Implications of mixed exponential occupancy distribution and patient flow models for health care planning

Gary W. Harrison

Department of Mathematics, University of Charleston, S.C. Charleston, SC 29456, USA
E-mail: harrsong@cofc.edu

There is considerable evidence that the distribution of the length of time that a patient occupies a bed in a hospital department is best described by a sum of two or three exponential terms, because of the presence of acute care, rehabilitation, and possibly long term care patients in the department. The patient flow models implied by these mixed exponential distributions are presented and fitting them to observed data when the admission rate fluctuates is discussed. Unlike single exponential distributions, mixed exponential distributions imply that the average length of stay of patients currently resident in the department is much longer than the average length of stay of a group of patients discharged over a period of time, so that the latter way of measuring will not correctly indicate what portion of the resources are being used by rehabilitation and long term care patients. Also, the expected additional length of stay increases dramatically with the time already spent in the department. Applications to predicting the effects of policy changes and to long term monitoring of hospital departments are presented. Two American hospitals are analyzed. The occupancy times in the government supported hospital follow a mixed exponential distribution similar to those found in the United Kingdom, but in the private hospital they fit a single exponential distribution, indicating markedly different management practices.

Keywords: occupancy distribution, mixed exponential distribution, flow models, conditional expected length of stay

1. Introduction

Occupancy times are an important measure of the functioning of any hospital department. It clearly matters to the patient how many days he or she has been in the hospital. It matters to the medical staff because the type of treatment shifts from acute care to rehabilitation to long term care as the stay lengths. And it matters to the hospital administrators because the length of stay is a major factor in determining the costs and the number of patients that can be treated in a given facility. In 1989 Millard [1] and Millard et al. [2] observed that occupancy times for geriatric departments could not be adequately described by a single number such as the mean, but that the observed distributions matched a mixed exponential distribution; the number of patients with length of occupancy greater than or equal to \( x \) equals \( Ae^{-Bx} + Ce^{-Dx} \). In 1991 Harrison and Millard [3] presented a model that explains this data based on the concept that patients “flow” into the department to receive acute care and from there to rehabilitation or long term care. Since then it has been found that mixed exponential distributions of occupancy times occur in many other types of departments and services and even entire hospitals (e.g., [4–6]). Often, however, three exponential terms are needed to give a good fit to the data, leading Harrison [7] to extend the original model to three compartments, which can be roughly identified with acute care, rehabilitation, and long term care [8]. Irvine [9] and Taylor et al. [10,11] have built on this framework to form more sophisticated stochastic models with a community component, which give similar results for the expected values but also allow computation of variances.

In spite of this growing body of research and evidence that mixed exponential distributions for length of occupancy occur in many and perhaps most hospital departments, neither mixed exponential distributions nor the models they imply have been widely used by hospital managers, probably because the implications of these occupancy patterns have not been thoroughly discussed. This paper explores some of these implications, hoping that managers and care givers can better understand how knowing the distribution of occupancy time can improve their decision making.

Section 2 first discusses two patient flow models that could produce mixed exponential distributions, their relationship and their differences. Section 3 discusses how to find the parameters (discharge and conversion rates) that make the models best fit the observed distribution in a particular department. It answers what has so far been the most common practical difficulty, how to estimate these parameters when the admission rate is not constant. Section 4 shows that the distribution of lengths of stay of all patients discharged in a period of time is different from the distribution of lengths of stay of all patients currently in the department. Section 5 computes the probabilities and expected value of the additional length of stay, both conditioned on the time that the patient has already been in the department and for all current occupants, showing how dramatically the expected additional stay increases with current occupancy time. The results of both of these sections do not depend on the model of patient flow used, only on the mixed exponential nature of the distribution. Section 6 illustrates how the models can be used to analyze the effects of changes in policy or practice, by examining the effect of a 20% reduction in the number of patients that are not rehabilitated and require long term
care. Section 6 discusses how long term monitoring of the occupancy distribution can reveal changes in how resources are being used or the effectiveness of rehabilitation. Section 7 examines the occupancy distribution of two American hospitals. The government funded hospital is similar to those found in the United Kingdom but the private hospital is markedly different.

