Modelling the Duration of Hypopnea

1Nur Zakiah Mohd Saat and 2Abdul Aziz Jemain
1Department of Biomedical Sciences, Faculty of Allied Health Sciences,
University Kebangsaan Malaysia, 50300 Jln Raja Muda Abdul Aziz, Kuala Lumpur, Malaysia
2School of Mathematical Sciences, Faculty of Science and Technology, Universiti Kebangsaan Malaysia,
43600 Bangi, Selangor DE, Malaysia

Abstract: In this study, we fit a parametric model to the duration of hypopnea. Hypopnea is a partial apnea. A total of two hundred and seven events of duration of hypopnea is considered in this study. The parametric model that has been used is gamma, Weibull, log-normal and log-logistic distribution. The most appropriate distribution to fit the duration of hypopnea is analyzed according to sleep stages. The results indicated that Weibull distribution is the best model for light, deep and Rapid Eye Movement (REM) sleep stages.

Key words: Hypopnea, gamma distribution, Weibull distribution, log-normal distribution, log-logistic distribution

INTRODUCTION

Apnea describes the process of complete cessation of airflow during sleep which last between 10 and 60 sec (Tsai et al., 1999). Notably, this type of sleep disorder has been associated with factors such as increasing age, obesity, gender and smoking habits (Guilleminault and Dement, 1976; Worsnop et al., 1998). Prior research on the modelling of apnea has focused on the occurrence of apnea during sleep and wake duration (Lo et al., 2004). On the other hand, a hypopnea is a partial apnea (Moser et al., 1994; Guilleminault et al., 2001). Hypopnea is defined as 50% cessation of airflow (Tsai et al., 1999). Also, research on the modelling of sleep disorder has extended to examine sleep wake transitions. For instance, study by Yang and Hensch (1973) gathered that the suitability of Markov model is depending on sleep stages i.e., light sleep, deep sleep and Rapid Eye Movement (REM) and age group.

Patients that are suffering from apnea usually experienced symptoms such as fatigue and sleepiness during the day (Black, 2003). To ensure appropriate treatment of the illness, the medical practitioners would refer these patients to a sleep clinic. At the clinic, the patients will be observed during their sleep in order to determine the severity of their apnea. Study on apnea duration is important as apnea is related to hypertension, cardiovascular disease, automobile accident and excessive day time sleepiness (Illa et al., 1994; Nieto et al., 2000; Hossain and Shapiro, 2002). Consequently, it is essential to study the distribution of apnea and hypopnea. Parametric model is one of the methods to identify the distribution of apnea and hypopnea. More importantly, it will allow us to understand the patterns of apnea.

Parametric models have been used in the study of rainfall in terms of rain rate (Kedem et al., 1994). The probability distribution that has been applied to model the rainfall data is gamma and lognormal distribution. It was found that the lognormal distribution was sensible for the rain rate in Darwin and Florida. Furthermore, Onoz and Bayazit (1995) employed pareto, log-normal and Generalized Extreme Value (GEV) distribution to fit the data of flood samples. It was found that GEV is the sensible model compared the other model selected. A number of authors applied parametric model in medical research. Atienza et al. (2008) used a mixture model which is gamma, Weibull and log-normal to model the length of stay in the hospital.

The distribution of apnea is truncated because an apnea is a cessation of airflow for at least 10 sec. It indicates that cessation of airflow that less than 10 sec is discarded from the recording of apnea event. Therefore, it is interesting to determine the effect of truncation towards the probability distribution of apnea. Using a probability distribution approach is essential in determining the distribution of apnea. A parametric model is one of the methods to identify the distribution of apnea. This will lead us to understand the process of apnea.
Although, there are many studies explaining regarding the duration of sleep (Hasler et al., 2004; Rechtschaffen et al., 1999; Wehr et al., 1998), there are limited studies on the duration of apnea. In this study, we consider the duration of hypopnea. The objective of this study is to determine the best model to fit the data of hypopnea. This is essential in order to understand the pattern of apnea and to find the best model for the duration of apnea.

