INSIGHTS FOR DEMAND FOR EMERGENCY DEPARTMENT RESOURCES, BASED UPON A PROBABILISTIC MODEL

Donald J. Kelly, Ph.D.
Department of Computer Science and Mathematics
Marist College
Poughkeepsie, New York 12601

ABSTRACT

This paper reports a probabilistic model for the demand arriving at a hospital's emergency department, and applies it in order to gain insights into why certain enhancements work and others do not. The full work is reported in [1].

The emphasis in this work is on demand characterization and on demand management, and not on service modeling. The demand is characterized in terms of client (patient) illness categories and emergency department resource parameters, using the literature and a resource group of medical professionals as a base.

The analytic form of the model is used to gain certain insights, and a spreadsheet is used for additional insights. These include comments on the variability of the resource delivery, provision of storage as a means of improving throughput, referral policies by which the demand variability is reduced, and the effects of shifts in the client base. The end result of this is a tool and a set of insights to understand and explain the nature of the demand so that practitioners can see <u>why</u> certain measures are effective. The spreadsheet and results can also provide the mechanism or conversation "enabler" by which training or planning sessions can be held on data planning needs, and on evaluations of scenarios.

1. INTRODUCTION

Emergency rooms are dynamic environments, the nature of the facility being random arrivals with a variety of client needs. The randomness gives rise to scheduling and staffing problems, with great challenges for an operations manager.

Visits to a number of facilities have established that each facility is almost unique, because of variations in the population served, types of emergencies seen, volume of case load, physical layout of the facility, proximity (or not) of supporting services, and other factors. The challenge is to understand when various innovations can be applied, in what combinations, with what success.

The preliminary work has also established that although there is literature on emergency services modeling, it is sparse and (1) most successful innovations were experiments introduced by creative operations managers, and (2) the data base on service times and characteristics is woefully lacking. Moreover, the typical facility manager is overwhelmed by the possibility that data needed for future planning must be very detailed, and that its collection is likely to be intrusive on the provision of health care.

2. LITERATURE

As a result of the analysis of the existing literature and of interviews conducted at ten facilities in the N.Y. Metropolitan Area, several issues have surfaced that illustrate the fundamental findings of both the literature search and of the supplemental visits:

> Item 1: The purpose of the emergency room is being refined in the 1990's and is fundamentally different from the 1950's. However, the recognition level of this evolution varies from facility to facility. Even in the literature, assessments sometimes reflect the older view of the mission of the emergency facility;
>
> Item 2: The need to disaggregate the client population seems to be of growing importance in assessing the demand for resources. Specifically, certain segments of the population (the young, the old, etc.) require different mixes of services, have different services durations and sometimes need specialty equipment sizes or designs;
>
> Item 3: An understanding of the distribution of customer illnesses identified by the population segment affords knowledge of resources needed for a given time of day, week, and season. Such information will serve as an efficient base for the analytic work of this dissertation which addresses the demand put upon the resources of the Emergency Room/Department both in terms of averages and in terms of a probabilistic model.

A matrix for patient illness categories and categories of resources was created through review of references, including those by Jenkens, Loscalzo [2], Gillies, et al [3] and Eliastam et al [4] and through discussions with a resource group of local experts and practitioners. The compilation of client complaints with their clusters of illnesses and array of resources was based on the authoritative medical text of Eliastam et al [4] along with the lengthy discussions held with the panel of experts. Service distributions were estimated in consultation with experts at several facilities visited.

The two dimensional matrix of patient illness and resource categories provides a tool with which to estimate frequency and level of demand for resources and to assist at a later date in the modeling of a 3-ply matrix of population segment, customer illness category, and resources, as shown in Figure 1.

