrix},$$

$$H_{i_1,\ldots,i_k} = \begin{pmatrix} H & \nabla_x g^+ & \nabla_x g_{i_1} & \ldots & \nabla_x g_{i_k} & \nabla_x h \\ (\nabla_x g^+)' & 0 & & \ldots & & 0 \\ (\nabla_x g'_{i_1}) & \vdots & & & & \vdots \\ \vdots & \vdots & & \ldots & & \vdots \\ (\nabla_x g'_{i_k}) & 0 & & \ldots & & 0 \\ (\nabla_x h)' & 0 & & \ldots & & 0 \end{pmatrix}, \qquad (4)$$

where $i_1, \ldots, i_k \in I^=$. We assume H is positive definite. Then H_0 and $H_{i_1,\ldots,i_k}$ are invertible under the linear independence assumption on ∇g_i and ∇h_j. Partition H_0^{-1} and $H_{i_1\ldots i_k}^{-1}$ as follows

$$H_0^{-1} = \begin{pmatrix} M_0 & R_0 & S_0 \\ R_0' & A_{22} & A_{23} \\ S_0' & A_{32} & A_{33} \end{pmatrix}, \quad (H_{i_1\ldots i_k})^{-1} = \begin{pmatrix} M_{i_1\ldots i_k} & R_{i_1\ldots i_k} & S_{i_1\ldots i_k} \\ R'_{i_1\ldots i_k} & B_{22} & B_{23} \\ S'_{i_1\ldots i_k} & B_{32} & B_{33} \end{pmatrix} \qquad (5)$$

where M_0 and M are $n \times n$-matrices,the column numbers of R_0 and S_0 correspond to the column numbers of $\nabla_x g^+$ and $\nabla_x h$ respectively, R and S are similarly defined. As H is symmetric,so are H_0 and $H_{i_1\ldots i_k}$ and their inverses. Thus (5) is a reasonable form of a partition of H_0^{-1} and $(H_{i_1\ldots i_k})^{-1}$.

Obviously the optimal solution of program (3) lies either in the relative interior or on the relative boundary of the feasible solution set D. The relative boundary of D consists of its faces and their intersections. Denote by D_0, D_{i_j} and $D_{j_1\ldots j_k}$ the relative interior, faces and the intersections of faces, respectively. Thus

$$
\begin{aligned}
D_0 &= \{z : \nabla_x g_i' z < 0,\ i \in I^=;\ \nabla_x g_i' z = 0,\ i \in I^+;\ \nabla_x h_j z = 0,\ j = p+1, \dots, q\} \\
D_{i_j} &= \{z : \nabla_x g_i' z < 0,\ i \in I^= \setminus i_j;\ \nabla_x g_i' z = 0,\ i \in I^+ \cup i_j;\ \nabla_x h_j' z = 0,\ j = p+1, \dots, q\} \\
D_{j_1 \dots j_k} &= \{z : \nabla_x g_i' z < 0,\ i \in I^= \setminus \{j_l,\, l = 1, \dots, k\}; \\
&\qquad \nabla_x g_i' z = 0,\ i \in I^+ \cup \{j_l,\, l = 1, \dots, k; = \bigtriangledown h_j' z = 0,\ j = p+1, \dots, q\}.
\end{aligned}
$$

The distribution of z^* in each part of D can be given below.

Theorem 1. Assume H is positive definite and the vectors $\{\nabla_x g_i,\ i \in I^= \cup I^+; \nabla_x h_j,\ j = p+1, \dots, q\}$ are linearly independent. Then

1) z^* in D_0 has the expression $z^* = M_0 \eta$;

2) z^* on $D_{i_1 \dots i_k}$ has the expression $z^* = M_{i_1 \dots i_k} \eta$.

