Université Libre de Bruxelles
Institut de Recherches Interdisciplinaires
et de Développements en Intelligence Artificielle

Instances Generator for the Vehicle Routing Problem with Stochastic Demand

Paola Pellegrini and Mauro Birattari

IRIDIA – Technical Report Series

Technical Report No.
TR/IRIDIA/2005-010

May 2005

IRIDIA – Technical Report Series
ISSN 1781-3794

Published by:
IRIDIA, *Institut de Recherches Interdisciplinaires et de Développements en Intelligence Artificielle*
Université Libre de Bruxelles
Av F. D. Roosevelt 50, CP 194/6
1050 Bruxelles, Belgium

Technical report number TR/IRIDIA/2005-010

Instances Generator for the Vehicle Routing Problem with Stochastic Demand

Paola PELLEGRINI paolap@pellegrini.it
Mauro BIRATTARI mbiro@ulb.ac.be

IRIDIA, Université Libre de Bruxelles, Brussels, Belgium

July 26, 2005

Abstract

In the following, an instances generator for the Vehicle Routing Problem with Stochastic Demand is presented. Data concerning real cities are used as guideline in the development of the generator. Two examples of instances are presented.

1 Introduction

In order to perform an empirical analysis of the behavior of an algorithm, a class of reference instances must be defined. The best option would be to consider a sample of real instances supplied by firms dealing with the specific problem tackled. This cooperation between industry and research would be extremely useful, but nowadays it is still very difficult to reach. Typically, keeping secret the database of customers and the corporate strategies is considered a source of competitive advantage, and then shearing this information with scientists is often seen as generating more drawbacks than benefits. As a consequence it is very hard to have the possibility of handling a large set of real instances. The solution to this problem is to consider artificial instances as benchmark, using the ones proposed in literature whenever convenient.

The work presented here has been motivated by the willing of investigating the behavior of different approaches when applied to the Vehicle Routing Problem with Stochastic Demand (*VRPSD*). This problem is pretty well known in literature, but a set of standard benchmark instances is not available. We describe, then, a procedure for generating instances specific for the *VRPSD*, which will be used in future studies. The main objective is to formalize and make publicly available a generator of instances as close as possible to what from our point of view can be real situations.

In Section 2 the Vehicle Routing Problem with Stochastic Demand is shortly described and in Section 3 the logic behind the procedure we use for the generation and the main framework of the algorithm are presented. Section 4 contains two elementary examples of instances obtainable following this framework.

2 The Vehicle Routing Problem with Stochastic Demand

The Vehicle Routing Problem with Stochastic Demand is defined on a graph $G = (V, A)$, where $V = \{v_0, v_1, v_2, ..., v_n\}$ represents the set of nodes and $A = \{(v_i, v_j) : v_i, v_j \in V\}$ represents the set of arcs connecting them. To the generic element $(v_i, v_j) \in A$ a cost c_{ij} is associated, often representing the distance between nodes v_i and v_j. A fleet of vehicles with fixed capacity, that in the standard configuration are identical, is placed in the depot v_0. Each node $v_i \neq v_0$ represents a customer, that requires the delivery of a certain amount of good. The objective of the problem is serving all the customer using the available vehicles at the minimum total cost.

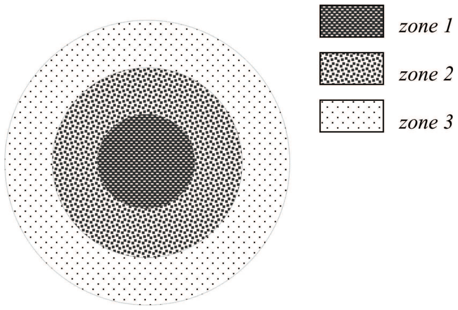

Figure 1. Representation of the division of a city in different *zones*

The peculiarity of the *VRPSD* consists in the fact that only the probability distribution of the demand of each customer is known before reaching him, which make impossible to establish a-priori a set of routes that minimizes the total cost for any possible realization. The objective typically considered is then the minimization of the expected total cost.

Many authors have studied this problem, for a deeper analysis we refer the reader to [1, 2, 3, 5, 6].

3 Instances Generation

The main purpose of this work is to propose a procedure for generating instances for the *VRPSD*, which will be used in future works.

