A MAXIMUM ENTROPY APPROACH TO UPDATE AIRLINES DEMAND DISTRIBUTIONS

Ioana C. Bilegan
LAAS-CNRS, 7, Avenue du Colonel Roche,
ENAC/DGAC, Air Transportation Department,
31077 Toulouse, France
bilegan@laas.fr

Sergio González-Rojo
LAAS-CNRS, 7, Avenue du Colonel Roche,
31077 Toulouse, France
ITCH 2909 Av. Tecnológico, Chih., Mexico
grojo@laas.fr

Carlos A. N. Cosenza
COPPE/UFRJ,
Rio de Janeiro, RJ, Brazil
cosenza@pep.ufrj.br

Félix Mora-Camino
LAAS-CNRS, 7, Avenue du Colonel Roche,
ENAC/DGAC, Air Transportation Department,
31077 Toulouse, France
mora@laas.fr

ABSTRACT

In order to design an efficient market-reactive revenue management system for airlines, it is necessary to have at hand a representative probability distribution of demand by period, by fare class and by order of arrival. For this, a dual geometric programming problem can be formulated according to the principle of maximum entropy. Making use of the corresponding primal form of the geometric program, the large scale convex optimization problem is transformed into a non-constrained non-convex minimization problem. The solution strategy proposed consists of two main steps: first, solve the primal form of the geometric program and second, compute by geometric inversion the updated probability distributions. A numerical solution of the non-convex primal geometric program is obtained using an ad-hoc designed genetic algorithm. Its performances had been evaluated under different simulation scenarios involving various fare classes, several forecasting periods and different demand profiles, showing satisfactory results.

INTRODUCTION

The aim of airline revenue management systems is to maximize the revenues of airlines obtained from selling the available seats on their scheduled flights. Advanced revenue management systems present four main components: forecasting, overbooking, seat inventory control and pricing. A revenue management system works on-line, gathering the most recent updates provided by the demand forecasting functions and the current state of reservations, to proceed with the treatment of new requests. In order to take into account the highly stochastic nature of booking requests, forecasts should be updated with the latest information available. When a feedback control loop can be established between an inventory control module and a demand forecasting updating process, this can lead to an efficient market-reactive revenue management system, as reprepresented in the figure:
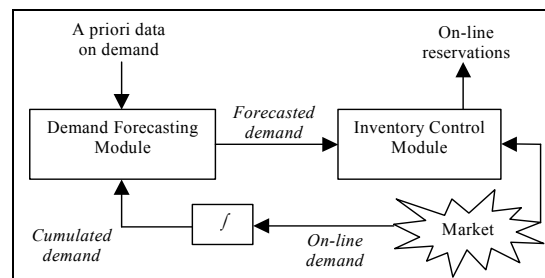

Figure 1. Structure for a market-reactive airline RM systems.

THE DEMAND PROBABILITY DISTRIBUTION UPDATING PROBLEM

Classical probabilistic distributions, such as Poisson distribution, binomial distribution or gamma distribution and variations or combinations of them have been proposed to model the dynamic process of booking, including cancellations, no-shows and go-shows. Empirical studies showed that the normal distribution

provides an acceptable continuous approximation for the distribution of the total demand for a flight, but researchers argue that this distribution becomes inappropriate when temporal distributions of demand are needed [1].

Several stochastic models of booking requests have been proposed in the literature: homogenous and non-homogenous Poisson processes, leading to the cumulative Poisson process and the censored Poisson process [2]. Data available in the records of past reservations which are used to provide demand forecasts are usually biased by the existence of booking limits for each fare class: in general booking requests exceeding the limits of these fare classes, are not recorded.

Once a prior probabilistic distribution of demand along the whole booking horizon is turned available, an on line adaptation process devoted to the updating, according to current data, of the temporal probability distributions of demand, should be started [3]. This need leads to the formulation of a new optimization problem, which must be solved at each time step of the booking process.