Proof. By the Kuhn-Tucker conditions the points z in D_0 being optimal to program (3) must satisfy the equations

$$
\begin{aligned}
&Hz - \eta + \sum_{i \in I^+} \lambda_i \nabla_x g_i + \sum_{j=p+1}^{q} \mu_j \nabla_x h_j = 0, \\
&\nabla_x g_i' z = 0, \qquad i \in I^+, \\
&\nabla_x h_j' z = 0, \qquad j = p+1, \dots, q.
\end{aligned}
$$

The matrix form of this equation system can be written as

$$
H_0 \begin{pmatrix} z \\ \lambda \\ \mu \end{pmatrix} = \begin{pmatrix} \eta \\ 0 \\ 0 \end{pmatrix}.
$$

The solution of this system is

$$
\tilde{z} = M_0 \eta, \qquad \tilde{\lambda} = R_0' \eta, \qquad \tilde{\mu} = S_0' \eta.
$$

This is derived from the necessary optimality condition. The objective function of program (3) is a positive definite quadratic function, thus $\tilde{z}$ is the optimal solution in D_0 of (3) if and only if the following feasibility conditions can be satisfied, i.e.,

$$
\nabla_x g_i' \tilde{z} = \nabla_x g_i' M_0 \eta < 0, \qquad i \in I^=, \qquad \tilde{\lambda}_i = R_0' \eta \geq 0, \qquad i \in I^+. \tag{6}
$$

The expression $\tilde{z} = M_0 \eta$ is the first assertion of the theorem.

If, for some sample values of η, $M_0 \eta$ can not satisfy the constraints (6), the optimal solution must lie on the relative boundary of D. Let us find the expression of the optimal solution on the face $D_{i_1 \dots j_k}$. By the Kuhn-Tucker conditions the optimal solution of program (3) located on $D_{i_1 \dots j_k}$ must satisfy the equations

$$
\begin{aligned}
&Hz - \eta + \sum_{i \in I^+ \cup i_1, \dots, i_k} \lambda_i \nabla_x g_i + \sum_{j=p+1}^{q} \mu_j \nabla_x h_j = 0, \\
&\nabla_x g_i' z = 0, \qquad i \in I^+ \cup i_1, \dots, i_k, \\
&\nabla_x h_j' z = 0, \qquad j = p+1, \dots, q.
\end{aligned} \tag{7}
$$

The solution $(\tilde{z}, \tilde{\lambda}, \tilde{\mu})$ of system (7) can be expressed as

$$\begin{pmatrix} \tilde{z} \\ \tilde{\lambda} \\ \tilde{\mu} \end{pmatrix} = H^{-1}_{i_1 \ldots i_k} \begin{pmatrix} \eta \\ 0 \\ 0 \end{pmatrix},$$

where $\mu = (\mu_{p+1}, \ldots, \mu_q)'$, or equivalently,

$$\tilde{z} = M_{i_1 \ldots i_k}\eta, \qquad \tilde{\lambda} = R'_{i_j}\eta, \qquad \tilde{\mu} = S'_{i_1 \ldots i_k}\eta. \tag{8}$$

As H is positive definite, program (3) is a convex program. Then $\tilde{z}$ in (8) is the optimal solution and $\tilde{\lambda}$, $\tilde{\mu}$ are the corresponding Lagrangian multipliers if and only if the following feasibility conditions are satisfied:

$$\begin{aligned} &\nabla_x g_i' M_{i_1 \ldots i_k}\eta \le 0, \qquad && i \in I^= \setminus \{i_1, \ldots, i_k\}, \\ &R'_{i_1 \ldots i_k}\eta \ge 0, && i \in I^+ \cup \{i_j, \ldots, i_k\}. \end{aligned}$$

Then the optimal solution of program (3) on $D_{i_1 \ldots i_k}$ has the expression $z^* = M_{i_1 \ldots i_k}\eta$. Thus the proof of Theorem 1 is completed. □

The results in Theorem 1 are one-step further than earlier results because the explicit expressions of the asymptotic distribution of z^* in each part is available now. But these results are still not enough for making statistical inference because the index sets $I^+, I^=$ are still unknown (since the true solution x^* is unknown). The next section is devoted to show how to overcome this difficulty.

3 Statistical Inference for a Linear form $v'x^*$

In this paper we will investigate two types of statistical inference on x^*. One is that for the linear form $v'x^*$ of x^*, this is the usual form in statistics and if the objective function is linear in x, the confidence bounds for $v'x^*$ is just the confidence bounds for the optimal value of program (1). The other is the confidence region for x^*. The latter may be of more interest for stochastic programming, because it asserts how far the x^* can be from the obtained estimator x_N (in probabilistic sense).