The first problem we focused on consists in detecting the characteristic that the localization of the customer should have in order to be able to be realistic. For the Vehicle Routing Problem in general, most real applications are represented by firms distributing (or collecting) goods to a set of customers. These customers can be either geographically concentrated in a small area or spread in a larger region. We decided to consider the case of customers distributed in one or more cities, with a higher concentration in the center than in the suburbs. Moreover, the number of customers in each city depends on the dimension of the city itself, that we consider proportional to the number of inhabitants. This structure has been inspired by the localization of retail stores (e.g. clothing or shoe stores) in typical European cities, but many other examples can be found.

The foundation for the creation of instances is a database containing the population of different cities and the distance between them. The data we consider are relative to real cities in Denmark, England, Finland, France, Germany, Italy and Norway. The distances are the real ones, given the highways system at the end of 2004. Clearly other databases can be used in order to obtain different instances.

Each city is represented as a set of concentric circles, the first with ray equal to a fixed constant r, the second with ray equal to $2r$ and so on. The number of this circles in each city C (n_C) depends on its population (the higher the population, the higher the number of circles). In particular, this number is currently computed in the following way: given the population p, one circle has been create if $p \leq 100000$, two circles if $p \leq 300000$, three circles if $p \leq 600000$ and so on. We will refer to the areas obtained in this way will as to *zones*, being *zone 1* the area of the first (inner) circle, *zone 2* the one between the first and the second, and so on (Figure 1). Once the number of retail stores is fixed, the nodes are uniformly distributed in each *zone* and the depot is located in the most external circle of a randomly selected city.

The distance between the nodes is expressed in travel time, associating to each circle a coefficient, representing the speed that can be used to go from one to the other, for which the Euclidean distance must be multiplied. To reproduce the situation of real cities, where this speed decreases while one approaches the center, the coefficient becomes smaller while moving from a circle to a more internal one according to the following rule: two coefficients equal for all the

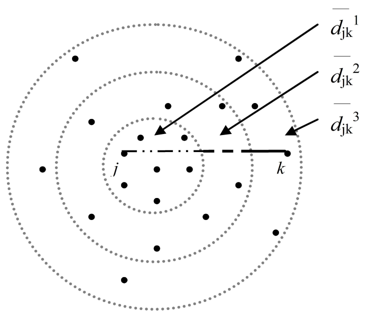

Figure 2. Example of nodes distribution in the generic city C

cities ($c_{1_C} = z$ and $c_{n_C} = Z$) are associated to the inner and the outer circles, while coefficients $c_{i_C} = c_{i-1_C} + \frac{Z-z}{n_C-1}$ are assigned to all the other circles i's in city C.

Then, to compute the travel time needed to go from node j to node k belonging to city C (d_{jk}), after calculating the Euclidean distance between the two ($\bar{d}_{jk}$), we estimate with a Montecarlo procedure the length of the segment that falls in each circle i ($\bar{d}^i_{jk}$) and we divide this measure for the proper coefficient c_{i_C}:

$$d_{jk} = \sum_{i=1}^{n} \frac{\bar{d}^i_{jk}}{c_{i_C}}.$$

In the example of nodes distribution in the generic city C reported in Figure 2, the distance between node j and node k is obtained computing the Euclidean distance between them, estimating the length of the segments $\bar{d}^1_{jk}, \bar{d}^2_{jk}$ and $\bar{d}^3_{jk}$ and then calculating the summation $\frac{\bar{d}^1_{jk}}{c_{1_C}} + \frac{\bar{d}^2_{jk}}{c_{2_C}} + \frac{\bar{d}^3_{jk}}{c_{3_C}}$.

Transferring this procedure in real world measures, we can think of the coefficients for two circles as 30 Km/h and 40 Km/h, then if connecting two nodes implies running 10 Km in the first circle and 5 Km in the second, the travel time distance will be

$$d = \frac{10 \text{ Km}}{30 \text{ Km/h}} \cdot \frac{5 \text{ Km}}{40 \text{ Km/h}} \approx 0.46 \text{ h}.$$

The distance between node a in city A and node b in city B, finally, depends on the circle to which they belong: if a is in the most external circle of A (n_A) and b is in the most external circle of B (n_B) (Figure 3 (a)), d_{ab} is equal to the distance (in travel time) between A and B. Otherwise, suppose b is in circle n_B while a is in circle $n_a < n_A$ (Figure 3 (b)); then we add to the distance between A and B the value $\sum_{i=n_a}^{n_A} \frac{r}{c_{i_A}}$. This measure is not the real distance from node a to the border of city A, but it is an easy, and eventually overestimated, representation of it. The same procedure is applied if also node b does not belong to n_B (Figure 3 (c)).

