NOTES IN OPERATIONS RESEARCH -- 4

Operations Research Center

University of California, Berkeley

November 1964 ORC 64-30

This research has been partially supported by the Office of Naval Research under Contract Nonr-222(83) and the National Science Foundation under Grant GP-2633 with the University of California. Reproduction in whole or in part is permitted for any purpose of the United States Government.

STATISTICAL INFERENCE IN MARKOV-RENEWAL PROCESSES

By Daniel P. Heyman

I. Introduction

The usual procedure in operations research studies is to model a system and then analyze system behavior from the model. In applying and testing these models, one is faced with the problem of estimating the parameters and testing assumptions.

The Markov-Renewal process has been used by Jewell [5] to solve a special structure dynamic programming problem, and it can be used to describe a queueing process with arbitrary service time distributions. This paper is addressed to the problem of finding estimators for the parameters of the process and to making inferences about stochastic service systems.

The inference problem for renewal processes is identical to that of making inferences about random variables; much work has been done to develop a large sample theory for Markov process (see Billingsley [2]). It will be shown that the combination of these two methods respectively suffice to find maximum likelihood estimators for the Markov-Renewal process and to show that with large samples, chi-square methods can be used to test composite hypotheses.

To illustrate the use of these methods, a procedure for estimating the parameters of the arrival and service time distributions of a queueing process with Poisson distributed arrivals and general service times will be outlined. Also, it will be described how to test the hypotheses that interarrival and service time distributions are independent of the number in the queue.

II. The Markov-Renewal Process

A Markov-Renewal process (M.R.P.) is a generalization of a Markov process in which the transition time from state i to state j is a sample from a probability distribution that may depend on both i and j . The first investigations of processes of this type were made by Levy, Smith, and Takacs; in references [7] and [8]. Pyke summarizes the initial work and presents original contributions. Since these two articles provide a uniform set of notations, definitions, and theorems, they will be used as a basic source. An extensive list of references will be found in [7].

The following notations, definitions, and theorems will be used:

$J_n \triangleq$ the state of the system after n transitions,[1]

$X_n \triangleq$ the length of time the process is in state J_n ; i.e., the length of time between the $(n)^{th}$ and $(n+1)^{st}$ transition; the distribution of X_n is not degenerate at zero. J_n and X_n are both random variables. The process can enter m states, where m is either a positive finite integer or infinity.

---

[1] If the system is in state i heading towards state j , during the $(n+1)^{st}$ transition, then $J_n = i$.

DEFINITION 2.1: $Q = (Q_{ij})$ is the matrix of functions satisfying (i) $Q_{ij}(t) = 0$ for $t \leq 0$, and (ii) $\sum_{j=1}^{m} Q_{ij}(\infty) = 1$, $(1 \leq i < m+1)$.

DEFINITION 2.2: The m vector $A = (a_1, a_2, \ldots, a_j, \ldots)$ is the vector of initial probabilities; i.e., $a_k = \Pr\{J_0 = k\}$.

DEFINITION 2.3: The (J, X) process is a two-dimensional stochastic process $\{(J_n, X_n); n \geq 0\}$ that satisfies

(i) $\Pr\{J_0 = k\} = a_k$

(ii) $X_0 = 0$ a.s., and

(2.1) $$\Pr\{J_n = k, X_n \leq x | J_0, J_1, X_1, J_2, X_2, \ldots, J_{n-1}, X_{n-1}\} \overset{a.s.}{=} Q_{J_{n-1}}(x) \text{ for all } x \in (-\infty, \infty) \text{ and } 1 \leq k < m+1 .$$

From this definition, we can interpret $Q_{ij}(t)$ as the conditional probability that the next transition is from state i to state j, that the transition will occur in t or less time units, given that the system is currently in state i.

$S_n \triangleq \sum_{i=0}^{n} X_i$ for $n \geq 0$, S_n is the elapsed time for the first n transitions.

LEMMA 2.1: The two-dimensional (J, S) process is a Markov process, and the J-process is a Markov chain. Let $P_{ij} = Q_{ij}(\infty)$, then

(2.2) $$\Pr\{J_n = k, S_n \leq y | J_0, J_1, S_1, \ldots, J_{n-1}, S_{n-1}\} \overset{a.s.}{=} Q_{J_{n-1},k}(y - S_{n-1})$$

and

(2.3) $\Pr\{J_n = j \mid J_0, J_1, \ldots, J_{n-1} = i\} = P_{ij}$.

Since the Q-functions need not be probability distribution functions, we define $\Phi_{ij} = p_{ij}^{-1} Q_{ij}$,① which "normalizes" the Q_{ij} , if $p_{ij} \neq 0$, to the probability distribution function of the transition time from i to j . If $p_{ij} = 0$, $\Phi_{ij} \stackrel{\Delta}{=} U_1$ (the unit step at 0). Let $\phi_{ij}(x) = \frac{d}{dx} \Phi_{ij}(x)$.

$H_i(t) = \sum_{j=1}^{n} Q_{ij}(t)$ is the probability density function of the time spent in state i; $\eta_i = \int_0^\infty t \, dH_i(t)$ is the mean time in state i . The mean transition time between states i and j is given by $b_{ij} = \int_0^\infty t \, d\Phi_{ij}(t)$. Several useful relations are available from the definitions already presented.

$$\Pr\{X_n \leq x \mid J_0, \ldots, J_{n-1}\} = H_{J_{n-1}}(x)$$

$$\Pr\{X_n \leq x \mid J_0, \ldots, J_{n-1}, J_n\} = \Phi_{J_{n-1}, J_n}(x)$$

$$\eta_i = \sum_{j=1}^{n} b_{ij} p_{ij}$$

DEFINITION 2.4: The integer-valued random variable $N(t) \stackrel{\Delta}{=} \sup\{n \geq 0, t \geq 0: S_n \leq t\}$ is the counting distribution for the number of transitions in time t . The integer-valued random variable $N_j(t)$ is the number of transitions into state j before $(N(t) + 1)$ transitions.

It is obvious that $N(t) = \sum_{j=1}^{m} N_j(t)$.

DEFINITION 2.5: Set $\eta(t) = \{N_1(t), N_2(t), \ldots, N_j(t), \ldots\}$.

---

① Φ_{ij} is denoted by F_{ij} in [7].

The stochastic process $\{\eta(t), t \geq 0\}$ is called the Markov-Renewal Process determined by (m, A, Q).

This means that a Markov-Renewal process is a vector-valued process which describes the probability of having n_i observations of state k in N changes of state, given the number of states, the initial state probabilities, and the probability law governing changes of state.

LEMMA 2.2: If $m < \infty$, then for all initial states i, $Pr\{N(t) < \infty, \forall t \geq 0\} = 1$.

This lemma is of great importance because it allows one to fix an interval of observation and then record information about the changes of state, and it implies that $N_j(t) < \infty$.

DEFINITION 2.6: The Markov chain determined by the J-process alone is called the corresponding Markov chain, C.M.C.

Many relationships between the M.R.P. and the C.M.C. can be shown, but it is of primary importance that the conditional probabilities, p_{ij}, are completely determined by the C.M.C. If the times between transitions from i to j are observed, they will form an embedded renewal process with "failure" distributions $\Phi_{ij}(t)$. There are m^2 embedded renewal process in the M.R.P., and these, the C.M.C., and the vector of initial probabilities are sufficient to describe the $Q_{ij}(t)$, and hence the M.R.P.

III. Limit Theorems from Pyke [6].

Define a function f on (J_{n-1}, J_n, X_n), and for $t \geq 0$ define $W(t) = \sum_{n=1}^{N(t)} f(J_{A-1}, J_n, X_n)$. Here, $f(.,.,.)$ is a function on the transitions and the transition times. By picking $f(J_{n-1}, J_n, X_n) = X_{ij}$, when $J_{n-1} = i$ and $J_n = j$, and zero otherwise, $f(\cdot, \cdot, \cdot)$ represents the transition times from i to j. If $f(J_{n-1}, J_n, X_n) = J_{n-1}, J_n$, $f(\cdot, \cdot, \cdot)$ rerepresents the n^{th} transition; hence, $W(t)$ can represent the C.M.C. or the embedded renewal processes.

The random variable $\{U_s^{(j)} : s > 0\}$ is $f(J_{n-1}, J, X_n)$ summed over the states after the s^{th} realization of state j, up to and including, the $(s+1)^{st}$ occurrence of state j. $E\{U_s^{(j)}\} = m_j$, μ_{jj} is the mean of the first passage time distribution in M.R.P., and μ_{jj}^* is the mean of the first passage time distribution in the C.M.C.

THEOREM 3.1 (Strong Law of Large Numbers). Under certain regularity conditions,

$$W(t)/t \rightarrow m_j/\mu_{jj} \text{(a.s.)}$$

the limit being zero if $\mu_{jj} = \infty$, and $\frac{m_j}{\mu_{jj}}$ does not depend on j.

THEOREM 3.2 (Weak Law of Large Numbers). If U_s satisfies the Strong Law of Large Numbers; and if μ_{jj} is finite, then

$$W(t)/t \xrightarrow{P} m_j/\mu_{jj}$$

THEOREM 3.3 (Central Limit Theorem). For $\mu_{jj} < \infty$,

$$\frac{W(t) - E\{W(t)\}}{[\mathrm{Var}\{W(t)\}]^{1/2}} \xrightarrow{L} N(0,1) .$$

IV. Statistical Inference in Markov Processes [1].

Consider an s state Markov chain with transition probabilities p_{ij} when n realizations of the process are observed, let

f_{ij} = the relative frequency of transition from i to j and f_i = the number of transitions out of state i. Define the random variable $\xi_{ij} = (f_{ij} - f_i p_{ij})/f_i$, then

$$E\{\xi_{ij}\} = 0 \quad \text{and} \quad p \lim \frac{f_i}{n} = p_i$$

THEOREM 4.1 In a stationary, ergodic, s state Markov chain,

$$\sum_{j\ p_{ij} > 0} (f_{ij} - f_i p_{ij})^2 / f_i p_{ij} \xrightarrow{L} \chi^2_{d_i - 1}$$

where d_i is the number of the non-zero p_{ij}.

