Distributionally robust workforce scheduling in call centers with uncertain arrival rates

S. Liao[1], C. van Delft[2], J.-P. Vial[3,4]

[1] Ecole Centrale, Paris, France
[2] HEC. Paris, France
[3] Prof. Emeritus, University of Geneva
[4] Ordecsys, scientific consulting, Geneva, Switzerland

Econometrics & OR Department, Tilburg University
April 13, 2011

ORDECSYS

Contents

ORDECSYS

1. Staffing a call center
2. Stochastic programming for uncertain mean arrival rates
3. Distributionally robust stochastic programming
4. Validation by simulation

Preliminaries

$$F_{\lambda,AWT}(N) = P\{WT \leq AWT \mid \lambda\}(N) =$$

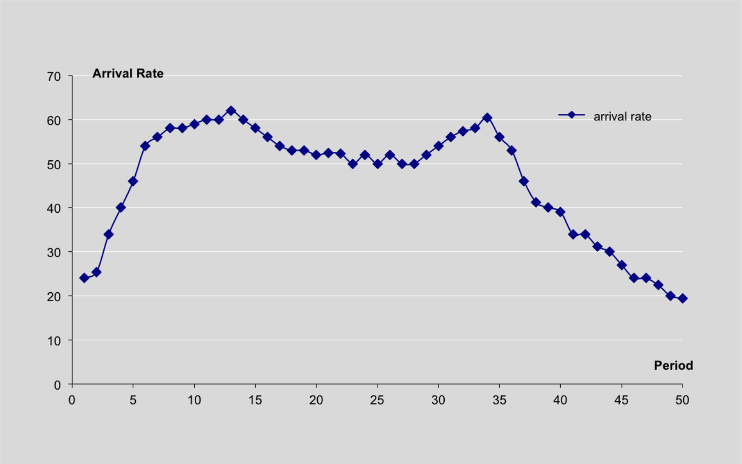
$$1 - \left(\sum_{j=0}^{N-1} \frac{(\lambda/\mu)^j}{j!} + \frac{(\lambda/\mu)^N}{N!\,(1 - \frac{\lambda/\mu}{N})} \right)^{-1} \frac{\lambda/\mu)^N}{N!\,(1 - \frac{(\lambda/\mu)}{N})}\, e^{-(N\mu-\lambda)\,AWT}$$

$SL = F_{\lambda,AWT}(N)$ is the service level guaranteed by the staff of N operators. The required number of operators to secure the service level SL for the average waiting time AWT is

$$N = F^{-1}_{\lambda,AWT}(SL)$$

Classical service level (80/20 rule) : the waiting time should be less than 20 seconds with probability at least 0.8.

The intra-day seasonal variations



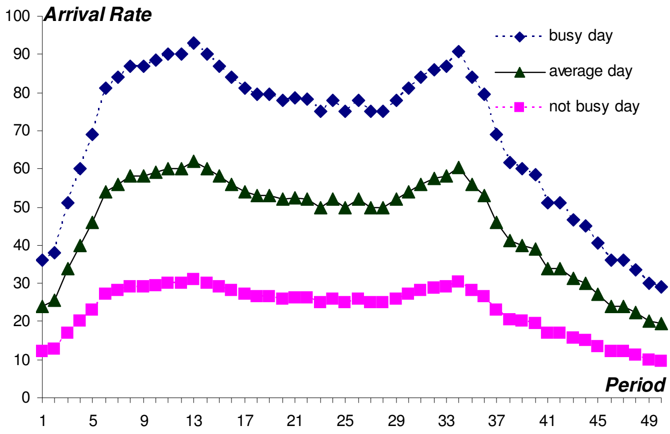
Staffing with seasonal variations

The day starts at 8:00 am, finishes at 8:30 pm, and is divided into $|I| = 50$ periods of 15 minutes each. The arrival rate at period i is λ_i and the target staff level is N_i.
The working schedule is organized is $|J|$ schedules. Each schedule j is represented by a Boolean vector $a_{\bullet j}$ with $a_{ij} = 1$ if this schedule includes period i and 0 otherwise.

