Distributionally robust workforce scheduling in call centers with uncertain arrival rates

S. Liao\textsuperscript{1}, C. van Delft\textsuperscript{2}, J.-P. Vial\textsuperscript{3,4}

\textsuperscript{1} Ecole Centrale, Paris, France
\textsuperscript{2} HEC. Paris, France
\textsuperscript{3} Prof. Emeritus, University of Geneva
\textsuperscript{4} Ordecsys, scientific consulting, Geneva, Switzerland

Econometrics & OR Department, Tilburg University
April 13, 2011
# Contents

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

2. Stochastic programming for uncertain mean arrival rates

3. Distributionally robust stochastic programming

4. Validation by simulation
Preliminaries

\[ F_{\lambda,AWT}(N) = P\{\text{WT} \leq AWT \mid \lambda\}(N) = \]

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

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

\[ N = F_{\lambda,AWT}^{-1}(SL) \]

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 $\lambda_i$ and the target staff level is $N_i$. The working schedule is organized into $|J|$ schedules. Each schedule $j$ is represented by a Boolean vector $a_{ij}$ with $a_{ij} = 1$ if this schedule includes period $i$ and 0 otherwise.

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

Arrival Rate

Period

busy day
average day
not busy day
The random mean arrival time at period $i$ is given by

$$\lambda_i = \theta f_i$$

- $f_i$: seasonal factor
- $\theta$: busyness factor


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

\[
\begin{align*}
\min & \quad \sum_{j \in J} c_j x_j \\
\text{s.t.} & \quad \sum_{j \in J} a_{ij} x_j \geq N_{i\ell}, \quad i \in I, \ell \in L \\
& \quad x_j \in \mathbb{Z}^+, \quad j \in J
\end{align*}
\]
Staffing a call center

Stochastic Programming

Distributionally Robust SP

Validation by simulation

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 \(\theta_\ell\), and \(\rho\) is a penalty parameter per unit of understaffing.

Instead of a penalty, we can put an\( \bar{U} \) on the total expected understaffing

\[
\begin{align*}
\min & \quad \sum_{j \in J} c_j x_j \\
\text{s.t.} & \quad \sum_{\ell \in L} q_{\ell} \sum_{i \in I} U_{i\ell} \leq \bar{U} \\
& \quad x_j \in \mathbb{Z}^+, \ j \in J
\end{align*}
\]

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

The bound \( \bar{U} \) could be a fraction (1\%, 2\%, \ldots) 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 \( q_{\ell} \) are positive.
Staffing a call center

Stochastic programming for uncertain mean arrival rates

Distributionally robust stochastic programming

Validation by simulation
Uncertainty on the probabilities

$q_\ell$ is an estimator of an unknown probability $p_\ell$. 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 $p$

\[
\begin{cases}
\sum_{\ell \in L} p_{\ell} U_{\ell} \leq \bar{U} \\
\sum_{\ell \in L} p_{\ell} = 1, \quad p \geq 0 
\end{cases}
\quad \Leftrightarrow \quad
\begin{cases}
\sum_{\ell \in L} p'_{\ell}(U_{\ell} - \bar{U}) \leq 0 \\
p' \geq 0, \quad p' \neq 0.
\end{cases}
\tag{2}
\]

We can adopt the following uncertainty model

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

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

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

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

ORDECSYS
Bound on the probability of satisfaction

Proposition

Assume $\xi_\ell, \ell \in L$ 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)

$$\text{Prob}(\sum_{\ell \in L} p_\ell U_\ell \geq \bar{U}) \leq e^{-\frac{k^2}{2}}.$$
Towards a mixed linear programming version

Because our problem involves integer variables, it is computationally more efficient to replace the ellipsoidal uncertainty set by one 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 $\Xi$ 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

\[ \sum_{\ell \in L} q_\ell U_\ell + k \sqrt{|L|} z + \sum_{\ell \in L} w_\ell \leq \tilde{U} \]
\[ z + w_\ell \geq q_\ell (U_\ell - \tilde{U}), \quad \ell \in L \]
\[ z + w_\ell \geq q_\ell (\tilde{U} - U_\ell), \quad \ell \in L \]
\[ w \geq 0, \quad z \geq 0, \]

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 \( U_{i\ell} \geq N_{i\ell} - \sum_{j \in J} a_{ij} x_j \).
The mixed integer LP problem

\[
\begin{align*}
\min & \quad \sum_j c_j x_j \\
\sum_{l \in I} q_\ell U_\ell + k \sqrt{|L|} z + \sum_{\ell \in L} w_\ell & \leq \bar{U} \\
z + w_\ell & \geq q_\ell (U_\ell - \bar{U}), \quad \ell \in L \\
z + w_\ell & \geq q_\ell (\bar{U} - U_\ell), \quad \ell \in L \\
U_\ell & = \sum_{i \in I} U_{i\ell}, \quad \ell \in L \\
U_{i\ell} & = \max\{ N_{i\ell} - \sum_{j \in J} a_{ij} x_j, 0 \}, \quad i \in I, \ \ell \in L \\
x & \text{ positive integers}, \quad w \geq 0, \quad z \geq 0.
\end{align*}
\]
The set of probability distributions taken into consideration in the definition of the uncertainty set can be expressed as

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

(5)

This is reminiscent of the Pearson’s dispersion measure

\[
\sum_{\ell \in L} \left( \frac{p_\ell - q_\ell}{q_\ell} \right)^2.
\]

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

\[
\mathcal{P}_\beta = \{ p \geq 0 : \sum_{\ell \in L} \left( \frac{p_\ell - q_\ell}{q_\ell} \right)^2 \leq \beta, \sum_{\ell \in L} p_\ell = 1 \}.
\]
Pearson’s dispersion measure leads to a slightly different equivalent robust counterpart.

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

(The $\sqrt{L}$ factor is hidden in the immunization factor $\beta$.)
1. Staffing a call center
2. Stochastic programming for uncertain mean arrival rates
3. Distributionally robust stochastic programming
4. Validation by simulation
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 $\Theta$ follows a Gamma distribution. We discretize this distribution on $|L| = 41$ states, to yield the $\theta_\ell$ and $q_\ell$.

Simulations. Generate a distribution on $\theta_1, \ldots, \theta_{41}$ as follows: perform 400 random experiments to choose the values $\theta$ according to the probabilities $q$. The distribution $p$ is made equal to the relative occurrences of each $\theta$ 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

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

$\bar{U} = 0$ 44957 0 - -

**Table**: Distributively robust solution with increasing immunization factor
Tradeoff curve

% of cost increase to decrease the frequency of constraint violation
Other tradeoff curves

Understaffing vs cost increase (max value and conditional expectation)
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 \( \pi_{ki} \), \( k \in K_i \), \( \sum_{k \in K_i} \pi_{ki} = 1 \). 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.
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 \( \pi_{ik} \), \( \sum_{k \in K_i} \pi_{ik} = 1 \) for each \( i \in I \).