LEMMA 4.1. $2 \sum_{i,j} f_{ij} \ln(f_{ij}/f_i p_{ij}) \xrightarrow{L} \chi^2_{s(s-1)}, \; p_{ij} > 0$.

The likelihood functions is $L = \sum_{i,j} p_{ij}^{f_{ij}}$. To maximize $\ln L$ subject to the conditions that

$\sum_j p_{ij} = 1$ for each i, form the Lagrangian

$$\Lambda(\ln L, \lambda_i) = \sum_{ij} f_{ij} \ln p_{ij} + \sum_i \lambda_i (1 - \sum_j p_{ij}) .$$

$$\frac{\partial}{\partial p_{ij}} \Lambda(\ln L, \lambda_i) = 0 \Longrightarrow \frac{f_{ij}}{p_{ij}} - \lambda_i = 0$$

$$\frac{\partial}{\partial \lambda_i} \Lambda(\ln L, \lambda_i) = 0 \Longrightarrow \lambda_i = f_i$$

Therefore, the maximum likelihood estimator of p_{ij} is

$\hat{p}_{ij} = f_{ij}/f_i$ and $\ln L_{max} = \sum_{i,j} f_{ij} \ln(f_{ij}/f_i)$.

The statistical analysis of Markov chains can be accomplished by applying theorem 4.1. A test for goodness of fit of $\hat{p}_{ij}$ can be made by substituting f_{ij}/f_i for p_{ij}, and lemma 4.1 (which follows from theorem 4.1) gives the Neyman-Pearson criterion.

V. Maximum Likelihood Estimators for the Markov-Renewal Process

The probability that a transition from i to j occurs in $(x, x + dx)$ is $q_{ij}(x)dx = dQ_{ij}(x) = p_{ij}\phi_{ij}(x)dx$. In observing the system, f_{ij} samples from $\Phi_{ij}(x)$ will be taken. Denote the aggregate of these sample values as $x_{ij}^{[f]}$ and let $\theta_{ij} = (\theta_1, \theta_2, \ldots, \theta_k, \ldots, \theta_n)$, $n < \infty$, be the parameters of $\phi_{ij}(x)$, where θ_k may be different in all θ_{ij}. The problem at hand is to find the maximum likelihood estimators for the p_{ij}'s and the vector components of all θ_{ij}.

The "likelihood" function[1] of the observations is:

$$L = \prod_d [p_{ij}\phi_{ij}(x_{ij}^{[f]}, \theta_{ij})]^{f_{ij}}$$

where d is the set (i, j) for which $f_{ij} > 0$. Taking logarithms,

$$\ln L = \sum_d \{f_{ij} \ln p_{ij} + f_{ij} \ln \phi_{ij}(x_{ij}^{[f]}, \theta_{ij})\}$$, and maximizing $\ln L$ subject to the restraint that $\sum_j p_{ij} = 1$ for each i by use of the Lagrangian

$$\Lambda(\ln L, \lambda_i) = \sum_d \{f_{ij} \ln p_{ij} + f_{ij} \ln \phi_{ij}(x_{ij}^{[f]}, \theta_{ij}) + \sum_i \lambda_i (1 - \sum_j p_{ij}) ,$$

---

[1] The function L is called the "liklihood" function because it is asymptotically equivalent to the Radon-Nikodym derivative of the process.

we get

$$\frac{\partial}{\partial p_{ij}} \Lambda(\ln L, \lambda_i) = \frac{f_{ij}}{p_{ij}} - \mu_i = 0$$

$$\frac{\partial}{\partial \lambda_{ij}} \Lambda(\ln L, \lambda_i) = 1 - \sum_j p_{ij} = 0$$

$$\frac{\partial}{\partial \theta_{ij,k}} \Lambda(\ln L, \lambda_i) = \frac{\partial}{\partial \theta_{ij,k}} (f_{ij} \ln \phi_{ij}(X_{ij}^{[f]}, \theta_{ij}) = 0$$

The first two equations yield $\hat{p}_{ij} = \frac{f_{ij}}{f_i}$, and the third is the maximum likelihood estimator of the function $\phi_{ij}(X, \theta_{ij})$ <u>considered only as a renewal density function assuming the values</u> $X_{ij}^{[f]}$.

This analysis points out some major properties of Markov-Renewal processes; that the C.M.C. and the imbedded renewal processes formed by the state-to-state transitions are asymptotically mutually independent, that knowledge of each of them separately implies knowledge of the complete process and vice-versa.

A simple example in computing maximum likelihood estimators will illustrate the results of this section. Assume that $\Pr\{\text{transition from } i \text{ to } j \leq t\} = 1 - e^{-\theta_{ij} t}$. Then

$$L = \prod_d p_{ij}^{f_{ij}} \theta_{ij}^{f_{ij}} e^{-\theta_{ij} \sum_f x_{ij}}$$

where the summation is taken over those x_{ij} representing transition times from i to j . The Lagrangian is

$$\Lambda(\ln L, \lambda_i) = \sum_d \{ f_{ij} \ln p_{ij} + f_{ij} \ln \theta_{ij} - \theta_{ij} \sum_f X_{ij} \} + \sum_i \lambda_i (1 - \sum_j p_{ij})$$

Equating the partial derivatives with respect to p_{ij} and λ_i to zero and solving simultaneously gives $\hat{p}_{ij} = f_{ij}/f_i$.

$$\frac{\partial}{\partial \theta_{ij}}(\ln L, \lambda_i) = \frac{f_{ij}}{\theta_{ij}} - \sum x_{ij} = 0 \Longrightarrow \hat{\theta}_{ij} = f_{ij} \Big/ \sum x_{ij}$$

The preceding result shows that the maximum likelihood estimator for the exponential transition time parameter between states i and j is the number of $i \rightarrow j$ transitions divided by the sum of the transition times. This is the maximum likelihood estimate of the exponential distribution parameter when the phenomena generating the observations is not specified.

VI. Estimation in Queues with Poisson Arrivals and General Service Time Distributions

If the customers queued before a service facility are observed at the points of departure, and J_n is the number in the queue immediatedly after the n^{th} departure $(n = 0, 1, 2, \ldots)$, with X_n defined as the time between J_n and J_{n-1}, the queue has been described as a Markov-Renewal process. The validity of the preceeding statement is established in the following manner: Let ξ_n be the number of arrivals during the service time of the nth customer, then $J_n = J_{n-1} + \xi_n - 1$ if the system was not empty after the $(n-1)^{st}$ customer completed service; if $J_{n-1} = 0$, $J_n = \xi_n$. Letting $U(x)$ denote Heaviside's Unit Function, the two relationships are combined as

$$J_n = J_{n-1} - U(J_{n-1}) + \xi_n .$$

Since Poisson arrivals have the property that ξ_n depends only on X_n, then J_n depends only on J_{n-1} and the J-process is a Markov chain. The implicit assumption that service times are independent samples of a positive random variable suffices to identify the imbedded renewal processes. If the process has been operating for a long time, the stationary probabilities are the initial probabilities, and by the comment following the definition in 2.6, the sufficient conditions to define an M.R.P. are met.

The service time distribution may depend on i and j; indeed it is the null hypothesis that this is not the case, which may be of primary interest when making statistical tests of the model. The state spaces of the C.M.C. consist of the nonnegative integers, since that is the range of possible queue lengths (number in queue does not include the number in service).

Notice that the proof of the existence of the C.M.C. is not invalidated if the Poisson parameters is dependent on i, the value of J_n. It is well known that Poisson arrivals with parameter λ_i have interarrival times that are expoentially distributed with parameter λ_i. Observing the interarrival times when the system is in state i, we have from Section V that $\hat{\lambda}_i$ = the number of arrivals divided by the sum of the interarrival times. The chi-square or Kolmogorov-Smirnov tests can be used to test the goodness of fit of $\hat{\lambda}_i$ and the null hypothesis that $\lambda_1 = \lambda_2 = \ldots = \lambda_k$ (i.e., that the arrival rate doesn't depend on the number in queue). If the hypothesis is rejected, alternate hypotheses about the dependence of the arrival rate and i can be tested with chi-square or Kolmorogov-Smirnov statistics.

In the preceding section it was shown how to find estimators for $\phi_{ij}(x, \theta_{ij})$ and p_{ij}. In this case, $\phi_{ij}(x, \theta_{ij})$ is the service time distribution when $i \geq 1$. It is the convolution of the interarrival time distribution and the service time distribution when $i = 0$, and the data for the estimators is generated by looking at the i-to-j transitions at the points of departure from the service facility. If $W(t)$ is taken so it represents the i-to-j transition times, and applying the Central Limit Theorem (Theorem 3.3), the chi-square test can be used to test hypothesis about the service times, such as $\phi_{ij} = \phi_i$, all the ϕ_{ij} are equal, or ϕ_{ij} depends on i and j in a specified manner. For the $i = 0$ case, λ_0 must be known so the estimators for ϕ_{ij} and chi-square test "theoretical value" items may be calculated.

