

# An Approach to the Representation of Gradual Uncertainty Resolution in Stochastic Multiperiod Planning

Vicente Rico-Ramirez<sup>a</sup>, Ignacio E. Grossmann<sup>b</sup>, Bora Tarhan<sup>b</sup>, Salvador Hernández-Castro<sup>c</sup> and Juan G. Segovia-Hernández<sup>c</sup>

<sup>a</sup>*Instituto Tecnológico de Celaya, Av. Tecnológico y Garcia Cubas S/N, Celaya, Gto., 38060, Mexico, vicente@iqcelaya.itc.mx*

<sup>b</sup>*Carnegie Mellon University, Pittsburgh, PA, 15213, USA*

<sup>c</sup>*Universidad de Guanajuato, Facultad de Química, Guanajuato, Gto., 36050, Mexico*

## Abstract

This work focuses on the modeling of multistage stochastic problems with endogenous (decision dependent) uncertainties. We assume that the probability distributions of the uncertain parameters are discrete, so that a scenario tree representation can be used. As the main contribution, the paper describes an approach to represent the gradual resolution of endogenous uncertainties after an investment in information is made; partial resolution of uncertainty through time is defined in terms of a percentage of variance reduction. The approach is based on the concepts of posterior and revelation distributions and on the practical propositions of the theory of conditional expectations. A mining production planning problem with endogenous uncertainty in ore quality is used as a case-study to show the scope of the proposed representation as well as to evaluate the effect of the gradual resolution of uncertainties on the optimal solution.

**Keywords:** Multiperiod stochastic planning, uncertainty resolution

## 1. Introduction

Typical engineering applications are decision problems subject to the combination of inherent, modeling and statistical uncertainties. Uncertainty in planning problems can be divided into two classes: exogenous (or market) uncertainty and endogenous (or technical) uncertainty [1]. Problems where stochastic processes are independent of decisions are said to have exogenous uncertainty, whereas problems where stochastic processes are affected by decisions are said to possess endogenous uncertainty. For example, in a multiperiod oilfield development project, the actual size and the initial deliverability of a reserve is generally uncertain [2]; however once a capital investment decision is made (interpreted as an investment in information) regarding exploration and/or production, the uncertainty will eventually be resolved (after a “learning time”). Hence, uncertainties in actual size and initial deliverability are endogenous and their resolution through time is influenced by the decisions made in a given time period.

Literature reports some previous approaches to handle decision dependent uncertainties. Jonsbraten et al. [3] assume that decisions affecting uncertainty resolution occurs at the first time period. Similarly, Goel and Grossmann [1,2] consider immediate resolution of endogenous uncertainties after an investment in information is made, and apply their approach to gas field development problems. More recently, Tarhan and Grossmann [4] analyze the synthesis of process networks with time-varying uncertain yields in which investment in pilot plants can be considered to reduce uncertainty of the yields.

The problem is formulated as a multistage stochastic program with decision dependent elements where investment strategies are considered to reduce uncertainty, and time-varying distributions are used to describe uncertainty. Nevertheless, it is assumed that full resolution of the uncertainty is achieved after one time step (period).

In this work we describe an approach to represent the gradual resolution of endogenous uncertainties after an investment in information is made; our resolution strategy allows partial resolution of uncertainty through several time periods, so that the resolution at each time period is defined in terms of a percentage of variance reduction. The approach has been incorporated into a multistage stochastic program with application to mining production planning. The model includes gradual resolution of ore quality (uncertain parameter) and the concept of indistinguishability. The following sections describe the gradual resolution of uncertainty approach and the multiperiod stochastic model. Further, a case-study is used to show the scope of the proposed representation as well as to evaluate the effect of the gradual resolution of uncertainties on the optimal solution.

## 2. Gradual Resolution of Endogenous Uncertainty

Dias [5] developed a strategy to model uncertainty reduction as the result of an investment in information. The mathematical representation is based on four propositions from the theory of conditional expectation. Also, the resolution process involves three different probability distributions; namely, prior distribution, posterior distribution and revelation distribution. The author applied his approach by assuming continuous probability distributions and expressed the reduction of uncertainty as a percentage of variance reduction.