In this section we make statistical inference for a linear form $v'x^*$ under the contemporary assumption that the constraints in program (1) are linear (in fact, nonlinearity does not cause essential difficulty. Here we impose linearity assumption just for technical simplicity). Then the interested programs take the following form

$$\begin{aligned} \min \quad & Ef(x, \xi) \\ \texttt{s.t.} \quad & x \in S = \{x : Ax \le b,\ Cx = d\}, \end{aligned} \tag{9}$$

where A is an $m \times n$-matrix and C is an $m_1 \times n$-matrix,

$$\begin{aligned} \min \quad & N^{-1}\sum_{i=1}^{N} f(x, t_i) \\ \texttt{s.t.} \quad & x \in S, \end{aligned} \tag{10}$$

and

$$\begin{aligned} \min \quad & \frac{1}{2}z'Hz - z'\eta \\ \texttt{s.t} \quad & z \in D = \{A_i'z \le 0,\ i \in I^=;\ A_i'z = 0,\ i \in I^+;\ Cz = 0\}, \end{aligned} \tag{11}$$

which correspond to programs (1), (2) and (3) respectively. Their optimal solutions are again denoted as x^*, x_N and z^* respectively. We will construct the confidence intervals.

$\boxed{3.1}$ Confidence Intervals of $v'x^*$

Suppose the optimal solution x_N and the associated Lagrangian multiplier vector λ_N have been calculated. We try to give an interval estimation of $v'x^*$, where v is a given vector in R^n. We study this problem in two different cases: x_N is in the relative interior or is on the boundary of S.

$\boxed{3.1.1}$ $x_N \in S^0$

In this case it holds that

$$\begin{aligned} &A_i' x_N < b_i, \qquad i = 1, \ldots, m, \\ &C x_N = d. \end{aligned}$$

Then for any $i \in \{I^+ \cup I^=\} = \{i : A_i x^* = b_i\}$ we have

$$\begin{aligned} &A_i N^{\frac{1}{2}}(x_N - x^*) < 0, \\ &C N^{\frac{1}{2}}(x_N - x^*) = 0. \end{aligned}$$

This implies that $N^{\frac{1}{2}}(x_N - x^*)$ is in the relative interior D_0 of D.

As z_N converges in distribution to z^*, then $v'z_N$ converges in distribution to $v'z^*$. By Theorem 1 the optimal solution z^* in D_0 is distributed as $M_0\eta$. Thus $v'z_N$ is approximately distributed as $v'M_0\eta$, i.e., normally distributed.

For a given confidence level α we can find a critical value t_α such that

$$P(\mid v'M_0\eta \mid \leq t_\alpha) = 1 - \alpha. \tag{12}$$

Since η is a normal random vector, the critical value t_α can be found easily. Then approximately we have the following probability statement

$$P(\mid v'N^{\frac{1}{2}}(x_N - x^*) \mid \leq t_\alpha) = 1 - \alpha. \tag{13}$$

From (15) a $(1-\alpha)100\%$-confidence interval for $v'x^*$ can be constructed as

$$v'x_N - N^{-\frac{1}{2}} t_\alpha \leq v'x^* \leq v'x_N + N^{-\frac{1}{2}} t_\alpha. \tag{14}$$

However this interval estimation is not ready for use. The reason for this is: the inferred x^* must be a feasible point of program (11) and thus t_α and v can not be chosen arbitrarily. Now we study how large the critical value t_α can be and what kind of vectors v should be.

From (15) we see that t_α is the bound of $v'N^{\frac{1}{2}}(x_N - x^*)$ with x^* in set S. Let L be the segment restricted in S of a line,going through x_N with direction v. Assume the contours $v'N^{\frac{1}{2}}(x_N - x^*) = u$ intersects L at $x^*(u)$. The point $x^*(u)$ on L can be expressed as $x^*(u) = x_N + \tau v$. Then the bound of τ determines the bound of $v'N^{\frac{1}{2}}(x_N - x^*)$. In order to keep $x^*(u)$ feasible, we should have

$$Cv = 0, \qquad \mid \tau \mid \leq \bar{\tau} = \min_{1 \leq i \leq m} \tau_i, \tag{15}$$

$$\tau_i = \begin{cases} \mid \frac{b_i - A_i' x_N}{A_i' v} \mid, & \text{if} \quad A'v \neq 0; \\ \infty, & \text{if} \quad A'v = 0. \end{cases}$$