At this point the interarrival time distributions and service time distributions have been estimated and tested. If the queue discipline is the same as assumed, the model should accurately predict q_i, the probability that there are i customers in the queue at steady state. Assuming that observations are taken on a steady state queue, q_i = the time spent in state i divided by the observation time T. In terms of the M.R.P.,

$$q_i = N_i(t)\eta_i/T = N_i(t)\sum_{j=1}^{n} b_{ij}p_{ij}/T.$$

The estimate of the $\hat{p}_{ij}$ is $p_{ij} = \frac{f_{ij}}{f_i}$,

$$\hat{q}_i = N_i(t)\sum_{j=1}^{m} b_{ij}/T$$

This result should agree with the calculated q: If this is not so, further investigation should be made as to the nature of the queue discipline and those characteristics of the process which were not included in the model.

REFERENCES

[1] Billingsley, P., "Statistical Methods in Markov Chains," Ann. Math. Stat., Vol. 32, No. 1 (1961).

[2] ________, Statistical Inference for Markov Processes, University of Chicago Press, Chicago, 1961.

[3] Cox, D. R. and W. L. Smith, Queues, John Wiley and Sons, Inc., New York, 1961.

[4] Freund, J. E., Mathematical Statistics, Prentice-Hall, Inc., Englewood Cliffs, New Jersey, 1962.

[5] Jewell, W. S., "Markov-Renewal Programming," Research Report No. 37, Operations Research Center, University of California, Berkeley, October, 1962.

[6] Parzen, E., Stochastic Processes, Holden-Day, Inc., San Francisco, 1962.

[7] Pyke, R., "Markov-Renewal Processes: Definitions and Preliminary Properties," Ann. Math. Stat., Vol. 32, No. 4, (1961).

[8] ________, "Markov-Renewal Processes with Finitely Many States," Ann. Math. State., Vol. 32, No. 4 (1961).

[9] ________, "Limit Theorems for Markov-Renewal Processes," Technical Report No. 6, University of Washington, 1963.

[10] Wolff, R. W., "Problems of Statistical Inference for Birth and Death Queueing Models," ORC 63-3, Operations Research Center, University of California, Berkeley, March 1963.

A NOTE ON DEFICIT, EXCESS, AND SPREAD IN A MARKOV RENEWAL PROCESS

by

William S. Jewell

1. Introduction

Consider a Markov Renewal Process (MRP) with a finite number of states, started in state i at time $t = 0$. Let $N(t)$ be the number of transitions in $(0, t]$, $S_{N(t)} \leq t$ the time at which the last previous transition occurred, and $S_{N(t)+1} > t$ the time at which the next following transition will occur. Random variables of interest are: (i) the deficit, $\tau_D(t) = t - S_{N(t)}$; (ii) the excess, $\tau_E(t) = S_{N(t)+1} - t$; and (iii) the spread, $\tau_S(t) = S_{N(t)+1} - S_{N(t)}$. The purpose of this note is to show the distributions of these r.v.s., particularly in the limit as $t \to \infty$. We shall use the definitions, notations, and conventions of PYKE [2][3] , who has previously found the distributions of excess, shown in slightly different form as (3) and (6) below.

2. Distributions of Deficit, Excess, and Spread

We first find the probability that the last previous transition was into state j during the interval $[t - y , t]$, and the next following transition is into state k during the interval $(t , t + x]$. This probability is

$\Phi_{jk}^{(i)}(y , x ; t)$ equals

$$P[Z_t = j, J_{N(t)+1} = k, t - y \leq S_{N(t)} \leq t < S_{N(t)+1} \leq t + x \mid Z_0 = i] ,$$

for $i, j, k = 1, 2, \ldots, m$; $0 \leq y \leq t$; $0 < x \leq \infty$; and $t \geq 0$.

By a straightforward use of the definitions of a MRP, this may be expressed as

$$\Phi_{jk}^{(i)}(y, x; t) = \delta_{ij}[Q_{jk}(t + x) - Q_{jk}(t)]U_0(y - t) \tag{1}$$
$$+ \int_{t-y}^{y} [Q_{jk}(t + x - z) - Q_{jk}(t - z)]dM_{ij}(z) .$$

The desired marginal distributions of deficit, excess, and spread are then easily found to be:

$$D_{jk}^{(i)}(y; t) = P[Z_t = j, J_{N(t)+1} = k, \tau_D \leq y \mid Z_0 = i] \quad 0 \leq y \leq t$$

$$= \delta_{ij}[p_{jk} - Q_{jk}(t)]U_0(y - t) \tag{2}$$
$$+ \int_{t-y}^{t} [p_{jk} - Q_{jk}(t - z)]dM_{ij}(z) ;$$

$$E_{jk}^{(i)}(x; t) = P[Z_t = j, J_{N(t)+1} = k, \tau_E \leq x \mid Z_0 = i] \quad 0 < x \leq \infty$$

$$= \delta_{ij}[Q_{jk}(t + x) - Q_{jk}(t)] \tag{3}$$
$$+ \int_{0}^{t} [Q_{jk}(t + x - z) - Q_{jk}(t - z)]dM_{ij}(z) ;$$

and

$$S_{jk}^{(i)}(w;\ t) = P[Z_t = j,\ J_{N(t)+1} = k,\ \tau_S \leq w \mid Z_0 = i]$$

(4)

$$= Q_{jk}(w)[M_{ij}(t) - M_{ij}(t - w)] - \int_{t-w}^{t} Q_{jk}(t - z)\, dM_{ij}(z) \qquad 0 < w \leq t$$

$$= Q_{jk}(w)[M_{ij}(t) + \delta_{ij}] - \int_{0}^{t} Q_{jk}(t - z)\, dM_{ij}(z)\ . \qquad t < w \leq \infty$$

The lower limits on x and w may be extended to include zero, because of the convention that $F_{ij}(0) = Q_{ij}(0) = G_{ij}(0) = M_{ij}(0) = 0$.

3. A Theorem on the Limiting Distributions.

As in [3], pp. 1254-55, it is possible to use a renewal theorem due to SMITH to examine the limiting forms of (1) - (4) as $t \to \infty$, whenever state j is recurrent and $b_{jk} < \infty$. The result when state j is transient follows from the fact that $\lim_{t\to\infty} M_{jj}(t)$ is finite. Paralleling, then, the proof of Theorem 7.1 in [3] , we may show that:

THEOREM

(i) If state j is transient, then $\lim_{t\to\infty} D_{jk}^{(i)}(y;\ t) = \lim_{t\to\infty} E_{jk}^{(i)}(x;\ t) = \lim_{t\to\infty} S_{jk}^{(i)}(w;\ t) = 0$, for all i , k and all $w,\ x,\ y \geq 0$.

(ii) If state j is recurrent and $b_{jk} < \infty$, then

$$\lim_{t\to\infty} D_{jk}^{(i)}(y;\ t) = D_{jk}^{(i)}(y)$$

(5)
$$= G_{ij}(\infty)\ p_{jk}\mu_{jj}^{-1} \int_0^y [1 - F_{jk}(z)]\ dz\ .$$

(6)
$$\lim_{t\to\infty} E_{jk}^{(i)}(x;\ t) = E_{jk}^{(i)}(x) = D_{jk}^{(i)}(x) \quad ;$$

and

$$\lim_{t\to\infty} S_{jk}^{(i)}(w;\ t) = S_{jk}^{(i)}(w)$$

(7)
$$= G_{ij}(\infty)\ p_{jk}\mu_{jj}^{-1} \left\{\int_0^w [1 - F_{jk}(z)]dz - w[1 - F_{jk}(w)]\right\},$$

where it is understood that if G_{ij} is a lattice d.f. then both t and w, x, or y may only take on values equal to multiples of its span.

4. Remarks

In many instances, the identity of the last previous and the next following transitions are unknown. Summing over the indices j and k, we obtain

(8)
$$D^{(i)}(y) = \sum_j P_{ij} \int_0^y \eta_j^{-1} H_j^c(z)\ dz \quad ;$$

(9) $$E^{(i)}(x) = D^{(i)}(x) \quad ;$$

and

(10) $$S^{(i)}(w) = \sum_j P_{ij} \left[\int_0^w \eta_j^{-1} H_j^c(z)\, dz - w\eta_j^{-1} H_j^c(w) \right] \quad ;$$

where the P_{ij} are the stationary (time-average) probabilities of the MRP, and $H_j^c(.) = 1 - H_j(.)$. Various conditional distributions can also be obtained.

In the event that the underlying Markov chain has a single recurrent class which is positive, it is well known that $P_{ij} = P_j$ for every j in that class, zero otherwise. In other words, formulas (8) - (10) become independent of the starting rate, and provide generalizations of the distributions of deficit, excess, and spread of renewal theory. (cf., for example, [1]).

5. Applications

The above results arose in the investigation of a model of human performance, where it was necessary to determine how random sampling of the length of job in progress biased the actual distribution of lengths of jobs of a certain type.

The results also have application in calculating certain terminal rewards which might arise in Markov-renewal programming [4].

REFERENCES

1. Cox, D. R., Renewal Theory, Methuen Monograph, J. Wiley and Sons, Inc., N. Y., 1962.

2. Pyke, R., "Markov Renewal Processes: Definitions and Preliminary Properties," Ann. Math. Stat., Vol. 32, No. 4, Dec. 1961, pp. 1231-1242.

3. ________, "Markov Renewal Processes with Finitely Many States," Ann. Math. Stat., Vol. 32, No. 4, Dec. 1961, pp. 1243-1259.

4. Jewell, W. S., "Markov-Renewal Programming: I and II," Operations Research, Vol. 11, No. 6, 938-971 (November-December, 1963).

NEAR-PARALLEL CONSTRAINED OBJECTIVES IN INTEGER PROGRAMS

by

David W. Matula

Some algorithms for the solution of integer programming problems (e.g. [1], [2]) search for the optimal integer valued solution in a way that necessitates examining all feasible extreme point solutions where the value of the objective exceeds its value at the optimal integral solution.

