"Science Stays True Here"
Journal of Mathematics and Statistical Science (ISSN 2411-2518), Vol.5, 146-152 | Science Signpost Publishing

Deriving, and Learning About, Spatial Demand Distributions in Location Models

Yigal Gerchak
Department of Industrial Engineering, Tel-Aviv University, Tel-Aviv, 69978, Israel.
Fax: 972-3-640-7669 E-mail: yigal@post.tau.ac.il

Abstract

Location models typically assume that the spatial demand distribution or weights of demand clusters is/are given. However, when a service is new, or considerably altered, such distributions are, in fact, learned gradually as more demands are realized. Recently, location analysts proposed robust optimization models which deal with ranges of weights. We argue that, in principle, uncertain weights can be viewed as mixtures of distributions, and are thus similar to ordinary weights. However, location analysts need a procedure, hopefully a simple one, to revise spatial probability distributions as more information is obtained. We provide a Bayesian model which accomplishes that in a sensible manner. We then show that this theory-based model is, in fact, equivalent to a very simple "physical" mechanism. As the spatial demand distribution evolves with experience, the home bases of servers can be adjusted accordingly, if feasible and desirable.

Keywords: Location, Learning, Bayesian, Robust Optimization.

1. Introduction

When a system or service, or the circumstances under which they operate, are new or have changed considerably, their effective control and management require learning over time. Initial parameters need to be chosen with care and predictions continuously revised in light of previous realizations. As a case in point, a system which dispatches servers to respond to requests originating from a set of locations (as well as a system where customers visit the servers) should, at least in the early stages, revise beliefs about relative magnitudes of future demands originating from various locations in light of where previous calls came from. Since the optimal home-base locations of the servers are a function of the spatial demand distribution, those could be recomputed, and the servers repositioned if the change in distribution is deemed large. The systems we have in mind include emergency services (e.g., Anderson and Fontenot 1992, Jamil, Batta and Malon 1994); automated material-handling systems (e.g., Egbelu 1993, Johnson and Brandeau 1996, Lu and Gerchak 1998a, 1998b); disk systems where a read/write head moves among tracks (King 1990, Vickson, Gerchak and Rotem 1995, Gerchak and Lu 1996); and elevators (Gamse and Newel 1982). Revising spatial distributions will be relevant in these and other systems regardless of number of servers and dimensions, and whether the servers visit customers or vice versa. Yet the location theory literature seems rather silent on the issue of learning and revising beliefs.

The goals of this work are to propose a (Bayesian) scheme for revising spatial distributions as experience accumulates, and to show that a very simple and easy to implement procedure is equivalent to that scheme. Before any demand is observed, the planner's beliefs can be expressed in the form of a prior distribution on the parameters of the predictive distribution. If experience with similar systems is limited, that prior might be rather diffused. At the outset, location decisions will be based on this initial predictive distribution. Much of location theory views what we refer to as predictive probabilities as *weights*. Some location analysts were interested in problems with "uncertain weights" (Drezner and Scott 1999, Averbach and Berman 1997, Demir et al. 2005). Tools from robust optimization (e.g., Kouvelis and Yu 1997) were then utilized. Had "weights" been viewed as probabilities, uncertain weights would become mixtures of probabilities, and hence themselves constitute probabilities (see eq. (1) below and discussion there). Thus "uncertain weights" are, in principle, not really a more general concept than "given weights (probabilities)". In that sense, one of our contributions is in providing a probabilistic framework for this line of research.

The parameters of the predictive distribution will be successively revised, in a Bayesian fashion, as observations accumulate. At first, when experience is limited, the origin of each new request for service will have a significant effect on the distribution (though the origin of the last request is not considered more informative than its predecessors' - the process is assumed stationary.) Eventually, the effect of each additional piece of information will be quite small. Although Bayesian techniques as such are well developed, our setting is, as we will argue, not quite similar to those in which these techniques are usually deployed. The multivariate distribution which we utilize for describing beliefs is the Dirichlet distribution. This is a rich and interpretable family and, in our setting, the posterior will also be Dirichlet, with slightly altered parameters. The predictive distribution is particularly simple.