### *Our approach to resolve endogenous uncertainty*

Here we have extended the approach described by Dias [5] to cases where the probability distributions are discrete, so that we can apply a scenario tree representation to multiperiod stochastic planning problems. The uncertainty resolution process is as follows. We assume that a *prior* (original) discrete distribution of the endogenous parameter is known, so that the probabilities  $\pi_i$  of  $n$  probable values ( $z_i$ ) are given ( $i=1\dots n$ ). Then, after an investment in information is made at any time period, the resolution process for the parameter starts and it will continue for a learning time involving  $r$  resolution steps, with a reduction of variance at each resolution step (the variance reduction is expressed in terms of a fraction  $V$ ). We also assume that, at each resolution step, we may receive  $m$  messages (each with probability  $\theta_j$ ,  $j=1\dots m$ ) so that the original distribution changes, resulting in  $m$  *posterior distributions*. The time varying profile for the endogenous parameter is finally given by using the *revelation distribution* at each resolution step. The revelation distribution is defined by the mean values (each with probability  $\theta_j$ ) of the posterior distributions. Figure 1 shows a representation of the three different probability distributions for the case of three messages. In the practical implementation of the approach, our main assumptions are: *i*) the number of messages is equal to the number of probable values ( $m=n$ ), *ii*) all of the messages are equally probable ( $\theta_j = 1/m$ ) and *iii*) each of the posterior distributions shows a high probability ( $p_z$ ) for one of the probable values. We derived a general expression to calculate the probability  $p_z$  of each posterior distribution which results in a reduction of variance  $V$ . The derivation is not presented here, but it is based on the theoretical proposition which states that the reduction of variance of the prior distribution is equal to the (expected) mean variance of the posterior distributions.

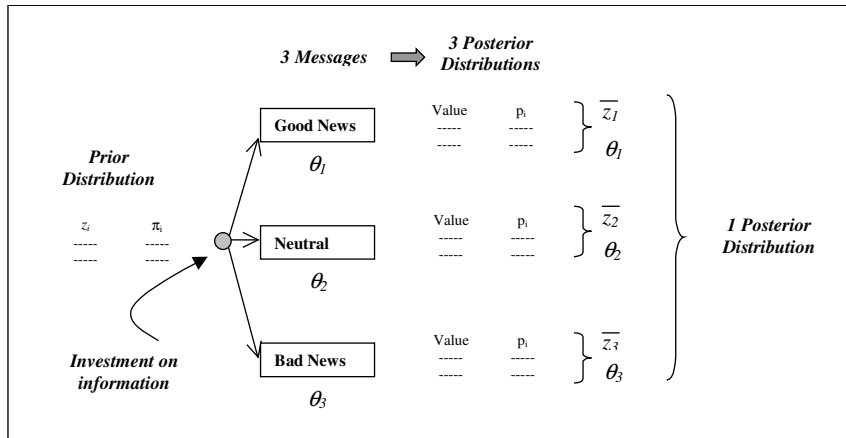


Figure 1. Prior, posterior and revelation distributions

After applying the assumptions enlisted above, the general expression reduces to Eq. 1.

$$\sum_{i=1}^n (z_i - \bar{z})^2 \left[ (1-V) + \frac{n^2(1-p_z)^2}{(n-1)^2} \right] = \frac{(1-p_z)}{n-1} \sum_{j=1}^n \sum_{i=1, i \neq j}^n (z_i - z_j)^2 \quad (1)$$

where  $\bar{z}$  is the mean of the prior distribution. Also, in Figure 1,  $\bar{z}_i$  is the mean of each posterior distribution. Eq. 1 can be solve to obtain  $p_z$ .

### 3. Multiperiod Model for Mining Production Planning

This section describes the case-study and the multiperiod model which incorporates the gradual resolution of endogenous uncertainty. Seeking simplicity due to space limitations, only one example and the main elements of the model are described.

#### Case-study and gradual resolution profiles

The case-study consists of a mining production planning problem involving three mines (adapted from Williams [6]). The mineral produced by each of the mines is mixed in order to satisfy the demand and ore quality at each time period. Royalties are paid if a mine is open for production. Optimal decisions includes the mines production profiles and, previous to that, whether a mine should be open (and royalties paid) or not. The ore quality of one of the mines is known (equal to 1), but it is uncertain for the other two. Regarding the mines with uncertain ore quality, it is also assumed that, after a mine starts operation (an investment in information is made through the payment of royalties) the ore quality will resolve gradually for one mine, but it will immediately be resolved for the other one (either 0.8 or 0.9). The prior distribution for the endogenous ore quality consists of two equally probable values ( $n=m=2$ ;  $\pi_1=\pi_2=0.5$ ;  $z_1=0.6$ ;  $z_2=0.7$ ) and the learning time includes three resolution steps (the variance reduction is of 33% at each step). Table 1 shows the mean values of the posterior distributions (revelation distribution) at each resolution step calculated by using Eq. 1; the investment in information is at time  $t$ . It can be observed that there is a successive reduction at the expected value of variance until full resolution is achieved.