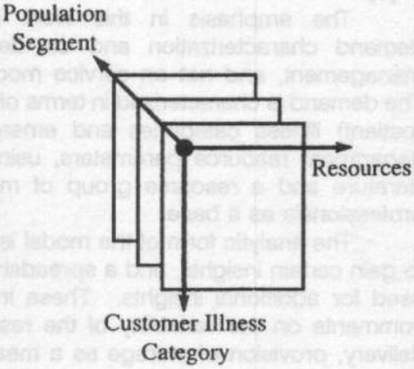


FIGURE 1

3. THE BASIC PROBABILISTIC MODEL OF SERVICE NEED

The basic model does not focus on the efficiency with which the demand is serviced, such as a queueing analysis or simulation might. Rather, it focuses on the "demand for service", its variability, and means of alleviating that variability. Some supplemental queueing modeling is done to address the efficiency with which the demand is processed, but this is secondary to the prime purpose: providing an understanding of the conditions and parameter ranges under which some strategies work and others do not.

Consider that there are N customer/patient demand categories ($i = 1,2,.....,N$) and M services ($j = 1,2,.....,M$). The customer/patient categories might range from cuts to respiratory distress to trauma cases. The resources might range from triage to x-ray to endoscopes. Not all services are necessarily invoked for a given customer category.

a) N Customer Categories and M Services, Poisson Demand The number of customers Y_i which arrive in a given demand category "i" is a random variable, with $Pr(Y_i=y) = p_i(y)$. The number of people showing up to be treated in a given time period can often be modeled as following a Poisson distribution with reasonable accuracy. Each Poisson distribution has mean λ_i. The mean may vary over time, but we will consider it fixed for the time period of the analysis.

For a given category "i", the patient is provided a set of resources ($j=1,2,...,M$), the duration of each of which is a random variable. The mean, variance, and distribution of the duration differs with both "i" and "j". The probability density distribution of the j^{th} resource for the i^{th} customer group is denoted $f_{ij}(t)$ in this work. The corresponding random variable is denoted T_{ij} in what follows.

Let ST_{ij} denote the "time need for the resource" for all Category "i" who arrive on the shift (i.e. the analysis period) needing resource "j". Because both the number of arrivals and the durations are random, it is clear that the ST_{ij} are also random variables.

If there were exactly "y" customers of the i^{th} type, the demand for resource "j" would have a distribution $g_{ij}(t|y)$ which is simply the convolution of $f_{ij}(t)$ with itself "y" times. This is so because the random variable $ST_{ij} = T^{n=1}_{ij} + T^{n=2}_{ij} + \cdots + T^{n=y}_{ij}$ where the "n" denotes individual customers, from n=1 to n=y (i.e. individual samples from the $f_{ij}(t)$ distribution).

Because $g_{ij}(t|y)$ is the result of a convolution of the same distribution with itself several times, the Central Limit Theorem can be invoked and we can say that $g_{ij}(t|y)$ is normal with mean $y\mu_{ij}$ and variance $y\sigma^2_{ij}$. This is only valid for larger values of "y", the number of customers. However, as long as the $f_{ij}(t)$ are relatively smooth, the normality of the $g_{ij}(t|y)$ can be used in practice, even for small "y".

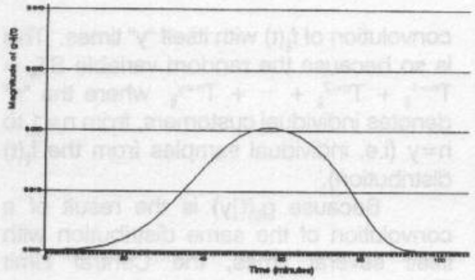
For those cases in which the "y" is too small and the $f_{ij}(t)$ too sharply defined, the next step will minimize the effect. *The net result is that the exact form of the $f_{ij}(t)$ distribution is rather unimportant, and they can be characterized adequately by their mean and variance.*

The distribution of $g_{ij}(t)$ is another matter, however. It is not simply the sum of several random variables. Rather, it takes on several different shapes, depending upon the number of customers "y". It is the composite of these shapes, weighted by their relative probabilities; let $P_r(Y_i = y) = p_i(y)$. Therefore,

