DISCUSSION PAPER NO. 218

OPTIMAL SAMPLING AND THE RIGHT-HAND-SIDE OF A LINEAR PROGRAM: SENSITIVITY ANALYSIS REVISITED*

by

Andrew F. Daughety

March 1976

GRADUATE SCHOOL OF MANAGEMENT
NORTHWESTERN UNIVERSITY

*The author expresses thanks to V. Balachandran, D.P. Baron, S. Deshmukh, S. Meyer and D.J. Roberts, who are in no way responsible for remaining errors.

Optimal Sampling and the Right-Hand-Side of a Linear Program: Sensitivity Analysis Revistied

ABSTRACT

This paper examines right-hand-side sensitivity analysis in linear programming as a problem in optimal sampling. Specifically, the insensitivity of a solution is defined as the point at which the expected gain from increased accuracy in the prediction of a resource level is equal to the expected cost of procurring the information. The problem is structured using the rudiments of optimal statistical decision theory.

Introduction

This paper is concerned with sensitivity analysis in linear programs. Usually, when we think of a solution's sensitivity, we find (for example) the range that the right-hand-side could change before a change in basis was required. The implicit motivation is that if the right-hand-side changed and the range of values over which the optimal basis was applicable was in some sense "small," then we would be operating with the wrong solution. In this paper we will attempt to formalize this notion.

We will argue that the real problem is to find the value for the right-hand-side that best trades-off potential losses (for using an incorrect value) and the cost of obtaining the improved solution. We will limit ourselves to changing only one element of the right-hand-side which will be a sample mean instead of the unknown, true value.

Two problems will be discussed. The first assumes some data has been used to form an estimate and we formulate the function to be minimized that indicates how much more data should be gathered so as to "desensitize" the optimal solution. We then will examine the problem of finding the overall sample size in order to produce an insensitive solution.

The Basic Problem

Consider the following linear program:

$$\begin{aligned} (LP): \quad & \max \quad px \\ & S.T. \quad Ax \leq b \\ & \qquad x \geq 0 \end{aligned}$$

where p and x are k-vectors, b is a non-negative m-vector and A is a m x k matrix. We will assume that p and A are perfectly known, as are all elements of b except for b_i, which will have to be estimated. Finally, we assume that n data points are used to construct an unbiased estimate of b_i, which we shall symbolize as $\hat{b}_i(n)$. While we have in mind the sample mean as the estimator, it will be seen that other estimators (such as regression coefficients) could also be used. However, for simplicity we take

$$\hat{b}_i(n) \sim N(b_i, \frac{\sigma^2}{n}) = \frac{\sqrt{n}}{\sqrt{2\pi}\sigma} \exp\left(\frac{-n}{2} \left(\frac{\hat{b}_i(n) - b_i}{\sigma}\right)^2\right)$$

where b_i is the true value of the amount of the i^{th} resource on hand. Let the optimal solution to (LP) when $\hat{b}(n)$ is used for b_i be $\hat{x}$ with associated basis matrix $\hat{B}$. Further let the optimal solution to (LP) when the true value of the i^{th} resouce (namely b_i) is used be x* with associated basis matrix B*.

