

Hospital Resource Consumption Modelling

Marco Papi (Corresponding author)

School of Engineering, Campus Bio-Medico University

Via Alvaro del Portillo, 21 - 00128 Rome (Italy)

E-mail: m.papi@unicampus.it

Luca Pontecorvi

School of Engineering, Campus Bio-Medico University

Via Alvaro del Portillo, 21 - 00128 Rome (Italy)

E-mail: l.pontecorvi@unicampus.it

Received: December 3, 2013 Accepted: December 25, 2013 Published: February 7, 2014

doi:10.5296/rbm.v1i1.5081

URL: <http://dx.doi.org/10.5296/rbm.v1i1.5081>

Abstract

Health service researchers frequently study length of hospital stay (LoS) as a health outcome. LoS is a valid proxy to estimate the consumption of hospital resources. Average LoS, however, albeit easy to quantify and calculate, can be misleading if the underlying distribution is not symmetric. Generally originating from heavily skewed distributions, LoS data can be difficult to model with a single parametric model. Mixture models can be quite effective in dealing with such data. In this paper we proposed a generalization of the phase-type distribution in order to capture all the statistical features observed in real situations. We also propose an analysis of the discharge rate and the admission rate. Those proxies provide information on the efficiency of departments and give operational guidelines for medical staff.

Keywords: Length of Stay, Bed Planning, Discharge Rate, Single Factor Risk Model

1. Introduction

The cost of healthcare is increasing, this common phenomenon in industrialized countries could be explained by the evolution of medical knowledge and biomedical technology, the escalating proportions of elderly people and the different perceptions of health. In 2011 health expenditures in Italy neared 114 billion euro with an increase of 17% from 2001 and it covered 7.2% of the GDP (gross domestic product).

The provision of healthcare services is one of the largest and most complex industries worldwide. As one of the essential necessities to sustain life, it faces the consequences of increasing demand in times of limited financial resources and competing social needs. Providing the appropriate medical care involves decision-making in terms of planning and management of healthcare resources. Therefore is important to have a decision support system for health institution that provides a quantitative estimation of operational management information.

Mathematically speaking, a hospital corresponds to a complex stochastic system so that the common deterministic approach for planning and managing the system can be expected to be inadequate Millard (1994), Shahani (1991). All of these features point towards a need for sophisticated hospital capacity models.

There is considerable scope for Operational Research models to be widely used for that purpose. Indeed, the need to create and utilize models for a wide range of scenarios in hospital resource modelling has spawned a vast array of different approaches. Within bed modelling alone, a number of operational approaches have been utilized, including queueing models Huang (1995), Kao (1991), integer programming Ruth (1981), forecasting Lin (1989) and simulation Dumas (1984), Gove (1995). A similar array of operational methods have been used in operating theatre modelling Gordon (1998), Wright (1996) and workforce planning Wolfe (1963), Dowsland (2000).

Extensive research has been conducted to identify the characteristics that are associated with hospital length of stay (LoS). LoS represents a reliable and valid proxy for measuring the consumption of hospital resources Fackrell (2008), Faddy (1999), Marshall (2003), Millard (1994). Average LoS, however can be misleading if the underlying distribution is not symmetric. Therefore, models based on the average LoS, cannot describe the underlying distribution of patients Costa (2003), Harper (2002). In fact, there are numerical and practical reasons that using the average LoS is inappropriate for use in the development of models, most notably that the typical length of stay distribution is highly skewed distribution and not well summarized by its mean value.

Moreover, from the point of view of individual elderly people, the identification of risk factors that are associated with transfer, early death and long-term survival is of considerable importance Bebbington (2001), Dale (2001), Rothera (2002). However, for planning purposes, care managers and budget holders need to know the overall pattern of LoS in long-term care.

Other models in the literature complement the existing research in providing a full picture of the overall behavior of LoS in residential and nursing home care. Methods that explicitly

model the LoS in care of elderly people have been proposed in Harrison (1991), McClean (1993), Taylor (1998) and Taylor (2000).

Compartmental models offer an alternative approach. A compartmental model describes the flow of something, such as patients, through a system, where the system is comprised of a finite number of homogeneous subsystems known as compartments; Godfrey (1983) is a standard text on these models. In Harrison (1991) compartmental models are introduced in order to describe patient flow: a two-compartment discrete-time deterministic model has been developed in order to describe such movements of patients. An extension of this model to its continuous-time stochastic analogue can be found in Taylor (1996-1998) and in Taylor (2000). Such compartmental models may all be regarded as phase-type distributions which describe the time to absorption of a finite Markov chain in continuous time, when there is a single absorbing state and the stochastic process starts in a transient state.

Compartmental models have been shown to provide a valid description of patient movements through the hospital system. Such models may be used to plan health and social care for an ageing population where compartments may correspond to short (acute), medium (rehabilitative) or long stay care.

Given the wealth of work that has already been done in this area, it is both surprising and disappointing that it has not found greater application in the Italian healthcare system. The existing literature highlights many concerns regarding the adopted methodologies and stated assumptions of various proposed models that necessarily require a deep knowledge of healthcare institutions where such support systems have to be implemented.

In this paper is to demonstrate how standard density function cannot describe the underlying distribution and is proposed the use of a generalization of the phase-type distribution in order to capture all the statistical features observed in real situations. The generalization proposed uses few number of parameters and supplies a phenomenological interpretation of data. Furthermore to afford an operational guideline for planning and management of hospital beds we propose a method to evaluate the forecasting ability.

Moreover we propose an analysis of the discharge rate (DR) and the admission rate (AR). Those proxies provide information on the efficiency of departments and give operational guidelines for medical staff. In fact the distribution of discharge rate supplies the probability that a given number of patients is discharged in one day; likewise the admission rate provides the probability that a given number of patients is admitted in one day.

In this paper we present models that could be used to estimate those probabilities. Our case of study is based on a data set originated from the Tor Vergata Cardiology department.