To state the corresponding optimization problem at time step n, the notations presented in the Nomenclature are adopted. In the simple case where demand suffers only temporal shifts, δ_n is given by:

$$\delta_n = D - d_n \qquad (1)$$

Otherwise, complex estimation techniques, including qualitative reasoning to take advantage of expert knowledge, must be used to get an updated mean value of the remaining demand δ_n. Consistency implies that the following relation holds:

$$\delta_n - \sum_{k=n}^{N} \sum_{j=0}^{J} j \cdot p_{jk}^n = 0 \qquad (2)$$

In the proposed approach, the a priori distribution probabilities have to be corrected at the beginning of each decision period *n*, over the remaining decision periods $k = n$ to *N*, taking into account the most recent available information (the expressed demand as well as other newly foreseen effects) obtained during the last decision period. To update the demand probability distribution, its potential changes (between the predicted and the real recorded demand) have to be spread over the remaining decision periods to meet the new

consistency condition (2). This leads to the minimization of the information gain criterion given in the following expression:

$$\min \sum_{k=n}^{N} \sum_{j=0}^{J} p_{jk}^n \cdot \log\left(\frac{p_{jk}^n}{p_{jk}^{n-1}}\right) \qquad (3)$$

which corresponds also to the maximization of a relative entropy between a measure of the distance between the $(n-1)^{th}$ and the nth demand probability distributions estimations.

Here, p_{jk}^{n-1} denotes the previous probabilities computed at the beginning of the $(n-1)^{th}$ decision period, hence available for the nth decision period and the p_{jk}^n are the new probabilities to be obtained from the solution of the optimization problem whose set of constraints (2) is completed by :

$$\sum_{j=0}^{J} p_{jk}^n = 1 \qquad (4)$$

for all *k=n* to *N*, and

$$p_{jk}^n \geq 0 \qquad (5)$$

for $j = 1$ to *J* and *k=n* to *N*.

The updating process of the p_{jk}^n probabilities of upcoming demand makes use of a dual geometric programming formulation [4] for an optimization criterion of the information gain type (3), under positivity (6), normality (7) and orthogonality (8, 9) constraints:

$$\frac{p_{jk}^n}{N-n} \geq 0 \qquad (6)$$

$$\sum_{k=n+1}^{N} \sum_{j=0}^{J} \frac{p_{jk}^n}{N-n} = 1 \qquad (7)$$

$$\sum_{k=n+1}^{N} \sum_{j=0}^{J} \left(j - \frac{\delta_n}{N-n}\right) \cdot p_{jk}^n = 0 \qquad (8)$$

$$\sum_{j=0}^{J} p_{j(n+1)}^n - \sum_{j=0}^{J} p_{jk}^n = 0 \qquad (9)$$

where $k \in \{n+2, n+3, \dots, N\}$, *n* is the decision period, *j* is the number of demands and p_{jk}^{n-1} are

the initial probability distributions, before updating.

The problem is transformed into a non-constrained non-convex minimization problem, making use of the corresponding primal form of the geometric program, relation (10), for which genetic algorithms are used to get a numerical solution.

$$\min \phi(\bar{t}) = \sum_{j=0}^{J} p_{j(n+1)}^{n-1} \cdot t_{n+1}^{j\frac{\delta_n}{N-n}} \cdot \prod_{k=n+2}^{N} t_k + \sum_{k=n+2}^{N} \sum_{j=0}^{J} p_{jk}^{n-1} \cdot t_{n+1}^{j\frac{\delta_n}{N-n}} \cdot t_k^{-1} \quad (10)$$

with $t_k > 0$, for $k \in \{n+1, n+2, ..., N\}$.

Then, using inverse geometric duality relations, the updated probability distributions are computed and made available for further use, according with relations (11) and (12):

for $k=n+1$:

$$p_{j(n+1)}^{n^*} = \frac{(N-n) \cdot p_{j(n+1)}^{n-1} \cdot t_{n+1}^{*\frac{j\delta_n}{N-n}} \cdot \prod_{k=n+2}^{N} t_k^{*}}{\sum_{j=0}^{J} p_{j(n+1)}^{n-1} \cdot t_{n+1}^{*\frac{j\delta_n}{N-n}} \cdot \prod_{k=n+2}^{N} t_k^{*} + \sum_{k=n+2}^{N} \sum_{j=0}^{J} p_{jk}^{n-1} \cdot t_{n+1}^{*\frac{j\delta_n}{N-n}} \cdot t_k^{*-1}} \quad (11)$$

for $k=n+2$ to N :

$$p_{jk}^{n^*} = \frac{(N-n) \cdot p_{jk}^{n-1} \cdot t_{n+1}^{*\frac{j\delta_n}{N-n}} \cdot t_k^{*-1}}{\sum_{j=0}^{J} p_{j(n+1)}^{n-1} \cdot t_{n+1}^{*\frac{j\delta_n}{N-n}} \cdot \prod_{k=n+2}^{N} t_k^{*} + \sum_{k=n+2}^{N} \sum_{j=0}^{J} p_{jk}^{n-1} \cdot t_{n+1}^{*\frac{j\delta_n}{N-n}} \cdot t_k^{*-1}} \quad (12)$$