Practitioners who need to revise beliefs over locations may intuitively come up with' the following simple physical procedure. In n urns, corresponding to the n locations, place balls whose relative quantities reflect our beliefs about the relative magnitude of the location demands. Subsequently, a ball is added to each urn which corresponds to a location that generated a request. At each stage, the distribution of balls among the urns reflects the current beliefs about where the next request will come from. We show that this simple scheme is equivalent to the specific Bayesian procedure just described. Thus implementing the Bayesian procedure is very easy, as this can be done via a simple, easy-to-program, algorithm.

In our Concluding Remarks we summarize our observations, and, for concreteness, specify their use in relocating servers on a line. We comment on the finding of medians of the predictive distributions, and on modeling dependence between locations of successive requests.

2. The Bayesian Framework

Label the locations which generate demand $\{1, \ldots, n\}$. Depending on the type of application, these locations can be in one or more dimensions. Predictions depend on a parameter-vector θ, whose value is unknown to us. At the outset, before any demands are realized, beliefs about this parameter are captured through the prior joint density function $f'(\theta)$ (prior distributions will be denoted by a single prime, while posterior distributions by double primes). The parameter θ affects the location k of next demand through the probability mass function $P(k|\theta)$, k = 1, ... ,n, referred to as the likelihood function. Let R_1 be the unknown location of the first request. The prior distribution and the likelihood generate the prior predictive distribution

$$P(R_1 = k) = \int_{-\infty}^{\infty} P(k|\theta)f'(\theta)d\theta, \qquad k = 1, \dots, n \tag{1}$$

It is the predictive distributions that we seek to obtain, since the location of the server's home base is derived from them. Note that, we see from (1) that, as argued in the Introduction, uncertain ("weighted") probabilities [the $P(k|\theta)$'s weighted by $f'(\theta)$] are themselves probabilities. In contrast, the uncertain-weights robust-optimization approach (Drezner and Scott, Averbach and Berman, Demir et al.) assumes that the unknown weight of location *k*, *k* = 1, ... , *n*, is within a range $[w_k^L, w_k^H]$. Nothing is assumed about the distribution over this range, a state of knowledge to which the Bayesians would refer as a diffused prior, and thus most naturally described by the uniform distribution. In our setting (e.g. (1)) that case would correspond to $f'(\theta)$ being uniform on $[w_k^L, w_k^H]$.

Suppose that one request has now arrived, and it was in location *j*. We now first revise the distribution of the parameter θ:

$$f"(\theta|R_1 = j) = \frac{f'(\theta)P(j|\theta)}{\int_{-\infty}^{\infty} f'(\theta)P(j|\theta)d\theta} \tag{2}$$

The new predictive distribution is

$$P(R_2 = k|R_1 = j) = \int_{-\infty}^{\infty} P(k|\theta)f"(\theta|j)d\theta, \qquad k = 1, \dots, n \tag{3}$$

where R_2 is the random location of the second request. Whether the source of R_1 will have a significant effect depends on how tight was the prior distribution (i.e., how much past experience with similar systems it reflected in it). After *m* successive requests, the predictive distribution uses the most recent update of the distribution of θ, $f^m(\theta|r_1, \dots, r_m)$, where r_i is the origin of the *i*-th request. After many requests, the impact of the prior distribution will become very small.

Formally, the above is a fairly standard Bayesian framework. However, in our setting the likelihood function pertains to the actual demand process, and (location) decisions are made, possibly, after each realization. In most Bayesian settings, the likelihood pertains to a sample which is taken solely for inference purposes (e.g., Raiffa and Schlaifer 1961, Winkler 2003).

3. A Specific Bayesian Model

Denote the probability of a request originating from location *i* by P_i, *i* = 1, ... ,*n*. This might be viewed as a multinomial distribution with a single trial. However, the values of the P_i*'s* are not known to us. We shall assume that, at the outset, this vector of probabilities has the Dirichlet joint distribution, the density of which is

$$f'_{P_1,\dots,P_n}(p_1, \dots, p_n) = \frac{(\sum_{i=1}^n k'_i - 1)!}{\prod_{i=1}^n (k'_i - 1)!} \prod_{i=1}^n p_i^{k'_i - 1} \tag{4}$$