PARAMETRIC MODELS

The data on apnea is obtained from a polysonography which was originated from Rigney et al. (1994). This database involves collaboration between Massachusetts Institute of Technology and Beth Israel Hospital (MIT-BIH) in Boston. The database consists of 14 subjects records of polysonography for 14 males of age between 32 and 56 years. The database was extracted from www.physiobank.net (Goldberger et al., 1994).

A hypopnea event is defined as reduction in breathing that remain for at least 10 sec (Guilleminault et al., 1976, 2001). However, reductions in breathing that last for less than 10 sec are not recorded. Let X denote the duration of a periods of reduced breathing. We assume that the duration of successive events are independent random variables from a distribution with probability density function (pdf), f(x; θ), where, θ is a vector of parameters to be estimated. The observed data are independent observations from this distribution left-truncated at X = 10. Therefore the pdf of the duration of an observed hypopnea is:

\[ f_0(x; \alpha, \beta) = \frac{f(x; \theta)}{\Pr(X \geq 10)} \]

Where:

\[ \Pr(X \geq 10) = \int_{10}^{\infty} f(x; \theta) \, dx \]

Let \( x_1, x_2, \ldots, x_n \) denote the durations of a sequence of observed hypopneas. Because of the assumption of independence, the log-likelihood function of \( \theta \) is:

\[ L(\theta) = \sum_{i=1}^{n} \log f_0(x_i; \theta) \]

The following distributions of durations have been investigated:

**Gamma distribution**

\[ f(x; \alpha, \beta) = \frac{x^{\alpha-1} \exp \left( -\frac{x}{\beta} \right)}{\Gamma(\alpha) \beta^\alpha}, \quad x > 0 \]

where, \( \alpha > 0 \) and \( \beta > 0 \)

In reference to Hogg and Tanis (2005), the mean and variance of the distribution is \( \alpha \beta \) and \( \alpha \beta^2 \), respectively.

**Log-normal distribution**

\[ f(x; \mu, \sigma) = \frac{1}{\sigma \sqrt{2\pi}} \exp \left( -\frac{(\log x - \mu)^2}{2\sigma^2} \right), \quad x > 0 \]

where, \( -\infty < \mu < \infty \) and \( \sigma > 0 \)

The mean of the distribution is \( \exp[\mu+(\sigma^2/2)] \) and the variance is \( \exp[2\mu+\sigma^2] \left[ \exp(\sigma^2)-1 \right] \) (Hosmer and Lemeshow, 1999).

**Weibull distribution**

\[ f(x; \kappa, \lambda) = \frac{\kappa}{\lambda} x^{\kappa-1} \exp \left( -\frac{x}{\lambda} \right)^\kappa, \quad x > 0 \]

where \( \kappa > 0 \) and \( \lambda > 0 \). The mean of the distribution is \( \lambda \Gamma(1+1/\kappa) \) and the variance is (Hosmer and Lemeshow, 1999).

**Log-logistic distribution**

\[ f(x; \mu, \sigma) = \frac{\exp \left( \log \frac{x - \mu}{\sigma} \right)}{\sigma \left[ 1 + \exp \left( \frac{\log (x - \mu)}{\sigma} \right) \right]^2} \]

where, \( -\infty < \mu < \infty \) and \( \sigma > 0 \)

The mean of the distribution is \( \exp(\mu) \left[ \Gamma(1+\sigma) \Gamma(1-\sigma) \right] \Gamma(1+\sigma) \Gamma(1-\sigma) \) (Law and Kelton, 1982).

RESULTS

**Distribution of hypopnea event:** There are 207 hypopnea events, with 108 during REM sleep. The mean duration of a hypopnea event is 15.88 sec and maximum is 41.20 sec. A histogram of duration for all sleep stages categories was constructed and specifically focusing on the hypopnea events as it has most events compared to other events(arousal, false oxygen saturation, obstructive sleep apnea and central sleep apnea). Figure 1 shows that hypopnea events are skewed to the right with most of the hypopnea duration are 10 to 20 sec. A standard survival
models which is gamma, log-normal, log-logistic and Weibull was fitted to the data. The statistical analysis was performed using SPLUS 6.0 for Windows, Insightful Corp (2000).