$$\begin{aligned} \text{min} \quad & \sum_{j \in J} c_j x_j \\ \text{s.t.} \quad & \sum_{j \in J} a_{ij} x_j \geq N_i, \; i \in I \\ & x_j \in \mathbb{Z}^+, j \in J \end{aligned}$$

1. Staffing a call center
2. Stochastic programming for uncertain mean arrival rates
3. Distributionally robust stochastic programming
4. Validation by simulation

Variation of the mean arrival rates



ORDECSYS

Probabilistic model for the mean arrival rates

The random mean arrival time at period *i* is given by

$$\lambda_i = \theta f_i$$

- f_i : seasonal factor
- θ : busyness factor

AVRAMIDIS, A., DESLAURIERS, A. and L'ECUYER, P. (2004). “Modeling daily arrivals to a telephone call center”. *Management Science*, 50 :896–908.

WHITT, W. (1999).“Dynamic staffing in a telephone call center aiming to immediately answer all calls”. *Operations Research Letters*, 24 :205–212.

Staffing with uncertain busyness factor

From historical data, the distribution of the busyness factor is approximated by the discrete distribution

$$\theta_\ell \text{ with probability } q_\ell, \ \ell \in L$$

An ultra-conservative deterministic formulation

$$\begin{aligned} \text{min} \quad & \sum_{j \in J} c_j x_j \\ \text{s.t.} \quad & \sum_{j \in J} a_{ij} x_j \geq N_{i\ell}, \ i \in I, \ \ell \in L \\ & x_j \in \mathbb{Z}^+, j \in J \end{aligned}$$

A stochastic programming formulation

$$\min \left\{ \sum_{j \in J} c_j x_j : + \rho \sum_{\ell \in L} q_\ell \sum_{i \in I} U_{i\ell} : x_j \in \mathbb{Z}_+,\ j \in J \right\}$$

where

$$U_{i\ell} = \max\{N_{i\ell} - \sum_{j \in J} a_{ij} x_j, 0\}$$

is the understaffing in period *i* under busyness factor θ_ℓ, and ρ is a penalty parmeter per unit of understaffing.

LIAO, S., VAN DELFT, C., KOOLE, G. and JOUINI, O. (2010). “Staffing a call center with uncertain non-stationary arrival rate and flexibility”. Working paper. Ecole Centrale Paris.

Instead of a penalty, we can put abound $\bar{U}$ on the total expected understaffing

$$\begin{aligned} \min \quad & \sum_{j \in J} c_j x_j \\ \text{s.t.} \quad & \sum_{\ell \in L} q_\ell \sum_{i \in I} U_{i\ell} \leq \bar{U} \\ & x_j \in \mathbb{Z}_+, \; j \in J \end{aligned} \tag{1}$$

$U_{i\ell} = \max\{N_{i\ell} - \sum_{j \in J} a_{ij} x_j, 0\}$.
The bound $\bar{U}$ could be a fraction (1%, 2%, ...) of the expected workforce $\sum_i \sum_\ell q_\ell N_{i\ell}$ needed to meet the targeted service level.

LP equivalence

The SP problem is a plain LP, because the q_ℓ are positive.

ORDECSYS

1. Staffing a call center
2. Stochastic programming for uncertain mean arrival rates
3. **Distributionally robust stochastic programming**
4. Validation by simulation

Uncertainty on the probabilities

q_ℓ is an estimator of an unknown probability p_ℓ. The true constraint in the SP problem

$$\sum_{\ell \in L} p_\ell \sum_{i \in I} U_{i\ell} \leq \bar{U}$$

has now uncertain coefficients. To handle this uncertain constraint we need to define an uncertainty model and a uncertainty set for the p. This construct leads to a distributionally robust version of the SP problem (1).