Table 1. Revelation distribution for the endogenous ore quality

| Time step | Posterior Distributions |              | E(variance) | Variance reduction |
|-----------|-------------------------|--------------|-------------|--------------------|
|           | Low quality             | High quality |             |                    |
| $t$       | 0.65                    | 0.65         | 0.00253     | 0                  |
| $t+1$     | 0.62127                 | 0.67872      | 0.00167     | 33%                |
| $t+2$     | 0.60938                 | 0.69062      | 0.00083     | 66%                |
| $t+3$     | 0.6                     | 0.7          | 0.00000     | 100%               |

As theoretically anticipated, the revelation distribution at full resolution is equal to the prior distribution. Both of the posterior distributions are assumed as equally probable. Notice that, due to the number of resolution steps, 16 combinations of values for the uncertain parameters result ( $8 \times 2 = 16$ ;  $2^3 = 8$  for the endogenous parameter and 2 for the immediately resolved uncertain parameter). Hence, if both of the mines with stochastic ore quality are operated, 16 different scenarios should be considered.

#### Stochastic Multiperiod Model

The MILP stochastic multiperiod model is defined by Eqs. 2 through 6. The time horizon is divided in  $N$  time periods.  $T$  is the set of time periods ( $1 \dots N$ ),  $I$  is the set of mines ( $1 \dots M$ ),  $S$  represents the set of scenarios,  $J$  is the set of mines with endogenous uncertain parameters,  $K$  is the set of mines without exogenous uncertainties and  $R$  is the number of resolution steps. Binary variables  $y$  represent decisions about opening and operation of a mine at time  $t$  and scenario  $s$ .

$$f = \sum_{s \in S} p^s \left[ \sum_{t=1}^N \alpha_t q_t - \sum_{i=1}^M \sum_{t=1}^N \rho_i y_{i,t}^{open} - \sum_{i=1}^M \sum_{t=1}^N \eta_i x_{i,t}^s \right] \quad (2)$$

$$\begin{aligned}
x_{i,t}^s &\leq U_i y_{i,t}^{oper,s} && \forall i \in I, \forall t \in T, \forall s \in S \\
y_{i,t}^{oper,s} &\leq y_{i,t}^{open,s} && \forall i \in I, \forall t \in T, \forall s \in S \\
y_{i,t+1}^{open,s} &\leq y_{i,t}^{open,s} && \forall i \in I, \forall t \in T, \forall s \in S \\
\sum_{i=1}^M P_{i,t}^s &\geq C_t q_t^s && \forall t \in T, \forall s \in S \\
P_{k,t}^s &= Q_k^s x_{k,t}^s && \forall k \in K, \forall t \in T, \forall s \in S \\
w_{j,t}^s &= \left( \sum_{\tau=1}^{t-1} y_{j,\tau}^{oper,s} = 0 \right) && \forall j \in J, \forall t \in T, \forall s \in S \\
xp_{j,t}^s &\Leftrightarrow y_{j,t}^{oper,s} \wedge w_{j,t}^s && \forall j \in J, \forall t \in T, \forall s \in S \\
\sum_{i=1}^M x_{i,t}^s &= q_t^s && \forall t \in T, \forall s \in S \\
\sum_{i=1}^M y_{i,1}^{open,s} &= M && \forall s \in S
\end{aligned} \quad (3)$$

$$Z_t^{s,s'} \Leftrightarrow \left[ \bigwedge_{\tau=1}^t (\neg y_{i,\tau}^{oper,s}) \right]_{r=1\dots R(j)} \vee \left[ \left( \sum_{\tau=1}^t y_{j,\tau}^{oper,s} = r+1 \right) \wedge (j,s,s') \in M_r(j,s,s') \right] \quad (4)$$

$$\forall (s,s'), s < s', \forall t \in T$$

$$\left[ \begin{array}{l} Z_t^{s,s'} \\ x_{i,t}^s = x_{i,t}^{s'} \\ y_{i,t+1}^{oper,s} = y_{i,t+1}^{oper,s'} \\ y_{i,t+1}^{open,s} = y_{i,t+1}^{open,s'} \end{array} \right] \vee [\neg Z_t^{s,s'}] \quad \forall (s,s'), s < s' \forall t \in T \quad (5)$$

$$x_{i,1}^s = x_{i,1}^{s'} \quad \forall i \in I, \forall t \in T, \forall s \in S$$

$$y_{i,1}^{oper,s} = y_{i,1}^{oper,s'} \quad \forall i \in I, \forall t \in T, \forall s \in S$$

$$y_{i,1}^{open,s} = y_{i,1}^{open,s'} \quad \forall i \in I, \forall t \in T, \forall s \in S$$

$$\left[ \begin{array}{l} xp_{j,t}^s \\ P_{j,t+r-1}^s = x_{j,t+r-1}^s Q_{r,j}^s \quad \forall r = 1\dots R(j)+1 \\ P_{j,t+r}^s = x_{j,t+r}^s Q_{R(j),j}^s \quad \forall r = R(j)+1\dots T-t \end{array} \right] \vee [\neg xp_{j,t}^s] \quad (6)$$