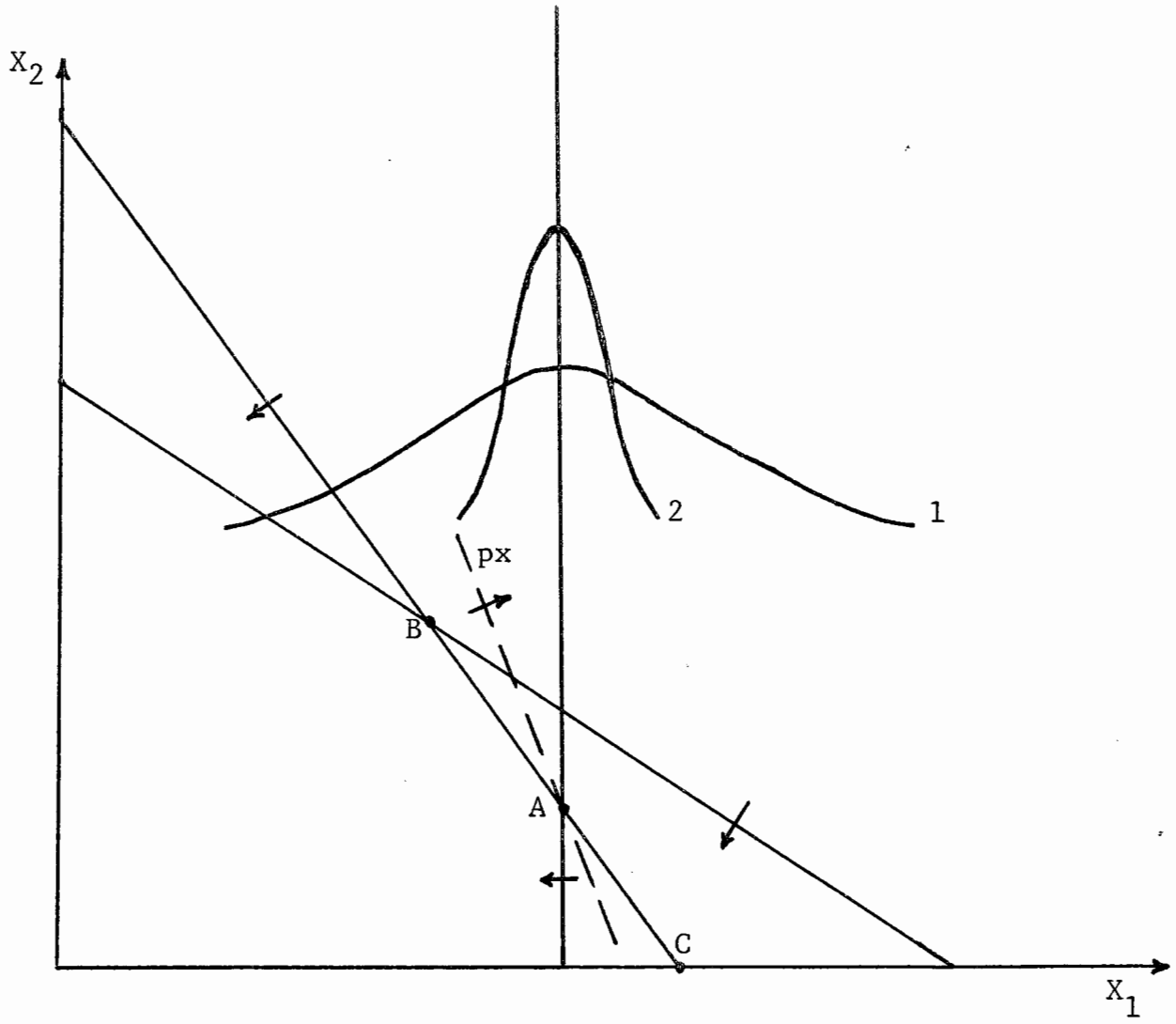
In general, our intuitive feeling is that a solution is sensitive to the right-hand-side value if "small" changes in the

value might require a change of basis. We will assume that the changes in the value of $\hat{b}_i(n)$ came about from changes in the data used to estimate b_i as n is changed.

The following diagram illustrates the problem to be discussed. The solid lines represent constraints while the dashed line represents the objective function. Two bell-shaped curves are drawn, centered about the true value of the right-hand-side of the third constraint. If this value were known then the optimal solution would occur at point A. Let the curve labelled '1' represent the density function of the sample mean given that n data points have been used to form the sample. For this right-hand-side the solution is sensitive in the sense that the sample mean might easily place the constraint considerably to the left or right of the true location, e.g. it might end up at B or C.

There are basically two approaches to the problem. One method would be to essentially chance-constrain the program. This would include increasing the sample size until the estimated variance of the sample mean were in some sense "small," i.e. the probability of a sample mean occurring outside of some interval about the true value would be less than or equal to α. This is represented in Figure 1 by the curve labelled '2'. This requires the setting of α at some arbitrary level and thus really does not clarify the sensitivity problem at all. Now the solution is sensitive to α, which was arbitrary.

A second approach is available if the cost of gathering data is expressible in the same units as the objective function,


X2
2
1
px
B
A
C
X1

e.g. dollars. We will examine this approach in detail in the rest of the paper. We will assume that a data cost function can be obtained and that it faithfully reflects the costs of acquiring, processing, storing and providing sample mean estimates. We assume that costs are a function of the amount of data only and thus c(n) is the cost for providing $\hat{b}_i(n)$.

In what follows we will develop loss functions and state the sensitivity problem as finding the sample size that minimizes expected loss plus cost. Thus, rather than setting an arbitrary parameter, we will consider a solution to be insensitive when the marginal expected gain from extra information is just balanced by its marginal cost.