Model for the uncertainty on p

$$\left\{\begin{array}{l} \sum_{\ell\in L} p_\ell U_\ell \leq \bar{U} \\ \sum_{\ell\in L} p_\ell = 1,\ p \geq 0 \end{array}\right. \quad \Leftrightarrow \quad \left\{\begin{array}{l} \sum_{\ell\in L} p'_\ell (U_\ell - \bar{U}) \leq 0 \\ p' \geq 0,\ p' \neq 0. \end{array}\right. \tag{2}$$

We can adopt the following uncertainty model

$$\left\{\begin{array}{l} p'_\ell = q_\ell(1+\xi_\ell),\ \forall l \in L \\ p = p' / \sum_\ell p'_\ell \\ \xi_\ell \in [-1, 1],\ \forall i \in I. \end{array}\right. \tag{3}$$

With this model of uncertainty, the condition on the uncertain constraint (2) becomes

$$\sum_{\ell\in L} p'_\ell (U_\ell - \bar{U}) = \sum_{\ell\in L} q_\ell (U_\ell - \bar{U}) + \sum_{\ell\in L} \xi_\ell q_\ell (U_\ell - \bar{U}) \leq 0.$$

Equivalent robust counterpart

We define the uncertainty set

$$\Xi = \{\xi : ||\xi||_\infty \leq 1, \; ||\xi||_2 \leq k\}.$$

The robust counterpart of the uncertain constraint is thus

$$\sum_{\ell \in L} q_\ell U_\ell + \sum_{\ell \in L} \xi_\ell q_\ell (U_\ell - \bar{U}) \leq \bar{U}, \; \forall \xi \in \Xi.$$

The equivalent robust counterpart is the inequality

$$\sum_{\ell \in I} q_\ell U_\ell + k||Q(U - \bar{U}) + w||_2 + ||w||_1 \leq \bar{U}, \; \text{for some } w, \quad (4)$$

where Q is a diagonal matrix with main diagonal $(q_\ell)_{\ell \in L}$.

Bound on the probability of satisfaction

Proposition

Assume ξ_ℓ, $\ell \in L$ *are independent random variables with range* $[-1, 1]$ *and common expectation* $E(\xi_\ell) = 0$. *Then for any solution to the equivalent robust counterpart (4)*

$$Prob(\sum_{\ell \in L} p_\ell U_\ell \geq \bar{U}) \leq e^{-\frac{k^2}{2}}.$$

Towards a mixed linear programming version

Because our problem involves integer variables, it is computationally more efficient to replace the ellipsoidal uncertainty set by on in the l_1-norm. Because the following inequalities hold for any $a \in \mathbb{R}^{|L|}$

$$\frac{1}{\sqrt{|L|}}||a||_1 \leq ||a||_2 \leq \sqrt{|L|}||a||_\infty$$

we can replace Ξ by the larger uncertainty set

$$\{\xi : ||\xi||_\infty \leq 1,\ ||\xi||_1 \leq k\sqrt{|L|}\} \supseteq \Xi$$

and the equivalent robust counterpart (4) by the stricter inequality

$$\sum_{l \in I} q_\ell U_\ell + k\sqrt{|L|}\,||Q(U - \bar{U}) + w||_\infty + ||w||_1 \leq \bar{U}, \text{ for some } w.$$

Robust counterpart as a set of linear inequalities

$$\begin{array}{rcl} \sum_{l \in I} q_\ell U_\ell + k\sqrt{|L|}z + \sum_{\ell \in L} w_\ell & \leq & \bar{U} \\ z + w_\ell & \geq & q_\ell (U_\ell - \bar{U}),\ \ell \in L \\ z + w_\ell & \geq & q_\ell (\bar{U} - U_\ell),\ \ell \in L \\ w \geq 0,\ z \geq 0, & & \end{array}$$

where

- $w \in \mathbb{R}^{|L|}$ and $z \in \mathbb{R}$ are auxiliary variables
- $U_\ell = \sum_{i \in I} U_{i\ell}$
- $U_{i\ell} \geq 0$ and $U_{i\ell} \geq N_{i\ell} - \sum_{j \in J} a_{ij} x_j$.