Therefore, as the bound of $v'N^{\frac{1}{2}}(x_N - x^*)$, t_α should satisfy

$$t_\alpha = \max_{x^*} \mid v'N^{\frac{1}{2}}(x_N - x^*) \mid = \max_{\tau} \mid v'N^{\frac{1}{2}}\tau v \mid \leq N^{\frac{1}{2}}\bar{\tau}\| v \|^2.$$

Without loss of generality we assume that $\| v \|= 1$. Thus we should restrict t_α in (16) by

$$t_\alpha \leq N^{\frac{1}{2}}\bar{\tau}, \tag{16}$$

where $\bar{\tau}$ is defined by (17) and v must satisfy $Cv = 0$.

$\boxed{\textbf{3.1.2}}$ $x_N \in \partial S$

In this case x_N may be on a face or on a intersection of some faces of S. We study our inference problem in detail only for the case that x_N is on a face of S. The problem for other cases can be solved in a similar way. Without loss of generality we assume

$$\begin{aligned} &A_1 x_N = b_1, \\ &A_i x_N < b_i, \qquad i = 2, \ldots, m \\ &C x_N = d. \end{aligned} \tag{17}$$

To construct confidence intervals for $v'x^*$, we have to further distinguish two subcases that $\lambda^*_{N1} > 0$ or $\lambda^*_{N1} = 0$, where λ^*_{N1} is the Lagrangian multiplier corresponding to $A_1 x_N = b_1$.

The case $\lambda^*_{N1} = 0$. In this case all nonnegative numbers are possible values of λ^*_1. This implies that both cases

$$A_1 x^* = b_1 \qquad \textit{and} \qquad A_1 x^* < b_1$$

are possible. Hence we can make statistical inference for a linear form $v'x^*$ for those v such that $A'_1 v \leq 0$.

Note that $A'_1 v < 0$ implies that the angle between direction v and the inward normal of $A'_1 x = b_1$ is an acute angle. Again in order to see the distribution of $v'N^{\frac{1}{2}}(x_N - x^*)$ with x^* in S, it is enough to know the distribution of $v'N^{\frac{1}{2}}(x_N - x^*)$ with x^* on the segment L of the line starting from x_N with direction v restricted in S. The point on L can be expressed as

$$x^* = x_N + \tau v, \quad \tau > 0. \tag{18}$$

Then (under the assumption $Cv = 0$) for $i \in I(x^*)$ (note that $1 \notin I(x^*)$, because $A_1 x^* = A_1 x_N + \tau A'_1 v < b_i$ by (19) and (20)), it holds that

$$\begin{aligned} A_i N^{\frac{1}{2}}(x_N - x^*) < 0, \\ C N^{\frac{1}{2}}(x_N - x^*) = 0. \end{aligned}$$

Thus in this case it still holds that $z_N = N^{\frac{1}{2}}(x_N - x^*) \in D_0$. Then by Theorem 1 $v'z_N$ is approximately distributed as $v'M_0\eta$ is. It should be noticed that under the condition $x^* \in L$ ($\tau \geq 0$ in (20)). $v'z_N$ is approximately distributed as $v'M_0\eta$ under the condition $v'M_0\eta < 0$. For a given probability level α, we can find t_α such that

$$P(\mid v'M_0\eta \mid \leq t_\alpha, \;\; v'M_0\eta < 0) = 1 - \alpha.$$

Then approximately we have

$$P(\mid v'N^{\frac{1}{2}}(x_N - x^*) \mid \leq t_\alpha \mid x^* \in L) = 1 - \alpha. \tag{19}$$

From the probability statement (21) a $(1-\alpha)100\%$-confidence interval for $v'x^*$ can be given by

$$v'x_N \leq v'x^* \leq v'x_N + N^{\frac{1}{2}}t_\alpha, \tag{20}$$

$$t_\alpha \leq \bar{t}'_\alpha,$$

where $\bar{t}'_\alpha$ is the bound of t_α determined, as in 3.1.1, by

$$\begin{aligned} \bar{t}'_\alpha &= N^{\frac{1}{2}}\bar{\lambda}, \\ \bar{\tau} &= \min\{\bar{\tau}_i,\ i = 2, \ldots, m\}, \\ \tau_i &= \begin{cases} \mid \frac{b_i - A'_i x_N}{A'_i v} \mid, & \text{if} \quad A'_i v \neq 0; \\ \infty, & \text{if} \quad A'_i v = 0. \end{cases} \end{aligned}$$

