RUTCOR
RESEARCH
REPORT

A Data Mining Problem in Stochastic Programming

András Prékopa [a] Xiaoling Hou [b]

RRR 5-2003, February, 2003

RUTCOR
Rutgers Center for
Operations Research
Rutgers University
640 Bartholomew Road
Piscataway, New Jersey
08854-8003
Telephone: 732-445-3804
Telefax: 732-445-5472
Email: rrr@rutcor.rutgers.edu
http://rutcor.rutgers.edu/~rrr

[a]RUTCOR, Rutgers University, Piscataway, NJ, 08854, Email: prekopa@rutcor.rutgers.edu
[b]RUTCOR, Rutgers University, Piscataway, NJ, 08854, Email: xhou@rutcor.rutgers.edu

Rutcor Research Report
RRR 5-2003, February, 2003

A Data Mining Problem in Stochastic Programming

András Prékopa Xiaoling Hou

Abstract. In this paper we consider a linear programming problem where some or all technology coefficients are deterministic but their values are unknown. Samples are taken to estimate these coefficients and the problem is to determine the optimal sample sizes. If we replace the unknown coefficients by their estimations, then we obtain a random linear programming problem the optimum value of which is also random. We want to find sample sizes such that the confidence interval, created for the unknown deterministic optimum value, by the use of the samples, should cover it by a prescribed large probability, and, subject to this constraint, the total cost of sampling should be minimum.

Acknowledgements: The partial support of DIMACS of the second author is gratefully acknowledged

1 Preliminaries

Consider the following random linear programming problem:

$$
\begin{array}{llll}
\max\{c_1x_1 + c_2x_2 + \cdots + c_nx_n\} & & & \\
\text{subject to} & & & \\
a_{11}x_1 + a_{12}x_2 + & \cdots & +a_{1n}x_n & \leq b_1 \\
a_{21}x_1 + a_{22}x_2 + & \cdots & +a_{2n}x_n & \leq b_2 \\
 & \vdots & & \\
a_{m1}x_1 + a_{m2}x_2 + & \cdots & +a_{mn}x_n & \leq b_m \\
x_1, x_2, \cdots, x_n \geq 0, & & &
\end{array}
\tag{1}
$$

where the a_{ik}, $i = 1, 2, \cdots, m$, $k = 1, 2, \cdots, n$ are independent random variables, and $c_k > 0$, $k = 1, 2, \cdots, n$, $b_i > 0$, $i = 1, 2, \cdots, m$ are constants.

We may assume, without loss of generality, that $c_k = 1$, $k = 1, 2, \cdots, n$, $b_i = 1$, $i = 1, 2, \cdots, m$, because we can divide the ith constraint by b_i, $i = 1, 2, \cdots, m$, and introduce new variables, replacing c_kx_k by x_k, $k = 1, 2, \cdots, n$. If the optimum value of problem (1) is finite with probability 1, then it is also positive with probability 1.

Assuming the existence of the expectations $a_{ik}^{(0)} = E[a_{ik}]$, together with the random linear programming problem (1) we consider the deterministic problem:

$$
\begin{array}{llll}
\max\{c_1x_1 + c_2x_2 + \cdots + c_nx_n\} & & & \\
\text{subject to} & & & \\
a_{11}^{(0)}x_1 + a_{12}^{(0)}x_2 + & \cdots & +a_{1n}^{(0)}x_n & \leq b_1 \\
a_{21}^{(0)}x_1 + a_{22}^{(0)}x_2 + & \cdots & +a_{2n}^{(0)}x_n & \leq b_2 \\
 & \vdots & & \\
a_{m1}^{(0)}x_1 + a_{m2}^{(0)}x_2 + & \cdots & +a_{mn}^{(0)}x_n & \leq b_m \\
x_1, x_2, \cdots, x_n \geq 0, & & &
\end{array}
\tag{2}
$$

where $c_k = 1$, $b_i = 1$, $a_{ik}^{(0)} = E[a_{ik}]$, $i = 1, 2, \cdots, m$, $k = 1, 2, \cdots, n$. If the optimum value of problem (2) is finite, then it is also positive.

It is known (see Kuhn and Quandt, 1963, Prékopa, 1972, 1995, Kabe, 1983, Cohen and Newman, 1989) that if $m, n \to \infty$, subject to some conditions, then the difference between the random optimum value μ of problem (1) and the optimum value $\mu^{(0)}$, that corresponds to the expectations, goes to 0 (in probability or with probability 1, depending on our conditions). This fact is due solely to the increase of the problem sizes m and n, we do not assume that the "order of magnitude" of the random variables a_{ik}, $i = 1, 2, \cdots, m$, $k = 1, 2, \cdots, n$ change while $m, n \to \infty$.

Some of the theorems in the above cited literature state convergence in probability, others convergence with probability 1. We recall only one of the theorems.

Lemma 1.1 (Prékopa, 1972). *Suppose that for every m, n, the random variables ξ_{ik}, $i = 1, 2, \cdots, m$, $k = 1, 2, \cdots, n$ are independent, have 0 expectations, their fourth moments exist and are uniformly bounded. Let $m, n \to \infty$ in such a way that the following condition is satisfied*

$$0 < \alpha \leq \frac{m}{n} \leq \beta < \infty,$$

where α and β are constants. Under these conditions we have ($\Rightarrow$ means convergence in probability):

$$\max_{1 \leq i \leq m} \frac{\xi_{i1} + \xi_{i2} + \cdots + \xi_{in}}{n} \Rightarrow 0,$$

$$\max_{1 \leq k \leq n} \frac{\xi_{1k} + \xi_{2k} + \cdots + \xi_{mk}}{m} \Rightarrow 0.$$

In what follows we will use the notation $\xi_{ik} = a_{ik} - a_{ik}^{(0)}$, $i = 1, 2, \cdots, m$, $k = 1, 2, \cdots, n$.

Theorem 1.1 (Prékopa, 1972). *Suppose that in connection with problem (1) the following conditions are satisfied:*

(a) There exist positive integers m_0, n_0 such that for every $m \geq m_0, n \geq n_0$ the random linear programming problem (1) has a finite optimum value μ with probability 1; also, problem (2) has a finite optimum value $\mu^{(0)}$ and $\mu^{(0)} \leq \delta$, where δ does not depend on m or n.

(b) The random variables ξ_{ik}, $i = 1, 2, \cdots, m$, $k = 1, 2, \cdots, n$ satisfy the conditions of Lemma 1.1.