The following example shows that this search can be a prohibitively costly method.

Example:

Maximize $z =$

$$\left(1 + \frac{1}{3}\right)x_1 + \left(1 + \left(\frac{1}{3}\right)^2\right)x_2 + \left(1 + \left(\frac{1}{3}\right)^3\right)x_3 + \ldots + \left(1 + \left(\frac{1}{3}\right)^{n-1}\right)x_{n-1},$$

subject to $x_i = 0, 1, \quad i=1,\ldots,n,$

$$2x_1 + 2x_2 + 2x_3 + \ldots + 2x_{n-1} + x_n = n ,$$

where n is an odd integer, $n = 2k + 1$.

Since n is odd, x_n must equal unity in any feasible integral solution. Observe that the constraint is symmetric in the first $n-1$ variables and that exactly $\frac{n-1}{2} = k$ of them must be unity in any feasible integral solution. It is clear that z will have a maximal value, z^0 , over the integral solutions if the k x_i's with largest coefficients in the objective are picked as the unity valued x_i's .

The optimal integral solution is therefore:

$$x_1 = x_2 = x_3 = \dots = x_k = 1, \; x_n = 1$$

$$x_{k+1} = x_{k+2} = \dots = x_{n-1} = 0$$

$$z^0 = k + \sum_{i=1}^{k} \left(\tfrac{1}{3}\right)^i$$

Now for any $1 \leq j_1 < j_2 < j_3 < \dots < j_k < n-1$, $1 \leq j_0 \leq n-1$, where $j_0 \neq j_i \quad i=1,\dots,k$, let

$$x^0_{j_i} = 1 \qquad i=1,\dots,k$$

$$x^0_{j_0} = \frac{1}{2}$$

$$x^0_m = 0 \quad m \neq j_i \qquad i=0,1,\dots,k$$

To show that this is a basic feasible solution observe that it satisfies the constraint and that if it were the midpoint of two solutions p , q , both p and q would have to have unit coordinates for variables j_1 , $j_2,\dots,j_k$, zero coordinates for variables $m \neq j_i$, $i=0,1,\dots,k$, and to satisfy the constraint $x_{j_0} = \frac{1}{2}$. But then $p=q$ and hence we have an extreme point solution.

The value, z'' , of the objective for this solution satisfies

$$z'' \geq k + \frac{1}{2} = k + \sum_{i=1}^{\infty} \left(\tfrac{1}{3}\right)^i > k + \sum_{i=1}^{k} \left(\tfrac{1}{3}\right)^i = z^0$$

Hence this solution is a basic feasible solution with a value of the objective exceeding its value for the optimal integral solution. But there are $k\binom{2k}{k}$ ways of picking the j_i above, each yielding a different z''. I.e., with only fifteen variables $(n=15)$, 24,024 such basic feasible solutions would exist.

The essential difficulty in the above integer program is the condition that the objective is "nearly parallel constrained"; that is, the coefficients of the objective are approximately proportional to the coefficients of a constraint. The example given here, where all the coefficients in the objective (except one) are themselves about equal, was chosen only to highlight the difficulty, and its very special structure should not cause it to be considered a mere pathological example.

Indeed if any subset of the coordinates of the objective, independent of their own relative size, are all nearly the same constant multiple of the corresponding coordinates of a constraint (or linear combination of constraints), all basic feasible solutions to the program that differ only among these variables will yield approximately the same value of the objective. This cluster of basic solutions might all have to be searched before arriving at the optimal integral solution. As there are in general many constraints and many subsets of variables that could be nearly parallel constrained, this phenomenon should not be dismissed as unlikely in a real problem.

REFERENCES

[1] Szwarc, W., The Mixed Integer Linear Programming Problem When the Integer Variables are Zero or One, School of Industrial Administration, Carnegie Institute of Technology, May 7, 1963.

[2] Elmaghraby, Salah E., An Algorithm for the Solution of the "Zero-One" Problem of Integer Linear Programming, Department of Industrial Administration, Yale University.

REVIEW:

PROGRAMMING UNDER NONLINEAR CONSTRAINTS BY UNCONSTRAINED MINIMIZATION: A PRIMAL-DUAL METHOD

By A. V. Fiacco and G. P. McCormick, Research Analysis Corporation Technical Paper RAC-TP-96, September, 1963.

This paper is concerned with the mathematical programming problem: Minimize $f(x)$ subject to $g_i(x) \geqq 0$ $(i = 1,\ldots,m)$. The method of solution follows a heuristically presented idea of C. W. Carroll--The Created Response Surface Technique for Optimizing Nonlinear Restrained Systems, Operations Research 9(1961), 169-185--in which an effort is made to transform the original problem into a sequence of unconstrained minimization problems in order that existing techniques (e.g., steepest descent) might be brought to bear. Carroll proposed the minimization of functions

$$P(x,\ r) = f(x) + r \sum_{i=1}^{m} W_i/g_i(x)$$

where r and the W_i are positive constants and the sequence of minimizations arises from the stipulation that $r = r_1, r_2, \ldots$ where $\{r_k\}$ is a strictly decreasing sequence converging to 0. (The function $P(x\ ,\ r)$ is the created response surface.) It was conjectured that for each r_k, an x_k would be determined such that $\lim_k x_k = \bar{x}$, where $\bar{x}$ solves the original problem.

The authors (Fiacco and McCormick) make the proposal rigorous by stating hypotheses clearly and proving the necessary theorems. It is shown that the constants W_i in $P(x, r)$ may all be taken as 1 . A considerable amount of computational experience is included.

It is worthwhile listing the hypotheses which are made in the course of the development:

1. $R^o = \{x \mid g_i(x) > 0 , i = 1,\ldots,m\}$ is non-empty.
2. f , $g_i (i = 1,\ldots,m)$ are twice continuously differentiable.
3. For every real number k , the set of feasible points such that $f(x) \leq k$ is bounded.
4. $f(x) \geq v_0 > -\infty$ for all feasible x .
5. A method exists for obtaining an element of R^o .
6. f is a convex function.
7. g_i is a concave function $(i = 1,\ldots,m)$.
8. $P(x, r)$ is strictly convex for all $x \in R^o$ and all $r > 0$.

The proof of convergence of the method to an optimal solution includes conditions 6 and 7 which constitute the assumptions for convex programming. The authors contend that conditions 1 - 5 are "not very restrictive." Certainly condition 1 is necessary to the method, and is not unusual. Without condition 4, there would be no problem. Fiacco showed [Operations Research 9 (1961), 184-185] that if R^o is non-empty, but none of its elements is at hand, then one can be determined by use of the method itself. Hence, condition 5 is not particularly troublesome. Condition 8 can obviously be met in several ways, one of which is by the inclusion of the constraints $x_i \geq 0$ $(i = 1,\ldots,n)$ into the problem. Conditions 2 and 3 are probably more restrictive than the others. Note that the latter precludes any horizontal asymptotic behavior on the part of f, and hence the optimal solution will occur at a finite point.

It appears that the sense in which this is a "primal-dual method" is that the successive minimizations of $P(x, r)$ yield feasible solutions to both the primal and dual problems and hence valuable information on convergence.

Richard W. Cottle

REVIEW:

La Méthode du Chemin Critique - Application aux Programmes de Production et d'Études de la Méthode P.E.R.T. et de ses Variantes (The Critical Path Method - Application of the PERT Method and its Variants to Programs of Production and Planning), A. Kaufmann and G. Desbazeille, Dunod, Paris, 1964, 180 pages, 24 francs.

This monograph provides an introduction to the now-classical methods of time-only (PERT, etc.) and cost-time (CPM, etc.) scheduling of projects, in which a network is used to describe the partial ordering between the jobs or activities of the project. The scope of the book is fairly indicated by its Table of Contents, and some of the sub-topics:

Chapter 1. Graphs and Orderings. A Graph, Strict Order Relations in a connected graph without circuits [cycles], Decomposition of a graph into levels, Finding an ordering.

Chapter 2. Establishing a Research or Production Program. Representation of a program by a graph, Critical path, Float intervals and job margins, Finding the critical path [two algorithms and one numerical example], Two examples of application, Jobs with random duration.

Chapter 3. Generalization of the PERT Method, [á la H. Eisner, Opns. Res., Vol. 10, No. 1, 1962].

Chapter 4. Optimization of the Cost Economic Function. Decreasing the total cost of a program, Acceleration of a program at least cost [numerical example], Arbitrary variation of operating cost as a function of the duration, Linear variation and the parametric linear program, Fulkerson's [and Kelley's] algorithm, Details of the iterative procedure, Example, Optimization of a program when the duration of the jobs are random variables [not really].

Appendices. The β Distribution.
Mean Value of a Critical Path [á la D.R. Fulkerson, Opns. Res., Vol. 10, No. 6, 1962].
Activity Subdivision and Realization of a Schedule. [á la T.L. Healy, Opns. Res., Vol. 9, No. 3, 1961].

As in other works by Kaufmann, the main strengths of this monograph are that it is timely, the explanations are clear, and the authors do not hesitate to give detailed numerical examples. Certainly the exercises involved in searching all of the current literature to extract the pertinent points was no doubt very tedious, and the authors are to be commended on eliminating much of the sophomoric philosophy and black magic present in most popularized accounts of the subject. The reviewer is in complete agreement with E. Ventura, who states in the preface:

> "I am inclined to think that [the critical path mathod's] success is, in a large part, due to the fact that, in contrast to the other methods of O.R., it does not require a previous elaborate mathematical training, and makes use of concepts which are fairly intuitive."

and with the authors' conclusion that:

> "The critical path method in application favors 'exactness', 'clairvoyance', and also 'audacity'; it is a good example of the use of mathematics in management." .