Goodness of fit test which is to discriminate between continuous distribution was done based on empirical distribution function Kolmogorov-Smirnov (K-S). We compare between distributions of light sleep and deep sleep to determine the differences between both distributions. There is slightly difference for cumulative density function (cdf) between NREM and REM. Based on K-S test NREM and REM are not significantly different (K-S = 0.181, p = 0.052) and there was no significantly different for cdf of light sleep and deep sleep (K-S = 0.191, p = 0.538). This indicated that there are small difference of distribution when we divide NREM to two groups because the p-value is almost less than 0.05. In order, to investigate a better understanding about the differences between distribution of light sleep and deep sleep Maximum Likelihood Estimation (MLE) is used.

Maximum Likelihood (ML) estimates of parameters of two parameter left-truncated probability distributions were obtained. The data is left truncated because it is truncated at 10 sec. The objective of the study is to fit the four distribution (gamma, log-normal, Weibull and log-logistic) to the duration of the hypopnea. The particular distributions were selected because the data set are positive distribution and skewed to the right. Another purpose of the study is to fit the distribution of duration of hypopnea according to sleep stages.

**FITTING A PARAMETRIC DISTRIBUTION FOR DURATION OF HYPOPNEA**

The log-likelihood is used to select the best model that described the duration of hypopnea according to sleep stages. The log-likelihood of fitting the data of duration of hypopnea are shown in Table 1. The results indicated almost similar value of log-likelihood for the four distributions. The log-likelihood according to NREM sleep stages is shown in Table 2. Generally, the log-likelihood is decreased compared to the log-likelihood of overall duration of hypopnea. This is due to the reduction of sample size in NREM sleep stages compare to overall duration of hypopnea. The log-likelihood for Weibull distribution is the lowest while the log-logistic distribution show the highest log-likelihood indicating that log-logistic is poorly fit for duration of hypopnea during NREM.

Table 3 shows the log-likelihood of fitting the data during REM sleep stages. Most of hypopnea events occurs during REM, with the number of hypopnea events is 108. Similarly, the log-likelihood of the four distribution is almost equal with the most difference is less than two. The parameter of gamma, log-normal and Weibull decreased in REM group compared to the NREM group. However the parameter of $\mu$ and $\sigma$ from the logistic distribution is slightly increased.

In Table 4, it is showed the log-likelihood values for light sleep stage group. There was higher increased in the shape parameter which is parameter $\alpha$ compared to REM and NREM sleep stages. Similarly, there was a slight
increased in parameter scale for Weibull distribution. However, there was a decrease in parameter log-normal and log-logistic distribution. There was a reduction almost 100 in terms of log-likelihood compared to log-likelihood in REM sleep stages. Results about deep sleep stages is shown in Table 5. Hypopnea events were less in deep sleep stages compared to REM and light sleep stages. The value of log-likelihood is almost similar for all the distributions. Moreover, there was an increase in the value of scale parameter for Weibull distribution. The K-S test has been used to determine the goodness of fit test. The results from Table 1 until Table 5 show that there was no appropriate distribution from the four distributions to fit the data. It is due to the limitation of the K-S test. K-S test is not preferable for the skewed data. Another limitation, the K-S test is not suitable for the distributions that the parameter is estimated from the data as it can determine the goodness of fit better if the parameter is estimated from the simulation data. It is believe due to the limitation in K-S test, the distribution that is appropriate for the data may not be determined (D’Agostino and Stephens, 1986).

The likelihood ratio in given by LR = -2 ln(θ). θ is l_1/l_2, l_1 is the log-likelihood for the first model and l_2 is the log-likelihood for the second model (Hosmer and Lemeshow, 1999). The value of LR is compared with the Chi-square with 5% significance level and degree of freedom is two. The results obtained in Table 6 comparing between light and deep sleep showed a reduction in maximized likelihood, for example for gamma distribution from 563.14 to 562.10. The LR test showed that there were significantly different between light and deep sleep stages for Weibull distributions (test statistic = 7.66, critical value = 5.99, p<0.05). However, for gamma, log-normal and log-logistic the LR test is similar between light and deep sleep stages. Interestingly, the decreased in maximized likelihood is due to subdivide the NREM to light and deep sleep. Log-logistic distribution showed a clearly higher maximized likelihood values compared to other distributions. Numerical optimization was used to maximize the log-likelihood which is non-linear least squares functions (Venables and Ripley, 2002).