(c) For every $m \geq m_0, n \geq n_0$ problem (2) and its dual have an optimal solution pair $x^{(0)} = (x_1^{(0)}, x_2^{(0)}, \cdots, x_n^{(0)})^T$, $y^{(0)} = (y_1^{(0)}, y_2^{(0)}, \cdots, y_m^{(0)})^T$ such that

$$\frac{n x_k^{(0)}}{x_1^{(0)} + x_2^{(0)} + \cdots + x_n^{(0)}} = \frac{n x_k^{(0)}}{\mu^{(0)}} \leq L_1, i = 1, 2, \cdots, m,$$

$$\frac{m y_i^{(0)}}{y_1^{(0)} + y_2^{(0)} + \cdots + y_m^{(0)}} = \frac{m y_i^{(0)}}{\mu^{(0)}} \leq L_2, k = 1, 2, \cdots, n,$$

where L_1, L_2 are constants (do not depend on m or n).

Then $\mu - \mu^{(0)} \Rightarrow 0$, when $m, n \to \infty$.

2 Main Results

The problem that we consider in this paper has deterministic but unknown technology coefficients. In order to obtain information, we take sample for each of them and estimate each unknown coefficient by a sampling mean. We assume that the samples are taken independently of each other. Let μ and $\mu^{(0)}$ designate the optimum values corresponding to the random and deterministic linear programming problems with coefficients a_{ik} and $a_{ik}^{(0)}$, respectively, where a_{ik} is the sampling mean used to estimate $a_{ik}^{(0)}$, $i = 1, 2, \cdots, m$, $k = 1, 2, \cdots, n$. We want to determine the sample sizes based on some principle. Our principle is that the sample sizes N_{ik} $i=1,2,\cdots,m$, $k=1,2,\cdots,n$ should be large enough such that for given small positive real numbers ε_1, ε_2, δ_1 and δ_2, we have

$$P(\mu^{(0)}-\mu\le\varepsilon_1\mu)\ge 1-\varepsilon_2$$

$$P(\mu^{(0)}-\mu\ge-\delta_1\mu)\ge 1-\delta_2$$

and subject to these constraints the total cost of the sampling should be minimum.

The pair of inequalities $(1-\delta_1)\mu\le\mu^{(0)}$, $\mu^{(0)}\le(1+\varepsilon_1)\mu$ gives rise to the confidence interval $[(1-\delta_1)\mu,(1+\varepsilon_1)\mu]$, where, however there are two confidence levels, $1-\varepsilon_2$, $1-\delta_2$, corresponding to the two inequalities, respectively. A lower bound for the joint probability of the two inequalities can be given as:

$$\begin{aligned} & P((1-\delta_1)\mu\le\mu^{(0)}\le(1+\varepsilon_1)\mu) \\ = \; & P(\mu^{(0)}-\mu\le\varepsilon_1\mu,\mu^{(0)}-\mu\ge-\delta_1\mu) \\ \ge \; & 1-\varepsilon_2-\delta_2. \end{aligned} \tag{3}$$

We take the cost of a sample proportional to the sample size and designate by C_{ik} the price to obtain elements in the sample taken to estimate $a_{ik}^{(0)}$, $i=1,2,\cdots,m$, $k=1,2,\cdots,n$. Let $\alpha_{ik}^{(j)}$, $j=1,2,\cdots,N_{ik}$ be the sample to estimate $a_{ik}^{(0)}$ and $a_{ik}=\frac{1}{N_{ik}}\sum_{j=1}^{N_{ik}}\alpha_{ik}^{(j)}$, $i=1,2,\cdots,m$, $k=1,2,\cdots,n$.

Our sample size optimization problem can be formulated in the following way:

$$\begin{aligned} & \min \textstyle\sum_{i=1}^{m}\sum_{k=1}^{n}C_{ik}N_{ik} \\ & \text{subject to} \\ & P(\mu^{(0)}-\mu\le\varepsilon_1\mu)\ge 1-\varepsilon_2 \\ & P(\mu^{(0)}-\mu\ge-\delta_1\mu)\ge 1-\delta_2 \\ & N_{ik}\ge 1, \quad i=1,2,\cdots,m, \quad k=1,2,\cdots,n. \end{aligned} \tag{4}$$

Assume that all objective function coefficients are positive.

The next step is to replace the constraints in (4) by mathematically more tractable ones. In order to do that we need the following:

Theorem 2.1. *Suppose that all variances $Var(\alpha_{ik}^{(j)})$ exist and let $Var(\alpha_{ik}^{(j)})=\sigma_{ik}^2$ for each j, $j=1,2,\cdots,N_{ik}$, $i=1,2,\cdots,m$, $k=1,2,\cdots,n$. Suppose that all conditions of Theorem 1.1 hold and ξ_{ik} $i=1,2,\cdots,m$, $k=1,2,\cdots,n$ have normal distribution. If for the N_{ik}, $i=1,2,\cdots,m$, $k=1,2,\cdots,n$, we have*

$$\begin{aligned} & \Pi_{i=1}^{m}\Phi\Big(\frac{\varepsilon_1}{\sqrt{\sum_{k=1}^{n}\frac{\sigma_{ik}^2(x_k^{(0)})^2}{N_{ik}}}}\Big)\ge 1-\varepsilon_2 \\ & \Pi_{k=1}^{n}\Phi\Big(\frac{\delta_1}{\sqrt{\sum_{i=1}^{m}\frac{\sigma_{ik}^2(y_i^{(0)})^2}{N_{ik}}}}\Big)\ge 1-\delta_2 \\ & N_{ik}\ge 1, \quad i=1,2,\cdots,m, \quad k=1,2,\cdots,n, \end{aligned}$$

then the constraints in (4) are satisfied with the same sample sizes.

Proof. We have the relations (see Prékopa, 1972,1995):

$$\min_{1\le k\le n}\frac{1}{m}\sum_{i=1}^{m}\xi_{ik}\frac{my_i^{(0)}}{\mu^{(0)}}\le\frac{1}{\mu}-\frac{1}{\mu^{(0)}}\le\max_{1\le i\le m}\frac{1}{n}\sum_{k=1}^{n}\xi_{ik}\frac{nx_k^{(0)}}{\mu^{(0)}}.$$

This inequality can be written as

$$\min_{1\le k\le n}\sum_{i=1}^{m}\xi_{ik}y_i^{(0)}\le\frac{\mu^{(0)}}{\mu}-1\le\max_{1\le i\le m}\sum_{k=1}^{n}\xi_{ik}x_k^{(0)}.$$

Since $a_{ik}=a_{ik}^{(0)}+\xi_{ik}$ and

$$a_{ik}=\frac{1}{N_{ik}}\sum_{j=1}^{N_{ik}}(\alpha_{ik}^{(j)}-a_{ik}^{(0)})+a_{ik}^{(0)},$$

it follows that

$$\xi_{ik}=\frac{1}{N_{ik}}\sum_{j=1}^{N_{ik}}(\alpha_{ik}^{(j)}-a_{ik}^{(0)}).$$

We have the relations

$$E[\xi_{ik}]=0$$

$$Var(\xi_{ik})=\frac{\sigma_{ik}^2}{N_{ik}},\quad i=1,2,\cdots,m,\quad k=1,2,\cdots,n.$$