In section 2 are reported a short description of Tor Vergata hospital and some related statistics are summarized. Section 3 is dedicated to the description of the proposed models. The paper then reports, in section 4, the methods used to fit the data and, in Section 5, the main results obtained. Finally, in Section 6 some general conclusions are drawn regarding the practical implications of these methods.

2. Data

Tor Vergata Hospital is a big size structure within the context of Italian health care system. Moreover this hospital offers a high quality service compared to the national health system. This work focuses on the flow of patients within the department of Cardiology. All the data are related to a period of 5 years going from 1st January 2008 to 16th May 2013. There have been considered 11510 admissions. The basic descriptive statistics of the patients LoS are reported in Table 1.

Table 1. Descriptive statistics of Cardiology department LoS

	Mean	Variance	Skewness	Kurtosis	N
LoS	3.33	4.86	3.46	24.40	11510

The values show that the LoS density function is asymmetric, it can only take positive values, and it exhibits a leptokurtic form (execute acute peak around the mean and have fatter tails). The long gradual tail to the right of the distribution is given by a very small number of patients that stay in hospital for a considerable amount of time.

We remark the importance of longer stay patients, in fact although the majority of patients is discharged after a short period, beds are mostly occupied by patients that stay for a longer period.

The discharge rate is the probability that in one day are discharged n patients; it is estimated from real data as the number of day in which are discharge n patients divided by the total number of considered days. Likewise the admission rate is defined as the number of days in which are admitted n patients divided by the total number of considered days. In Table 2 are reported the main characteristics of the admission rate and the discharge rate.

Table 2. Cardiology department statistics of admission and discharge rate

	Mean	Variance	Skewness	Kurtosis	Max
Admission	5.83	6.62	0.06	2.59	14
Discharge	5.83	6.96	0.17	2.93	16

3. Models

3.1 LoS Model

In this work to describe the empirical distribution of LoS it is used a generalization of *phase-type distribution (PH-distribution)*. PH distribution is a very versatile class of distributions. Introduced by Neuts in Neuts (1975), phase-type distributions are defined via a

continuous time homogeneous Markov chain $X(t)$ on a finite state space $\{0, 1, \dots, n\}$, $n \geq 0$, see [10]. The state 0 is absorbing and Markov chain is assumed to be irreducible. Then the random variable representing the first hitting time of 0 is

$$\tau = \inf\{t \geq 0: X(t) = 0\}$$

which is finite almost everywhere, and its distribution is called a phase-type distribution. Let the initial state probability distribution be $\alpha \in \mathbb{R}^n$ such that

$$\sum_{i=1}^n \alpha_i = 1, \quad \alpha_i \geq 0 \quad \forall i = 1, 2, \dots, n.$$

Let $T = [\lambda_{ij}]$ for $i, j = 1, 2, \dots, n$ be a $n \times n$ matrix such that for $i \neq j$, $\lambda_{ij} \geq 0$, satisfying

$$\lambda_{ii} \leq - \sum_{j=0, j \neq i}^n \lambda_{ij}.$$

The PH distribution is said to have a representation (α, T) of order n if the probability density function of τ is as follows Fackrell (2008), Ross (1995):

$$ph(t) = -\alpha e^{Tt} T e, \quad t > 0 \quad (1)$$

Where e is a $n \times 1$ vector of ones.

In this paper we shall consider an extension of (1). Given the density function $\gamma(t; k, \theta)$ of a Gamma distribution with parameters k, θ and a diagonal matrix $T = -diag(\lambda_1, \lambda_2, \dots, \lambda_n)$, with $\lambda_i \geq 0$ for $i = 1, 2, \dots, n$, we define

$$f(t) = \sum_{i=1}^n \alpha_i \lambda_i \gamma(t \lambda_i; k, \theta), \quad \text{for } t > 0, \quad (2)$$

that is

$$f(t) = \frac{\theta^k}{\Gamma(k)} \sum_{i=1}^n \alpha_i \lambda_i^k t^{k-1} e^{-\theta \lambda_i t}.$$

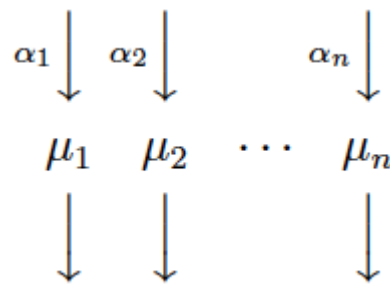
In the rest of the paper we refer to (2) as the density function of *phase-type Gamma* distribution (PHG).

Remark We observe that equation (2) may be generalized by using the following:

$$h(t) = -\alpha g(Tt) T e.$$

Which is a generalized version of (1) see the Appendix.

The density function (2) provides a useful tool to describe LoS data; the standard PH-distribution can not describe leptokurtic densities; furthermore the use of the PHGamma



distribution allows to reach the peak and fit the fat tail of the data. Moreover the PHGamma can be described using the following state transition diagram with n states in parallel: where the average time spent μ_i in each state is evaluated by:

$$\mu_i = \frac{1}{\lambda_i} \frac{k}{\theta}$$

3.2 AR/DR Model

To evaluate the discharge and the admission distributions it is used an approach based on *single factor model* Jang (2002). In general this model is used in financial mathematics to estimate credit risk of a portfolio asset Palombini (2009). The basic idea behind this model consists into creating an analogy between the default of a financial institution and discharge of a patient. The aim is to come up with a simple and closed-form formula for the distribution function of the discharge rate. Deriving a closed form solution requires making a set of simplifying assumptions. We will progressively introduce these assumptions and their implications for the model.

To derive the DR and the AR, knowing the individual probabilities of the patients is not enough; we also need to know their correlation structure.