As for the rest, the decision of this method of computation was based on the observation of reality: most cities nowadays are equipped with a ring road, and it's not too unrealistic to suppose that the time needed to move from a city to another is not strongly influenced by the point of the ring road from which one starts traveling.

To conclude, to each customer some parameter are assigned that characterize its (discrete) demand: the shape of distribution considered is the same for all the customers and the parameters are randomly drawn between bounds defined by the user; in these intervals, the probability distribution considered is uniform. The possible distributions of the demand are

- uniform, in which case the parameters attached are *average* and *spread*[1]: all the values k (between $average - spread$ and $average + spread$) are equally likely with probability $p(k) = \frac{1}{2 \cdot spread + 1}$ of realization;

[1] Being these parameters randomly selected between an upper and a lower bound, after each drawn the program checks if *spread* is smaller than or equal to *average*. If this is not the case, *spread* is set equal to *average*.

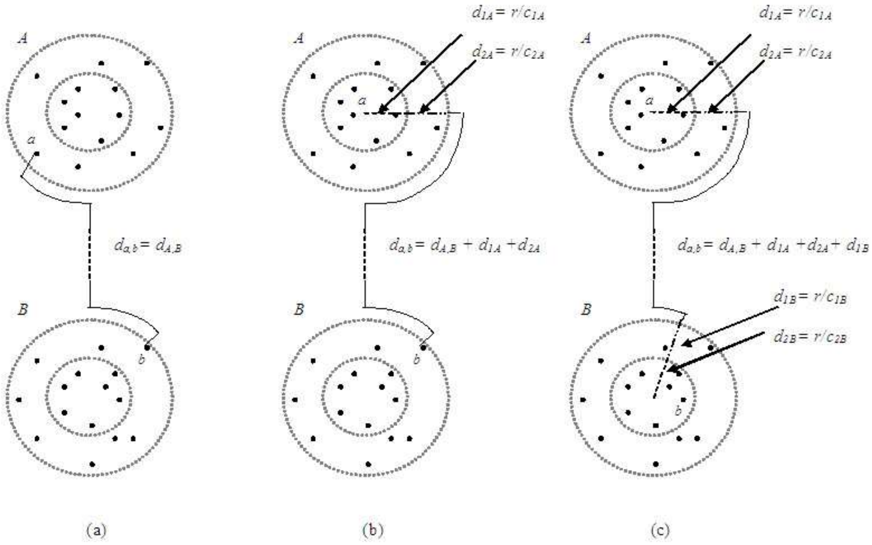

Figure 3. Distance computation for nodes belonging to different cities: Representation of the distances considered in case: (a) node a is in the most external circle of city A and b is in the most external circle of city B; (b) node b is in the most external circle of city B while a is in an interior circle of city A; (c) nodes a and b are in interior circles of cities A and B respectively.

- bernoulli, in which case the parameters attached are n, p and *shift*: all the values of the demand k (between *shift* and $shift+n$) and have probability $p(k) = \binom{n}{k - shift} p^{k-shift}(1-p)^{n-k+shift}$ of realization.

The capacity of the vehicle is defined by the user.

4 Example of instances

In order to give an example of instances that can be created using this procedure, we considered a geographical structure based on two Italian cities, Florence (population 375000) and Genoa (population 632000); the distance between them, along part of the Genova-Rosignano and part of the Firenze-Pisa highways (Figure 4), is 225 km.Following the previously exposed procedure, it is easy to calculate that Florence will be represented with two circles while Genoa with three.

The time distances are considered in minutes, and for ease of visualization integer values are reported. The speed considered in order to compute the coefficients are 15 km/h for the most internal circle of each city, 40 km/h for the most external one, and 80 km/h for the route connecting the two cities. This choice follows the observation of the typical average speed of vehicles in Italian cities and is confirmed by [4]. For allowing a proper visualization, the instances comprehend only fifteen nodes, and the distances are reported as integer values.

In the instance reported in Table 1 (Appendix), the demand of each customer is uniformly distributed, with average comprised between 15 and 25 and spread between 5 and 10. In the one in Table 2 (Appendix), it follows a bernoulli distribution, with the parameters in the following intervals: n between 15 and 20, p between 0.5 and 0.9, $shift$ between 5 and 15.