Constructing a Loss Function

We will consider two loss functions that could be constructed for the problem at hand. Both will be piecewise linear functions of the simplex multipliers and the resource vector.

Consider a decision maker who wishes to solve (LP) with $\hat{b}(n)$ as the right-hand-side. He may take one of the following approaches:

1) He may choose to produce the activities associated with $\hat{B}$ on the assumption that $\hat{b}(n)$ is available. In this case he will either not use $b-\hat{b}(n)$ resources or he will run out early (when $b < \hat{b}(n)$). At this point he would re-evaluate his optimal strategy.

2) He may choose to produce the activities associated with $\hat{B}$ with whatever is available. In this case he never runs out or has a waste. However, as before, if $\hat{b}(n) \neq b$ then he may use the basis $\hat{B}$ for a level of resources which is not optimal.

In the first case an asymmetry will exist in terms of whether the resource has been overestimated or underestimated.

Let us consider the first case. Let b be the <u>true</u> vector of resource levels with i^{th} element b_i and $\hat{b}$ be the vector with $\hat{b}_j = b_i$, $j \neq i$ and $\hat{b}_i = \hat{b}_i(n)$. We define the loss associated with using $\hat{b}$ instead of b as $L_1(b,\hat{b})$ and will take it to be:

$$L_1(b,\hat{b}) = \begin{cases} \{\max px | Ax \leq b,\ x \geq 0\} \\ \qquad - \{\max px | Ax \leq \hat{b},\ x \geq 0\} & \text{if } b \geq \hat{b} \\ \{\max px | Ax \leq b,\ x \geq 0\} \\ \qquad - \{\max px | Ax \leq b,\ x_{\hat{B}} \geq 0,\ x_{\hat{NB}} = 0\} & \text{if } b \leq \hat{b} \end{cases}$$

where $(x_{\hat{B}} \geq 0,\ x_{\hat{NB}} = 0)$ means that only those vectors represented by $\hat{B}$ are free to be non-negative. The loss function indicates that when $b \geq \hat{b}$ the loss will be taken to be the difference between the profit that would have been obtained if it were known that the resource were actually b and the profit that will be obtained by using the program associated with $\hat{B}$ and resource level $\hat{b}$. The left-over resources will be considered waste and will be charged to the solution as loss. If $b \leq \hat{b}$ then we have overestimated the amount of resource. The loss that will be incurred is the difference in profit between what would have been obtained if the resource had been properly estimated, and what will be obtained given that we will use the basis matrix $\hat{B}$ but have only b in actual resources. Thus there will be a shortfall in production. The tacit assumption is that no

adjustment is made by the operator once the error in estimation is perceived. They use the prescribed program (i.e. they pursue the basic activities) and either run out of resources early or have some left over.

Let $\pi_{\cdot}$ be the vector of simplex multipliers associated with the optimal feasible basis $(\cdot)$. Then the loss function can be written as:

$$L_1(b,\hat{b}) = \begin{cases} \pi_{B*}b - \pi_{\hat{B}}\hat{b} & b \geq \hat{b} \\ \pi_{B*}b - \pi_{\hat{B}}b & b \leq \hat{b} \end{cases}$$

This may be further rewritten as

$$(1) \qquad L_1(b,\hat{b}) = \begin{cases} (\pi_{B*} - \pi_{\hat{B}})b + (b-\hat{b})\pi_{\hat{B}} & b \geq \hat{b} \\ (\pi_{B*} - \pi_{\hat{B}})b & b \leq \hat{b} \end{cases}$$

Thus, the loss function can be viewed as being composed of the loss due to using the wrong basis plus the loss due to overestimation of resource availability.

From the above discussion it should be clear that the second loss function is

$$L_2(b,\hat{b}) = (\Pi_{B*} - \Pi_{\hat{B}})b$$

Thus, the second loss function only penalizes for using the wrong basic solution. In the following we will assume that the decision maker has chosen an approach and thus a loss function and will therefore delete the subscript on the function.