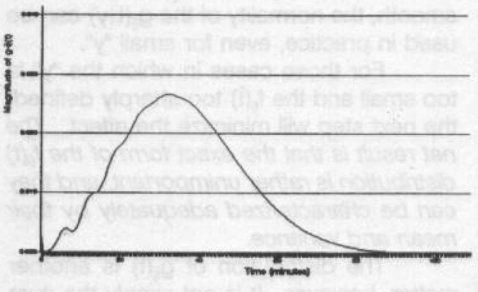
$$g_{ij}(t) = \sum p_i(y)\ g_{ij}(t|y)$$

where the summation is over "y". Further, the summation is for $y \geq 0$, so that there is a concentration of probability at $y = 0$. The $g_{ij}(t)$ is the simple addition of the weighted components (multiply each $g_{ij}(t|y)$ by its corresponding $p_i(y)$ value and then add).

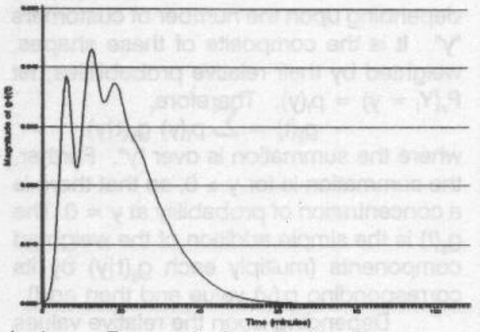
Depending upon the relative values of the means μ_{ij}, variances σ^2_{ij}, and the demand distributions, the resultant ST_{ij} distributions can be rather smooth (Figure 2a), show some irregularity (Figure 2b), or have pronounced irregularities (Figure 2c). These cases are all shown for Poisson demand, with various λ_i and $\mu_{ij} = 6$



a) High Demand (λ = 10); Rather Smooth Curve



b) Lower Demand (λ = 6); Some Irregularities



c) Low Demand (λ = 3); Pronounced Irregularities

FIGURE 2
DIFFERENT ST_{ij} DISTRIBUTIONS, DEPENDENT UPON DEMAND FOR VARIOUS λ_{ij}, WITH μ_{ij}=6 AND σ_{ij}=1.6

minutes, σ_{ij} = 1.6 minutes.

After some reflection, it is clear that the irregular distributions occur when the component distributions do not overlap.

Note that the tail is relatively smooth even in Figure 2c even though the principal components are not, because the $\sigma_{ij}\sqrt{y}$ are larger compared to the step; for y=10 and σ_{ij} = 0.7 minutes, the $\sigma_{ij}\sqrt{y}$ = 2.2 minutes and the even the "2σ" spread is ± 4.4 minutes about the center of $y\mu_{ij}$ = 10(6) = 60 minutes for this value of "y", thereby overlapping the y=9 and y=11 distributions (with the $p_i(y)$ providing further tapering).

b) Convolution to Generate the Resource Need Distributions TT_j The ST_{ij} distributions are not the total measure of demand for a particular resource, but simply components of that demand. Define TT_j as the total demand for a resource "j", and note

$$TT_j = \sum ST_{ij}$$

where the summation is over "i", the various customer categories.

Let $h_j(t)$ be the probability density function associated with TT_j. It is the result of a convolution of the $g_{ij}(t)$, for each of the N "i" terms. The mean and variance of TT_j is simply computed in terms of the ST_{ij} means and variances, assuming independence.

The convolution of the ST_{ij} does not result in a simple invocation of the Central Limit Theorem in this case, however. The ST_{ij} distributions are each different from the others, and it is not at all a case of adding several identical distributions. The TT_j is not necessarily normal, although it is frequently smooth and appears to be normal. The smoothness can occur because the summation (i.e. the convolution) has this effect.