To this reviewer, it was interesting to see how Anglo-Saxon terms become transposed into French. To avoid the Academic conflict of "minimer" versus "minimaliser", the authors use [MIN] throughout; for "labelling", they use, variously, "etiquetage", "affichage", and "marquage". Students of graph and network flow theory should be warned that "arc", "arête" [edge], "chemin" [path], "chaîne", and "circuit" are sometimes used differently than in the U.S. literature. "Ordonnancement" can mean simply [arrangement in regular order, or in O.R. contexts can mean "sequencing" or "scheduling", or both. Technical French is changing, also; in addition to the obvious coinings "bétonnage" [concreteing], "habillage" [making-ready], "compactage" [compacting], it is also apparently proper to refer to "grutage" ["craneing"]. Sometimes the French is better; how marvelous it would be to have a title like "organisateur", or "planifacteur"! The "float" or "slack" associated with jobs becomes "marge libre", "marge totale", or "marge certain", which makes a happy distinction with the slack in an event, which is simply "intervalle de flottement" (except in a marvelous typographical error on p. 42, where it becomes "intervalle de frottement"!).

There is some uneveness in the presentation. The very first sentence of the book is, approximately,:

> "Without abusing the theoretical aspects of finding ordinal relations in a graph, this concept being, as we shall see, none other than a set of elements between which oriented connections exist, we think that it is suitable to recall certain terms from graph theory,and to show how to use various procedures permitting one to achieve ordinal relations".

Even assuming the authors wish to avoid philosophy, it seems a small amount of motivation would be welcomed at this point by the non-technical reader.

Surely not all the graph-theoretic concepts in Chapter I and the set-theoretic notation in Sections 2, 10, 18, 27, 28 and Appendix II are needed to state the algorithms, since no proofs are given, and the examples presented do not require it.

Also, giving a separate Chapter to Eisner's Generalization, and Section 20 to The Entropy Function seem maladroit, when the space could better be used to explore cost-time methods, explain parametric studies, discuss resource loading, give a list of available computer codes, or discuss other important omitted topics.

This reviewer also feels there are some pedagogical limitations in the monograph. For example, the reader is assumed to know about linear programming, duals, and network-flow theory, but this knowledge is never used, since nothing is proven. In fact, the two examples of cost-time solutions only show cases where the jobs are contiually compressed (or remain fixed) as the project time decreases. Pity the poor reader who tries an example of his own, and finds some job being extended (bounded flow variable re-entering the basis)!

As a personal bias, the reviewer feels that the Beta distribution is, as usual, over-emphasized. The authors are never quite clear as to what a random job duration means (a one-shot "personal probability"?), whether the scheduling is to be performed before or after the job durations are known, what corrective action will be taken if the scheduling is incorrect, and so on. The analysis of Section 32 , which purports to carry out cost-time optimization with random job durations, is in fact misleading, unless one reads very carefully between the lines to find out the real assumptions of the analysis. Thus, although the monograph provides a good survey and clarification of the existing literature, the feeling persists that the authors have not completely integrated the material with their own experience, or added research results of their own.

Finally, the reviewer had hoped that a little chauvanism would be shown in discussing Fench research in the critical path area. Although little known outside France, the contributions of B. Roy and others, beginning from graph theory and yet motivated by real scheduling problems, provided a parallel and coincident stream of independent scheduling methods, quite similar to the American developments described in this book. (Cf. , for example, B. Roy, Contribution de la Theorie des Graphes a l'Etude de Problemes d'Ordonnacement, Proceedings Second International Conference on Operational Research, Aix-en-Provence, France, September, 1960, Dunod, Paris).

Summary of Tasting: A faithful pressing by a well-known chateau of a famous stock imported from California. Body: adequate. Color: clear, some sediment. Bouquet: heady. Alcoholic content: high, some residual fermentation. Price: popular. Conclusion: needs more maturing.

William S. Jewell

REVIEW:

Les Problèmes d'Ordonnancement - Applications et Méthodes (Scheduling Problems - Applications and Methods), Number 2 in Monographies de Recherche Operationelle, A.F.I.R.O., "by a group of specialists led by B. Roy", Dunod, Paris, 1964, 151 pages, no price given.

This monograph may conveniently be regarded as a supplement to La Methode du Chemin Critique (see Review, this issue); it provides further elaboration of critical path mathods, showing how the theory may be extended somewhat to include other problems of scheduling, and providing several clear, interesting examples of French application of these methods. To a certain extent, this monograph answers the reviewer's objection that La Methode did not sufficiently stress early French contributions to the area of project scheduling. Part of this seeming reluctance is, of course, due to the fact that our emphasis on "publish or perish" has yet to reach the Continent.

The monograph's raison d'etre is given in the Forward:

> "It is observed, when scanning specialized journals of all types, that scheduling problems are beginning to take on increasing importance. Not only the mathematicians work increasingly each day on theoretical plans, but even industrialists hesitate less and less in confiding to operational chercheurs the difficulties they meet with in their planning.
>
> These works remaining, nevertheless, very dispersed, the French Society of Operations Research [So.F.R.O.] judged it desirable to assemble into a single volume, a certain number of texts, for the most part already published on the occasion of seminars, colloquia, or congresses, which would form a sample of French work achieved during recent years, in the domain of industrial applications, as well as in theoretical research. Thus, we hope to offer to the reader, in a relatively limited number of pages, the means to get an idea of the current state of the entire subject."

Because of the variety of the various contributions to the monograph, we shall discuss each Chapter separately.

Chapter I. Physignomy and Features of Scheduling Problems (B. Roy)
This introductory article describes the general problem of scheduling "something" by decomposing it into "tasks", and then choosing a starting time and duration for these tasks. This choice is limited by three principle types of constraints, which Roy calls "potential" (PERT-type), "disjunctive" (mutually exclusive use of facilities), and "cumulative" (resource-loading). The article concludes with a comparison of the two types of graphs which may be used to describe the partial orderings of the tasks: "Task-Potential" graphs (those used by Roy and co-workers), and "Stage Event-Potential" graphs, (the familiar "arrow" diagram used in the U.S.A.).

Chapter 2. Operations Research in Construction (P. Pacaud).
This Chapter is primarily an extended example of the use of critical path methods in constructing an industrial building with five bays. Some attention is devoted to the possible savings in decomposing the various "critical" operations for the entire building into jobs for the individual bays, thus reducing total time. There is also discussion of indirect and direct costs in building as a function of the total job duration.

Chapter 3. Scheduling of Operations and Personnel in the "Coffering-Tunnels" (J. deRosinski).

An industrial construction example is described which consists of repeated use of "coffering-tunnels" in building up successive stories of the main framework of an apartment building. The primary limitation in using critical path methods was the need for "craneing" during "rotation", when the "tunnels" are put in place on the new "cell" to be concreted. This imposes a "disjunctive" constraint, which was solved by heuristic means.

Another limitation on possible schedules was the need to maintain a proper "rhythm" of "work cadence", i.e. to schedule the manpower shifts, number of men in a given area, lunchtimes, etc. in a realistic, repetetive manner.

Chapter 4. Maintenance Scheduling in a Refinery (D. Carré).

A general discussion of the critical-path method as used to schedule planned maintenance in a refinery, attempting a realistic evaluation of advantages and disadvantages of the method.

Chapter 5. Scheduling Outfitting Operations at Chantiers del'Atlantique (Ph. Guitart)

The principal engineer of Chantiers d'Atlantique describes the problems of scheduling outfitting on the liner "France", carried out with the aid of the Société d'Économie et de Mathématiques Appliquées (S.E.M.A.) and Compagnie des Machines Bull. In this application the technological sequence problem was secondary to the problem of maintaining a smooth manpower-loading curve of the various manpower "corporations": "metalwork", "woodwork", "ventilation", "piping", "painting", "electricity", and "sail-making". Because of the many internal and external constraints on the "starting date" and "elasticity" [rate of work] of each job, the schedule was found by means of a "segmented" form of linear programming, which permitted rapid visualization of the manpower loading curves, and selection of a desirable schedule. Even though it was at first thought possible to find a "best" solution directly

> "....indeed, it was impossible to define a numerical indicator which permits one to satisfactorily classify the various possible schedulings relative to each other. For, whatever the definition of this indicator, there will always correspond a well-defined value for a given schedule, but, unfortunately, the converse is not true."

Chapter 6. Contribution of the Theory of Graphs to the Study of Scheduling Problems (B. Roy).

This article is essentially the same paper presented to the Second International Conference on Operational Research at Aix-en-Provence in 1960. It is of interest, primarily because it indicates the early French developments in scheduling theory, which were suprisingly parallel, and in some sense, better than the early American P.E.R.T. developments.

Chapter 7. Using the Fortran-Compiled-List-Processing-Language in the Solution of a Scheduling Problem (P. Darnaut and G. Sandier)

This article describes the use of a list-processing language in developing a computer program for solving PERT-type problems; it is stated that this type of language is advantageous in retaining a compact description of the network in rapid-access memory. The program written for an IBM 7090 handles 3,000 jobs, and gives the starting dates "wedged to the right and to the left", ie. both "earliest" and "latest".

Chapter 8. Scheduling Problems with Disjunctive Constraints (Ph. Tuan Nghiem).

The last Chapter studies the problem of "disjunctive" constraints in scheduling problems when jobs belonging to certain sets are required to be carried out during disjoint intervals of time, as for example, jobs using the same resource.

> "The set of minimal schedulings is studied and a procedure is found which permits one to move over the interior of this set, which has the structure of an "arborecence" [forest], by passing from one schedule to all of its successors by a simple change in the order of carrying out the tasks subject to the constructive constraints."