In Table 7, we reported a comparison between NREM and REM. The LR test revealed that the log-likelihood in combining NREM and REM are significantly higher compared to the full model (p< 0.05). It shows that gamma, log-normal, Weibull and log-logistic distribution is significantly different between NREM and REM sleep stages. The likelihood ratio test presents that Weibull is appear to provide the best fit for the NREM and REM because it has the highest LR test, compared to other three distributions. Therefore, Weibull distribution is selected as the best model to fit the REM and NREM sleep stages. This indicated that there was a significantly different between REM and NREM for the duration of hypopnea, especially when Weibull is fitted to the data.

**DISCUSSION**

Many studies have shown that apnea have a relationship with cardiovascular disease, obesity and hypertension (Ichimaru et al., 1990). Therefore study on apnea is important in order to understand the underlying process of apnea. Researcher has been searching a statistical analysis that is appropriate to determine patients that has severe apnea. Measurement that has been used include percentage of apnea in REM, light sleep and deep sleep stages. Study by Goh et al. (2000) has found that most of apnea event is during REM. This is consistent with present study as most of hypopnea event is also in REM sleep stages. Earlier investigators has considered other variables such as Apnea-Hypopnea Index (AHI) to compare between age group, gender and race. Most of the study conclude that AHI is differ according to demographic factors (Redline et al., 1997; Bixler et al., 2001; Stradling and Davies, 2004). However, there was very few study on the distribution of apnea according to subjects, age group, gender and other demographic factors. Consequently, the appropriate model to describe the distribution of apnea is essential to give more explanations about the comparison. This will allow the researcher to distinguish the apnea between the sleep stages rather than percentage results only. This study highlighted the distribution of hypopnea, which is 50% reduction of airflow (Tsai et al., 1999). In this study, a variety of parametric distributions, particularly hypopnea is used in order to find a model that is suitable in presenting the duration of hypopnea distribution. The histogram of the duration of hypopnea is skewed to the
right with a long tail. However, which distributions would provide the best fit was not obvious from the histogram. The normal distribution was not considered in this study because it may give a poor fit because the data is skewed. Present results demonstrate that most of the hypopnea event is during REM sleep stages and very seldom during deep sleep stage. This finding consistent with the study by Goh et al. (2000), who founds that most of the hypopnea events, is during REM compared to NREM sleep stages and very few during deep sleep stage.

The most important result of this study showed that an almost similar log-likelihood value for gamma, lognormal, Weibull and log-logistic distribution. Secondly, the appropriate distribution for the overall data is lognormal distribution and Weibull distribution is the best fit for light, deep and REM sleep stages. Finally another important finding is that there was improvement in fit when the data is subdivided into light, deep and REM sleep stages from the LR test. Models that is fit badly such as log-logistic model gave an almost similarly loglikelihood for all the three categories of sleep stages. This result may be due to the truncation as hypopnea is an event that at least 10 sec. Therefore further studies on simulation of hypopnea data is suggested to considering the effect of truncation to the duration of hypopnea.

The results that are obtained may also be influenced by other factors such as age group, overweignt, smoking habits, race and alcohol consumption. Many studies have shown that there were more frequent apnea per hour in patients that male, older adults, overweight and smokers (Sakakibara et al., 1999; Redline et al., 1997; Stradling et al., 2004). Therefore it is also suggested that further study to determine the best model to describe the duration of apnea according to age group, gender, overweignt, smoking habits and alcohol consumption.

ACKNOWLEDGMENTS

The authors acknowledge the Ministry of Higher Education Malaysia for financial support. The corresponding author thanks M.S. Ridout for the comments on the results.

REFERENCES


