Surgical Patient Flow and Bed Requirement

Revlin ABBI

School of Computer Science,
University of Westminster,
London, UK
abbir@westminster.ac.uk

Abstract. Various techniques have been proposed that use patient length of stay (LOS) data to derive patient flow models, which in turn help to better understand the temporal characteristics of the patients cared for by a health care facility. An approach complementary to flow modelling and developed by the authors is concerned with deriving the case-mix of patients from LOS observations, and building a LOS patient classification model. Based on this perspective, the population consist of various groups, staying in hospital according to the statistical properties of the fitted GMM. In this paper, based on an average number of admissions, we illustrate how the grouping can theoretically describe the patients entering into hospital and how this knowledge can further be used to determine the bed requirement.
Keywords: length of stay(LOS), surgical patients flow, bed requirement, Gaussian mixture model
Math. Subject Classification 2000: 62H30; 65C20

Introduction

Various techniques have been proposed that use patient length of stay (LOS) data to derive patient flow models, which in turn help clinicians and managers to better understand the temporal characteristics of the patients cared for by a health care facility (Millard P.H, 1988, Harrison G.W and Millard P.H, 1991, Millard P.H, 1992, Harrison G.W, 1994, Faddy M.J and McClean S.I, 1999). An approach complementary to flow modelling and developed by the authors is concerned with deriving the case-mix of patients from LOS observations, and building a LOS patient classification model (Abbi R et al., 2007b, Abbi R et al., 2008). In summary, the methodology comprises of several processing steps (Abbi R et al., 2007b, Abbi R et al., 2008), where the optimal Gaussian mixture model (GMM), based on the minimum description length criterion (Rissanen J, 1978), represents various groups of patients according to their LOS.

The GMM of a given patient population describes patients as belonging to one of the m groups defined. Based on the perspective that the population consist of m groups, we can view the population as staying in hospital according to the statistical properties of the fitted GMM. In this paper, based on an average number of admissions, we illustrate how the GMM can theoretically describe the groups of patients entering into hospital and how this knowledge can further be used to determine the bed requirement.

The LOS of a patient population is a very important factor when determining the bed requirement for a given health care facility because it can often be diverse depending on the case-mix. Such diversity has direct implications on bed requirements and can make it difficult to understand the way in which beds are being used. Knowledge of the case-mix of patients, their LOS, and the corresponding number of beds used, helps to provide a greater understanding and insight as to the way in which the various groups of patients treated are occupying the beds for the given health care facility. This can be used to fully understand the relationship of the likely bed-usage for each group, as well as the implications on the number of patients treated.

Methods

The Gaussian Mixture Model

The GMM is a probability density function comprising of m normally distributed component functions (Titterington D.M et al., 1985, McLachlan G.J and Peel D, 2000). These normally distributed components are combined together to form the overall density model, flexible enough (depending on m) to approximate almost any distribution (Bishop C.M, 2006).

We use the GMM to approximate the LOS distribution, where each normally distributed component is used to model a LOS group, described using three parameters: mean, variance (or standard deviation) and mixing coefficient. The mean μ_j for component j expresses the most likely LOS for patients belonging to LOS group j, whilst the variance σ_j^2 (or standard deviation σ_j) quantifies the variation within the LOS group. The mixing coefficient ω_j for component j is used to describe the likely proportion of the overall patient population belonging to the group j.

The probability $p(x)$ of a patient staying x days, without any knowledge of the patient and any possible seasonal or weekly fluctuations is defined as the sum of the probabilities from each component, $p(x) = \sum_{j=1}^{m} P(j)p(x|j)$. In this case, $P(j)$ is the prior probability of belonging to group j, equivalent to the mixing coefficient ω_j, and $p(x|j)$ is the conditional probability of LOS observation x belonging to a Gaussian function, parameterised according to component j, $p(x|j) = \frac{1}{\sqrt{2\pi\sigma_j^2}} exp(\frac{(x-\mu_j)^2}{2\sigma_j^2})$. The probability of a LOS observation belonging to a LOS group j is derived using the Bayes rule, $p(j|x) = \frac{p(x|j)P(j)}{p(x)}$.

Fitting the Gaussian Mixture model to LOS data

The expectation maximisation (EM) algorithm (Dempster A.P et al., 1977) is used to estimate the parameters of the GMM from the LOS data (a process also known as fitting the model to the data). Instead of randomly initialising the EM algorithm and to reduce computation, the k-means clustering algorithm

(MacQueen J. B, 1967) is employed. In this cas , the $(\frac{100}{m+1} * j)^{th}$ percentile value of the LOS data is used as the initial cluster centre c_j as inputs for k-means, where m is the number of groups.