We of course do not know b. We therefore assume that the decision maker or analyst has some prior information about b_i, namely that $b_i \sim N(\bar{b},\bar{\sigma}^2)$. Thus, the posterior density function on b_i given $\hat{b}_i$ is:

$$(2) \qquad f(b_i|\hat{b}_i(n)) = N\left((n\hat{b}_i(n)\bar{\sigma}^2 + \bar{b}\sigma^2)/(n\bar{\sigma}^2 + \sigma^2),\ \bar{\sigma}^2\sigma^2/(n\bar{\sigma}^2 + \sigma^2)\right)$$

With the above in mind we are in a position to make the general statement in the introduction more concrete. We will consider a solution to (LP) to be insensitive relative to $\hat{b}_i(n+s)$ if the expected loss associated with $\hat{b}_i(n+s)$ plus cost of sampling s data points c(s) is minimal. Thus, we wish to choose s such that we minimize loss plus cost, i.e.

$$(3) \qquad \min_s \int L(b,\hat{b}(n+s))\ f(b_i|\hat{b}_i(n+s))db + c(s).$$

The solution to (3), s*, would be the amount of data to add so as to make the optimal solution to (LP) insensitive. However, in that we do not know how $\hat{b}_i(n+s)$ will change with s, we must use our best estimate of the mean, namely the mean of the posterior $E(b|\hat{b}(n))$. Thus we will find s* so as to minimize

$$(4) \qquad \int L(b,E(b|\hat{b}(n)))\ f_s(b_i|\hat{b}_i(n))db + c(s)$$

where

$$f_s(b_i|\hat{b}_i(n)) = N\left(E(b_i|\hat{b}_i(n)),\ \bar{\sigma}^2\sigma^2/((n+s)\bar{\sigma}^2 + \sigma^2)\right)$$

Minimizing (4) is a heuristic approach to solving (3). In (4) we have used the best estimate of the mean and are only altering

the variance of the posterior density function. Using a parametric programming routine on the linear program optimization package will yield the simplix multipliers for all values of b_i. Thus, as is typical in these formulations (see [1]), one would compute the value of (4) for different values of s and select that s that yields at least a local minimum.

A More General Problem

The above procedure raises a significant point, namely the fact that it assumes that n data points have already been sampled and used to form $\hat{b}_i(n)$. We would in fact like to find the total sample size in a one-shot approach, thereby avoiding the possibility of having oversampled to begin with.

We will assume here that both parameters of the independent normal process (i.e. b_i and σ_i^2) are unknown and must be treated as random variables. A sample of n data points, $(b_{i1},\ldots,b_{in})$ will be taken from which the following statistics could be computed:

$$m \equiv \frac{1}{n} \sum_j b_{ij}$$

$$v \equiv \frac{1}{n-1} \sum_j (b_{ij} - m)^2$$

$$\nu \equiv n-1$$

Thus, we wish to find n that again minimizes expected loss plus cost c(n). To formulate the expected loss we assume that the prior distribution of $(b_i, 1/\sigma_i^2)$ is Normal-gamma (see the Appendix) where $h_i \equiv 1/\sigma_i^2$ and the parameter of the prior is (m',v',n',ν'). Thus one might use previous estimates of b_i to form (m',v',n',ν') or use $m' = \bar{b}$, $v' = \bar{\sigma}^2$. As mentioned in the Appendix, the posterior distribution will also be Normal-gamma with parameters (m'',n'',v'',ν''):

$$m'' = (n'm' + nm)/(m' + n)$$

$$n'' = n'' + n$$

$$\nu'' = \nu' + \nu + \delta(n') + \delta(n) - \delta(n'')$$

$$v'' = ([\nu'v + n'(m')^2] + [\nu v + nm^2] - n''(m'')^2)/\nu''$$

$$\delta(x) = \begin{cases} 0 & \text{if } x = 0 \\ 1 & \text{if } x > 0 \end{cases}$$

The unprimed parameters reflect the sampling distribution values. Since we are most interested in the posterior associated with b, we will use the marginal of $f_{N\gamma}(b,h|m'',v'',n'',\nu'')$ which (see Appendix) is the Student's distribution $g(b|m'',v'',n'',\nu'')$. Therefore the expected loss is

$$E_L(n) = \int\int\int L(b,m'')g(b|m'',v'',n'',\nu'')D(m,v|m',v',n',\nu';n,\nu)dmdvdb$$

where $D(\cdot)$ is the sampling distribution (see Appendix) and the updating equalities above are used. Finally, this means that the optimal sample size n* is the number of data points that solves:

$$\min_n E_L(n) + c(n)$$

Evaluating $E_L(n)$, while tedious, will not be difficult especially since $L(b,m'')$ is piecewise-linear.

Other Issues

First, it should be obvious that a similar approach can be taken, via duality, for changes in one element of the price vector p. More important, however, is the problem of restricting the analysis to one element of the parameter set, be it b or p. One would normally expect to see a sensitivity discussion which considered changes in the entire vector, not just one element.