As can be seen, in both cases nodes 1, 2 and 3 are in the first and nodes 4 and 5 in the second circle of the first city (*Florence*). Then three nodes are placed in each of the circles of the second city (*Genoa*), starting from the smaller one. In the first example the depot (node 0) is placed in the most external circle of *Florence*, while in the second one it is placed in the most external circle of *Genoa*.

Figure 4. Map representing the north of Italy highlighting the itinerary connecting Genoa to Florence

The distance between nodes is 4 hours and 20 minutes in case we consider a node in the center of *Florence* and another in the center of *Genoa*; it decreases to 3 hours and 40 minutes in case one node is in the center of *Florence* and the other is in the middle circle of *Genoa*, and to 4 hours and 5 minutes if the first is in the suburb of *Florence* and the second is in the center of *Genoa*. It becomes 3 hours and 3 minutes for going from the center of *Florence* to the suburb of *Genoa*, and 3 hours and 25 minutes from the suburb of *Florence* to the intermediate area of *Genoa*. Finally the travel time arrives to 2 hours and 48 minutes for connecting the suburbs of the two cities. For moving inside the a city, the distance is comprised between 3 and 130 minutes.

The demand distributions of the customers are represented in Figures 5 and 6 (Appendix), for the first and the second instance respectively.

References

[1] Bertsimas D.J., "A Vehicle Routing Problem with Stochastic Demand", *Operations Research*, vol. 40, no. 3, 1992, pp. 574–585.

[2] Bertsimas D., Chervi P., Peterson M., "Computational Approaches to Stochastic Vehicle Routing Problems", *Transportation Science*, vol. 29, no 4, 1995, pp. 342–352.

[3] Bertsimas D.J., Simchi-Levi D., "A New Generation of Vehicle Routing Research: Robust Algorithms, Addressing Uncretainty", *Operations Research*, vol. 44, no. 2, 1996, pp. 286–304.

[4] Dahme K., Britton H., Epskamp K., "Results of the common indicators: Statistical indicators on local and regional passenger transport in 40 European cities and regions", European Commission, DG Energy and Transport, Brussels, 2002.

[5] Secomandi N., "Exact and Heuristic Dynamic Programming Algorithms for the Vehicle Routing Problem with stochastic Demand", PhD dissertation, College of Business Administration, University of Houston, 1998.

[6] Yang W-H., Mathur K., Ballou R.H., "Stochastic Vehicle Routing with Restocking", *Transportation Science*, vol. 34, no.1, 2000, pp. 99–112.

Appendix

Tables 1 and 2 represent two examples of instances. In the first line the number of nodes $\#N$ and the type of distribution of the demand (1 for uniform and 2 for bernoulli distribution) is reported. In the following $\#N$ lines there are the parameters related to the demand distribution of each customer (for uniform distribution we find the progressive number of the node, *average* and *spread* while for bernoulli one the progressive number of the node, *shift*, p and n). Then the matrix of distances is presented and finally the capacity of the vehicle.

Table 1. Example of instance with demand following uniform distribution

15	1													
0	0	0												
1	21	9												
2	18	7												
3	24	7												
4	18	6												
5	19	8												
6	22	8												
7	16	9												
8	23	6												
9	24	6												
10	20	7												
11	18	6												
12	16	9												
13	19	8												
14	23	5												
0	51	66	72	17	23	245	245	245	205	205	205	168	168	168
51	0	34	25	53	20	260	260	260	220	220	220	183	183	183
66	34	0	13	54	49	260	260	260	220	220	220	183	183	183
72	25	13	0	59	46	260	260	260	220	220	220	183	183	183
17	53	54	59	0	11	245	245	245	205	205	205	168	168	168
23	20	49	46	11	0	245	245	245	205	205	205	168	168	168
245	260	260	260	245	245	0	36	8	56	19	33	39	49	38
245	260	260	260	245	245	36	0	36	56	46	38	75	81	75
245	260	260	260	245	245	8	36	0	65	27	21	40	50	38
205	220	220	220	205	205	56	56	65	0	25	54	36	36	40
205	220	220	220	205	205	19	46	27	25	0	36	27	34	27
205	220	220	220	205	205	33	38	21	54	36	0	60	70	57
168	183	183	183	168	168	39	75	40	36	27	60	0	10	3
168	183	183	183	168	168	49	81	50	36	34	70	10	0	13
168	183	183	183	168	168	38	75	38	40	27	57	3	13	0
100														