for $p_i \geq 0$, $\sum_{i=1}^n p_i = 1$ and $k'_i \geq 1, i = 1, \dots, n$. This is a multivariate generalization of the beta distribution (e.g., Gelman et al. 2013). As demonstrated in several Bayesian decision analysis texts (e.g., Raiffa and Schlaifer 1961, Winkler 2003) the family of beta distributions (corresponding to *n* = 2 in (4)) can be left or right- skewed, uni- or bi-modal, and assume a rich array of forms. Also, beta "beliefs" can be interpreted (e.g., Winkler) as based on (possibly fictitious) past experience based on $k'_1 + k'_2$ Bernoulli trials of which k'_1 were "successes". Its Dirichlet multivariate extension possess the same richness (e.g., Gelman et al.) and can be interpreted as founded on a past experience based on $\sum_{i=1}^n k'_i$ multinomial trials,

k_i' of which, i = 1, ... , n, originated from location i. Note that the special case $k_i' = 1$, i = 1, ... ,n, corresponds to a (multivariate) uniform distribution on $[0,1]^n$, the joint density of which is $(n-1)!$. Thus the Dirichlet is a rich and flexible family, which can accomodate many patterns of prior beliefs. Also, the Dirichlet distribution can be easily generalized to any finite intervals. Thus the "uncertain weights" discussed above may be viewed as a special case.

Assessing the parameters of a Dirichlet prior distribution should not be too difficult, since its vector of means equals $k_i'/\sum_{j=1}^{n} k_j'$, i = 1, ... , n, so one can obtain the parameter values by asking about the means, with one additional question.

The predictive location distribution of R_1 is thus a Dirichlet mixture of (single-trial) multinomials. For general multinomials, such a mixture gives rise to a multivariate Polya-Eggenberger distribution (Berger 1985, Gelman et al. 2013). In our special case, however, it reduces to the very simple expression:

$$P(R_1 = i) = k_i'/\sum_{j=1}^{n} k_j' , \qquad i=1,\ldots,n \qquad (5)$$

Note that this discrete distribution is as simple to use as if the demand probabilities (weights) were "known" to start with.

Since the multinomial and Dirichlet distributions are known to be a conjugate pair (Gelman et al.), the posterior joint distribution of $P_1, \ldots, P_n$, following an arrival of a request originating from location m, is also Dirichlet, with parameters

$$k_i'' = k_i' , \quad i \neq m , \qquad k_m'' = k_m' + 1$$

The predictive distribution is thus

$$P(R_2 = i | R_1 = m) = k_i''/\sum_{j=1}^{n} k_j'' , \qquad i=1,\ldots,n \qquad (6)$$

which, again, is very simple to compute and use.

One can revise the probabilities after each request or, if the server's home base cannot be relocated so frequently, every several requests; the resulting posterior distribution will be the same. Eventually, the effect of the prior will fade away .

Example

Suppose that n = 7, and that $k_i' = 1$, i = 1, ... ,7 (i.e., the prior is a (diffused) multivariate uniform). The initial predictive distribution is then $P(R_1 = i) = 1/7$, i = 1, ... ,7. Now suppose that of the first five requests to arrive two originated from i = 1, two from i = 2 and one from i = 3. Then

$$k_1'' = k_2'' = 3, \quad k_3'' = 2 , \qquad k_4'' = k_5'' = k_6'' = k_7'' = 1$$

Thus the posterior predictive distribution (of R_6) is then (3/12,3/12,2/12,1/12,1/12,1/12,1/12).

4. An Equivalent Simple Direct Physical Mechanism

We shall now describe a seemingly unrelated "ad-hoc" physical mechanism, which is extremely simple to program and use. As we shall see, it turns out to be equivalent to the previous theory-based Dirichlet-multinomial model.

Suppose that initially we place a_j balls in urn j, j = 1, ... , n. The relative numbers of balls in the urns reflect our prior beliefs as to the relative magnitude of demands. The absolute amounts reflect our confidence in the prior beliefs. Subsequently, each time a request arrives from a location corresponding to some urn, we add a ball to that urn. This procedure is related to the Polya-Eggenberger Urn Model

(Johnson and Kotz 1977, Berg 2006). Thus if, out of *m* requests, location j, j = 1, ... ,*n*, requested service k_j times, the predictive distribution becomes

$$P(R_{m+1} = i) = \frac{a_i + k_i}{\sum_{j=1}^{n} a_j + m}, \qquad i=1,\dots,n \tag{7}$$

This is exactly the same as the predictive distribution obtained from the Bayesian model (6)! Not surprisingly, these probabilities converge to k_i/m. Now,

$$E(R_{m+1}) = \frac{\sum_{j=1}^{n} j(a_j + k_j)}{m + \sum_{j=1}^{n} a_j},$$

which makes intuitive sense. As *m* grows, that expectation converges to $\sum_{j=1}^{n} jk_j/m$.