All in all, this collection represents an interesting group of articles written at consistently intelligent, yet uncomplicated, level--- something which is not true of most American articles on scheduling. If the future finds the French publishing more of their applied research, we shall have and welcome a high-level competition in the field of O.R.

It is interesting to note that this monograph resulted from a study group organized by the Société Francaise de Recherche Opérationnelle (So.F.R.O.), but that this group merged in March 1964 with the Association Francaise de Calcul et de Traitement de l'Information (A.F.C.A.L.T.I.), so that the monograph has the imprimatur of Association Francaise d' Informatique et de Recherche Opérationnelle (A.F.I.R.O.), (7^{ter}, Rue de la Chaise, Paris, 7^{ieme}, France). Is this also a trend of the future?

Summary of Tasting: An interesting variety of France "amuse-gueule", very suitable for a "dégustation" with La Méthode du Chemin Critique.

William S. Jewell

REVIEW:

Scientific Decision Making in Business: Readings in Operations for Nonmathematicians, A. Shuchman, Rinehart and Winston, 1963.

The author describes his book as "a collection of the best writing I have been able to find that describes the aims, methods and tools of management science or operations research without recourse to technical jargon or complex mathematical symbolism," and he supplements the papers with connective discussion so as to present a unified whole.

Whether the articles represent the best selection of nontechnical papers available, the writer is not able to judge since he has not made a study of the literature with Shuchman's aim in mind. As a whole, the readings do present a cross section of viewpoint, and the book is the first comprehensive, nontechnical attempt at examining the role of operations research in business, which also presents a synthesis of many points of view.

Turning now to the specific articles in Shuchman's book, Shuchman is to be complimented on the general content and organization of the material. There are four major parts to the book: Part 1, What is Operations Research; Part 2, The Methodology of Operations Research: Models and Model Building; Part 3, The Methodology of Operations Research:

Techniques; and Part 4, Some Applications of Operations Research.

In Part 1, the first paper, by Dean R. Wooldridge, on the scientists' invasion of the business world is not a very penetrating discussion of this subject. It introduces the Operations Research Specialist as a scientist called in to make an obvious observation on a military operating problem, and gives the reader the impression that operations research is another form of traditional management consulting - now a respectable occupation for scientists who bring the virtues of objectivity, quantitativeness and a capacity for learning new fields. The interest of scientifically trained people in the operating and planning problems of business is, however, real and significant.

The next paper, by C. C. Herrmann and John F. Magee, under the title of Operations Research For Management, gives the usual party line of Operations Research as an introduction of the scientific method into business decisions, defining models and measures of effectiveness to assist in choosing among alternative courses of action. Here Operations Research is described as overlapping but not identical with Statistics, Accounting, Marketing Research and Industrial Engineering, with the latter field distinquished from Operations Research not in problem content but because "industrial engineers usually apply established methodologies to their problems-" a somewhat myopic observation. Equally near-sighted are the comments that "their work is generally restricted in scope to manufacturing activities" and "industrial engineering is not commonly characterized by the mental discipline and techniques of analysis that are commonly associated with the physical scientist."

The next three papers written by Melvin L. Hurni provide the best and most complete answer to the question, "What Is Operations Research" by discussing the developments that make Operations Research possible, the needs and opportunities for Operations Research and the basic processes of Operations Research. As a trilogy, this discourse argues that business decisions are rational, that the process involved is becoming too complex for judgement alone, with increasing interactions between its parts through automation, requiring more integration quantitatively in planning and control, and that business is a flow process requiring integrated planning in time, all of which provide challenge and opportunity for Operations Research to provide appropriate quantitative models for analysis and synthesis in decision making. Hurni's discussion of the basic processes of Operations Research follows the usual pattern of defining the problem, model building, evaluating alternatives and implimentation. It fails to make clear the families of Operations Research models which can be used in the more important decision making areas. The last paper in Part 1, by Akoff, is a didactic repetition of the other papers.

The papers of Part 2 on Models and Model Building are generally clear, useful and would be good reading for scientifically trained people as well as students of business. One paper by Robert S. Weinberg on The Uses and Limitations of Mathematical Models attempts to give a detailed description of the kinds of mathematical models used in Operations Research, particularly those related to marketing. This discussion fails to provide a clear definition of behavioral equations as probabalistic models, and, in fact, is misleading as to the character of a stochastic model. Weinberg's behavioral equations are simple statistical regressions without modeling of the underlying process, being at most statements about averages in which there is considerable ambiguity in identification of parameters, and, he fails to explain the hazards involved in extrapolation and manipulation known to most econometricians today, resorting finally to justification by stating that such models "help us to systematically organize our data for analysis." The principal deficiencies of this part of Shuchman's collection are: (a) it does not provide a good description of the activity analysis model of linear programming, despite the long paper by Vazsonyi on the Uses of Mathematics in Production and Inventory Control, (b) no good examples of stochastic models are given, such as those used in queueing and inventory theory. The last paper by Ackoff on Prototype Models provides a reasonably complete classification of Operations Research models, but is weak in clarifying the quantitative structures involved.

Part 3 on Operations Research Techniques is by far the most complete, comprising half the book in number of pages and carrying the reader through most of the recognized techniques. Shuchman classifies techniques into those for coping with Complexity, Variability, and Lack of Information, which is not bad, although one wonders why he includes Factor Analysis under Complexity rather than Variability where he covers statistical techniques. Dynamic programming is classified under Complexity rather than Variability, which does not emphasize the real significance of this method, but one must admit that all classification systems create arbitrary compartments.

Under Tools For Coping With Complexity, there are papers on Mathematical Programming, Dynamic Programming, Symbolic Logic and Factor Analysis.

The leading paper on Mathematical Programming by Henderson and Schlaifer is a good account of why and when linear programming may be used in a variety of single period, static allocation situations, involving many interacting variables not easily handled by trial and error methods. It does not indicate the full potential of the technique for multistage decisions over time, nor the limitations of uncertainty related thereto, and may be regarded by the "Executive" as a computer solution procedure for rather trite problems in the maze of his decision making. It would be useful, also, to add here a good nontechnical discussion of Network Flow Analysis extending into Critical Path Methods such as PERT. The second paper on programming is an anti-climax and does not add much to the discussion.

The three papers on dynamic programming emphasize the multistage character of the technique, but treat only deterministic problems and the discerning reader may rightfully surmise that mathematical programming could equally well be used to handle the class of problems discussed. Thus, the real significance of Dynamic Programming is lost, namely to provide optimal policies under situations of uncertainty, having the character of feed-back control, i.e., a policy stating do this or that depending upon the realized values of the random variables involved in the actual operation of the decision process studied. Shuchman's description of policy as terminology misses the point. The combinatorial limitations of Dynamic Programming could have been clarified in contrasting this technique with Mathematical Programming.

The single pa er on Symbolic Logic will undoubtedly lose the nonmathematical reader, although he may grasp the significance of propositional calculus.

The two papers on Factor Analysis are overly simplified discussions of Analysis of Variance which may be understoo. better as topics under Tools For Coping With Variability, and the statistical implications need clarification.

The section headed Tools For Coping With Variability starts with papers explaining the concepts of mathematical probability. The one by Warren Weaver is the best, conceptually. Some confusion is introduced by the paper of Kurnow, Glasser and Ottman, which defines mathematical probability as a relative frequency of an event in an infinite sequence of trials, rather than treating it as an idealized model for comparison with experimental trials - and the term probability measure is used without defining the space on which the measure is defined. These matters of rigor are not trivial, because a good deal of confusion about probability arises from the fact that the measure and event space are not carefully defined.

The next sub-section under Variability is devoted to Queueing Theory and consists of a single paper written by Shuchman which gives an adequate discussion of elementary concepts, primarily in terms of optimizing the number of channels in a service facility. Following this, the two papers presented on the subject of Decision Theory do not attempt any discussion of sequential decisions. Instead they are confined to subjective probabilities and criteria for optimizing decisions, with little or no development of model structures in risk analysis. The reader will leave this section with no clear conception of a theory of decision. Perhaps the subject itself is vague.

The final sub-section of Tools For Coping With Variability contains two papers on Game Theory, defined in the first paper by Shubik as a study of decision making in situations of conflict. Shubik's paper is a good discussion of games, strategies and the potential usefulness of game theory in operations research.

The last section of Part 3 on Techniques contains several papers under the heading Tools For Coping With Lack of Information, discussing mainly topics in Statistical Inference and Simulation suitable for a completely uninitiated reader.

Part 4 (Some Applications of Operations Research) has three classes of papers describing operations research in Production Management, Marketing Management and Financial Management, characterized by Shuchman as "not fully representative of the Operations Research work they are intended to represent," since limited mastery of techniques presumed for the reader has prompted him to select papers which he can at most expect to impart "a deeper appreciation and understanding of the operations research approach to business problems." The papers dealing with Production Management are confined to economic lot sizes and inventory control. Little or no examples are given showing how operations research has been applied to scheduling, production planning, distribution, congestion in work flow, allocation of facilities, etc. Those on Marketing Management are equally well limited in scope, discussing statistical analyses for measurement of relationships between sales and promotional effort to dollar volume, with one paper on a large simulation study to determine number and location of warehouses for a national distributor. No studies of brand switching or other marketing phenomena are given in which stochastic models are used. The papers on Financial Management deal with elementary sampling and statistical estimation in accounting and rudimentary use of game theory in capital budgeting and study of mergers and acquisitions. As Shuchman has warned, his illustrations of applications of operations research give the impression that uses of operations research in business are at a low level of rationalizing decision making.