$$\forall j \in J, \forall t \in T, \forall s \in S$$

$x$  represents the ore production of a mine and  $q$  is the total ore production.  $Q$  is the endogenous parameter (ore quality),  $P$  is the mine production of the valued mineral,  $U$  is the maximum achievable ore production of a mine and  $C$  is the minimum required quality of the produced material. Very important,  $Z$  is a binary variable representing indistinguishability of scenarios  $s$  and  $s'$ . Finally,  $xp$  and  $w$  are binary variables used just to represent the logical relationships among decisions. The objective function (Eq. 2) includes three terms; the profit, the royalty cost and the operation cost.

#### Results and discussion

For time horizon involving 8 time periods, the model contains 2881 continuous variables and 1984 binary variables, which shows the combinatorial complexity of the approach. As per the results, the difference between the deterministic and stochastic objectives (value of the stochastic solution, VSS) is almost 48% for the example, which shows the significance of incorporating uncertainties in the model parameters. Interestingly, the optimal decisions for all the scenarios and the deterministic case include the opening and operation of the mine with endogenous uncertainty since the first time period. However, because of poor ore quality, the deterministic case suggested stopping operation after 2 time periods. As an example of the numerical results, Table 2 shows the ore production ( $10^6$  ton/year) in 4 time periods of the mine with immediate resolution of ore quality in scenarios 1 through 8.

#### 4. Conclusions and Future Work

This paper describes an approach to model the gradual resolution of endogenous uncertainties represented by discrete probability distributions on the context of MILP multiperiod planning problems.

Table 2. Numerical results ( $x$ ) for mine with immediate resolution of uncertainty

| Time period | $s_1$  | $s_2$  | $s_3$  | $s_4$  | $s_5$  | $s_6$  | $s_7$  | $s_8$  |
|-------------|--------|--------|--------|--------|--------|--------|--------|--------|
| 1           | 2.5000 | 2.5000 | 2.5000 | 2.5000 | 2.5000 | 2.5000 | 2.5000 | 2.5000 |
| 2           | 1.0000 | 1.0000 | 1.0000 | 2.5000 | 1.0000 | 1.0000 | 2.5000 | 0.0000 |
| 3           | 2.5000 | 2.5000 | 1.0000 | 1.0000 | 2.5000 | 1.0000 | 2.5000 | 1.0000 |
| 4           | 0.8080 | 0.8080 | 0.8080 | 0.8080 | 0.8080 | 0.8080 | 0.8080 | 0.8080 |

The derivation is based on the theoretical proposition which states that the reduction of variance of the prior distribution is equal to the (expected) mean variance of the posterior distributions. In fact, partial resolution of uncertainty through time is defined in terms of a percentage of variance reduction. Due to combinatorial complexity, the number of scenarios rapidly increases with both the number of resolutions steps and the number of probable values of the uncertain parameters. Hence, the resulting MILP models can only be solved through an LP-based branch and bound for smaller instances. The difference between the deterministic and stochastic objectives (value of the stochastic solution, VSS) is almost 48% for the case study, which shows the significance of incorporating uncertainties in the model parameters. Duality-based branch and bound algorithms for solving larger (linear and nonlinear) problems are currently being developed and tested. Furthermore, a paper describing a generalized approach to model the time varying profiles of gradually resolved endogenous parameters is also in preparation.

## 5. Acknowledgements

V. Rico-Ramirez thanks the financial support provided by the Fulbright Scholarship program and by CONACYT, Mexico.

## References

- [1] Goel, V. and I. E. Grossmann. A class of stochastic programs with decision dependent uncertainty. *Mathematical Programming - Series B*, 108, 355-394, 2006.
- [2] Goel, V. and I. E. Grossmann. A stochastic programming approach to planning of offshore gas field developments under uncertainty in reserves. *Computers & Chemical Engineering* 28, 8, 1409–1429, 2004
- [3] Jonsbraten, T.W., R. J. B. Wets and D. L. Woodruff, A class of stochastic programs with decision dependent random elements. *Annals of Operations Research*, 82, 83-106, 1998.
- [4] Tarhan, B. and I. E. Grossman. A multistage Stochastic Programming Approach with Strategies for Uncertainty Reduction in the Synthesis of Process Networks with Uncertain Yields, *Computers & Chemical Engineering* 32, 4-5, 766-788, 2008.
- [5] Dias, M. A. G. Investment in information in petroleum: Real options and revelation, In the proceedings of the 6th Annual International Conference on Real Options, Cyprus, 1-47, 2002
- [6] Williams, H.P. *Model building in mathematical programming*, 4th Edition John Wiley and sons, London, 2002.