The case $\lambda^*_{N1} > 0$. As $N^{\frac{1}{2}}(\lambda^*_N - \lambda^*_0)$ is bounded in probability, thus in the present case most of possible values of λ^*_{N1} are positive. This implies that $A'_1 x^* = b_1$ and then $z_N = N^{\frac{1}{2}}(x_N - x^*)$ will be on the face P_1 of the limit set D. Then the confidence interval for $v'x^*$ can be of interest only for those v such that $A'_1 v = 0$, which implies that the segment L of the line through x_N with direction v restricted in S lies on the hyperplane $A'_1 x = b_1$. For any x^* on L we have the expression

$$x^* = x_N + \tau v.$$

Hence

$$\begin{aligned} &A_1 N^{\frac{1}{2}}(x_N - x^*) = 0, \\ &A_i N^{\frac{1}{2}}(x_N - x^*) < 0, \qquad i \in I^+ \cup I^= \setminus \{1\}. \end{aligned}$$

Thus $z_N = N^{\frac{1}{2}}(x_N - x^*)$ is on the face P_1 and, by Theorem 1, z_N is approximately distributed as $M_1\eta$. Then approximately we can have

$$P(\mid v'N^{\frac{1}{2}}(x_N - x^*) \mid \leq t'_\alpha) = 1 - \alpha, \tag{21}$$

where α is a probability level and t'_α satisfies

$$P(\mid v'M_1\eta \mid \leq t'_\alpha) = 1 - \alpha.$$

From (23) a $(1-\alpha)100\%$ confidence interval for $v'x^*$ can be made as

$$v'x_N - N^{-\frac{1}{2}}t'_\alpha \leq v'x^* \leq v'x_N + N^{-\frac{1}{2}}t'_\alpha, \tag{22}$$

$$t'_\alpha \leq \bar{t}'_\alpha,$$

where $\bar{t}'_\alpha = N^{\frac{1}{2}}\bar{\tau}'$ and

$$\bar{\tau}' = \min\{\mid \frac{b_i - A'_i x_N}{A'_i v} \mid,\ i = 2, \ldots, m,\ A'_i v \neq 0\}.$$

For any other cases of x_N the confidence intervals can be obtained in a similar way based on the distribution of z^* on the corresponding parts of D. The details will be omitted.

3.2 Hypothesis Test

With the obtained confidence intervals it is easy to get rejection principles for testing linear hypothesis

$$H_0 : v'x^* = a.$$

For example, in the case $x_N \in S^0$, we will accept H_0 if

$$v'x_N - N^{-\frac{1}{2}}t_\alpha \leq a \leq v'x_N^* + N^{-\frac{1}{2}}t_\alpha,$$

and reject H_0 otherwise. Under this rejection principle we may make first type of mistake with probability α.

In the case $A_1'x_N = b_1$ and $\lambda_{N1}^* = 0$ we will accept hypothesis

$$H_0 : v'x^* = a$$

with $A_1'v < 0$ if

$$v'x_N \leq a \leq v'x_N + N^{-\frac{1}{2}}t_\alpha,$$

and reject H_0 otherwise.

In the case $A_1'x_N = b_1$ and $\lambda_{N1}^* = 0$ we will accept hypothesis

$$H_o : v'x^* = a$$

with $A_1'v = 0$ if

$$v'x_N - N^{-\frac{1}{2}}t_\alpha' \leq a \leq v'x_N^* + N^{-\frac{1}{2}}t_\alpha',$$

and reject H_0 otherwise.

The t_α, t_α' used here are the same as that defined in 3.1 in the corresponding cases.

4 Confidence Regions of x^*

In many practical problems constructing confidence regions for the unknown solution x^* would be of more interest, because the decision makers want to know how far the true solution x^* is from the sample solution x_N with an acceptable probability level $1-\alpha$. Here we do it for general nonlinear estimation problems. Again we distinguish two cases: x_N is in the relative interior or on the relative boundary of S.