We assume that each patient status is describe by a random variable. Therefore we introduce the real value random variable X_1, \dots, X_N , where N is the number of inpatients defined on the same probability space $(\Omega, \mathcal{F}, \mathbb{P})$, which drive the patients discharge/admission probabilities. We also assume that the correlation coefficient of each pair of random variables X_n and X_m is $\rho_{n,m}$. The correlation coefficient $\rho_{n,m}$ between each pair of random variables

X_n and X_m is the same for any two patients:

$$\text{corr}(X_n, X_m) = \rho_{n,m} = \rho.$$

There exists a source of uncertainty affecting all patients. Moreover, the random variable X_1, \dots, X_N are described by (3)

$$X_n = \sqrt{\rho}Y + \sqrt{1 - \rho}\varepsilon_n, \tag{4}$$

for all $n = 1, \dots, N$ where $Y, \varepsilon_1, \dots, \varepsilon_N$ are independent random variables on $(\Omega, \mathcal{F}, \mathbb{P})$. We can interpret (4) as follows: each random variable X_n represents the disease state of patient, whose realization determines if the patient n is discharged. In fact, if the disease state is less

than a threshold value K the patient is discharged. X_n can be expressed as the sum of two factors: one common or systematic factor Y and an idiosyncratic factor ε_n that are i.i.d. and defined on the same probability space. Y affects all patients in the same way and has density function f with mean $\mathbb{E}[Y] = 0$ and variance $\mathbb{V}[Y] = \sigma^2$, the idiosyncratic factors are independent across patients and have a common density function g such $\mathbb{E}[\varepsilon_n] = 0$ and $\mathbb{V}[\varepsilon_n] = \sigma^2$. The covariance between X_n and X_m is

$$\begin{aligned} cov(X_n, X_m) &= \mathbb{E}[X_n X_m] \\ &= \mathbb{E}[(\sqrt{\rho}Y + \sqrt{1-\rho}\varepsilon_n)(\sqrt{\rho}Y + \sqrt{1-\rho}\varepsilon_m)] \\ &= \rho\sigma^2 + (1-\rho)v^2\delta_{n,m} \end{aligned}$$

where $\delta_{n,m}$ is the Kronecker delta.

Conditioning on the factor Y , the discharge probability of each patient, denoted by $p(Y)$, is easily computable:

$$\begin{aligned} p(Y) &= \mathbb{P}[X_n < K|Y] \\ &= \mathbb{P}[\sqrt{\rho}Y + \sqrt{1-\rho}\varepsilon_n < K|Y] \\ &= \mathbb{P}\left[\varepsilon_n < \frac{K - \sqrt{\rho}Y}{\sqrt{1-\rho}} \mid Y\right] \\ &= G\left(\frac{K - \sqrt{\rho}Y}{\sqrt{1-\rho}}\right) \end{aligned}$$

where G is the cumulative density function related to g . Hence the random variables X_1, \dots, X_N are conditionally independent given the random systematic factor Y .

Consider, for each patient n , the random variable L_n takes value 0 if the patient n has not discharged from the department under consideration and is 1 otherwise. Then define $L(N)$ as the number of discharged patients:

$$L(N) = \sum_{n=1}^N \mathbf{1}_{X_n < K_n}$$

In general N cannot be considered as deterministic. Actually, the number of patient hospitalized in each department is modelled as a discrete random variable N_p with probability mass function h . We also assume that N_p is defined on probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and is independent of ε_n and Y .

In our empirical study, the maximum number of patients that can be received in the department N_{max} is estimated from historical data and is reported in Table 2. Therefore the density function of $L(N_p)$ is formulated in light of the fact that $\mathbb{P}(L = m|N_p = n, Y = y)$ is binomial distributed with parameters n and $p(Y)$. The probability that $L=m$ is given by

$$\begin{aligned}
 \mathbb{P}(L = m) &= \sum_{n=m}^{N_{max}} h(n) \mathbb{P}(L = m | N_p = n) \\
 &= \sum_{n=m}^{N_{max}} h(n) \int_{-\infty}^{+\infty} \mathbb{P}(L = m | N_p = n, Y = y) f(y) dy \\
 &= \sum_{n=m}^{N_{max}} h(n) \int_{-\infty}^{+\infty} \binom{n}{m} p(y)^m [1 - p(y)]^{n-m} f(y) dy
 \end{aligned}$$

In a future work in preparation we will consider the dependence between the efficiency of the department and the number of patients.

4. Optimization Method

According to Asmussen (1996), Faddy (1999) and Fang (1994) maximum log-likelihood estimation is used in order to fit the parameters of LoS distributions. The patients number N is reported in Table 2 The log-likelihood objective function is given by

$$\text{Log}L(\Theta) = \operatorname{argmax}_{\Theta \in \mathcal{D}} \sum_{i=1}^N \log(f(t_i; \Theta)),$$

where $f(\cdot, \Theta)$ is the density function of the considered model depending on the set of parameters described by a vector Θ .

The optimization routine is implemented in the MATLAB involvement using existing algorithm based on the interior-point method. Parameters calibration starts from a Halton quasi-random sequence of $10 \times n$ initial points Fang (1994), Niederreiter (1992). In order to choose the optimal PHGamma, a sequential procedure is adopted whereby increasing numbers of phases n are tried starting with $n = 1$ until little improvements in the fit of the data can be obtained by adding a new phase. The number of the phases that allows the best compromise between model complexity and goodness of the fit is chosen. In Table 3 are reported the values of the objective function related to the PHGamma model. The analysis of the data shows that the use of a model with more than 3 states a negligible in the performance. Hence we fix the number of the phases to 3.

Table 3. Log-likelihood function values obtained in LoS density estimation using a n -states PHGamma model

	$n=1$	$n=2$	$n=3$	$n=4$	$n=5$	$n=6$	$n=7$
$\text{Log}L(\Theta)$	-1.9038	-1.8026	-1.7698	-1.7605	-1.6528	-1.5607	-1.5210

5. Estimation Results

5.1 LoS Model

The PHGamma distribution provides a very versatile tool to describe a wide class of density functions. Using a few parameters is possible to describe distributions that exhibit different characteristics. The performance of the model are compared with different standard density.