2. Observations and models

The occupancy time \( s \) of a patient in a hospital department is the number of days the patient has occupied a bed in that department since admission. As discussed in the introduction, occupancy times in many hospital departments follow a mixed exponential distribution given by

\[
N(s \geq x) = Ae^{-Bx} + Ce^{-Dx} + Ee^{-Fx},
\]

where \( x \) is the time in days, \( N(s \geq x) \) is the total number of current patients that have been in the department for greater than or equal to \( x \) days, and \( A, B, C, D, E, \) and \( F \) are constants. Frequently only two exponential terms are needed to fit the data, in which case \( E = 0 \), and occasionally a single exponential gives an adequate fit. Two or three exponential terms are needed in most European medical systems observed so far, because of the presence of rehabilitative and long term care patients.

There are two possible models to explain these observations, as shown in figure 1. Both divide the patients into three groups with numbers \( N_1, N_2, \) and \( N_3 \), each group having different fractional discharge rates (fraction of patients discharged per day). In the cascading flow model, first presented by Harrison and Millard [3] and extended to three compartments in Harrison [7], a cohort of patients of size \( A_0 \) arriving each day contains three types of patients; a fraction \( w_1 \) of these patients will require acute care, a fraction \( w_2 \) of them will require rehabilitation, and a fraction \( w_3 \) of them will require long term care. This model is described by simpler equations

\[
\begin{align*}
N_1(s + 1) &= N_1(s) - (v_1 + r_1)N_1(s), \\
N_2(s + 1) &= N_2(s) + v_1N_1(s) - (v_2 + r_2)N_2(s), \\
N_3(s + 1) &= N_3(s) + v_2N_2(s) - r_3N_3(s)
\end{align*}
\]

with initial conditions \( N_1(0) = A_0, N_2(0) = 0, \) and \( N_3(0) = 0 \).

In the Separate Flow Model the cohort of patients of size \( A_0 \) arriving each day contains three types of patients; a fraction \( w_1 \) of these patients will require acute care, a fraction \( w_2 \) of them will require rehabilitation, and a fraction \( w_3 \) of them will require long term care. This model is described by simpler equations

\[
\begin{align*}
N_1(s + 1) &= N_1(s) - b_1N_1(s), \\
N_2(s + 1) &= N_2(s) - b_2N_2(s), \\
N_3(s + 1) &= N_3(s) - b_3N_3(s)
\end{align*}
\]

with initial conditions \( N_1(0) = w_1A_0, N_2(0) = w_2A_0, \) and \( N_3(0) = w_3A_0 \).

Which model to use cannot be decided on the basis of mathematics alone, since both models produce a mixed exponential distribution of occupancy times (equations (4) and (8) below), but must be decided on the basis of which model best represents the reality of patient care. The separate flow model assumes that the mix of patients in the three groups is predetermined by the types of patients entering the department. The medical staff cannot control this mix; it only reacts with the appropriate treatment for each patient. In contrast the cascading flow model assumes that patients are converted from one group to the next. It may not be a physical move but does represent a shift in the type of care given the patient. This shift in care is made in response to the patient’s needs, but the model implies that the rates at which patients are converted from one group to another is at least partially controlled by the medical staff and that they try to cure and/or rehabilitate all patients before consigning them to long term care. Thus the author feels that the cascading flow model more accurately represents current medical practice.

The results in sections 4 and 5 are the same for both models, although the interpretation of the parameters may be different, but simulating the effects of changes in departmental practice in section 6 does depend heavily on which model of current practice is used.

The use of discrete time in these models instead of the continuous time implied by the exponential terms in equation (1) is mostly a matter personal choice. It is used here because hospital records are kept on a daily basis and the equations are easier to interpret.