4.1 $x_N \in S^0$

In this case we have

$$\begin{aligned} g_i(x_N) &< 0, \quad i = 1, \ldots, p, \\ h_j(x_N) &= 0, \quad j = p+1, \ldots, q. \end{aligned}$$

Thus under continuous differentiability of g_i and h_j, it holds that

$$\begin{gathered} 0 > g_i(x_N) = g_i(x^*) + \nabla_x g_i(x^*)'(x_N - x^*) + \circ(\|x_N - x^*\|), \\ 0 = h_j(x_N) = h_j(x^*) + \nabla h_j(x^*)'(x_N - x^*) + \circ(\|x_N - x^*\|), \\ i = 1, \ldots, p;\ j = p+1, \ldots, q. \end{gathered}$$

As the remainders in these expansions are of higher order of infinitesimal compared with $\|x_N - x^*\|$ and $\|x_N - x^*\|$ by consistency of x_N, (see Wets (1991)), we must have

$$\begin{aligned} &\nabla_x g_i'(x_N - x^*) < 0, \qquad i = 1, \ldots, p, \\ &\nabla_x h_j'(x_N - x^*) = 0, \qquad j = p+1, \ldots, q. \end{aligned} \tag{23}$$

Thus $N^{\frac{1}{2}}(x_N - x^*)$ is an interior point of D in program (3). Then by Theorem 1 $N^{\frac{1}{2}}(x_N - x^*)$ is approximately distributed as $M_0\eta$.

Note that $M_0\eta$ is a normal vector, distributed on the $n-q+p$ dimensional manifold $\overline{M} = \{z : \nabla h_j' z = 0,\ j = p+1, \ldots, q\}$. Then there is a matrix Q such that $(QM_0\eta)'(QM_0\eta)$ has a χ^2-distribution of freedom $(n-q+p)$. For a given probability level α there is a number r_α such that

$$P\{(QM_0\eta)'(QM_0\eta) \leq r_\alpha\} = 1 - \alpha.$$

Then approximately we have

$$P\{N(Q(x_N - x^*))'(Q(x_N - x^*)) \leq r_\alpha\} = 1 - \alpha.$$

From this probability statement we see that the confidence region R_1 at probability level $1-\alpha$ can be constructed as

$$R_1 = \{x : N(x_N - x^*)'Q'Q(x_N - x^*) \leq r_\alpha\}. \tag{24}$$

In order to guarantee x^* in (26) is inside S, r_α must be restricted by

$$r_\alpha \leq \bar{r} = \min\{r(i) = d(x_N, P_i),\ i = 1, \ldots, p\},$$

where P_i is the surface $P_i = \{x : g_i(x) = 0\}$.

$\boxed{4.2}$ The Case that x_N is on ∂S

We study the problem in the following case that only one constraint is active at x_N:

$$\begin{aligned} &g_1(x_N) = 0, \\ &g_i(x_N) < 0, \qquad i = 2, \ldots, p; \\ &h_j(x_N) = 0, \qquad j = p+1, \ldots, q. \end{aligned}$$

Further we distinguish two subcases that $\lambda_{N1}^* > 0$ and $\lambda_{N1}^* = 0$.

The case $\lambda_{N1}^* > 0$. In this case most of possible values of λ_{01}^* should be positive and this implies that for most of possible values of x^* it holds that $g_1(x^*) = 0$. Then $z_N = N^{\frac{1}{2}}(x_N - x^*)$ will be on the face P_1 of the set D. Hence z_N is approximately distributed as $M_1\eta$ and z_N is an $n-q+p-1$-dimensional random vector. There is a matrix R such that $N(x_N - x^*)'R'R(x_N - x^*)$ has a χ^2 distribution of freedom $n-q+p-1$. With a probability level α one can find a critical value t_α'' such that

$$P\{N(x_N - x^*)R'R(x_N - x^*) \leq t_\alpha''\} = 1 - \alpha.$$

Then the confidence region R_2 in this case can be constructed as

$$R_2 = \{(x : N(x_N - x)'R'R(x_N - x) \leq t_\alpha''\}. \tag{25}$$

The case $\lambda^*_{N1} = 0$. By boundedness in probability of $N^{\frac{1}{2}}(x_N - x^*)$ the most of possible values are around x_N, but they can be in the interior D_0 or P_1, according to $g_1' M_0 \eta$ is less than zero or not, as shown in Theorem 1. Thus by the analysis made above for a given probability level α one can find positive numbers $r(\alpha), t''_\alpha$ such that

$$P\{N(Q(x_N - x^*))' Q(x_N - x^*) \le r(\alpha),\ g_1(x^*) < 0\} = \frac{1}{2}(1-\alpha),$$
$$P\{N(x_N - x^*)' R' R(x_N - x^*) \le t''_\alpha,\ g_1(x^*) = 0\} = \frac{1}{2}(1-\alpha).$$