Although the TT_j distribution is frequently smooth, it is also true that it often has a tail, due to the long tail on one or more of the component ST_{ij}

distributions. At the same time, the TT_j and the comparable normal often match in the upper tail, even if they do not for lower values. The match in the upper tail has an interesting implication: decisions made based upon the 95th percentile point would not be far wrong if the normal distribution were used.

4. SUPPORTING COMPUTATIONAL TOOLS

In order to explore the implications of the above formulation, and to investigate ranges of parameters, a set of programs were written and interfaced with both a spreadsheet and a graphics package. Much of the same analysis could be done in spreadsheets, although it is not as convenient to accommodate an arbitrary $f_{ij}(t)$ distribution. A spreadsheet implementation leading to the $g_{ij}(t)$ was done for Poisson customer arrivals and normal $f_{ij}(t)$; it could be easily adapted to non-Poisson demands, and adapted with more difficulty to other $f_{ij}(t)$. This spreadsheet was used for extensive analysis of cases to gain insights.

Some additional spreadsheets were prepared for comparisons of two different parameter set analyses. Reference [1] contains the spreadsheet for a full implementation of the sixteen client categories and twenty-five resources cited.

5. SUPPORTING ANALYTIC WORK: μ_{TTj}, σ_{TTj}^2

Exercising the programs reinforced the early perception that there are too many parameters for efficient insights, whether by computation or by simulation. The parameter sets to consider include:

The number of customer categories, N;
The number of resources, M;
The demand rates, λ_i;
The mean service duration times, μ_{ij};
The service duration time standard deviations, σ_{ij}.

This does not even address the questions of the work flow and the queueing as patients are actually processed, which is not within the scope of this paper.

Rather than design a "complete block" simulation experiment over these parameters, and then search for insights, it was decided to pursue additional analytic development. The insights gained into the shapes of $g_{ij}(t|y)$, $g_{ij}(t)$, and $h_j(t)$ were used to focus on the expressions for mean and standard deviation, rather than distribution shape.

Reference [1] contains the derivations for the mean and variance of the ST_{ij} and the TT_{ij}. The relation for μ_{TTj} is:

$$\mu_{TTj} = \sum_i \mu_{ij}\, \mu_{Cust\,i} \qquad (1)$$

This states that the various results should be added up in order to determine the total average demand for service.

The variance is:

$$\sigma^2_{TTj} = \sum_i \{\sigma^2_{ij}\, \mu_{Cust\,i} + \mu^2_{ij}\, \sigma^2_{Cust\,i}\} \qquad (2)$$

The simplicity of these relations provides us with a powerful tool, which will actually drive much of the analysis.

In many cases, it is the second terms which dominate. The overall variance is caused by the product of the uncertainty in the demand ($\sigma_{Cust\,i}$) and mean service time (μ_{ij}). This is where attention must be focused. Of course, the overall analysis is complicated by the fact that several such terms are added to obtain the variance of TT_j.

Equations 1 and 2 are used in the spreadsheet "COMPARE".

a) Special Case: Poisson Demand

The above derivations were done for an arbitrary demand distribution. Further, they did not make use of any assumption on the normality of the $f_{ij}(t)$; these distributions were also of arbitrary form.

If the arrival demands can be

characterized by Poisson distributions, some special forms of Equations 1 and 2 result:

$$\sigma^2_{STIj} = \lambda_i \{\sigma^2_{ij} + \mu^2_{ij}\} \quad (3)$$

and

$$\sigma^2_{TTj} = \sum_i \lambda_i \{\sigma^2_{ij} + \mu^2_{ij}\} \quad (4)$$

This shows that the $f_{ij}(t)$ contribute primarily through their means, not their variances (unless there is exceptional variability in the service durations, compared to the means).