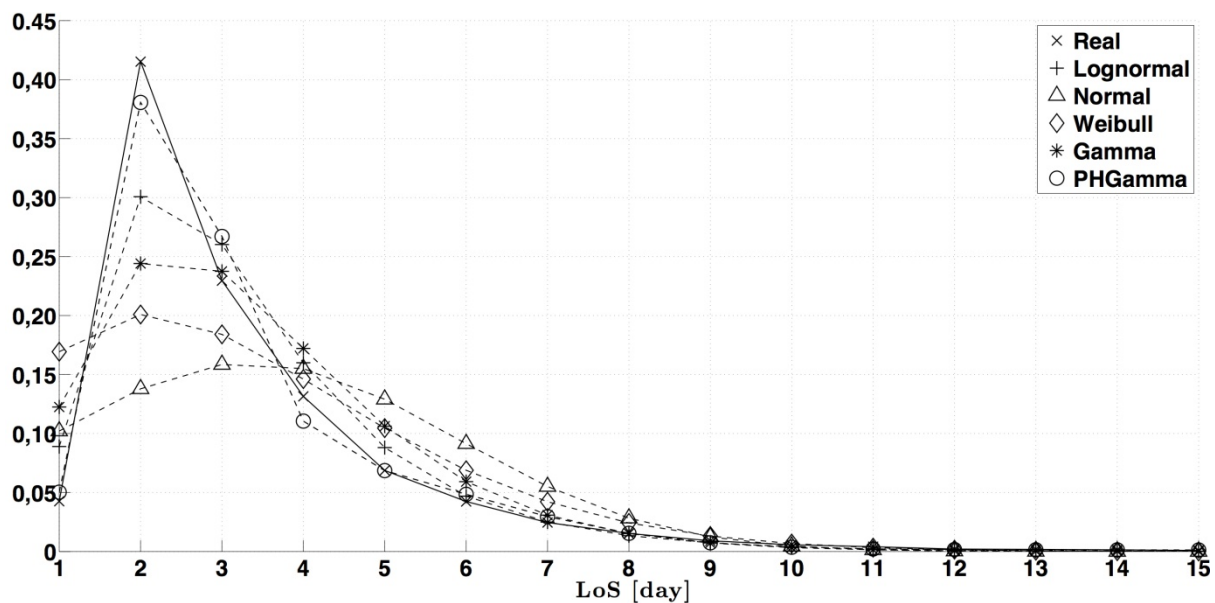


Figure 1. Probability density function

In Figure 1 Normal, Lognormal, Weibull, Gamma and 3-states PHGamma is estimated. The results show that the proposed model provides a better fit of the data. In Table 4 are reported the parameters estimated. In order to take measure of the goodness of the fit in Table 5 are reported the log-likelihood function for each model. 3-states PHGamma model performs the biggest log-likelihood value.

Table 4. 3-states PHGamma estimated parameters

α_1	α_2	α_3	λ_1	λ_2	λ_3	k	θ
0.0161	0.2538	0.7302	0.9794	2.8180	6.1106	9.5710	0.6430

Table 5. Log-likelihood function value of LoS estimation for each model

	Lognormal	Normal	Weibull	Gamma	PHGamma
$\text{Log}L(\Theta)$	-1.8082	-2.3131	-2.0131	-1.9038	-1.7698

Table 6. Summary of four discrepancy measures

	Lognormal	Normal	Weibull	Gamma	PHGamma
AAE	0.0050	0.1286	0.0385	0.0077	0.0039
APE	1.4761	10.3361	5.6001	2.4620	0.4702
ARPE	0.0224	0.1566	0.0849	0.0373	0.0071
RMSE	0.0120	0.1325	0.0441	0.0196	0.0051

To evaluate the performance of those models, we also use four discrepancy measures: the average prediction error (APE), the average absolute error (AAE), the root mean-square error (RMSE), and the average relative prediction error (ARPE). Let f be the density function and t_k for $k = 1, 2, \dots, K$ the evaluation time. Then the four error estimators are defined as follows:

$$AAE = \sum_{k=1}^K \frac{|f(t_k) - \hat{f}(t_k)|}{K}, \quad APE = \sum_{k=1}^K \frac{|f(t_k) - \hat{f}(t_k)|}{f(t_k)}, \quad ARPE = \frac{1}{K} \sum_{k=1}^K \frac{|f(t_k) - \hat{f}(t_k)|}{f(t_k)},$$

$RMSE = \sqrt{\sum_{k=1}^K \frac{[f(t_k) - \hat{f}(t_k)]^2}{K}}$. These measures are shown in Table . In particular we observe that the PHGamma distribution achieves the smallest errors for all these measures.

5.1.1 Forecast Ability

In order to provide an operational guideline in bed planning, it is important to estimate the probability that a patient is discharged before a given time. To validate the performance of the model and its stability over time is used the following procedure: we consider a calibration data set consisting of 1000 patients ordered by time to estimate the parameters of the model. Then a new sample is taken where the oldest 100 elements are replaced by new data. On this updated sample the empirical distribution is evaluated and is compared with the density previously estimated using the PHGamma model. The updated data set becomes the calibration data set and the procedure is repeated hundred times using as the starting point for the calibration the previously estimated parameters (for the first step of this procedure a Halton quasi random sequence of $10 \times n$ initial points is used). There have been considered more than 11000 different admissions from April 3rd 2008 to April 30th 2013.

To evaluate the stability of the model it is considered the time analysis of the parameters.