RESOLUTION OF THE PRIMAL GEOMETRIC PROBLEM

In order to solve the Primal Geometric Problem, which consists of the minimization of the following non-convex problem, under the positivity constraints, $t_k > 0$:

$$\min_{\{t_k\}} \sum_{j=0}^{J} x_{j(n+1)}^{n-1} \cdot t_{n+1}^{j\frac{\delta_n}{N-n}} \cdot \prod_{k=n+2}^{N} t_k + \sum_{k=n+2}^{N} \sum_{j=0}^{J} x_{jk}^{n-1} \cdot t_{n+1}^{j\frac{\delta_n}{N-n}} \cdot t_k^{-1} \quad (13)$$

where the t_k , $k = n+1$ to N, are the variables, strictly positive and where $\{\delta_n\}$ and $\{x_{jk}^{n-1}\}$ are positive constants, an optimization method based on Geometric Algorithms is developed and applied for obtaining the numerical solution.

The choice of this alternative for solving the Primal Geometric Problem is based on the complexity of the minimization problem, for which the Genetic Algorithms offer the advantage of a parallel exploration of the search space. The Genetic Algorithm offers a simple way of solution encoding, a fast way of finding good solutions and the capacity to escape from local minima (contrarily to the application of mathematical programming techniques for which the efforts of elaboration of analytic expressions of gradients etc. and for which the corresponding computing time may be prohibitive.

Figure 2 gives an example of a graphical representation of the objective function (13) in the case of two variables.
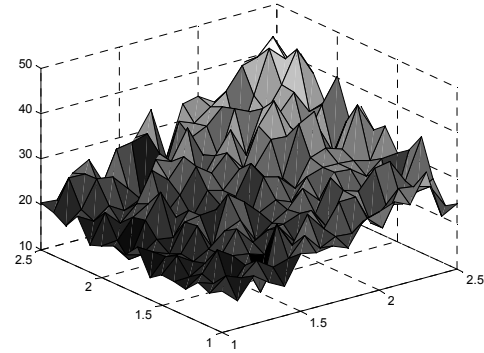

Figure 2. Three-dimensional representation of the objective function for the Primal Geometric Problem.

In order to solve this optimization problem applying a Genetic Algorithm, the following characteristics for its implementation have been chosen:

The fitness function consists of the minimization criterion itself (13) within the search space given by $\{t_k \in \Re^{+*}, k = 1$ to $N\text{-}n\}$.

The encoding retained consists of allocating to each variable t_k of the solution an equal number of binary genes (of value 0 or 1) such that each decimal value has a corresponding binary number. Then, the original value is obtained from:

$$t_k = \text{decimal value} \cdot (t_{k_{max}} / (2^8 - 1)) \quad (14)$$

where $t_{k,max}$ is the superior limit adopted for the variable t_k. This encoding allows, with a sufficient number of genes for the representation of each variable (i.e. 16, 32), to choose the precision level within the search space. For example, a vector $\underline{t}$, formed by four components (t_1, t_2, t_3, t_4), with

eight genes for the representation of each of them, is such that:

t_1	t_2	t_3	t_4
10011101	01110100	10001110	10110001

Figure 3. Binary representation of a solution vector.

The selection method chosen here for the genetic algorithm is of the following type: each individual x_i of the population within which the selection is performed will be assigned a probability p_i of being chosen, in accordance with its corresponding value computed by applying the fitness function. A random number r is generated uniformly between 0 and 1. Then, the k^{st} individual is selected if:

$$\sum_{i=1}^{k-1} p_i < r \leq \sum_{i=1}^{k} p_i \qquad (15)$$

Thus, the probability that x_k is selected is given by p_k. In the literature, this classical procedure is known as the *roulette wheel* selection method [5].