ORDECSYS

The mixed integer LP problem

$$
\begin{aligned}
\min \quad & \sum_{j} c_j x_j \\
& \sum_{l \in I} q_\ell U_\ell + k\sqrt{|L|}z + \sum_{\ell \in L} w_\ell \leq \bar{U} \\
& z + w_\ell \geq q_\ell (U_\ell - \bar{U}),\ \ell \in L \\
& z + w_\ell \geq q_\ell (\bar{U} - U_\ell),\ \ell \in L \\
& U_\ell = \sum_{i \in I} U_{i\ell},\ \ell \in L \\
& U_{i\ell} = \max\{N_{i\ell} - \sum_{j \in J} a_{ij} x_j, 0\}, i \in I,\ \ell \in L \\
& x \text{ positive integers },\ w \geq 0,\ z \geq 0.
\end{aligned}
$$

ORDECSYS

Dispersion measure

The set of probability distributions taken into consideration in the definition of the uncertainty set can be expressed as

$$\left\{ p : p = \frac{p'}{\sum_{\ell \in L} p'_l}, \sum_{\ell \in L} \left(\frac{p'_\ell - q_\ell}{q_\ell} \right)^2 \leq k^2, p' \geq 0 \right\}. \tag{5}$$

This is reminiscent of the Pearson's dispersion measure

$$\sum_{\ell \in L} \frac{(p_\ell - q_\ell)^2}{q_\ell}.$$

We can use the Pearson dispersion to build an uncertainty set on p

$$\mathcal{P}_\beta = \{p \geq 0 : \sum_{\ell \in L} \frac{(p_\ell - q_\ell)^2}{q_\ell} \leq \beta, \sum_{\ell \in L} p_\ell = 1\}.$$

Equivalent rob. counterpart with Pearson's dispersion

Pearson's dispersion measure leads to a slightly different equivalent robust counterpart.

$$\begin{aligned} \sum_{\ell \in L} q_\ell \, U_\ell + \sum_{\ell \in L} q_\ell w_\ell + \beta z &\leq \bar{U} \\ \sqrt{q_\ell} \, [U_\ell + v + w_\ell] &\leq z, \; \ell \in L \\ -\sqrt{q_\ell} \, [U_\ell + v + w_\ell] &\leq z, \; \ell \in L \\ w \geq 0. & \end{aligned}$$

(The $\sqrt{L}$ factor is hidden in the immunization factor β.)

ORDECSYS

1. Staffing a call center
2. Stochastic programming for uncertain mean arrival rates
3. Distributionally robust stochastic programming
4. Validation by simulation

Simulations

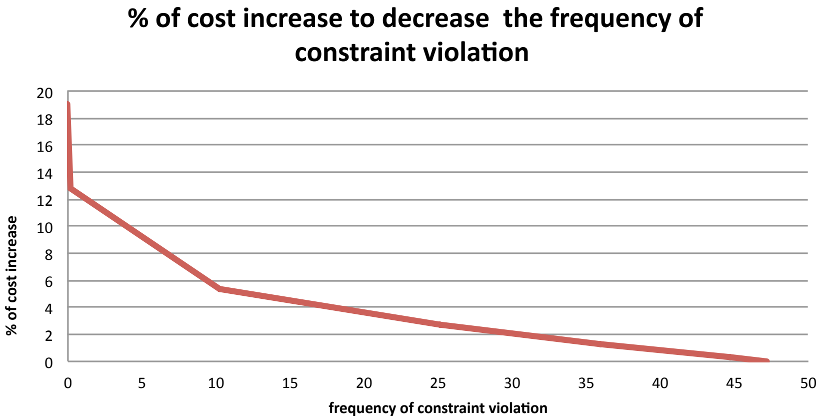
- *Seasonal factors*. The average rate of arrivals at each period of the day is supposed to have been estimated by statistical analysis on a record of $n = 400$ working days.
- *Busyness factor distribution*. Following Avramidis et al. we assume that Θ follows a Gamma distribution. We discretize this distribution on $|L| = 41$ states, to yield the θ_ℓ and q_ℓ.
- *Simulations*. Generate a distribution on $\theta_1, \ldots, \theta_{41}$ as follows : perform 400 random experiments to choose the values θ according to the probabilities q. The distribution p is made equal to the relative occurrences of each θ value. Repeat $K = 10000$ the simulation of p.
- *Performance*. Compute the understaffing for each simulation.
  - Frequency of $U > \bar{U}$ (constraint violation).
  - Conditional expectation of excess understaffing $\{(U - \bar{U}) \mid U > \bar{U}\}$.
  - Worst case for the excess understaffing $(U - \bar{U})$.