Since ξ_{ik} $i=1,2,\cdots,m,\ k=1,2,\cdots,n$ have normal distributions with expectation 0, it follows that $\sum_{k=1}^{n}\xi_{ik}x_k^{(0)}$ and $\sum_{i=1}^{m}(-\xi_{ik}y_i^{(0)})$ also have normal distributions with expectation 0 and

$$Var(\sum_{k=1}^{n}\xi_{ik}x_k^{(0)})=\sum_{k=1}^{n}\frac{\sigma_{ik}^2(x_k^{(0)})^2}{N_{ik}},$$

$$Var(\sum_{i=1}^{m}(-\xi_{ik}y_i^{(0)}))=\sum_{i=1}^{m}\frac{\sigma_{ik}^2(y_i^{(0)})^2}{N_{ik}}.$$

Also, using the independence of the ξ_{ik}, $i = 1, 2, \cdots, m$, $k = 1, 2, \cdots, n$ we obtain

$$\begin{aligned}
P(\max_{1\leq i\leq m}\sum_{k=1}^{n}\xi_{ik}x_k^{(0)}\leq\varepsilon_1) &= P(\sum_{k=1}^{n}\xi_{ik}x_k^{(0)}\leq\varepsilon_1,\quad i=1,2,\cdots,m)\\
&= \Pi_{i=1}^{m}P(\sum_{k=1}^{n}\xi_{ik}x_k^{(0)}\leq\varepsilon_1)\\
&= \Pi_{i=1}^{m}\Phi(\frac{\varepsilon_1}{\sqrt{\sum_{k=1}^{n}\frac{\sigma_{ik}^2(x_k^{(0)})^2}{N_{ik}}}})\\
P(\min_{1\leq k\leq n}\sum_{i=1}^{m}\xi_{ik}y_i^{(0)}\geq-\delta_1) &= P(\sum_{i=1}^{m}\xi_{ik}y_i^{(0)}\geq-\delta_1,\quad k=1,2,\cdots,n)\\
&= \Pi_{k=1}^{n}P(\sum_{i=1}^{m}\xi_{ik}y_i^{(0)}\geq-\delta_1)\\
&= \Pi_{k=1}^{n}P(\sum_{i=1}^{m}(-\xi_{ik}y_i^{(0)})\leq\delta_1)\\
&= \Pi_{k=1}^{n}\Phi(\frac{\delta_1}{\sqrt{\sum_{i=1}^{m}\frac{\sigma_{ik}^2(y_i^{(0)})^2}{N_{ik}}}}).
\end{aligned}$$

Thus, if

$$\Pi_{i=1}^{m}\Phi(\frac{\varepsilon_1}{\sqrt{\sum_{k=1}^{n}\frac{\sigma_{ik}^2(x_k^{(0)})^2}{N_{ik}}}})\geq 1-\varepsilon_2$$

$$\Pi_{k=1}^{n}\Phi(\frac{\delta_1}{\sqrt{\sum_{i=1}^{m}\frac{\sigma_{ik}^2(y_i^{(0)})^2}{N_{ik}}}})\geq 1-\delta_2$$

hold, then the confidence constraints in (4) are satisfied.

□

Based on Theorem 2.1, we formulate the new problem:

$$\begin{aligned}
&\min \sum_{i=1}^{m}\sum_{k=1}^{n}C_{ik}N_{ik}\\
&\text{subject to}\\
&\Pi_{i=1}^{m}\Phi(\frac{\varepsilon_1}{\sqrt{\sum_{k=1}^{n}\frac{\sigma_{ik}^2(x_k^{(0)})^2}{N_{ik}}}})\geq 1-\varepsilon_2\\
&\Pi_{k=1}^{n}\Phi(\frac{\delta_1}{\sqrt{\sum_{i=1}^{m}\frac{\sigma_{ik}^2(y_i^{(0)})^2}{N_{ik}}}})\geq 1-\delta_2\\
&N_{ik}\geq 1,\quad i=1,2,\cdots,m,\quad k=1,2,\cdots,n.
\end{aligned} \tag{5}$$

Problem (5) is a relaxed problem as compared to problem (4). Thus, the set of feasible solutions in the former one is somewhat smaller than that of the latter one. Hence the optimum value of problem (4) is smaller than or equal to the optimum value of problem (5).

We show that problem (5) is convex, i.e., the set of feasible solutions is a convex set (here we handle the N_{ik}, $i = 1, 2, \cdots, m$, $k = 1, 2, \cdots, n$ as continuous variables). We prove somewhat more: the constraining functions in the constraints of problem (5) are logconcave functions of the variables N_{ik}, $i = 1, 2, \cdots, m$, $k = 1, 2, \cdots, n$.

Theorem 2.2. *The functions* $\Pi_{i=1}^m \Phi(\frac{\varepsilon_1}{\sqrt{\sum_{k=1}^n \frac{\sigma_{ik}^2 (x_k^{(0)})^2}{N_{ik}}}})$ *and* $\Pi_{k=1}^n \Phi(\frac{\delta_1}{\sqrt{\sum_{i=1}^m \frac{\sigma_{ik}^2 (y_i^{(0)})^2}{N_{ik}}}})$ *are logconcave functions of* N_{ik}*,* $i = 1, 2, \cdots, m$*,* $k = 1, 2, \cdots, n$*.*

Proof. First we prove the concavity of the function

$$f(z_1, z_2, \cdots, z_n) = \frac{1}{\frac{1}{z_1} + \frac{1}{z_2} + \cdots + \frac{1}{z_n}} = \frac{z_1 z_2 \cdots z_n}{z_2 z_3 \cdots z_n + z_1 z_3 \cdots z_n + \cdots + z_1 z_2 \cdots z_{n-1}},$$

where $z_1 > 0$, $z_2 > 0$, $\cdots$, $z_n > 0$.

For simplicity, we introduce the notation $g = z_2 z_3 \cdots z_n + z_1 z_3 \cdots z_n + \cdots + z_1 z_2 \cdots z_{n-1}$. Since $\frac{\partial g}{\partial z_i} = \frac{g - z_1 \cdots z_{i-1} z_{i+1} \cdots z_n}{z_i}$, it follows that

$$\begin{aligned}
\frac{\partial f}{\partial z_1} &= \frac{(z_2 z_3 \cdots z_n) g - z_1 z_2 \cdots z_n \frac{g - z_2 z_3 \cdots z_n}{z_1}}{g^2} \\
&= (z_2 z_3 \cdots z_n)^2 g^{-2} \\
\frac{\partial^2 f}{\partial z_1^2} &= -2 (z_2 z_3 \cdots z_n)^2 g^{-3} \frac{g - z_2 z_3 \cdots z_n}{z_1} \\
\frac{\partial^2 f}{\partial z_1 \partial z_2} &= 2 z_2 (z_3 \cdots z_n)^2 g^{-2} - 2 (z_2 z_3 \cdots z_n)^2 g^{-3} \frac{g - z_1 z_3 \cdots z_n}{z_2} \\
&= 2 z_1 z_2 (z_3 \cdots z_n)^3 g^{-3} \\
&\vdots \\
\frac{\partial^2 f}{\partial z_1 \partial z_n} &= 2 z_1 z_n (z_2 z_3 \cdots z_{n-1})^3 g^{-3} \\
\frac{\partial^2 f}{\partial z_2 \partial z_1} &= 2 z_1 z_2 (z_3 \cdots z_n)^3 g^{-3} \\
\frac{\partial^2 f}{\partial z_2^2} &= -2 (z_1 z_3 \cdots z_n)^2 g^{-3} \frac{g - z_1 z_3 \cdots z_n}{z_2} \\
&\vdots
\end{aligned}$$