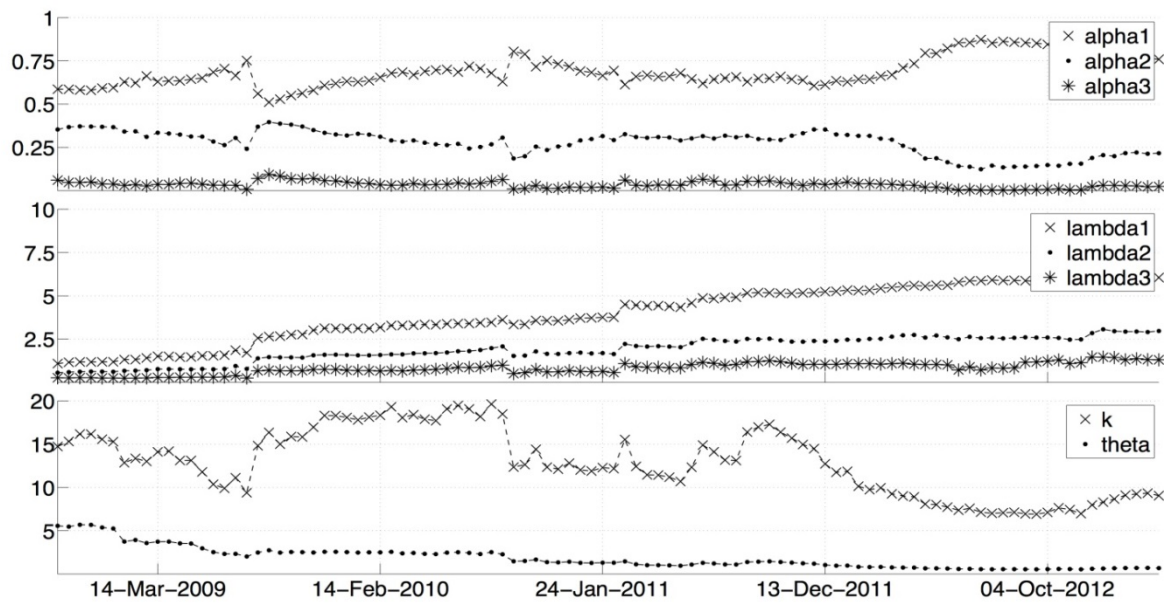


Figure 2. Parameters time trends

In Figure 2 is shown the time trend of estimated parameters that exhibit a regular pattern with a limited variation.

In order to evaluate the model performances in each scenario has been considered the following error estimator

$$\varepsilon = \frac{\sum_k t_k [f_r(t_k) - f_e(t_k)]}{\sum_k t_k f_r(t_k)}$$

where f_r is the real distribution and f_e is the estimated distribution.

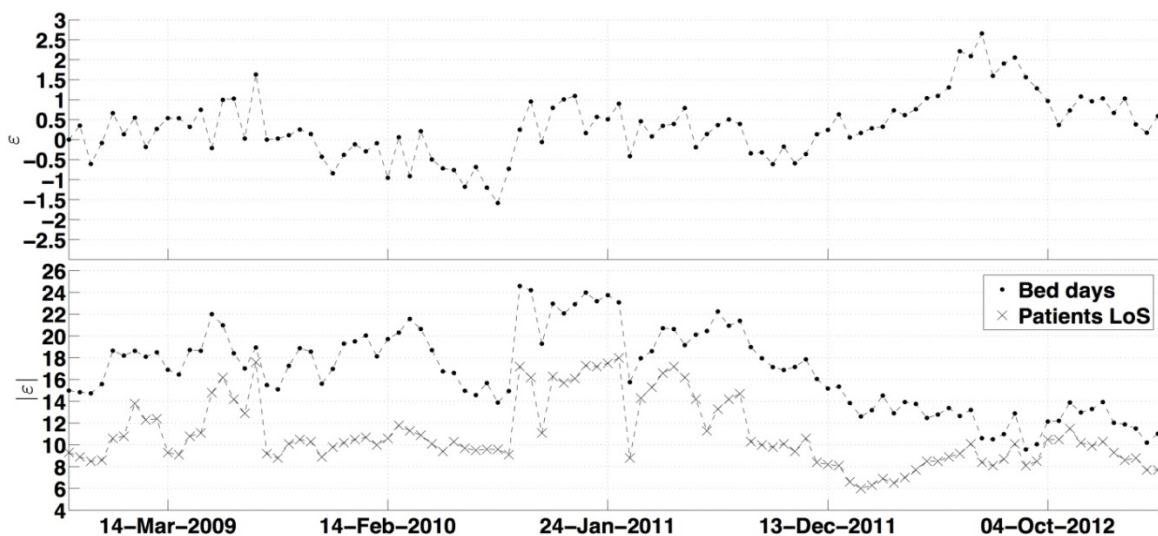


Figure 3. Error time trends

In the first plot of Figure 3 is reported the time trend of the error. Each scenario has 1000 patients and covers an average period of 17 days with a use of 3500 bed days in average. The error value, that is bounded in the range $\pm 3\%$, shows that the model has an average bed days error estimation lower than 15. In the second plot is reported the absolute error for the bed days and patients. The bed days error formula is

$$|\varepsilon_r| = \frac{\sum_k t_k |f_r(t_k) - f_r(t_k)|}{\sum_k t_k f_r(t_k)}$$

and patient LoS formula is

$$|\varepsilon_l| = \frac{\sum_k |f_r(t_k) - f_r(t_k)|}{\sum_k t_k f_r(t_k)}$$

These indicators evaluate the distance between the distributions and show underestimation and overestimation of bed days consumption and patient LoS in each scenario. In particular, the plot shows that an average error of 17% in the bed days evaluation consumption is made, and an average error of 10% for the patients LoS realized. The absolute error allows to highlight errors of overestimation and underestimation in contrast of using the error based on the simple difference that takes into account the cumulative consumption of resources.

5.2 AR/DR Model

An important proxies to evaluate the efficiency of a department are the admission rate and the discharge rate. Furthermore using the single factor model is provided a phenomenological interpretation. In our empirical study we shall consider the following models for the distribution densities:

$$\begin{cases} f(t) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{t^2}{2\sigma^2}} \\ g(t) = \frac{1}{v\sqrt{2\pi}} e^{-\frac{t^2}{2v^2}} \\ h(t) = \frac{z(t|\mu, \xi)}{\xi} e^{-z(t|\mu, \xi)}, \quad z(t|\mu, \xi) = e^{\frac{t-\mu}{\xi}} \end{cases}$$

Here the density h is selected according with a separated estimation procedure where we observe that the number of patients in the department is well described by the extreme value distribution, see Figure 4.