b) Insights From the Analytic Formulation Equation 4 for the Poisson arrivals may also be rewritten as

$$\sigma^2_{TTj} = \sum_i \lambda_i \mu^2_{ij} \{1 + CV^2_{ij}\} \quad (5)$$

where $CV_{ij} = \sigma_{ij}/\mu_{ij}$ is the coefficient of variation. It is reasonable to expect $CV_{ij} \le 0.20$ in most cases, so that the term in the brackets is generally less than 1.04. This emphasizes the importance of the $\lambda_i \mu^2_{ij}$ terms, and the reality that improvements to the σ_{ij} have little benefit in most cases: they do not contribute to the mean μ_{TTj} at all, and are generally negligible in σ^2_{TTj}.

The coefficient of variation for the TT_j can be written as

$$CV_{TT} = \sqrt{\frac{CV^2_{11}}{\lambda} + \frac{\sigma^2_1}{\lambda^2}} \quad (6)$$

where CV_{11} is the coefficient of variation of the $f_{ij}(t)$ distribution for i=1 and j=1, and where σ^2_1 is the variance of the customer demand.

Clearly, if $\sigma^2_1 = 0$, then the two coefficients of variation are directly proportional, and CV_{TT} is inversely proportional to the square root of the number of customers.

Just as clearly, whenever σ^2_1 is significant compared to the mean number of customers, the second term in Equation 8 quickly dominates <u>and</u> drives toward an asymptote determined by the σ^2_1/λ ratio.

What emerges is an illustration of the relative importance of different terms, under different conditions. Of course, if the effects of two customer categories with different ratios were considered concurrently, additional complexities are introduced.

It is also important to recognize that a stable CV ratio is not sufficient. Consider the scheduling of people or machines to meet the demand: both people and demand come in integer quantities. Even if CV is stable, the σ increases in proportion to the number of customers.

As the variance of the demand increases, the resource duration <u>needed</u> also increases. Clearly, as the demand becomes more uncertain, it might be necessary to bring on more staff or more equipment, simply to deal with the uncertainty.

At the same time, increasing demand generally decreases the <u>relative</u> variability. Nonetheless, the absolute variability grows. This is a situation in which the number of staff or equipment required becomes more uncertain on a fixed scale, despite being a smaller percentage problem.

Can an emergency department cope with more uncertainty because it is a smaller percentage of the business? The best answer is, it depends. Certainly it might be a surprise for some administrators to find that an uncertainty of ±1 person becomes an uncertainty of ±2 persons when the business goes up by a factor of four.

6. CASE STUDIES

Three case studies were considered in the complete work [1]:

1) the variance of TT_j is studied,

specifically in terms of the contributing factors;

2) the effects on the needed resources, when the population using the facility shifts, or when facilities in different areas (with different clientele) are considered;

3) the strategy of referring certain overflows to other facilities, thereby making the demand more regular at least one of the facilities, and the effect on both (or all) of the participating facilities.

From Equation 4, the point has already been made that σ^2_{ij} is generally negligible compared to μ^2_{ij}. Rather than it mattering that the σ_{ij}/μ_{ij} ratio is 0.20, substitution allows Equation 4 to be re-written as

$$\sigma^2_{TTj} = 1.04 \sum_i \lambda_i \mu^2_{ij} \qquad (7)$$

Simply put, the σ_{ij} simply do not contribute and their specific values are virtually irrelevant in this case.

Notice that if the μ_{ij} were all comparable, then that term could be taken out of the summation in Equation 9, and made a coefficient. The principal term would then be $\sum_i \lambda_i$, and the σ_{TTj} would be driven by this term.

What are some of the insights from these case studies?

From the first case study, three points: (1) the σ_{ij} simply do not contribute in the typical situation encountered, so that undue attention to decreasing them simply does not make sense; (2) the randomness of the Poisson demand shows up in the CV_{TTj} with full force, and any effort in mitigating the randomness of the input demand has significant benefit; (3) any resource which has one or more μ_{ij} which "stand out from the crowd" contributes directly to the CV_{TTj} and must be reviewed --- particularly if it is multiplied by a significant λ_i.