The cross-over operator retained here is the same as the one proposed by Osyczka [6]. It consists of a cross-over performed on a single variable t_k of the solution in order to maintain the consistence of the structure of the solution vector. To implement this operator, the binary string composing the chromosome is divided into (N-n) parts, each of them representing one component of the solution vector $\underline{t}$. In the following figure (Figure 4), the result obtained by applying the cross-over operator on the third variable (t_3) of the initial solutions (X et Y) is represented:

Figure 4. Crossover operator example

An example of the evolution of the objective function obtained by applying the genetic algorithm, with the following parameters, is given in Figure 5:

Population size = 50;
Mutation Probability = 0.1;
Cross-over probability = 0.6;
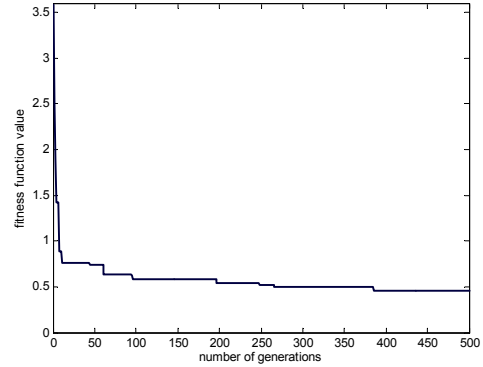Number of generations = 500.

Figure 5. Evolution of the fitness function value

SIMULATION SCENARIOS FOR THE DAILY DEMAND FORECASTS UPDATING APPROACH

Several numerical results obtained by applying the dynamic demand forecasts updating approach are presented here. Two qualitative evaluation indicators are proposed, in order to characterize the degree of accuracy of the a priori information about the demand.

A proposed first indicator (I_n) is given by:

$$I_n = \left| \sum_{k=1}^{n} \sum_{j=0}^{J} p_{jk}^{n+1} \log \frac{p_{jk}^{N+1}}{p_{jk}^{n+1}} \right| \qquad (16)$$

for $\forall n = N - 1, N - 2, ..., 1$.

where :

- n is the number of decision periods on which the updating process is applied, using the new estimates δ_n (the current forecast for the global demand to come during the remaining booking horizon),

- p_{jk}^{n+1} are the probabilities to have j booking requests for the considered fare class during the k^{st} period, updated for the n remaining periods,

- p_{jk}^{N+1} are the initial probabilities, computed at the beginning of the time horizon (for the N-1 remaining periods) as shown in Figure 6.

A second indicator is proposed (R), giving a *global* information on the accuracy of the degree of a priori knowledge characterizing the forecasts for the remaining booking horizon (N-1 periods).

$$R = \sum_{n=1}^{N-1} r_n = \sum_{n=1}^{N-1} \left| \sum_{j=0}^{J} p_{jn}^{n+1} \log \frac{p_{jn}^{N+1}}{p_{jn}^{n+1}} \right| \qquad (17)$$

Here, $\{p_{jn}^{n+1}\}$ represents the probability distribution updated at each remaining period $n+1$, $n= N$-1, …, 1.
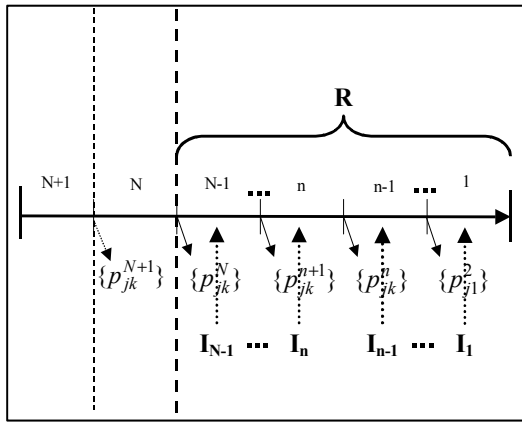

Figure 6. The time dimension of the updating process.

Several simulation scenarios have been considered in order to display the effectiveness of the proposed approach.

In the first scenario, the initial and updated forecasts are considered to be close to each other.

In the following three scenarios, significant differences have been considered. In the second scenario, the updated forecasts characterize a demand lower than the initially estimated one. In the third and fourth scenarios, an over-estimation and, respectively, an under-estimation of the initial forecasts are considered.

The values of the proposed indicators, I_n and r_n (relations 16 and 17), are showing good correlation with the accuracy of the a priori degree of knowledge characterizing the forecasts for the remaining booking horizon (10 days).

In the first scenario (figure 7) the value of both indicators is very low (close to zero),

reflecting a very high degree of a priori knowledge characterizing the forecasts.

In the second scenario (figure 8) the degree of a priori knowledge characterizing the forecasts is quite low, which implies high values for r_n and especially for the indicator *In*.

In scenarios 3 and 4 (figures 9 and 10) the values of the two indicators r_n and I_n show good correlation with the different degrees of a priori knowledge characterizing the forecasts: the values of the indicators are high when there is poor a priori knowledge about the forecasts and vice-versa.