The estimated parameters of the model are reported in Table 7.

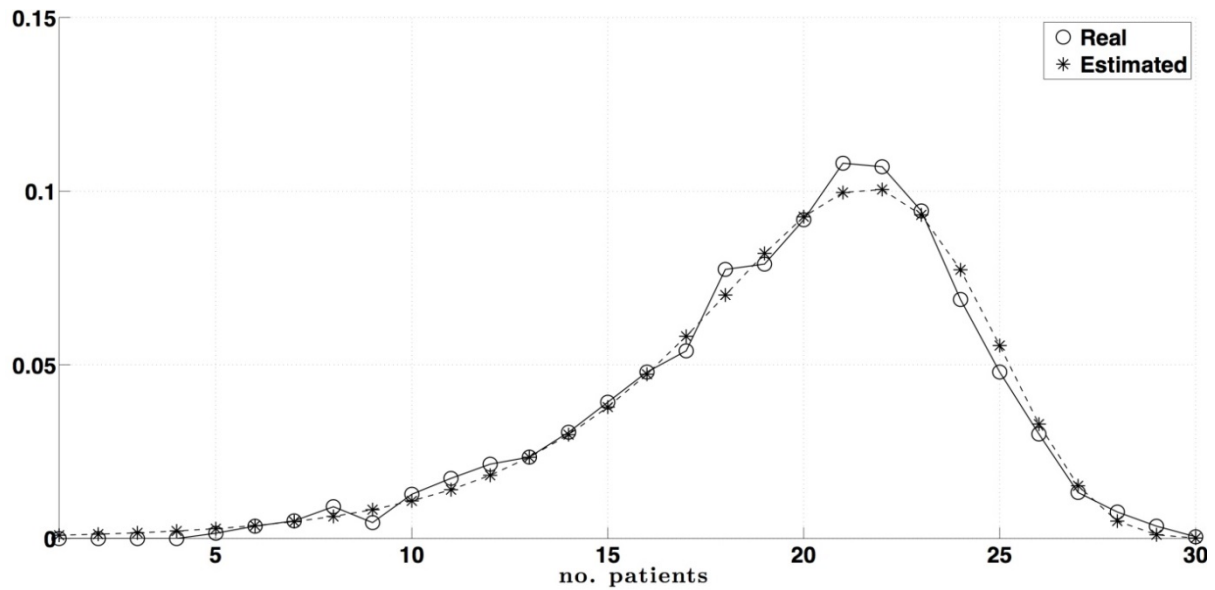


Figure 4 Number of patients distribution

Table 7. Estimated parameters for admission rate and discharge rate

	K	ρ	ν	μ	ξ	σ
<i>AR</i>	1.0813	0.8233	22.3929	11.6209	1.3980	4.5974
<i>DR</i>	-0.0386	0.9861	29.7045	12.9137	1.9190	1.1737

In Figures 5 and 6 are reported the probability functions estimated. The proposed model provide a good description of the empirical distribution.