Where do we stand, now, after this detailed review of Scientific Decision Making In Business? Are there major areas of operations research omitted? Evidently Reliability is not regarded as part of business decision making. In some industries, notably Electronics, reliability of performance is an over-riding consideration. The nature, role and problem areas of computer systems is more or less ignored, yet business is deeply engaged in such systems for data processing, record keeping, performance of clerical tasks and analysis. Here Information Theory deserves consideration. Also a more adequate discussion of control processes in business would be desirable. Replacement and Maintenance of Equipment is another important area not receiving adequate discussion.

Strangely, except for elementary notions of costs, modern economic-theoretic models and concepts in operations research are not seriously considered. Specifically the role of operations research in evaluation of opportunity costs and planning capital investments deserves more discussion - particularly linear economic models ought to be of interest. The whole area of forecasting, which is of great importance to business, is susceptible of operations research study and certainly more useful than abstruse discussions of Decision Theory.

It is difficult to escape the feeling that Shuchman's collection of best nontechnical writings is at too low a level, neither impressing the hard headed executive with the accomplishments of operations research, nor stimulating the brighter students of business. Yet the book may be the best compendium available.

Shuchman appeals to business executives and students of business administration for serious attention to the "new management technology" as being "concerned with the very core of a manager's job - decision making." One may conjecture from the first paper "Operations Research: The Scientists Invasion of the Business World" that Shuchman correctly assesses that scientists are now regarding business management as a fertile field of endeavor, but he does not pursue the full implications of this development. Rather he argues for the need of managers and students of business "to make an effort to understand the discipline" of operations research, which "need not be that of an expert, but it must be sufficient to enable the executive to know when and how the expert can assist him."

Shuchman's concluding statement of his introduction:

> "Accordingly, the advent of Operations Research may very well mean that managerial decision making is to be, in the future, less of an art and more of a science. If this is the possible significance of Operations Research, then students of business, both in school and on the firing line, must examine it carefully."

has implications which should be elaborated further.

In the reviewer's opinion, managerial decision making and operating control of our modern industries and services is in the process of passing from professionally-trained or experience-educated students of business to operations-research-educated applied scientists. This is a natural result of the growth of our complex modern technology, where a mere familiarity with operations research is not sufficient to meet the challenge.

Heretofore, the problems of business have attracted individuals with a set of talents peculiar to the quid pro quo activities of the business arena: heuristic, intuitive reasoning aptitudes, coupled with an extroverted personality, strongly developed in an ability to understand or dominate other humans. Students with these qualities have prospered in the past, but now face a challenge to which they are fundamentally ill adapted. The technology of our business society is rapidly being changed into highly integrated, automated systems, driven by the machine and not the man, and requiring an enormous amount of data handling. Quantitative models developed by operations-research-scientists are being integrated with data processing systems, so as to automatically perform the large bulk of detailed decisions heretofore regarded as being the responsibility of management. This impersonal technology will perform better than middle management, both in energy, capacity and quality of decisions, and will ultimately obsolete the need and major employment opportunity for business students, as well as replacing a large number of clerical employees.

In time, executives may very well be recruited from scientifically educated individuals trained in computer technology and operations research, with some education in the soft sciences, since it is this background which will be required to understand the decision making system.

At this time, it appears that new curricula in industrial engineering will be providing most of these students, both because operations research has become a major component of the Industrial Engineering curriculum and because the students traditionally study economics, business administration and psychology against a common background of engineering technology.

Thus, it seems that students of business need to re-assess their motivations, aptitudes and educational goals. Those with a flair for science should seek deeper education in the mathematical sciences and operations research, obtain a real understanding of computer technology and de-emphasize the many courses which they normally take in institutional subjects, moving so to speak in the direction of Engineering. The others should realize that a revised management role is needed for Business Administration. Clearly the technology of Scientific Management falls short of defining objectives initiating new enterprises and organizing human effort. Leadership is evidently needed for this guiding function. Can students of business undertake this responsibility? Do they need a more liberal education rather than a somewhat competence in Operations Research? Are not studies of behavioral science, industrial economy, social objectives and public institutions more important.

Ronald W. Shephard

Following are some abstracts of dissertations and theses submitted to the University of California at Berkeley on subjects pertaining to Operations Research.

A Simulation Model of Airport Landing Capacity

By

Ernest Farnsworth Bisbee II

M S., Industrial Engineering, 1961

ABSTRACT

The undesirable prospect of increasing congestion on the nation's airways has caused attention to be focused more and more on terminal areas. Two distinct efforts to improve traffic flow have been studied elsewhere. One, by means of careful control of aircraft separation along flight paths leading to the terminal runways attempts to reduce flow variability, while the other attempts to reduce service times on the runways by the use of high speed exit taxiways. Within this paper, a Monte Carlo simulation is proposed in order to discover relations between error of control and traffic flow for optimally located exits on a runway used exclusively for landing. The simulation process is discussed at length and an expression is found for minimum sample size with stated confidence level.

Limiting Velocity on a Two Lane Road

by

Ernest Farnsworth Bisbee II

Ph.D., Engineering, 1962

ABSTRACT

The limiting average velocity is found for an infinitely fast test vehicle on a two-lane road in terms of the average vehicle density in each lane and the relative velocity between them. The problem is solved under the conditions of an infinitely long, straight, road without intersections on which the traffic has a fixed velocity and an arbitrary distribution of spacings between successive vehicles for each lane. The solution of this problem leads to the expression of a bound on the fundamental relation between flow rate, density, and velocity.

The methods used are principally those of renewal theory and the central problem is shown to be a two-dimensional general renewal process which is solved by a reduction to one dimension through probability arguments. The form of the solution is a simple expression which indicated the influence of passing opportunities on the limiting velocity. The model emphasizes moderate to high densities and relative velocities between the two streams.

Four Investigations in Operations Research

William O. Blattner

M.S., Industrial Engineering, 1961

i) Solving Operating Problems with Electronic Computers

ii) A Possible Solution to a Special Case of the Machine Loading Problem

iii) Simulation of Soaking Pit Operations at Geneva Works

iv) Sensitivity Analysis Illustrated

Scheduling of Near Economic Lots

by

James Douglas Cumming

M.S., Industrial Engineering, 1962

ABSTRACT

Economic lot sizes cannot readily be scheduled as production lots when two or more items share the same production equipment. A simple procedure is developed for allocating the available machine time among the various items so that a feasible manufacturing cycle results with production lots approaching economic lot sizes.

A Delivery-Lag Inventory Model with Emergency

By

Klaus Hermann Daniel

Ph.D. in Statistics, 1961

ABSTRACT

In a recent article[1] E.W. Barankin investigated a one-period, one-commodity inventory model with a one-period-lag delivery of "regular" orders and with a built-in emergency provision which allowed the possibility of placing an "emergency" order with instantaneous delivery. The quantity to be purchased in the case of emergency was assumed to be a fixed amount m . The problem consisted in finding an optimal regular ordering policy and an optimal emergency policy so as to minimize a payoff function specified by the structure of the model. It turned out that, if the stock level lies below a certain critical number, an emergency situation occurs, which causes an emergency order of size m . This raises the initial stock level x , which can be any real number, to the starting stock level $x + m$. If then the starting stock level is smaller than another critical number, an optimal regular ordering policy calls for an additional purchase of items.

The model presented here differs slightly from the one proposed by Barankin. We do not restrict the emergency quantities $\{m_i\}$, $i=1,2,\ldots,n$

[1] Arrow, Karlin and Scarf, Studies in the Mathematical Theory of Inventory and Production," Stanford University Press, 1958.

to be constant, but only require that there is an upper bound m_u on the m_i's . Subject to this restriction we search for an optimal emergency policy of an n-stage problem. This modified approach results in an optimal emergency policy described by certain critical numbers. The critical number obtained for the one-stage problem differs now from the one establised in [2]. The optimal emergency quantity to be ordered becomes a piecewise linear function, defined on the set of initial stock levels.

Both models reflect certain assumptions which will be met in many practical problems. The model considered by Barankin reflects the situation that an order for immediate delivery interferes with scheduled deliveries of items carried out by a supplier. Although a higher unit cost is charged for such an order, compared to the unit cost for regular delivery, the supplier may not be albe or willing to deliver instantaneously any amount of items. If the size of the order is too small the delivery costs for the supplier may offset his revenue resulting from such an operation. If the ordered amount is too large it may exceed the quantity available at the supplier's storehouse. A compromise between these two limitations is then expressed by the assumption that only a fixed amount of items can be delivered without delay.

Our model does not take into account a positive lower bound on the emergency quantities to be purchased. Any amount, so long as it is limited from above by, say, m_u will be available in case of emergency. This less restrictive condition on the class of emergency orders will result in an optimal expected loss function (for any n-stage problem) which is smaller than or equal to the optimal expected loss function of the model stated in [2]. But what is even more important, is the fact that the extension to the two-stage problem of the model treated in [2] presents great difficulties, since the optimal loss function in the on-stage problem is the minimum of two convex functions, which is not necessarily a convex function. However, the modification outlined above reduces the optimal loss function to a convex function. The convexity property achieved in this way remains invariant for the general n-stage problem as will be shown in the treatment to follow.

This opens the way for the solution of the n-stage problem of our model after a proper class of cost functions entering the model has been chosen.

Optimization of the Design of Open Pit Mining Systems

by

Beverley Chew Duer

D. Eng., Industrial Engineering, 1962

ABSTRACT

The optimization of equipment procurement and operational policy dominantly reduces mining costs. Capital and operating expenditures for open pit mine materials handling depend upon the geometry of, and schedule of ore removal in the ore deposit; topography; mill, storage, and transportation facility locations; equipment operating characteristics and conditions; the financial capability of the mining company; and the schedule of customer demand. This analytic integrated systems study of open pit operations and equipment selection guarantees simultaneous evaluation of important variables, whose independent study is infeasible.