Also,

$$E(R_{m+1}^2) = \frac{\sum_{j=1}^{n} j^2(a_j + k_j)}{m + \sum_{j=1}^{n} a_j},$$

so

$$Var(R_{m+1}) = \frac{\left(m + \sum_{j=1}^{n} a_j\right)\sum_{j=1}^{n} j^2\left(a_j + k_j\right) - \left[\sum_{j=1}^{n} j\left(a_j + k_j\right)\right]^2}{\left(m + \sum_{j=1}^{n} a_j\right)^2}$$

As *m* grows, that variance converges to $\sum_{j=1}^{n} j^2 k_j/m$.

The choice of the a_j's scale will affect the rate at which the new information changes our beliefs.

This simple physical procedure updates the predictive distribution directly, rather than indirectly via a change in the parameters' distribution as the Bayesian procedure does. However, the results are equivalent.

5. Concluding Remarks

Having an up to date probabilistic forecast of where the next request will come from is important and useful. When one is learning the demand distribution over time, one may react to a significant enough (cumulative) change in the forecast by relocating servers' home bases, or reallocating resources to existing home bases. Thus a coherent, yet simple and fast, procedure for updating probabilistic forecasts as experience accumulates is needed. We provided such a simple physical procedure, and showed it is equivalent to a rigorous Bayesian framework.

Our probabilistic forecast revision framework applies to any finite number of demand sources, regardless of whether they are located on/in a line, plane or have some other topography. For concreteness, however, consider the implication for demands on a line - highway, elevator shaft, etc. Further, suppose one wishes to locate, and relocate, a single mobile server in such an environment.

It is well known that if the goal is to minimize the expected squared distance traveled to a random request on a line, a single server's home-base should be at the median of the distribution. So, for that purpose, it will be sufficient to revise the median of the distribution. After many requests have arrived, the median will simply be the location which had an equal number of requests on both sides. However, early on the prior will also have an effect. Due to the simple form of our predictive distribution, computing the median is very easy. The median is simply the value *M* for which $\sum_{i=1}^{M}(a+k)_i \big/ \sum_{i=1}^{n}(a+k)_i \approx 1/2$. An additional request can only alter the median by one position. Note the difference between our approach, which synthesize all information into one distribution (or median), and the robust optimization approach, which treats each possible distribution separately.

If there is more than one server, and/or the goal is to minimize travel time, which is nonlinear in the distance, the entire distribution becomes relevant (Anderson and Fontenot; Vickson et al., Lu and Gerchak 1998b). The simplicity of our predictive distribution is likely to help there too.

A different approach to the so called "K-server problem" has been discussed in the theoretical computer science literature (Koutsoupias and Papadimitriu 1995, and references therein). There the performance measure is the ratio of the "cost" of a particular server (re)location policy to the "cost" had there been a perfect foresight as to where the requests will come from.

For some systems, it has been observed that the locations of successive requests are not independent (and not only due to Bayesian learning). In databases, for example, the location of the last request is considered more likely to be the source of the next request (Mitrani 1992, p. 555). Mitrani proposes the model

$$P(R_{k+1} = j | R_k = i) = \begin{cases} p + (1-p)P_i & i = j \\ (1-p)P_j & i \neq j \end{cases}$$

where $P_l, \ldots, P_n$ is a' probability mass function over the set of locations and p is the probability that a local run continues. This is special case of a Markovian model. Now, if p is known, our scheme can be used to update the P_i *'s* and eq. (7) used to make prediction. If p is also unknown, and is to be revised simultaneously with the P_i*'s*, the problem is more challenging. In other situations one might argue that when a location generates a demand and this demand is supplied, its updated demand is "one less" than previously. Our model, which assumes independence over time, should be a useful starting point to analyze both the positive and negative dependence scenarios.