Table 2. Example of instance with demand following bernoulli distribution

15	2													
0	0	1	0											
1	19	0.83	9											
2	15	0.82	9											
3	16	0.73	9											
4	19	0.77	7											
5	16	0.57	14											
6	16	0.55	11											
7	19	0.71	8											
8	18	0.78	6											
9	16	0.59	7											
10	19	0.76	12											
11	16	0.88	10											
12	16	0.88	8											
13	16	0.55	9											
14	18	0.55	14											
0	190	190	190	168	168	86	90	109	96	125	128	34	36	38
190	0	18	37	37	38	235	235	235	209	209	209	190	190	190
190	18	0	48	33	28	235	235	235	209	209	209	190	190	190
190	37	48	0	12	76	235	235	235	209	209	209	190	190	190
168	37	33	12	0	49	213	213	213	187	187	187	168	168	168
168	38	28	76	49	0	213	213	213	187	187	187	168	168	168
86	235	235	235	213	213	0	32	37	84	48	43	89	59	61
90	235	235	235	213	213	32	0	17	53	63	41	74	77	82
109	235	235	235	213	213	37	17	0	48	51	25	90	92	97
96	209	209	209	187	187	84	53	48	0	76	51	70	121	130
125	209	209	209	187	187	48	63	51	76	0	25	135	85	85
128	209	209	209	187	187	43	41	25	51	25	0	117	100	102
34	190	190	190	168	168	89	74	90	70	135	117	0	45	50
36	190	190	190	168	168	59	77	92	121	85	100	45	0	6
38	190	190	190	168	168	61	82	97	130	85	102	50	6	0
100														

In Figures 5 and 6 the distributions of the demand of customers are reported, for the instance in Table 1 and 2 respectively. Only 14 graphs are shown since in both cases node 0 is not considered representing the depot, with null demand.