$$\begin{aligned}
\frac{\partial^2 f}{\partial z_2 \partial z_n} &= 2z_2 z_n (z_1 z_3 \cdots z_{n-1})^3 g^{-3} \\
&\vdots \\
\frac{\partial^2 f}{\partial z_n^2} &= -2(z_1 z_2 \cdots z_{n-1})^2 g^{-3} \frac{g - z_1 z_2 \cdots z_{n-1}}{z_n}.
\end{aligned}$$

Let D_j, $j = 1, 2, \cdots, n$ designate the jth principal minor of the following negative Hessian matrix of the function f:

$$\begin{pmatrix}
-\frac{\partial^2 f}{\partial z_1^2} & -\frac{\partial^2 f}{\partial z_1 \partial z_2} & \cdots & -\frac{\partial^2 f}{\partial z_1 \partial z_n} \\
-\frac{\partial^2 f}{\partial z_2 \partial z_1} & -\frac{\partial^2 f}{\partial z_2^2} & \cdots & -\frac{\partial^2 f}{\partial z_2 \partial z_n} \\
\vdots & \vdots & \ddots & \vdots \\
-\frac{\partial^2 f}{\partial z_n \partial z_1} & -\frac{\partial^2 f}{\partial z_n \partial z_2} & \cdots & -\frac{\partial^2 f}{\partial z_n^2}
\end{pmatrix}.$$

Since

$$\begin{vmatrix}
c_1 & -1 & \cdots & -1 \\
-1 & c_2 & \cdots & -1 \\
\vdots & \vdots & \ddots & \vdots \\
-1 & -1 & \cdots & c_j
\end{vmatrix}$$

$$= \Pi_{i=1}^{j}(1 + c_i)(1 - \frac{1}{1 + c_1} - \frac{1}{1 + c_2} - \cdots - \frac{1}{1 + c_j}),$$

we obtain

$$\begin{aligned} D_j &= (2g^{-3})^j (z_2 z_3 \cdots z_n)^2 (z_1 z_3 \cdots z_n)^2 (z_1 \cdots z_{j-1} z_{j+1} \cdots z_n)^2 \\ &\quad \begin{vmatrix} \frac{g - z_2 z_3 \cdots z_n}{z_1} & -\frac{z_1 z_3 \cdots z_n}{z_2} & \cdots & -\frac{z_1 \cdots z_{j-1} z_{j+1} \cdots z_n}{z_j} \\ -\frac{z_2 z_3 \cdots z_n}{z_1} & \frac{g - z_1 z_3 \cdots z_n}{z_2} & \cdots & -\frac{z_1 \cdots z_{j-1} z_{j+1} \cdots z_n}{z_j} \\ \vdots & \vdots & \ddots & \vdots \\ -\frac{z_2 z_3 \cdots z_n}{z_1} & -\frac{z_1 z_3 \cdots z_n}{z_2} & \cdots & \frac{g - z_1 \cdots z_{j-1} z_{j+1} \cdots z_n}{z_j} \end{vmatrix} \\ &= (2g^{-3})^j (z_1 z_2 \cdots z_n)^{3j} (z_1 z_2 \cdots z_j)^{-4} \\ &\quad \begin{vmatrix} \frac{g - z_2 z_3 \cdots z_n}{z_2 z_3 \cdots z_n} & -1 & \cdots & -1 \\ -1 & \frac{g - z_1 z_3 \cdots z_n}{z_1 z_3 \cdots z_n} & \cdots & -1 \\ \vdots & \vdots & \ddots & \vdots \\ -1 & -1 & \cdots & \frac{g - z_1 \cdots z_{j-1} z_{j+1} \cdots z_n}{z_1 \cdots z_{j-1} z_{j+1} \cdots z_n} \end{vmatrix} \\ &= (2g^{-3})^j (z_1 z_2 \cdots z_n)^{3j} (z_1 z_2 \cdots z_j)^{-4} \frac{g^j z_1 z_2 \cdots z_j}{(z_1 z_2 \cdots z_n)^j} \\ &\quad \left(1 - \frac{z_2 z_3 \cdots z_n}{g} - \frac{z_1 z_3 \cdots z_n}{g} - \cdots - \frac{z_1 \cdots z_{j-1} z_{j+1} \cdots z_n}{g}\right) \\ &= 2^j (z_1 z_2 \cdots z_n)^{2j} (z_1 z_2 \cdots z_j)^{-3} \frac{z_1 \cdots z_j z_{j+2} \cdots z_n + \cdots + z_1 z_2 \cdots z_{n-1}}{g^{2j+1}} \\ &> 0, \end{aligned}$$

for every $j = 1, 2, \cdots, n$. This proves the concavity of function f.

Since ε_1, σ_{ik}^2, $x_k^{(0)}$ are constants and $\sqrt{.}$ is an increasing and concave function, it follows that

$$\frac{\varepsilon_1}{\sqrt{\sum_{k=1}^n \frac{\sigma_{ik}^2 (x_k^{(0)})^2}{N_{ik}}}}$$

is a concave function of N_{ik}, $i = 1, 2, \cdots, m$, $k = 1, 2, \cdots, n$. Since $\log \Phi(.)$ is increasing and concave (see, e.g., Prékopa, 1995), and the product of logconcave functions is logconcave, it follows that $\prod_{i=1}^m \Phi(\frac{\varepsilon_1}{\sqrt{\sum_{k=1}^n \frac{\sigma_{ik}^2 (x_k^{(0)})^2}{N_{ik}}}})$ is logconcave. The proof of the logconcavity of $\prod_{k=1}^n \Phi(\frac{\delta_1}{\sqrt{\sum_{i=1}^m \frac{\sigma_{ik}^2 (y_i^{(0)})^2}{N_{ik}}}})$ is the same.