It is interesting to note that in the field of search theory (e.g. Stone 1989), Bayesian learning from earlier search results as to the target's location is quite prominent.

References

[1]. Anderson, L.R. and R.A. Fontenot (1992), "Optimal Positioning of Service Units Along a Coordinate Line", Transportation Science, 26, 346-351.

[2]. Averbakh, I. and O. Berman (1997) "Minmax Regret Robust Median Location on a Network under Uncertainty", Location Science, 15, 247-254.

[3]. Berg, S. (2006), "Urn Models", Vol. 14 in Encyclopedia of Statistical Sciences, S. Kotz and N.L. Johnson (eds.), Wiley.

[4]. Berger, J.O. (1985), Statistical Decision Theory and Bayesian Analysis, 2nd ed., Springer-Verlag, New York.

[5]. Demir, M.H., B.C. Tansel and G.F. Scheuenstahl (2005) "Tree Network 1-Median Location with Interval Data". IIE Transactions, 37, 429-439.

[6]. Drezner, Z. and C. Scott (1999) "On the Feasible Set for the Euclidean Weber Problem and Applications". European Journal of Operational Research, 118, 620-630.

[7]. Egbelu, P.J. (1993), "Positioning of Automated Guided Vehicles in a Loop Layout to Improve Response Time", European Journal of Operational Research, 71,32-44.

[8]. Fortini, S. and S. Petrone (2016). “Predictive Distribution (de Finetti’s view), Wiley Stats Ref: Statistical References Online.

[9]. Gamse, B. and G.F. Newel (1982), "An Analysis of Elevator Operation in Moderate Height Buildings-I",

Transportation Research-B, 16-4, 330-319.
[10]. Gelman, A. J. Carlin, H. Stern, D. Dunson, A. Ventari and D. Rubin (2013), Bayesian Data Analysis, 3rd ed. Chapman-Hall, London.
[11]. Gerchak, Y. and X Lu (1996) "Optimal Anticipatory Position of a Disk Arm for Queries of Random Length and Location" INFOR, 34, 251-262.
[12]. Jamil, M., R. Batta and D.M. Malon (1994), "The Traveling Repairperson Home Base Location Problem", Transportation Science, 28, 150-161.
[13]. Johnson, M.E. and M.L. Brandeau (1996), "Stochastic Modeling for Automated Material Handling System Design and Control", Transportation Science, 30, 330-350..
[14]. Johnson, N.L. and S. Kotz (1977), Urn Models and Their Applications: An Approach to Modern Discrete Probability Theory, Wiley.
[15]. King, R.P. (1990), "Disk Arm Movement in Anticipation of Future Requests", ACM Transactions on Computer Systems, 8, 214-229.
[16]. Koutsoupias, E. and C.H. Papadimitriu (1995), "On the K-Server Conjecture", Journal of the Association for Computing Machinery, 42, 971-983.
[17]. Kouvelis, P. and G. Yu (1997) Robust Discrete Optimization and its Application, Kluwer, Dodrecht.
[18]. Lu, X. and Y. Gerchak (1998a), "Minimizing the Expected Response Time to Next Call of an Idled Server on a Line", lIE Transactions, 30, 401-408.
[19]. Lu, X. and Y. Gerchak (1998b), "Multiple Anticipatory Moves of a Server on a Line" Location Science, 5, 269-287.
[20]. Mitrani, 1. (1992), "Computer System Models", Ch. 10 in Computing, E.G. Coffman Jr., J.K. Lenstra and A.H.G. Rinnooy Kan (eds.), Vol. 3 of Handbooks in OR & MS, North Holland.
[21]. Raiffa, H. and R. Schlaifer (1961), Applied Statistical Decision Theory, Harvard University Press.
[22]. Stone, L.D. (1989) The Theory of Optimal Search, 2nd ed., Topics in OR, IN- FORMS.
[23]. Vickson, R.G., Y. Gerchak and D. Rotem (1995), "Optimal Positioning of Read/Write Heads in Mirrored Disks", Location Science, 3, 125-132.
[24]. Winkler, R.L. (2003), Bayesian Inference and Decisions, 2nd ed. Probabilistic Publishing.

Published: Volume 2019, Issue 6 / June 25, 2019