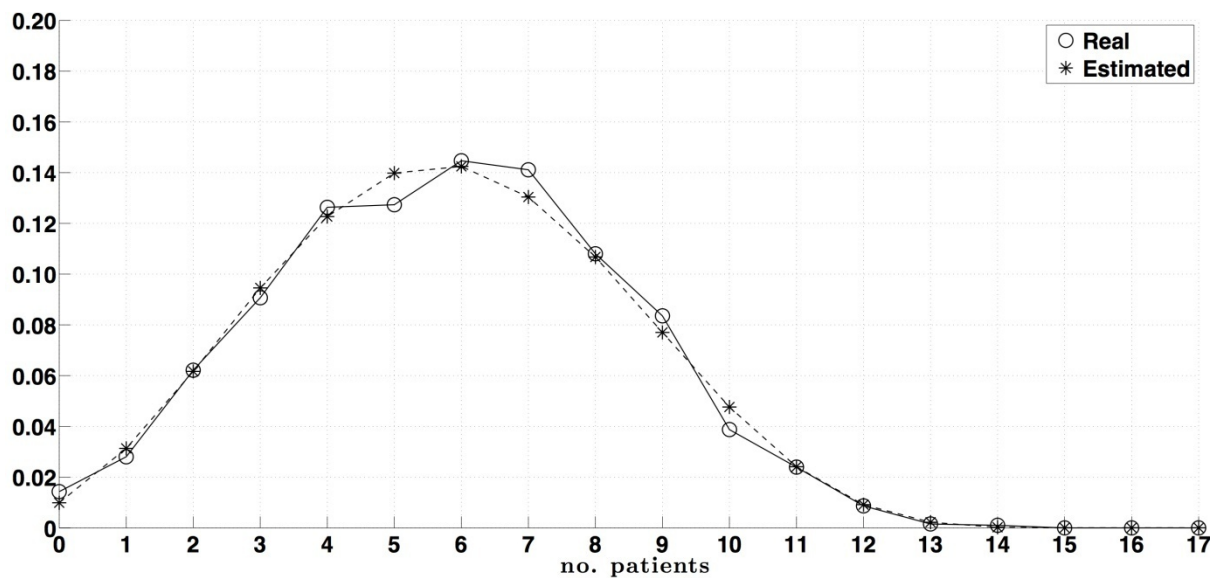


Figure 5 Admission rate distribution

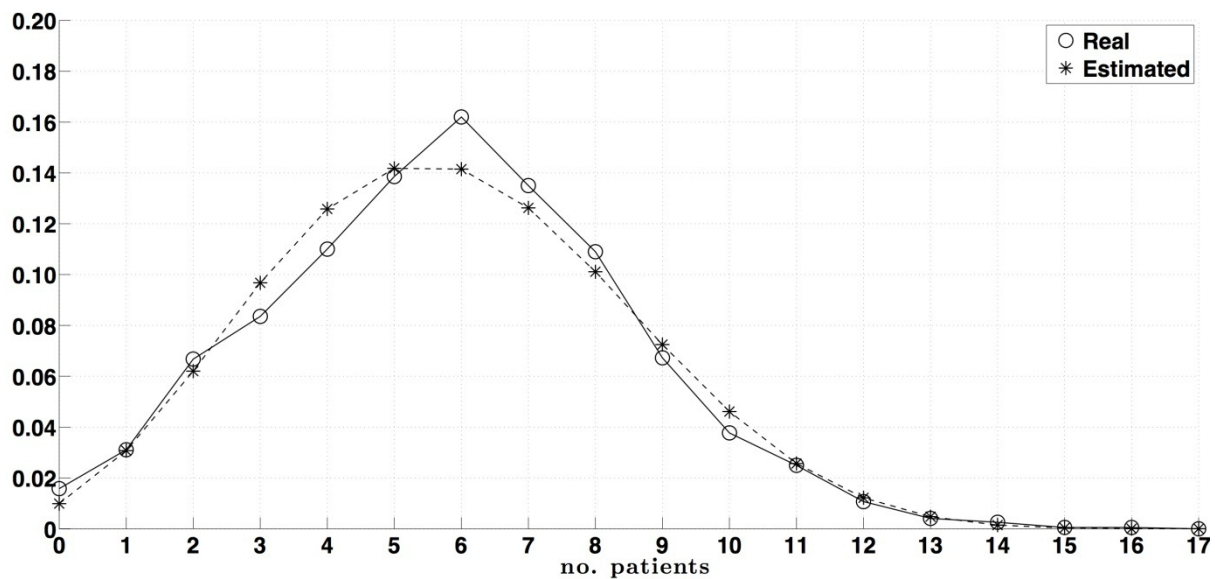


Figure 6 Admission rate distribution

6. Conclusion

In this paper we used a new type of PH distribution, we gave a brief overview of some properties and we presented a particular application to hospital LoS data. The generalization proposed uses a few number of parameters and it supplies a phenomenological interpretation of data. To provide an operational guideline in bed planning it is analysed the error estimators time trend and the it is validated the stability of the estimated parameters. We also introduced a new approach (single factor model) for this context which arises from financial mathematics. financial model, in order to describe the admission rate and the discharge rate of the department.

The following conclusions can be drawn:

- 1) PHGamma distribution provides a good fit of the Cardiology department data in fact it performs the smallest discrepancy measures (AAE, APE, ARPE, and RMSE).
- 2) The single factor model provides a good description of the admission rate and the discharge rate and supplies an adequate representation of the hospital operations department.
- 3) The forecast ability is validated using a sequential procedure based on the analysis of some error indicator. Moreover the stability of the estimated parameters is shown.

Acknowledgements

This research is supported by the Tor Vergata department of Cardiology.

References

Asmussen, S., Nerman, O., & Olsson, M. (1996). Fitting phase-type distribution via EM algorithm. *Scand j Statist*, 23, 419-441.

- Bebbington, A., Darton, R., & Netten, A. (2001). Care Homes for Older People. *Admissions, Needs and Outcomes* (vol 2). Canterbury: University of Kent.
- Bobbio, A., & Cumani. (1992). ML estimation of the parameters of a PH distribution in triangular canonical form. *Computer performance evaluation* (pp 33-46). Elsevier, Amsterdam.
- Costa, A. X. (2003). Mathematical modelling and simulation for planning critical care capacity. *Anaesthesia*, 58, 320327.
- Dale, M. C., Burns, A., Panter, L., & Morris, J. (2001). Factors affecting survival of elderly nursing home residents. *Int. J. Geriatr. Psychiatr.*
- Dowland, K. A., & Thompson, J. M. (2000). Solving a nurse scheduling problem with knapsacks, networks and tabu search. *Journal of the Operational Research Society*, 51.
- Dumas, M. B. (1984). Simulation modelling for hospital bed planning. *Simulation*, 43.
- Fackrell, M. (2008). Modelling healthcare system with phase-type distributions. *Health Care Manage Sci.*
- Faddy, M. J., & McClean, S. I. (1999). Analysing data on length of stay of hospital patients using phase-type distribution. *Appl Stoch Models Bus Ind*, 15, 311-317.
- Fang, K. T., & Yang, Y. (1994). *Number-Theoretic Methods in Statistics*. London: Chapman&Hall.
- Godfrey, K. (1983). *Compartmental Models and their Application*. London: Academic Press.
- Gordon, T. (1998). Surgical unit time utilization review: Resource utilization and management implications. *Journal of Medical Systems*.
- Gove, D, Hewett, D., & Shahani, A. K. (1995). Towards a model for hospital case-load decision support, *IMA Journal of Math. Applied to Medical Biology*, 12.
- Harper, P., & Shahani, A. (2002). Modelling for the planning and management of bed capacities in hospitals. *Journal Operational Resource Society*, 53.
- 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(3).
- Huang, X. (1995). A planning model for requirement of emergency beds, *IMA Journal of Math. Applied to Medical Biology*.
- Jang, J. W. (2002). The distribution of loan portafolio value. *Risk*.
- Kao, E. P. C., & Tung, G. G. (1991) Bed allocation in a public health care delivery service. *Management Science*, 27.
- Lin, W. T. (1989). Modelling and forecasting hospital patient movements: Univariate and multiple time series approaches, *IJF (The Netherlands)*, 5.