□

In problem (5) we have the values $x_k^{(0)}$, $k = 1, 2, \cdots, n$, $y_i^{(0)}$, $i = 1, 2, \cdots, m$ which are unknown, therefore we cannot use the problem in this form in practice. However, using past experience or other information, we may know some upper bounds for $x_k^{(0)}$, $k = 1, 2, \cdots, n$,

$y_i^{(0)}$, $i = 1, 2, \cdots, m$. If

$$\begin{aligned} x_k^{(0)} &\leq X_k^{(0)}, k = 1, 2, \cdots, n \\ y_i^{(0)} &\leq Y_i^{(0)}, i = 1, 2, \cdots, m, \end{aligned}$$

where the right hand side values are known, then we formulate the following sample size deterministic problem:

$$\begin{array}{l} \min \sum_{i=1}^m \sum_{k=1}^n C_{ik} N_{ik} \\ \text{subject to} \\ \Pi_{i=1}^m \Phi\left(\frac{\varepsilon_1}{\sqrt{\sum_{k=1}^n \frac{\sigma_{ik}^2 (X_k^{(0)})^2}{N_{ik}}}}\right) \geq 1 - \varepsilon_2 \\ \Pi_{k=1}^n \Phi\left(\frac{\delta_1}{\sqrt{\sum_{i=1}^m \frac{\sigma_{ik}^2 (Y_i^{(0)})^2}{N_{ik}}}}\right) \geq 1 - \delta_2 \\ N_{ik} \geq 1, \quad i = 1, 2, \cdots, m, \quad k = 1, 2, \cdots, n. \end{array} \tag{6}$$

The set of feasible solutions of problem (6) is part of that of problem (5). Hence the optimum value of problem (6) is greater than or equal to that of problem (5). If the variances in problems (5) or (6) are unknown, then we can use upper bounds for them, too, to obtain a completely specified optimization problem. Using bounds on statistical parameters to determine sample sizes is a well-known technique in statistics.

Problems (5) and (6) are discrete variable problems because the N_{ik}, $i = 1, 2, \cdots, m$, $k = 1, 2, \cdots, n$ mean sample sizes. Since these problems are convex nonlinear programming problems if we handle the N_{ik}, $i = 1, 2, \cdots, m$, $k = 1, 2, \cdots, n$ as continuous variables, to solve the problem for integer variables we advise a two stage solution procedure: (1) solve the continuous variable problem and then (2) make a search for the integer optimal solution N^* around the continuous optimal solution N or take simply $N^* = \lceil N \rceil$ where $\lceil N \rceil = (\lceil N_{ik} \rceil)$ and $\lceil N_{ik} \rceil$ is the smallest integer not smaller than N_{ik}, $i = 1, 2, \cdots, m$, $k = 1, 2, \cdots, n$.

3 The Case of General Inequality Constraints

In this section we allow that some of the constraints in the LP (1) are of "$\le$" type while some others are of "$\ge$" type. Our random linear programming problem is:

$$
\begin{array}{lllll}
\max\{c_1x_1 + c_2x_2 + \cdots + c_nx_n\} & & & & \\
\text{subject to} & & & & \\
a_{11}x_1 + a_{12}x_2 + & \cdots & +a_{1n}x_n & \le & b_1 \\
 & \vdots & & & \\
a_{r1}x_1 + a_{r2}x_2 + & \cdots & +a_{rn}x_n & \le & b_r \\
a_{(r+1)1}x_1 + a_{(r+1)2}x_2 + & \cdots & +a_{(r+1)n}x_n & \ge & b_{r+1} \\
 & \vdots & & & \\
a_{m1}x_1 + a_{m2}x_2 + & \cdots & +a_{mn}x_n & \ge & b_m \\
x_1, x_2, \cdots, x_n \ge 0. & & & &
\end{array}
\tag{7}
$$

We assume, as before, that $c_k = 1$, $k = 1, 2, \cdots, n$, $b_i = 1$, $i = 1, 2, \cdots, m$.

Together with the random linear programming problem (7) we consider the deterministic problem, written up with the expectations:

$$
\begin{array}{lllll}
\max\{c_1x_1 + c_2x_2 + \cdots + c_nx_n\} & & & & \\
\text{subject to} & & & & \\
a_{11}^{(0)}x_1 + a_{12}^{(0)}x_2 + & \cdots & +a_{1n}^{(0)}x_n & \le & b_1 \\
 & \vdots & & & \\
a_{r1}^{(0)}x_1 + a_{r2}^{(0)}x_2 + & \cdots & +a_{rn}^{(0)}x_n & \le & b_r \\
a_{(r+1)1}^{(0)}x_1 + a_{(r+1)2}^{(0)}x_2 + & \cdots & +a_{(r+1)n}^{(0)}x_n & \ge & b_{r+1} \\
 & \vdots & & & \\
a_{m1}^{(0)}x_1 + a_{m2}^{(0)}x_2 + & \cdots & +a_{mn}^{(0)}x_n & \ge & b_m \\
x_1, x_2, \cdots, x_n \ge 0. & & & &
\end{array}
\tag{8}
$$

Assume that there exist positive integers m_0, n_0 such that for $m \ge m_0$, $n \ge n_0$ problem (8) has finite optimum value and problem (7) has finite optimum value with probability 1. It follows that these optimum values are also positive (with probability 1 in case of problem (7)). As before, we want to determine the minimum cost sample sizes such that a confidence interval should cover the unknown optimum value by a prescribed probability. We prove the following:

Theorem 3.1. *Suppose all conditions in Theorem 2.1 are satisfied for problems (7) and (8). If for given positive numbers* δ_1 *and* δ_2*, we have the relations:*

$$\Pi_{k=1}^{n}\Phi\Big(\frac{\delta_1}{\sqrt{\sum_{i=1}^{m}\frac{\sigma_{ik}^2 (y_i^{(0)})^2}{N_{ik}}}}\Big) \geq 1-\delta_2$$
$$N_{ik} \geq 1, \quad i=1,2,\cdots,m, \quad k=1,2,\cdots,n,$$

then we also have:

$$P(\mu^{(0)}-\mu \geq -\delta_1\mu) \geq 1-\delta_2$$
$$N_{ik} \geq 1, \quad i=1,2,\cdots,m, \quad k=1,2,\cdots,n.$$

Proof. Since μ is the optimum value of the dual of problem (7):

$$\begin{array}{llll} \min\{y_1+y_2+\cdots+y_m\} & & & \\ \text{subject to} & & & \\ a_{11}y_1+a_{21}y_2+ & \cdots & +a_{m1}y_m & \geq 1 \\ & \vdots & & \\ a_{1n}y_1+a_{2n}y_2+ & \cdots & +a_{mn}y_m & \geq 1 \\ y_1,\cdots,y_r \geq 0, \quad y_{r+1},\cdots,y_m \leq 0, & & & \end{array} \tag{9}$$

it follows that

$$\frac{1}{\mu} = \max_{\gamma}\min_{1\leq k\leq n}\sum_{i=1}^{m} a_{ik}\gamma_i, \tag{10}$$

where $\gamma_i \geq 0, \quad i=1,2,\cdots,r, \quad \gamma_i \leq 0, \quad i=r+1,r+2,\cdots,m, \quad \sum_{i=1}^{m}\gamma_i = 1.$