A moments reflection will bring to light why this has not been pursued. In standard sensitivity analysis the right-hand-side is parametized on a scaler and ranging is performed by varying the scaler. This amounts to requiring the right-hand-side to expand along a ray, which does not make sense in the present application. On the other hand, while the loss function for such a procedure is well understood and easy to compute for various values (this being a direct extension of the material in the previous sections), the loss function to be used for the problem at hand (where b would not expand along a ray) is not known. In fact, we would not have the benefit of not having to solve the program for each $\hat{b}$, but would essentially have to solve an infinite number of programs in order to find the optimal $\hat{b}$.

The obvious heuristic is to vary one element at a time. This is clearly not optimal. It should be clear that each element of the right-hand-side would be a function of all other elements and that, for example, the "optimal" $\hat{b}$ would be sequence dependent, i.e. would change depending on the sequence of $\hat{b}_i$ examined.

Summary

This paper is an attempt to apply basic decision theory to sensitivity questions in linear programming. We have taken the position that a solution is sensitive only if the expected gain from improving the solution exceeds the cost of improvement. This would seem to be an obvious and intuitive approach, but it does not seem to be the approach commonly used in sensitivity analysis.

While closed form solutions do not appear possible, locally optimal solutions are clearly possible if the user has the facility to do basic sensitivity analysis and has access to (or will acquire) a computer routine for numerical integration.

Appendix

The following is shown in [1]:

1. If h is distributed Gamma-2 then we mean that $f_{\gamma 2}(h|v,\nu) =$

$$e^{-v\nu h/2}\ (\tfrac{1}{2}\ v\nu h)^{\frac{\nu}{2}-1}\qquad \frac{v\nu/2}{(\frac{\nu}{2}-1)!}$$

Thus, if (μ,h) is distributed Normal-gamma then

$f_{N\gamma}(\mu,h|m,v,n,\nu) = f_N(\mu|m,hn)f_{\gamma 2}(h|v,\nu)$.

2. If the prior distribution of (μ,h) is Normal-gamma with parameter (m',n',v',ν') and if a sample then yields a sufficient statistic (m,v,n,ν), the posterior distribution of (μ,h) is Normal-gamma with parameters (m'',n'',v'',ν''):

$$m'' = (n'm' + nm)/(n'+n)$$
$$n'' = n' + n$$
$$\nu'' = \nu' + \nu + \delta(n') + \delta(n) - \delta(n'')$$
$$v'' = ([\nu'v' + n'(m')^2] + [\nu v + nm^2] - n''(m'')^2)/\nu''$$

where

$$\delta(x) = \begin{cases} 0 & \text{if } x = 0 \\ 1 & \text{if } x > 0 \end{cases}$$

3. If the joint distribution of (μ,h) is Normal-gamma as defined above then the marginal distribution of μ is the Student distribution:

$$g(\mu|m,v,n,\nu) = \frac{\nu^{\nu/2}\,\Gamma(\frac{n}{2})}{\Gamma(\frac{1}{2})\Gamma(\frac{1}{2}\nu)} \left[\nu + \frac{n}{v}(\mu-m)^2\right]^{-n/2} \sqrt{\frac{n}{v}}$$

4. If a sample of size $n > 0$ is to be taken from an independent Normal process with parameter (μ,h) having a Normal-gamma distribution with parameter (m',v',n',ν'), with $(v',n',\nu' > 0)$, then the unconditional joint distribution of the statistic (m,v) is

$$D(m,v|m',v',n';n,\nu) =$$

$$= A(n)\,\frac{(\nu v)^{\nu/2\ -1}}{(\nu'v' + \nu v + n_u[m-m']^2)^{\nu''/2}}$$

where

$$n_u = n'n/(n+n')$$

$$\nu'' = \nu' + \nu + 1$$

$$A(n) = \frac{\Gamma(\nu''/2)}{\Gamma(\frac{n-1}{2})} \sqrt{\frac{n}{n+n'}\left(\frac{n-1}{2}\right)^{n-1}} \cdot \sqrt{\frac{n'}{2\pi}} \cdot \frac{\left(\frac{v'\nu'}{2}\right)^{\nu'/2}}{\Gamma(\nu'/2)}$$

$\Gamma(x)$ = gamma function of x.

References

1. Raiffa, H. and R. Schlaiffer, Applied Statistical Decision Theory, Cambridge, MIT Press, 1961.