- Marshall, A. H., & McClean, S. I. (2003) Conditional phase-type distributions for modelling patient length of stay in hospital. *International Transactions in Operational Research*, 10.
- McClean, S., & Millard, P. H. (1993). Modelling in-patient bed usage behaviour in a department of geriatric medicine. *Meth. Inform. Med.*
- Millard, P. H. (1988). *Geriatric medicine: a new method of measuring bed usage and a theory for planning*. MD Thesis. University of London, London.
- Millard, P. H. (1994). Modelling Hospital Resource Use: a Different Approach to the Planning and Control of Health Care Systems. *Royal Society of Medicine, London*, 2937.
- Neuts, M. F. (1975). Probability distributions of phase-type. In *Liber amicorum Prof. Emeritus H. Florin*. Department of Mathematics, University of Louvain, Louvain.
- Niederreiter, H. (1992). *Random Number Generation and Quasi-Monte Carlo Methods*. Philadelphia: SIAM.
- Palombini, E. (2009). *Factor models and credit risk of a loan portfolio*. MPRA, Munich.
- Ross, S. M. (1993). Introduction to probability models (5th ed.). San Diego: Academic.
- Ross, S. M. (1995). Stochastic Process (2nd ed.). John Wiley&Sons, Inc.
- Rothera, I. C., Jones, R., Harwood, R., Avery, A. J., & Waite, J. (2002). Survival in a cohort of social services placements in nursing and residential homes: factors associated with life expectancy and mortality. *Public Health*.
- Ruth, R. J. (1981). A mixed integer programming model for regional planning of a hospital inpatient service. *Management Science*, 27.
- Shahani, A. K. (1991). Reasonable averages that give wrong answers. *Teaching Statistics* 3.
- Taylor, G. J., McClean, S. I., & Millard, P. H. (1998) Continuous-time Markov models for geriatric patient behaviour. *Appl. Stochast. Models Data Anal*, 13.
- Taylor, G. J., McClean, S. I., & Millard, P. H. (2000). Stochastic models of geriatric patient bed occupancy behaviour. *J. R. Statist. Soc.*
- Taylor, G. J., McClean, S., & Millard, P. H. (1996). Geriatric patient flow-rate modelling. *IMA Journal of Math. Applied to Medical Biology*.
- Wolfe, H., & Young, J. P. (1963). Staffing the nursing unit, part I: Controlled variable staffing. *Nursing Research*, 14.
- Wright, I. H. (1996). Statistical modelling to predict elective surgery time: Comparison with a computer scheduling system and surgeon provided estimates. *Anaesthesiology*, 85.

Appendix

Function of a matrix. Given a matrix $n \times n$ matrix T with real entries and a scalar function

$g: \mathbb{R} \rightarrow \mathbb{R}$, we are interested in definition of $g(T)$ which specifies $g(T)$ to be a matrix of the same dimensions of T . It is a standard result that the matrix T can be expressed in the Jordan canonical form:

$$Z^{-1}TZ = \text{diag}(J_1, J_2, \dots, J_p), \quad (5)$$

$$J_k = J_k(\lambda_k) = \begin{bmatrix} \lambda_k & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \lambda_k \end{bmatrix} \in \mathbb{C}^{m_k \times m_k}$$

where Z is nonsingular and $m_1 + m_2 + \dots + m_p = n$. The Jordan matrix J is unique up to the ordering of the blocks J_i , but the transforming matrix Z is not unique. Denote by $\lambda_1, \lambda_2, \dots, \lambda_s$ the distinct eigenvalues of T and let n_i be the order of the largest Jordan block in which λ_i appears, which is called the *index* of λ_i . We need the following terminology.

Definition 1. The function g is said to be defined on the spectrum of A if the values exist. These are called the values of the function g on the spectrum of A . Here $D^j g$ denotes the j th derivative of g .

$$D^j g(\lambda_i), \quad j = 0, \dots, n_i - 1, \quad i = 1, \dots, s$$

In most cases of practical interest g is given by a formula. However, the following definition of $g(A)$ requires only the values of g on the spectrum of A ; it does not require any other information about g . It is only when we need to make statements about global properties (such as the continuity) that we will need to assume more about g .

Definition 2. (Jordan canonical form). Let g defined on the spectrum of $A \in \mathbb{C}^{n \times m}$ and let A have the Jordan decomposition form (5). Then

$$g(A) = Zg(J)Z^{-1} = Z\text{diag}(g(J_k))Z^{-1},$$

We observe that the definition yields a matrix $g(A)$ that can be independent of the particular Jordan canonical form that is used. Second, note that if A is diagonalizable, then the Jordan canonical form reduces to an eigenvalue decomposition $A = ZDZ^{-1}$, with $D = \text{diag}(\lambda_i)$ and the columns of Z are the eigenvectors of A . Therefore for diagonalizable matrices $g(A)$ has the same eigenvectors as A and its eigenvalues are obtained by applying g to those of A .

A function $g: \mathbb{R} \rightarrow \mathbb{R}$ is a probability density function of a continuous random variable if

- g is a non-negative Lebesgue-integrable function;
- $\int_{-\infty}^{+\infty} g(x)dx = 1$.

Then we have the following result:

Proposition 1. Let g be a probability density function. For every $\alpha_1, \alpha_2, \dots, \alpha_n \geq 0$ with $\alpha_1 + \alpha_2 + \dots + \alpha_n = 1$, and $\lambda_1, \lambda_2, \dots, \lambda_n > 0$ the function

$$h(\tau) = \sum_{i=1}^n \alpha_i g(\lambda_i \tau) \lambda_i$$

is a probability density function.

It is also possible evaluate the mean and the variance of a random variable distributed according to a PGH density function:

$$\mathbb{E}[X] = \frac{k}{\theta} \sum_{i=1}^n \frac{\alpha_i}{\lambda_i},$$

$$\mathbb{E}[(X - \mathbb{E}[X])^2] = \frac{k^2 + k}{\theta^2} \sum_{i=1}^n \frac{\alpha_i}{\lambda_i} - \left(\frac{k}{\theta} \sum_{i=1}^n \frac{\alpha_i}{\lambda_i} \right)^2,$$

$$\mathbb{E}[e^{Xt}] = \sum_{i=1}^n \alpha_i \left(1 - \frac{1}{\theta} \frac{t}{\lambda_i} \right)^{-k}$$

Here the moment generating function is well defined on the interval $t < \theta \min(\lambda_1, \lambda_2, \dots, \lambda_n)$.

Copyright Disclaimer

Copyright reserved by the author(s).

This article is an open-access article distributed under the terms and conditions of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by/3.0/>).