Numerical results are shown in the following four figures, where the different curves of the level of forecasted cumulated demand versus the remaining days before departure, as well as the corresponding curves with the computed values of the two indicators I_n and r_n are presented.
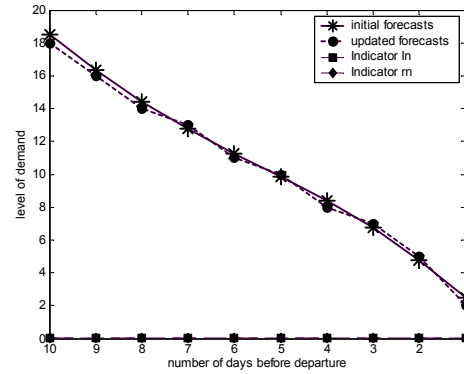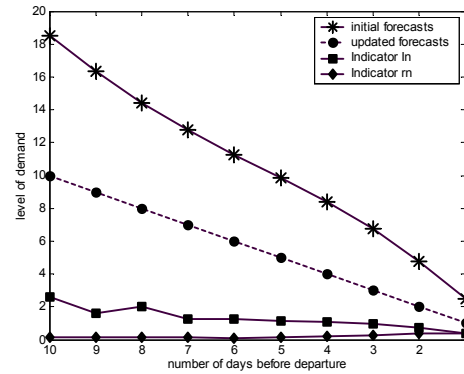

Figure 7. Scenario 1.
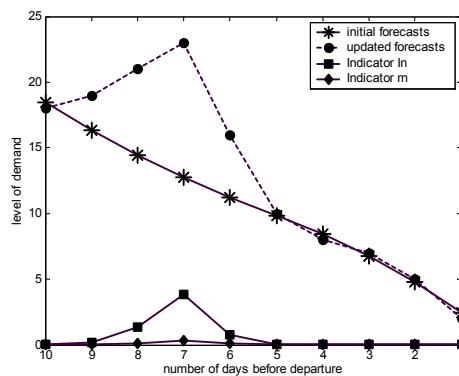

Figure 8. Scenario 2.
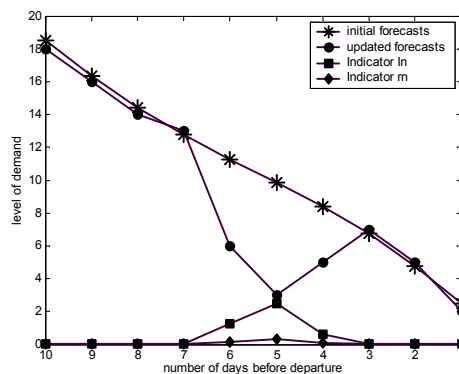
Figure 9. Scenario 3.

Figure 10. Scenario 4.

CONCLUSIONS

The present communication brings three main contributions. First, an on-line probability distribution updating process is proposed so that dynamic approaches for Revenue Management Decision Making can be turned market-reactive. This approach can be also applied to other fields related with perishable asset Revenue Management.

Second, the proposed solution approach mixes for the first time an advanced mathematical programming technique, geometric programming, and a stochastic numerical solution technique, genetic algorithms, to solve by geometric duality inversion a highly constrained non linear optimization problem.

Third, for each realization of the forecasts updating process, the a priori knowledge about the demand for a given fare class on a given flight, can be evaluated through proposed qualitative indicators. Then, these indicators can

play an important tactical part with respect to the commercial policy of an airline.

REFERENCES

1. **J.I. McGill and G.J. Van Ryzin, Revenue Management: Research Overview and Prospects,** *Transportation Science*, **33, pp. 233-256, (1999)**
2. **A.O. Lee, Airline Reservations Forecasting: Probabilistic and Statistical Models of the Booking Process, Ph.D. thesis, Flight Transportation Laboratory, M.I.T., Cambridge, MA,1990.**
3. **I.C. Bilegan, Développement de Systèmes d'Aide à la Décision pour les Opérations Aériennes,** *2éme Congrès des doctorants*, **Ecole Doctorale Systèmes, Toulouse, 2001.**
4. **F. Mora-Camino,** *Introduction à la Programmation Géométrique*, **COPPE, Rio de Janeiro, 1978.**
5. **D.E. Goldberg,** *Genetic Algorithms in Search, Optimization and Machine Learning*, **Addison-Wesley, 1989.**
6. **A. Osycsa, Evolutionary Algorithms for Single and Multicriteria Design Optimization, Physica-Verlag Heindenberg, New York, 2002, p. 41.**