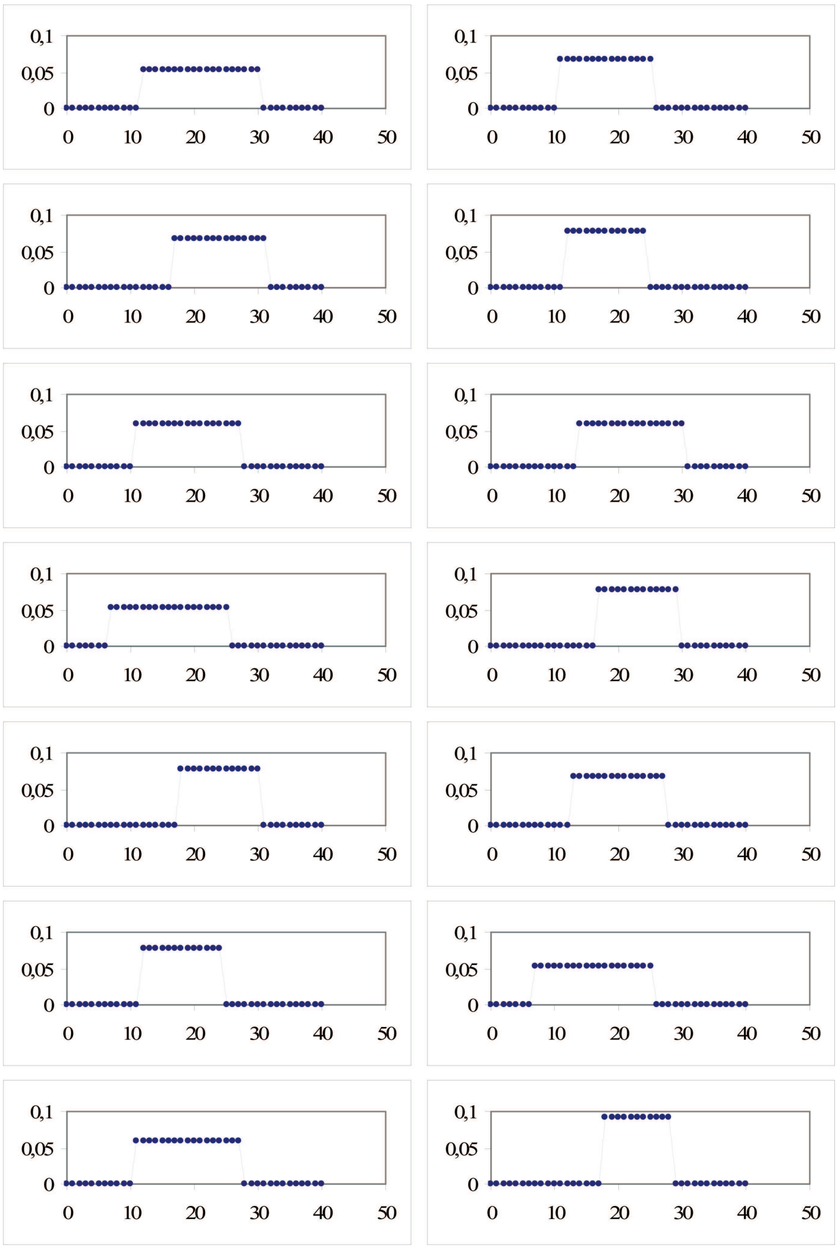

Figure 5. Demand distributions for the instance reported in Table 1

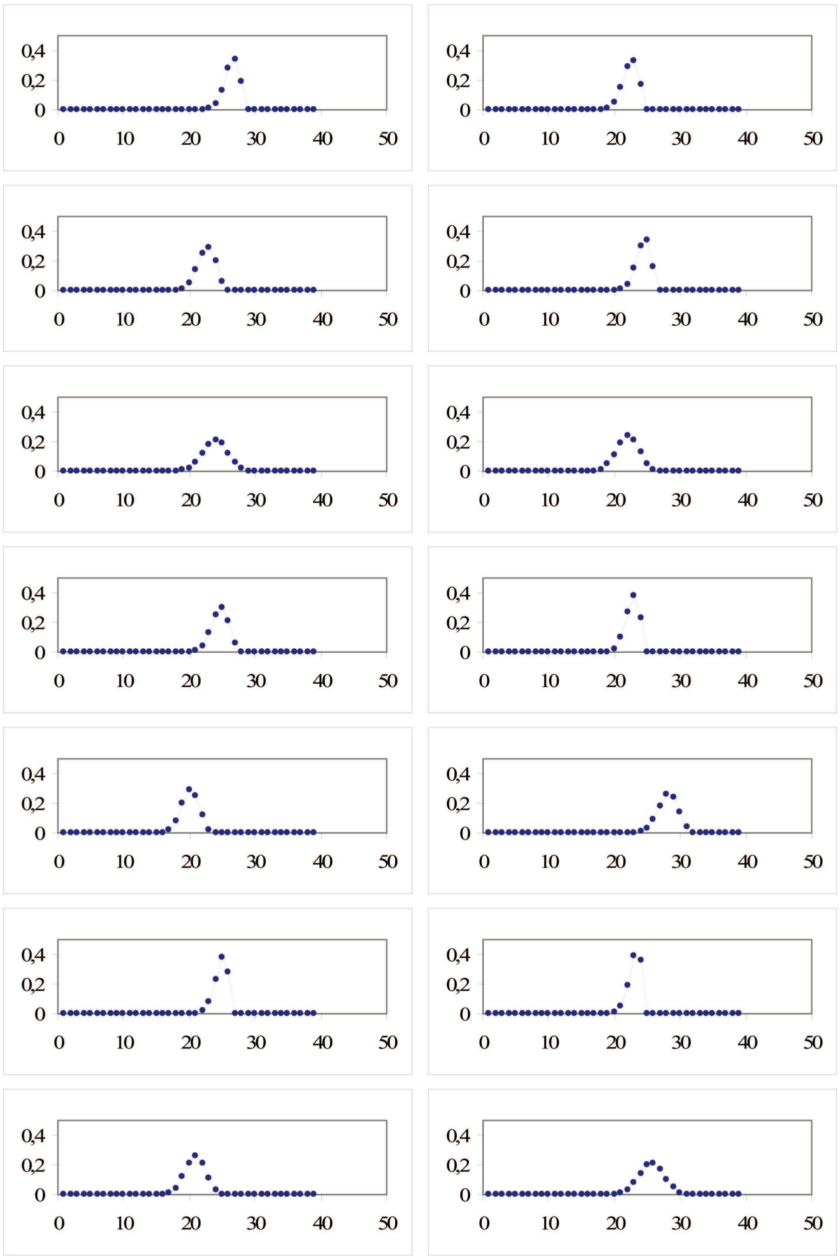

Figure 6. Demand distributions for the instance reported in Table 2