A hypothetical mine model, formed by quantization of the mineral deposit into geographical zones, levels and ore types and definition of the related haulage road net, is the basis for materials handling reduction. Stochastic analysis of open pit loading and haulage forms the basis for suboptimizing equipment type selection, road grade determination, and optimal balance between loading and haulage units for all locations in the mine model through the study of a network of independent, single-station cyclic queues with haulage units representing customers and the loading equipment, the servers. Finally, a smooth mine production sch-

edule (utilizing equipment parameters previously suboptimized, is developed by linear programming to optimize equipment utilization and determine capital equipment requirements and operating policy to meet customer demand. Production smoothing, order of ore removal, production capacity and schedule of customer demand constraints are utilized in this large-scale model.

Simulation analysis shows that unit mining costs can be reduced by 25-50% if the capacity and economic life of large scale equipment can be utilized. It is found by linear programming analysis that a 50-75% reduction in total equipment hour variations between adjacent scheduling periods may be accomplished within the set of all feasible smooth mine production schedules.

All required input data for this analysis are available or generatable economically, and all models are economically optimizable.

Network Flows, Graphs, and Integer Programming

By

Ellis Lane Johnson

Ph.D., Industrial Engineering, 1964

ABSTRACT

The connection is developed between graphical methods and the simplex method of linear programming for solution of network flows, flows with gains, and programming in an undirected graph. Determining dual variables, pricing out, and resolving degeneracy are shown to simplify for phase I of the simplex method applied to the network flow problem, and the resulting algorithm corresponds closely to the labeling procedure for network flows. Phase I of flows with gains does not simplify as much, but a special case, programming in an undirected graph, does simplify just as much as in network flows, and, hence, the primal-dual method has as much advantage. A solution of an integer program is given for phase I of programming in an undirected graph. The problem includes some previous work on degree-constrained subgraphs, matchings, and coverings. Alternating paths play a key role.

Matching and covering in a graph extends to an important class of integer programs: packing and covering in a family of sets. A discussion is given of two special cases of the packing problem: coloring a graph and finding a maximum set of independent vertices

of a graph. The general packing problem is reduced to the problem of finding a maximum set of independent vertices. No special device for integer programming is given for this class of problems, but a refinement of Gomory's cutting plane method for general integer programs is given. The algorithm is based on the simplex method with multipliers. Lexicographical ordering is replaced by an ad hoc perturbation technique, and a linear programming subroutine is used to assure finite convergence and to speed convergence.

An Investigation of the Burn In and Related Problems

By

Michael J. Lawrence

M.S., Engineering-Industrial, 1964

ABSTRACT

Two problems involving the derivation of bounds on a distribution with a decreasing failure rate (a D.F.R. distribution) are presented.

1. Given that an item has a decreasing failure rate, sharp upper and lower bounds on the burn in time to achieve a specified mean residual life are presented. The bounds rely only on the D.F.R. assumption and a knowledge of the first moment and a percentile of the failure distribution.

2. An early estimate of the five year survival proportion (commonly called the five year cure rate) is of great interest in assessing the value of a treatment for a mortal disease such as cancer. Assuming that the distribution of time to death is D.F.R. and assuming a knowledge of the mean and a percentile, sharp upper and lower bounds on the five year survival proportion are obtained.

In addition some bounds on the hazard rate and density of a D.F.R. distribution are stated and proved.

Stochastic Programming

By

S. M. Sinha

Ph.D., Statistics, 1962

ABSTRACT

This investigation is concerned with a linear programming problem, where the coefficients of the objective function, the constraint inequalities and the resources are random variables. Such a problem is known as a stochastic programming problem. The linear programming model for such a case however has no meaning and it is necessary to formulate a new model to deal with such cases. The uncertainty aspect of the problem suggests that it can only be solved probabilistically and hence one reasonable formulation of the stochastic programming problem requires that our activity levels should be such that with a certain preassigned (high) probability, the total quantities required for each item should not exceed the available quantities and at the same time should guarantee a maximum objective with a preassigned (high) probability.

The usual methods described in the literature for attacking this type of problem are not applicable because of the non-differentiability of expressions involving square roots of positive semi-definite forms.

Project Planning By Decomposition

By

Shailendra C. Parikh

M.S., Industrial Engineering, Feb. 1963

ABSTRACT

This study considers "critical path" networks which are used for the planning and scheduling of projects that consist of well defined sequences of individual activities. When the number of activities is large, it becomes difficult to prepare a network in the high-speed memory of a digital computer. Also, in the cases where there are two or more independent projects, which are weakly inter-related by common activities, the problem of efficient scheduling of all the projects becomes quite difficult.

This thesis presents a method to "tear" or "decompose" a project network into several subnetworks, schedule the subnetworks and then to put the subnetworks back together. Thus, it is possible to prepare subnetworks as individual units, to store the subnetworks in the memory of the computer and to schedule and revise (update) the projects as needed.

A computational algorithm is first given for time-only (PERT) networks; then, a computational algorithm is given for a cost-time (CPM) network of project subnetworks. The later algorithm is a generalization of an

algorithm due to Fulkerson, in order to handle piecewise-linear, convex, cost-time curves for each activity. Also, necessary conditions for solution of continuous, convex, cost-time curves for each activity are presented. In the appendices, an efficient time-only algorithm and a node re-ordering algorithm are given.

ABSTRACT

Analytical Models for Attack and Control of Wildland Fires

by

George Merriman Parks

Ph.D., Engineering, 1963

Decisions concerning the dispatch of suppression forces to attack and control a wildland fire must be based on the physical characteristics of the fire and the effectiveness of the forces sent against it. The behavior of a fire is affected by the fuel, topography, and weather conditions in the immediate area, and the effectiveness is determined by the type of force, the training, and the equipment. There is a dollar tradeoff between maintaining large and costly suppression forces to keep fires small and employing small forces but allowing fires to become large and cause extensive damages. These decisions have traditionally been based principally on qualitative criteria because of the lack of established quantitative relationships to describe the behavior of free-burning fires and the effects of various types of suppression activity thereon. The objective of this study was to mathematically describe these relationships to enable finding the size force such that the sum of these suppression and damage costs is minimum.

Three deterministic analytical models were developed to describe the basic tactics of fire fighting- direct, indirect, and parallel. The

size force necessary to keep total costs at a minimum was determined in each case under various assumptions concerning fire growth and suppression activity. This development was extended to include sensitivity analyses of the results, risk models where decisions must be made under uncertainty, the allocation of limited forces to simultaneous fires, and least-cost reinforcement policies. The relationship of elapsed time between the detection of a fire and the start of suppression activity and the size of force required for control at minimum cost was also investigated within the framework of each of these models. Two stochastic models were suggested for determining optimal strategies to employ in certain fire situations.

Comparison of actual suppression and damage costs for the calendar year 1959 on the Plumas National Forest with the results predicted by one of the above deterministic models indicated that significant dollar savings can be achieved in total fire-caused costs by increasing the size of the suppression force and/or reducing the time required for initial attack. The implications of these results for fire planning and budgeting were discussed and significant areas for further research in fire modelling were indicated.

Various statistical analyses were performed on data from approximately 14,000 actual wildland fires to determine historical values for some of the fire growth and suppression force effectiveness parameters employed by these models.

Recovery from Interrupted Service: A Queueing Model

By

Tracy F. Slaughter

M.S., Industrial Engineering, 1961

ABSTRACT

This thesis discusses, in physical and mathematical terms, the effect that a temporary interruption of service has on a servicing system where the steady state mode of operation is such that queues are small.

The mathematical model employed assumes Poisson arrival and exponential service distribution. Time dependent expressions are derived which describe the behavior characteristics of such a system following interruption of service. Specifically, it is possible to predict a) the average number of units in the system, b) the probability that a late arrival will encounter delay greater than a specified time, and c) the average delay that a unit entering the system will experience. These predictions are possible for the case in which the condition of the system at termination of interruption is not known as well as the conditional case in which queue length following interruption of service is known. The ability to predict such performance characteristics is of considerable value in both the initial design and the routine operation of such systems.

Programming under Uncertainty

By

Roger Wets

Ph.D., Industrial Engineering, 1964

ABSTRACT

The solution of a two-stage linear program under uncertainty can be obtained by solving an equivalent convex programming problem. We characterize the solution set of this equivalent convex programming problem as well as the functional to be optimized. Since we seek an here-and-now decision, this solution set and the functional are expressed in terms of the first stage decision variable. If the constraints of the initial problem, or the probability space of the random variable of the two-stage linear program under uncertainty possess certain properties, then the solution set of the equivalent convex program is a convex polyhedron. In this case, if the problem cannot be solved using existing solutions methods (e.g. linear programming with upper bounds, quadratic programming) one can use an algorithm developed in the last section of this thesis.

The constraints imposed on the here-and-now decision can be divided in two parts: first, the fixed (or technology) constraints where no uncertainty is involved; second, the induced constraints which limit our first stage alternative to those which allow a feasible solution to the second stage whatever be the values assumed by the random elements of the problem. The fixed constraints are expressed as

linear equations and inequalities in terms of the here-and-now decision variable. The induced constraints are usually not available in that form and one of the greatest difficulties one encounters when solving a linear program under uncertainty is to determine if a given decision satisfies or does not satisfy the induced constraints.

When there are no induced constraints one says the problem is "complete"; e.g. pure inventory models, transportation problems with uncertain demand. This problem leads to an equivalent separable convex programming problem. If the random elements of the problem assume some specific distributions functions, then the equivalent convex program is reducible to programming problems for which efficient algorithms exist. If the random elements are discretely distributed, then the equivalent convex program is a linear program with upper bounds; if the random variables are uniformly distributed, then it is a quadratic program. In general (when the random variables are not discrete or uniformly distributed) we suggest an algorithm for which an experimental computer program has been written.