Finding the number of groups

For human comprehension considerations (Miller G.A, 1956) and from own experimental analysis we only consider GMMs with between two and six components. In order to find the optimum number of components we employ the Minimum Description Length (MDL) criterion (Rissanen J, 1978).

Although MDL has shown to be effective for model selection (Walter M et al., 2001), it is also known for over-estimating the number of components (Walter M, 2002). As such, we assess the contribution of additional components based on the percentage decrease of the MDL value. The MDL criterion has also been validated against other commonly used criteria (Abbi R et al., 2007a), such as the Akaike information criterion (Akaike H, 1973) and the Bayesian information criterion (Schwarz G, 1978), and was found to suggest the same number of components.

To further aid the selection of an appropriate GMM, ten random samples of synthetic data are generated based on the parameters of each GMM. The 10^{th}, 25^{th}, 50^{th}, 75^{th}, 95^{th}, 99^{th}, 99.5^{th}, and 100^{th} percentile values for all samples are averaged, and used to make comparisons, measuring how well each GMM fits the LOS data.

Patient flow

Based on a derived GMM for a given patient population, we interpret patients entering a hospital as belonging to one of m groups, Figure 1. The probability of a patient admitted into hospital to belong to group j is determined by $p(j)$ i.e. the prior probability equivalent to the mixing coefficient ω_j.

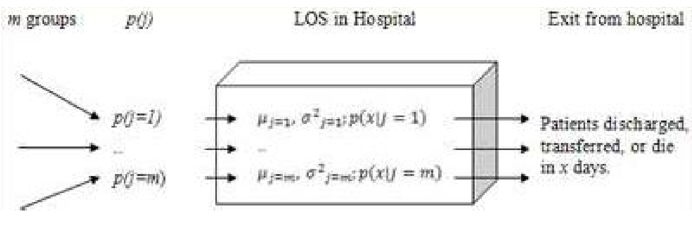

Fig. 1. Flow of m groups of patients, where group j stays in hospital for μ_j days depending on the variation σ_j^2, where the probability of a patient staying x days is given by $p(x|j)$.

As depicted within Figure 1, a patient belonging to group j is likely to stay μ_j days depending on the variation σ_j^2 of group j. However, as an alternative to using (the average LOS for a patient belonging to group j), a more accurate estimate of the likely LOS of a patient is to use $p(x|j)$ i.e. the probability of a patient belonging to group j being discharged in x days.

Bed requirement

To estimate the average bed requirement we take into account the average number of patients entering the hospital, the groups of patients within the population and their corresponding LOS in hospital. Amongst the average number of patients admitted each day, the calculation involves (1) the probability of a patient admitted into hospital to belong to group j, and (2) the probability of a patient belonging to group j being discharged in x days.

Based on probability theory, we calculate the likely bed requirement according to four core components of the bed requirement model. These are as follows:

1. Patients entering hospital: This is defined as an average number of daily admissions, denoted as a, where a_j of these patients, defined as $a_j = a \cdot w_j$, belong to group j.
2. Likely discharge of patients: The probability of a patient, belonging to group j, being discharged d days after admission, Equation 1.
   a. Using the cumulative Gaussian function $F_j(d) = P(x \leq d)$, defining the probability of being discharged in d days, and the average number of patients being admitted per day a_j ,we can calculate the likely number of patients to be discharged within d days, calculated as $[F_j(d) \cdot a_j]$.
   b. We can also derive the likely percentage of patients from a_j that are discharged within d days after admission, defined as $[F_j(d) \cdot 100]$.
   c. Note that the probabilities associated with negative LOS values refer to patients who stay zero days, discharged on the same day of being admitted.
3. Likely number of patients still in hospital: By defining patient admission and discharge percentage rates, we can derive, for each group, the likely number of patients in hospital.
   a. To calculate the likely occupancy levels for group j, we denote, n_j^d as the likely number of patients, from the admitted a_j, still in hospital d days after being admitted, Equation 2.
   b. In addition, we can also derive the percentage of those who are still in hospital d days after admission $[1 - F_j(d)] \cdot 100$.
4. Accumulation of patients still in hospital and bed requirement: Based on the above, we can derive the likely accumulation of patients and the bed requirement.
   a. We drive the number of days d' where it becomes most likely for all patients belonging to group j, to be discharged, i.e. $F_j(d') \geq 99.5$.

b. We then to calculate the bed requirement for group j by taking into account the accumulation of patients within hospital from $d = 1$ until day d', i.e. $\sum_{q=1}^{d'} n_j^q$, see Table 1.
c. The bed requirement for the overall population takes the sum for all m groups, i.e. $\sum_{j=1}^{m} \sum_{q=0}^{d'} n_j^q$.