The real issue in this last point is that any $\mu^2_{ij}\lambda_i$ term which is a significant part of the variance term σ^2_{ij} or dominates it must be addressed. On the other hand, terms which are so small that they "hide" amongst much larger terms need not be addressed at all, unless in concert with sets of such terms.

In the second case study, for minor shifts from a "typical" population to a heavier representation of "older" people with greater frequencies in some categories, there were increases in 20 of the 25 resource categories, representing an average percent increase of 14.3% within those experiencing an increase, and a total of 8.9% increase in needed service-hours for the same total number of patients. Of course, in addition to the shifts themselves, in most organizations a decrease in need within four categories (one had no change) does *not* result in an immediate shift in support of the 20 resources experiencing growth. This institutional reality exacerbates any problems which occur due to a different or changing clientele profile.

Needless to say, as the nature of the resource need changes, the facility must evolve. A number of redesigns are simply attempts to catch up, in one step, with past clientele changing patterns. The COMPARE sheet is one tool which can aid emergency department operators and designers articulate these changes, recognize them, and adjust.

The results of the third case study are more complex to present, and are to be reported in the literature separately.

7. CONCLUDING REMARKS

In the course of this work, it became clear that there is no real consensus on the client categories, nor on the estimates of the resource duration estimates, nor even on the major groups into which the population should be disaggregated. The very existence of a framework (i.e. the "COMPARE" spreadsheet) with labels and numbers will surely (a) make the assumptions clear and open to easy refinement, and (b) generate the discussion leading to such refinement.

The author is also satisfied with the insights gained by using the analytic formulation as a tool. Simulation, which can be a valuable tool, is sometimes used as a brute force approach. Innumerable runs would have been needed to "reveal" the relationships and sensitivities reported in this paper.

These lessons become more complicated to apply, and the effects more subtle, when several customer categories with significantly different resource needs are present at the same time. Nonetheless, they are valid, as could be seen in the case studies in [1].

One of the products of this work is a conceptual framework, namely the linkage from the disaggregated demand and the resource descriptors $f_{ij}(t)$ to the total needs for resources. This was the product of many discussions and reviews. It became clear that the framework itself --- implemented in the "COMPARE" spreadsheet --- was an effective vehicle for focusing the discussion.

Rather than the conversations diverging based upon different perspectives, people with disparate backgrounds found common ground: Are the client and resource lists detailed enough? Are important cases covered? Are the μ_{ij} and σ_{ij} reasonable? Are the results relevant?

The author found that discussions focusing around the set of resources needed, particularly in the face of a changing customer base, freed people to address these questions and start volunteering alternate scenarios (What if this happened? What about that?). Clearly, the spreadsheet became an "enabling technology" for people to focus on quantitative exercises they had previously found cumbersome or did not fully conceptualize.

Therefore, it is planned that in later work the spreadsheet will be used as the tool by which one or more teams address the underlying assumptions (which are explicit, even obvious) in workshops and/or focus groups.

It also became clear that the spreadsheet can be used to train or "bring up to speed" people who are uninitiated in thinking about trends, future planning, and redesign. This can be done in the format of a one-day training course, with worked problems.

REFERENCES

1. Kelly, Donald J., "Insights for Demand for Emergency Department Resources, Based Upon a Probabilistic Model", PhD Dissertation, Polytechnic University, Brooklyn NY, June 1993.

2. Jenkins J.L MD, Loscalzo J. MD, Manual of Emergency Medicine(Diagnosis and Treatment) Second Edition Little, Brown and Company. 1991

3. Gillies J,Marchall S, Ruedy J, On Call. Principles and Protocols. W.B. Saunders Co. 1989

4. Eliastam M, Sternback G, Bresler MJ, Manual of Emergency Medicine. Fifth Edition, Mosby Year Book