If $y^{(0)}$ and $\mu^{(0)}$ are optimal solution and optimum value, respectively, of the dual of problem (8), we also have the following equation:

$$\frac{1}{\mu^{(0)}} = \min_{1\leq k\leq n}\sum_{i=1}^{m} a_{ik}^{(0)}\frac{y_i^{(0)}}{\mu^{(0)}}. \tag{11}$$

Using inequation (10) and (11), we derive the inequality:

$$\begin{aligned} \frac{1}{\mu} &= \max_{\gamma}\min_{1\leq k\leq n}\sum_{i=1}^{m}(\xi_{ik}+a_{ik}^{(0)})\gamma_i \\ &\geq \min_{1\leq k\leq n}\sum_{i=1}^{m}(\xi_{ik}+a_{ik}^{(0)})\frac{y_i^{(0)}}{\mu^{(0)}} \\ &= \min_{1\leq k\leq n}\Big(\frac{1}{m}\sum_{i=1}^{m}\xi_{ik}\frac{my_i^{(0)}}{\mu^{(0)}}+\sum_{i=1}^{m}a_{ik}^{(0)}\frac{y_i^{(0)}}{\mu^{(0)}}\Big) \\ &\geq \min_{1\leq k\leq n}\frac{1}{m}\sum_{i=1}^{m}\xi_{ik}\frac{my_i^{(0)}}{\mu^{(0)}}+\frac{1}{\mu^{(0)}}. \end{aligned} \tag{12}$$

The rest of the proof is the same as that of Theorem 2.1. □

Based on Theorem 3.1 we formulate the sample size determination problem:

$$\begin{aligned} &\min \textstyle\sum_{i=1}^{m} \sum_{k=1}^{n} C_{ik} N_{ik} \\ &\text{subject to} \\ &\Pi_{k=1}^{n} \Phi\Big(\frac{\delta_1}{\sqrt{\sum_{i=1}^{m} \frac{\sigma_{ik}^2 (y_i^{(0)})^2}{N_{ik}}}}\Big) \geq 1 - \delta_2 \\ &N_{ik} \geq 1, \quad i = 1, 2, \cdots, m, \quad k = 1, 2, \cdots, n. \end{aligned} \tag{13}$$

In problem (13), we can replace the $y_i^{(0)}$, $i = 1, 2, \cdots, m$ by their upper bounds $Y_i^{(0)}$, $i = 1, 2, \cdots, m$ to get a more relaxed problem with completely specified parameters.

We can see that only the "$\leq$" inequality constraints contribute to the constraints in problem (13), the others are irrelevant from this point of view.

4 An Illustrative Example

The diet problem as a optimization problem was first formulated by Stigler (1945). There are various stochastic programming formulations of it, see, e.g., Armstrong and Balintfy (1975), Lancaster (1992), Prékopa (1995). The problem formulation and solution we are presenting in this section serve as illustration to the results of the former sections but may be useful in the practical application of the diet problem as well.

Suppose that we want to create a plan for the composition of a fruit cocktail to be served for everybody in a population of 25-50 years old females such that thiamin, riboflavin, niacin and ascorbic acid should be contained in on the DRI (Dietary Reference Intakes) level and its carbohydrate content be minimized. For possible inclusion into the cocktail we choose the fruits: apple, banana, cherry, grape, orange, peach, pear, plum, tangerine and watermelon. The nutrient contents of these fruits will be estimated by taking samples. The nutrient content data (see Gebhardt and Thomas, 2002) may serve only as a guideline. In Table 4.1 we present these data for the chosen fruits, together with the DRI [11] of these nutritional elements for a female, aged 25-50.

Fruit	*Vitamin*				
	Thiamin	Riboflavin	Niacin	Ascorbic Acid	Carbohydrate
Apple	0.02	0.02	0.1	8	21
Banana	0.05	0.12	0.6	11	28
Cherry	0.03	0.04	0.3	5	11
Grape	0.05	0.03	0.2	5	9
Orange	0.11	0.05	0.4	70	15
Peach	0.02	0.04	1.0	6	11
Pear	0.03	0.07	0.2	7	25
Plum	0.03	0.06	0.3	6	9
Tangerine	0.09	0.02	0.1	26	9
Watermelon	0.23	0.06	0.6	27	21
RDA	1.1	1.1	14	75	

Table 4.1

One Apple: Raw, unpeeled, $2\frac{3}{4}$" diam (about 3 per lb). *One Banana*: Raw, whole, medium, 7" - $7\frac{7}{8}$" long. *Ten Cherries*: Sweet, raw, without pits and stems. *Ten Grapes*: Raw, Seedless. *One Orange*: Raw, whole, without peel and seeds, $2\frac{5}{8}$" diam. *One Peach*: Raw, whole, pitted, $2\frac{1}{2}$" diam (about 4 per lb). *One Pear*: Raw, whole, cored, $2\frac{1}{2}$" diam. *One plum*: Raw, $2\frac{1}{8}$" diam. *One Tangerine*: Raw, without peel and seeds, $2\frac{3}{8}$" diam. *One Wedge of Watermelon*: Raw, one wedge is about $\frac{1}{16}$ of a melon (15" long, $7\frac{1}{2}$" diam).

This problem can be formulated as the following linear programming problem:

$$\begin{array}{lll}
\min\{21z_1 + 28z_2 + 11z_3 + 9z_4 + 15z_5 + 11z_6 + 25z_7 + 9z_8 + 9z_9 + 21z_{10}\} & & \\
\text{subject to} & & \\
0.02z_1 + 0.05z_2 + 0.03z_3 + 0.05z_4 + 0.11z_5 & & \\
+0.02z_6 + 0.03z_7 + 0.03z_8 + 0.09z_9 + 0.23z_{10} & \geq & 1.1 \\
0.02z_1 + 0.12z_2 + 0.04z_3 + 0.03z_4 + 0.05z_5 & & \\
+0.04z_6 + 0.07z_7 + 0.06z_8 + 0.02z_9 + 0.06z_{10} & \geq & 1.1 \\
0.1z_1 + 0.6z_2 + 0.3z_3 + 0.2z_4 + 0.4z_5 & & \\
+1.0z_6 + 0.2z_7 + 0.3z_8 + 0.1z_9 + 0.6z_{10} & \geq & 14 \\
8z_1 + 11z_2 + 5z_3 + 5z_4 + 70z_5 & & \\
+6z_6 + 7z_7 + 6z_8 + 26z_9 + 27z_{10} & \geq & 75 \\
z_1, z_2, z_3, z_4, z_5, z_6, z_7, z_8, z_9, z_{10} & \geq & 0.
\end{array} \tag{14}$$