Thus the confidence region of x^* in this case can be constructed as

$$\begin{aligned} R_3 &= \{x : N(x_N - x)' Q' Q(x_N - x) \le r(\alpha),\ g_1(x) < 0\} \\ &\cup\ \{x : N(x_N - x)' R' R(x_N - x) \le t''_\alpha,\ g_1(x) = 0\}. \end{aligned} \tag{26}$$

The confidence region in other subcases of $x_N \in \partial S$ can be obtained in a similar way. We omit the details. We see that in different cases the confidence regions have different shapes. This kind of confidence regions is a generalized version of the usual confidence intervals in classical statistics. This kind of inference directly give the indication of where x^* could be under a given probability level. It is of more practical interest for stochastic programming problems.

Remark. In the discussion on the inference we need H and the distribution of η. However as pointed out in section 1, H and the covariance matrix $\operatorname{cov} L(x^*, \xi)$ of η depend on the unknown solution x^*. This situation seems to cause an obstacle in making statistical inference on x^*. But this obstacle can be removed by replacing $H = EH(x^*, \xi)$, and $\operatorname{cov} L(x^*, \xi)$ by $N^{-1} \sum H(x_N, t_i)$, and

$$N^{-1} \sum [L(x_N, t_i) - N^{-1} L(x_N, t_i)][L(x_N, t_i) - N^{-1} L(x_N, t_i)]'$$

respectively. This kind of approximation is often used in statistics. The reasonability of doing so lies in the facts that

$$\begin{aligned} &\lim_{N\to\infty} N^{-1} \sum H(x_N, t_i) = H, \\ &\lim_{N\to\infty} N^{-1} \sum [L(x_N, t_i) - N^{-1} L(x_N, t_i)][L(x_N, t_i) - N^{-1} \sum L(x_N, t_i)]' \\ &\quad = \operatorname{cov} L(x^*, \xi) \end{aligned}$$

guaranteed by $x_N \to x^*$ and the weak law of large numbers.

References

[1] J. Birge and R.J-B. Wets, *Stochastic Programming. Part I and Part II*, Baltzer Science Publishers BV, Bussum, Annals of Operations, Research Vols. 30 and 31, 1991.

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

[3] A.J. King, *Asymptotic behavior of solutions in stochastic optimization: Nonsmooth analysis and the derivation of nonnormal limit distribution*, Ph.D Dissertation, Univ. of Washington, 1986.

[4] A.J. King and R.T. Rockafellar, Asymptotic theory for solutions in statistical estimation and stochastic programming, *Math. Oper. Res.* 18 (1993) 148–162.

[5] C.K. Liew, Inequality constrained least squares estimations, *J. Amer. Stat. Asso.* 71 (1976) 746–751.

[6] A. Shapiro, Asymptotic properties of statistical estimators in stochastic programming, *Ann. Statist.* 17 (1988) 841–858.

[7] A. Shapiro, Towards a unified theory of inequality constrained testing in multivariate analysis, *Intern. Statist. Rev.* 56 (1988) 49–62.

[8] A. Shapiro, Asymptotic behavior of optimal solutions in stochastic programming, *Math. Oper. Res.* 18 (1993) 829–845.

[9] T. Robertson, F. Wright and R. Dykstra, *Order Restricted Statistical Inference*, John Wiley, New York, 1988.

[10] R.J-B. Wets, Constrained estimation: consistency and asymptotics, *Appl. Stoch. Models and Data Anal.* 7 (1991) 17–32.

[11] J. Zhao, Approximate distribution of statistical estimators of stochastic programming, *J. of Nanjing Univ., Math. Biquaterly* 14 (1997) 154–160.

Manuscript received 9 July 2004
revised 27 October 2004
accepted for publication 29 October 2004

Jinde Wang
Department of Mathematics, Nanjing University, Nanjing, 210093, P.R. China
E-mail address: wangjd@nju.edu.cn