$$F_j(d) = P_j(x \leq d) = \int_{-\infty}^{d} f_j(x)dx \qquad (1)$$
$$n_j^d = [1 - F_j(d)] \cdot a_j \qquad (2)$$

Table 1. The accumulation of patients belonging to group j

Day (d)	# of patients admitted	# of patients in hospital	# of patients discharged
1	a_j	n_j^1	$(a_j - n_j^1)$
2	a_j	$n_j^2 + n_j^1$	$(a_j - n_j^1) + (a_j - n_j^2)$
3	a_j	$n_j^3 + n_j^2 + n_j^1$	$(a_j - n_j^1) + (a_j - n_j^2) + (a_j - n_j^3)$
.	a_j		
d'	a_j	$n_j^d + n_j^{d-1} + ... + n_j^1$	$(a_j - n_j^1) + ... + (a_j - n_j^{d-1}) + (a_j - n_j^d)$

The table above mathematically depicts the accumulation of patients as time passes from day 1 to day d'. For instance, on day $d = 1$, a_j patients belonging to group j are admitted into hospital. Amongst the a_j of patients admitted, n_j^1 stay in hospital, are discharged on the same day of being admitted, and n_j^2 stay in hospital overnight and are still in hospital on day $d = 2$. Thus the next day ($d = 2$), a further a_j are admitted, where n_j^1 stay in hospital, n_j^2 are still in hospital from day $d = 1$. Moreover, $a_j - n_j^1$ are discharged on the same day of being admitted and $a_j - n_j^2$ are discharged after being in hospital for two days.

This cycle of patients entering into the hospital system and leaving continues and the number of patients in hospital accumulates until day d' where the system reaches a stable state. In other words the time required to obtain a consistent bed occupancy level, whereby the build up of patients is taken into account according to the admission rate and LOS of patients.

Dataset

The data used is a surgical administrative discharge dataset. It consists of 7,723 records detailing the spells of patients undergoing some form of surgery in a tertiary hospital in Adelaide, Australia, discharged between 1^{st} July 1997 and 30^{th} June 1998. The variables describing each patient spell include the dates of hospital admission and discharge, the LOS in days, the gender of the patient, whether the patient was treated in a public or private hospital, whether the patient was admitted as an emergency case, and finally the diagnosis information - coded using major diagnostic categories (MDC). The MDC coding system consists of 25 categories, each of which corresponds to a single organ system. Each MDC for a patient is determined by the primary disease or condition for which a patient is hospitalised or treated.

Results

The Surgical data was split into two subsets and various GMMs were fitted to the LOS data using the EM algorithm. Based on the MDL criterion, percentile analysis and the statistical criteria the four component model was selected. The parameters of the four component model indicate that patients stay in hospital according to one of the four patterns of stay, Table 2. Most of the patients belonging to group one will all be discharged within four days of being admitted, Table 3. In addition, patients belonging to group two will be discharged within 12 days for group two, 29 days for group three, and 125 days for group four, Table 4.

Table 2: Mean, variation, and proportion for each of the four groups of surgical patients

	Patient group				
	f=1	f=2	f=3	f=4	Total
Mean LOS (days)	2.2	5.4	13.4	39.2	-
Standard Deviation	0.5	2.2	6.0	26.0	-
Proportion of patients	38.3%	39.4%	18.8%	3.5%	-

Table 3: Percentage of patients belonging to group 1 being discharged in d days

Days (d)	% of patients discharged
0	0
1	0.8
2	34.5
3	94.5
4	99.9

Table 4: Number of days when it becomes most likely for patients to be discharged

Group	d'
1	4 days
2	12 days
3	29 days
4	125 days

To determine the likely bed occupancy we derived the average daily number of surgical patient admissions, which was 21 admissions per day. In this case, from the 21 admissions, we estimated that eight patient admissions would exhibit the LOS properties of group one, eight patient admissions of group two, four patient admissions of group three, and 1 patient admission per day to group four.

According to the average admissions for group one, on day d', 21.5 patients are in hospital, thus we define the bed requirement for group one as 22, Table 5, and we derived the average number of beds that were most likely to be required for treating the surgical patient population, Table 6.