To obtain an LP with all 1's in the objective function coefficient vector and on the right hand sides, we introduce the new variables $y_1 = 21z_1$, $y_2 = 28z_2$, $y_3 = 11z_3$, $y_4 = 9z_4$, $y_5 = 15z_5$, $y_6 = 11z_6$, $y_7 = 25z_7$, $y_8 = 9z_8$, $y_9 = 9z_9$ and $y_{10} = 21z_{10}$, and divide each constraint by its right hand side value. Then the dual of this new linear programming problem is:

$$
\begin{aligned}
&\max\{x_1 + x_2 + x_3 + x_4\}\\
&\text{subject to}\\
&8.7\times 10^{-4}x_1 + 8.7\times 10^{-4}x_2 + 3.4\times 10^{-4}x_3 + 5.1\times 10^{-3}x_4 \le 1\\
&1.6\times 10^{-3}x_2 + 3.9\times 10^{-3}x_2 + 1.5\times 10^{-3}x_3 + 5.2\times 10^{-3}x_4 \le 1\\
&2.5\times 10^{-3}x_1 + 3.3\times 10^{-3}x_2 + 1.9\times 10^{-3}x_3 + 6.1\times 10^{-3}x_4 \le 1\\
&5.1\times 10^{-3}x_1 + 3.0\times 10^{-3}x_2 + 1.6\times 10^{-3}x_3 + 7.4\times 10^{-3}x_4 \le 1\\
&6.7\times 10^{-3}x_1 + 3.0\times 10^{-3}x_2 + 1.9\times 10^{-3}x_3 + 6.2\times 10^{-2}x_4 \le 1\\
&1.7\times 10^{-3}x_1 + 3.3\times 10^{-3}x_2 + 6.5\times 10^{-3}x_3 + 7.3\times 10^{-3}x_4 \le 1\\
&1.1\times 10^{-3}x_1 + 2.5\times 10^{-3}x_2 + 5.7\times 10^{-4}x_3 + 3.7\times 10^{-3}x_4 \le 1\\
&3.0\times 10^{-3}x_1 + 6.1\times 10^{-3}x_2 + 2.4\times 10^{-3}x_3 + 8.9\times 10^{-3}x_4 \le 1\\
&9.1\times 10^{-3}x_1 + 2.0\times 10^{-3}x_2 + 7.9\times 10^{-4}x_3 + 3.9\times 10^{-2}x_4 \le 1\\
&1.0\times 10^{-2}x_1 + 2.6\times 10^{-3}x_2 + 2.0\times 10^{-3}x_3 + 1.7\times 10^{-2}x_4 \le 1\\
&x_1, x_2, x_3, x_4 \ge 0.
\end{aligned} \tag{15}
$$

Using *Excel*, we get the optimal solutions of problem (15) and its dual problem:

$$
\begin{aligned}
x^{(0)} &= (56, 102, 87.4, 0)^T,\\
y^{(0)} &= (0, 0, 0, 0, 0, 105.8, 0, 82.2, 0, 57.3)^T.
\end{aligned} \tag{16}
$$

The optimal solutions (16) have been obtained on the basis of the nutrient contents presented in Table 4.1. Since the data in that table are not necessarily the same as those contained in those foods that we want to serve the calculation of the optimal sample sizes, by the use of $x^{(0)}$ and $y^{(0)}$, provide us only with a guideline. Let $\varepsilon_1 = \varepsilon_2 = \delta_1 = \delta_2 = 0.1$. Suppose $\alpha_{ik}^{(j)}$, $i = 1, 2, \cdots, 10$, $k = 1, 2, 3, 4$ have independent normal distributions, $\mu_{ik} = E[\alpha_{ik}^{(j)}] = a_{ik}$, $i = 1, 2, \cdots, 10$, $k = 1, 2, 3, 4$, then $\sigma_{ik} \approx 0.10\mu_{ik}$, $i = 1, 2, \cdots, 10$, $k = 1, 2, 3, 4$. So

$$
\sigma = (\sigma_{ik}) = \begin{pmatrix}
8.7\times 10^{-5} & 8.7\times 10^{-5} & 3.4\times 10^{-5} & 5.1\times 10^{-4}\\
1.6\times 10^{-4} & 3.9\times 10^{-4} & 1.5\times 10^{-4} & 5.2\times 10^{-4}\\
2.5\times 10^{-4} & 3.3\times 10^{-4} & 1.9\times 10^{-4} & 6.1\times 10^{-4}\\
5.1\times 10^{-4} & 3.0\times 10^{-4} & 1.6\times 10^{-4} & 7.4\times 10^{-4}\\
6.7\times 10^{-4} & 3.0\times 10^{-4} & 1.9\times 10^{-4} & 6.2\times 10^{-3}\\
1.7\times 10^{-4} & 3.3\times 10^{-4} & 6.5\times 10^{-4} & 7.3\times 10^{-4}\\
1.1\times 10^{-4} & 2.5\times 10^{-4} & 5.7\times 10^{-5} & 3.7\times 10^{-4}\\
3.0\times 10^{-4} & 6.1\times 10^{-4} & 2.4\times 10^{-4} & 8.9\times 10^{-4}\\
9.1\times 10^{-4} & 2.0\times 10^{-4} & 7.9\times 10^{-5} & 3.9\times 10^{-3}\\
1.0\times 10^{-3} & 2.6\times 10^{-4} & 2.0\times 10^{-4} & 1.7\times 10^{-3}
\end{pmatrix}. \tag{17}
$$

The optimal fruit composition problem, as a special case of problem (5), takes the form:

$$\begin{array}{l} \min \sum_{i=1}^{10} \sum_{k=1}^{4} N_{ik} \\ \text{subject to} \\ \Pi_{i=1}^{10} \Phi\left(\frac{0.1}{\sqrt{\sum_{k=1}^{4} \frac{\sigma_{ik}^2 (x_k^{(0)})^2}{N_{ik}}}}\right) \geq 0.9 \\ \Pi_{k=1}^{4} \Phi\left(\frac{0.1}{\sqrt{\sum_{i=1}^{10} \frac{\sigma_{ik}^2 (y_i^{(0)})^2}{N_{ik}}}}\right) \geq 0.9 \\ N_{ik} \geq 1, \quad i = 1, 2, \cdots, 10, \quad k = 1, 2, 3, 4, \end{array} \tag{18}$$

where $y_i^{(0)}$, $x_k^{(0)}$, σ_{ik}, $i = 1, 2, \cdots, 10$, $k = 1, 2, 3, 4$ are given by (16) and (17). Using *Excel*, we compute the minimum sample sizes as continuous variables for this diet problem, they are given in:

$$N = (N_{ik}) = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1.056682 & 1 & 1 & 1 \\ 1 & 1.75022 & 3.423089 & 4.103629 \\ 1 & 1 & 1 & 1 \\ 1.222637 & 2.722797 & 1.057123 & 3.886176 \\ 1.396225 & 1 & 1 & 1 \\ 3.028301 & 1 & 1 & 5.175706 \end{pmatrix}.$$