Example of results

Immun-	Salary cost	Constr. viol. (%)	Expectation $(U-\bar{U} \mid U>\bar{U})$	Worst case $U-\bar{U}$
Upper bound $\bar{U}=120.77$ is 1% of total required workforce				
0	27481	47.19	29.73	162.76
0.05	27849	35.95	24.72	142.81
0.1	28220	25.11	21.42	126.20
0.2	28953	10.25	16.03	94.16
0.5	30998	0.13	7.11	24.62
0.8	32713	0	-	-20.22
1	33667	0	-	-39.90
$\bar{U}=0$	44957	0	-	-

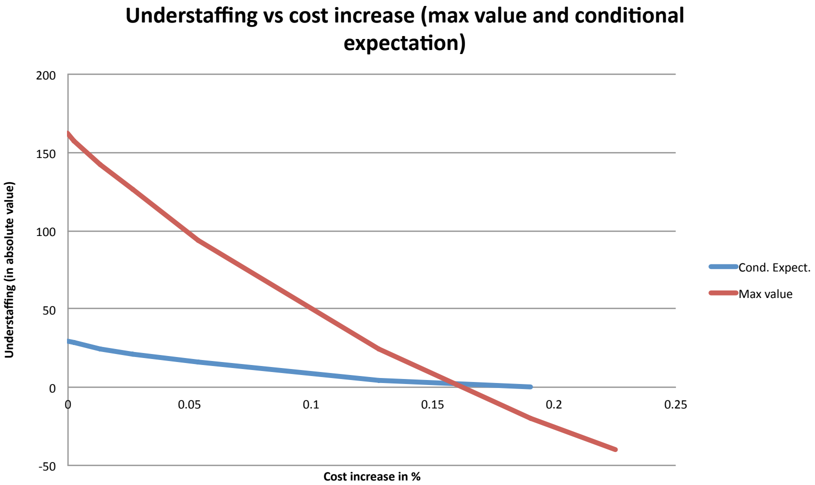
TABLE: Distribut$^{\text{lly}}$ robust sol. with increasing immunization factor

Tradeoff curve



ORDECSYS

Other tradeoff curves



Conclusions

Main results

- The DR approach realistically accounts for the imperfect knowledge of the busyness factor.
- DR solutions can be viewed as an enhancement of stochastic programming.

Extension

- The intra-day seasonal factors are also uncertain. The same methodology can be used to handle this uncertainty.

Uncertain seasonal factors with known probabilities

f_i can take values f_{ki} with probabilities π_{ki}, $k \in K_i$, $\sum_{k \in K_i} \pi_{ki} = 1$. The idea is to assume that the uncertainty on the seasonal factors are independent with respect to one another and with respect to the busyness factor. We compute the required workforce for each of these instances $N_{ki\ell}$, and define the understaffing quantity

$$U_{ki\ell} = \max\{N_{ki\ell} - \sum_{j \in J} a_{ij} x_j, 0\}.$$

Then

$$U_{i\ell} = \sum_{k \in K_i} \pi_{ki} U_{ki\ell}.$$

The results displayed in this presentation were done under the assumption that the seasonal factors could take three values : a nominal value half-way of two extreme values with probabilities 0.25, 0.5 and 0.25.

Uncertain seasonal factors with uncertain probabilities

It is possible to consider that these probabilities are themselves uncertain. If we relax the equality constraint to

$$U_{i\ell} \geq \sum_{k \in K_i} \pi_{ki} U_{ki\ell}$$

we can apply the same DR approach to make the whole problem distributionally robust with respect to the distributions π_{ik}, $\sum_{k \in K_i} \pi_{ik} = 1$ for each $i \in I$.