Table 5. The accumulation of patients belonging to group j

Day (d)	# of patients admitted	Accumulation of patients in hospital
0	8	8
1	8	7.9 + 8.0
2	8	5.2 + 7.9 + 8.0
3	8	0.4 + 5.2 + 7.9 + 8.0
d'= 4	8	0.0 + 0.4 + 5.2 + 7.9 + 8.0

Table 6: Average number of surgical patient admissions and occupied beds for each group

patient group (j)	Average admissions per day (21)	Average number of beds	% of beds used
1	8	22	14.1
2	8	49	31.4
3	4	55	35.3
4	1	30	19.2
Total	21	156	100

According to the model, the first group, representing 38.3% of the population, who stay on average 2.2 days, occupy 14.1% of the beds. The second group, representing 39.4% of the population who stay on average 5.4 days occupy 31.4% of the beds. The third group, representing 18.8% of the population staying on average 13.4 days occupy 35.3% of the beds. Lastly, the fourth group representing 3.5% of the population staying on average 39.2 days occupy 19.2% of the beds.

Discussion

In this paper, we described how the GMM can be seen as representative of the way in which patients flow through a given hospital or health care facility. In addition, based on this concept of patient flow, we described how we can estimate the average occupancy to determine the bed requirement for a given population. The calculations involved in determining the bed requirement were obtained using an Excel spreadsheet package.

Such a model can also be used to perform a what-if type of analysis to assess the impact of different strategies on the bed requirement and answer questions related to improving bed use. For instance, what would be the impact on the bed requirement if we reduced the mean LOS of the longer stay surgical patients by ten days?

Interestingly, we found that the third group use the largest proportion of beds, which was surprising since they represent only 18.8% of the population.

References

1. **Abbi R, El-Darzi E, Vasilakis C. & Millard P.** (2007a), Length of stay based grouping and classification methodology for modelling patient flow. Journal of Operations and Logistics (Submitted).
2. **Abbi R, El-Darzi E, Vasilakis C & Millard P.** (2008), Length of stay based grouping and classification methodology for modelling patient flow. Journal of Operations and Logistics (To be published).
3. **Abbi R, El-Darzi E, Vasilakis C & Millard P.** (2007b), Intelligent methods for modelling patient flow. The International Conference on Industrial Engineering and Systems Management (IESM 2007). Beijing CHINA.
4. **Akaike H.** (1973), Information theory and an extension of the maximum likelihood principle. Petrov BN, Csaki F (eds). Second International Symposium on Information Theory. Akademia Kiado, Budapest.
5. **Bishop C.M.** (2006), Pattern Recognition and Machine Learning, New York, USA, Springer.
6. **Dempster A.P, Laird N.M & Rubin D.B.** (1977), Maximum likelihood from incomplete data via the EM algorithm. Journal of the Royal Statistical Society. Series B (Methodological), 39, 1-38.
7. **Faddy M.J & Mcclean S.I.** (1999), Analysing data on length of stay of hospital patients using phase-type distributions. Applied Stochastic Models in Business and Industry, 15, 311-317.
8. **Harrison G.W.** (1994), Compartmental models of hospital patient occupancy patterns. In: Millard P.H, McClean S.I (eds). Modelling hospital resource use: a different approach to the planning and control of health care systems. Royal Society of Medicine: London, 53-61.
9. **Harrison G.W & Millard P.H.** (1991), Balancing acute and long term care: the mathematics of throughput in departments of geriatric medicine. Methods of Information in Medicine, 30, 221-228.
10. **Macqueen J. B.** (1967), Some Methods for classification and Analysis of Multivariate Observations. Proceedings of 5-th Berkeley Symposium on Mathematical Statistics and Probability. Berkeley, University of California Press.
11. **Mclachlan G.J & Peel D.** (2000), Finite Mixture Models, New York, John Wiley & Sons.
12. **Millard P.H.** (1988), Geriatric medicine: a new method of measuring bed usage and a theory for planning. St. George's Hospital Medical School. University of London.

13. **Millard P.H.** (1992), Flow rate modelling: a method of comparing performance in departments of geriatric medicine. St George's Hospital Medical School. University of London.
14. **Miller G.A.** (1956), The Magical Number Seven, Plus or Minus Two. The Psychological Review, 63, 81-97.
15. **Rissanen J.** (1978), Modelling by the shortest data description. Automatica, 14, 465-471.
16. **Schwarz G.** (1978), Estimating the Dimension of a Model. The Annals of Statistics, 6, 461-464.
17. **Titterington D.M, Smith A.F.M & Makov U.E.** (1985), Statistical Analysis of Finite Mixture Distributions, New York, John Wiley & Sons.
18. **Walter M.** (2002), Automatic Model Acquisition and Recognition of Human Gestures. Harrow School of Computer Science. London, University of Westminster.
19. **Walter M, Psarrou A & Gong S.** (2001), Data Driven Model Acquisition using Minimum Description Length. British Machine Vision Conference. Manchester, UK.