If we take $N^* = \lceil N \rceil$, then

$$N^* = (N_{ik}^*) = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 2 & 1 & 1 & 1 \\ 1 & 2 & 4 & 5 \\ 1 & 1 & 1 & 1 \\ 2 & 3 & 2 & 4 \\ 2 & 1 & 1 & 1 \\ 4 & 1 & 1 & 6 \end{pmatrix}.$$

Otherwise, we can make a search around the optimal solution of the continuous problem to get a better integer solution if any exists. We were able to find one in our example which is exhibited below:

$$N^{**} = (N_{ik}^{**}) = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 2 & 3 & 4 \\ 1 & 1 & 1 & 1 \\ 2 & 3 & 1 & 4 \\ 2 & 1 & 1 & 1 \\ 3 & 1 & 1 & 5 \end{pmatrix}.$$

Suppose now that we do not know the nutrient contents of the fruits, that we plan to include into the fruit cocktail, but we know upper bounds on the components of the optimal solutions $x^{(0)}$, $y^{(0)}$. Let the upper bounds be given by the components of the vectors $X^{(0)}$, $Y^{(0)}$: $X^{(0)} = (60, 110, 90, 10)^T$, $Y^{(0)} = (10, 10, 10, 10, 10, 110, 10, 90, 10, 60)^T$. We also suppose that the variances σ_{ik}, $i = 1, 2, \cdots, 10$, $k = 1, 2, 3, 4$ are exactly known (if they are not, then we may have upper bounds on them as well). Then, using the same ε_1, ε_2, δ_1 and δ_2, as before, the solution of problem (6) gives the following fractional values:

$$N = (N_{ik}) = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & 1.239755 & 1 & 1 \\ 1 & 1.051974 & 1 & 1 \\ 1 & 1.067844 & 1 & 1 \\ 1.361261 & 1.117383 & 1 & 1 \\ 1 & 1.972729 & 3.805863 & 4.476521 \\ 1 & 1 & 1 & 1 \\ 1.424193 & 3.237028 & 1.233966 & 4.466415 \\ 1.781922 & 1 & 1 & 1 \\ 3.479575 & 1.16157 & 1 & 5.686951 \end{pmatrix}.$$

Then we either use $N^* = \lceil N \rceil$ to get

$$N^* = (N_{ik}^*) = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & 2 & 1 & 1 \\ 1 & 2 & 1 & 1 \\ 1 & 2 & 1 & 1 \\ 2 & 2 & 1 & 1 \\ 1 & 2 & 4 & 5 \\ 1 & 1 & 1 & 1 \\ 2 & 4 & 2 & 5 \\ 2 & 1 & 1 & 1 \\ 4 & 2 & 1 & 6 \end{pmatrix}$$

or search for a better integer sample sizes . We succeeded to find such an N^{**} and we present it below:

$$N^{**} = (N^{**}_{ik}) = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & 2 & 1 & 1 \\ 1 & 2 & 1 & 1 \\ 1 & 2 & 1 & 1 \\ 2 & 2 & 1 & 1 \\ 1 & 2 & 4 & 5 \\ 1 & 1 & 1 & 1 \\ 2 & 4 & 2 & 5 \\ 2 & 1 & 1 & 1 \\ 4 & 2 & 1 & 6 \end{pmatrix}.$$

5 Application in Game Theory

Consider a two-person zero-sum game (see, e.g., Neumann and Morgenstern, 1944) with finite strategy sets of m, n elements, respectively and payoff matrix $A = (a_{ik})$. Let μ designate the value of the game. We can obtain it as the reciprocal value of the optimum of any of the problems in the primal dual pair of linear programs:

$$\begin{array}{l} \frac{1}{\mu} = \max\{x_1 + x_2 + \cdots + x_n\} \\ \text{subject to} \\ \sum_{k=1}^{n} a_{ik} x_k \le 1, \quad i = 1, 2, \cdots, m \\ x_k \ge 0, \quad k = 1, 2, \cdots, n, \end{array} \tag{19}$$

$$\begin{array}{l} \frac{1}{\mu} = \min\{y_1 + y_2 + \cdots + y_m\} \\ \text{subject to} \\ \sum_{i=1}^{m} a_{ik} y_i \ge 1, \quad k = 1, 2, \cdots, n \\ y_i \ge 0, \quad i = 1, 2, \cdots, m. \end{array} \tag{20}$$

These are exactly the problems (1) and (2) if we replace there $b_i = 1$, $i = 1, 2, \cdots, m$ and $c_k = 1$, $k = 1, 2, \cdots, n$. Thus, the results of Sections 2, 3 can be applied to the LP's (19), (20) without change.

We assume that the elements of the payoff matrix are deterministic but unknown. We designate these values by $a_{ik}^{(0)}$, $i = 1, 2, \cdots, m$, $k = 1, 2, \cdots, n$ and take independent samples to estimate the unknown values. In order to find an optimal collection of sample sizes we formulate and solve problem (5) or, if we replace the values $y_i^{(0)}$, $i = 1, 2, \cdots, m$, $x_k^{(0)}$, $k = 1, 2, \cdots, n$ by their upper bounds, we solve problem (6).

References

[1] Armstrong, R. D. and J. L. Balintfy. 1975. A Chance Constrained Multiple Choice Programming Algorithm. *Operations Research* **23**, 494-510.

[2] Cohen, J. E. and C. M. Newman. 1989. Host-Parasite Relations and Random Zero-Sum Games: The Stabilizing Effect of Strategy Diversification. *The American Naturalist* **133**, 533-552.

[3] Kabe, D. G. 1983. Laws of Large Numbers for Random Payoff Games. it The Journal of the Industrial Mathematics Society **33** (Part I), 73-86.

[4] Kuhn, H. W. and R. E. Quandt. 1963. An Experimental Study of the Simplex Method. *Proceedings of the Symposium on Applied Mathematics* **15**, 107-124.

[5] Lancaster, L. 1992. The Evolution of the Diet Model in Managing Food Systems. *Interfaces* **22**, No.5, 59-68.

[6] Neumann, J., von and O. Morgenstern. 1944. *Theory of Games and Economic Behavior*. Princeton University Press, Princeton, New Jersey.

[7] Prékopa, A. 1972. Laws of Large Numbers for Random Linear Programs. *Mathematical Systems Theory* **6**, 277-288.

[8] Prékopa, A. 1995. *Stochastic Programming.* Budapest, Kluwer.

[9] Stigler, G. J. 1945. The Cost of Subsistence. *Journal of Farm Economy* **27**, No.2, 303-314.

[10] Gebhardt, S. E. and R. G. Thomas. 2002. *Nutritive Value of Foods.* United States Department of Agriculture, Agriculture Research Service, Nutrinent Data Laboratory, Beltsville, Maryland.

[11] DRI. *Dietary Reference Intakes: Vitamins.* http://www.iom.edu/iom/iomhome.nsf/WFiles/webtablevitamins/$files/ webtablevitamins.pdf.