Research Paper

Volume : 4 | Issue : 5 | May 2015 • ISSN No 2277 - 8179

Gamma Frailty Model to Evaluate Heterogeneity for A Survival Data

Statistics

KEYWORDS : Gamma frailty model, heterogeneity, right censored, survival analysis.

S. Deepapriya	M.Sc.(Maths), M.Phill. (Maths), M.Sc. (Stats), Research Scholar, Presidency college, Chennai.
Dr.. R Ravanan	M.Sc., M.Phil., Ph.D., Head, Department of statistics, Presidency college, Chennai.

ABSTRACT

A frailty model is a random effect model for time to event data. In the application of survival analysis, unobserved individual heterogeneity also called as frailty is considered in this paper. The marginal proportional hazards frailty model for multivariate failure time data characterizes the intracluster dependency with the frailty model while formulating the marginal distributions with the proportional hazards model. The gamma frailty distribution has been widely used to model intracluster dependency because of its simple interpretation and mathematical tractability, Glidden (2000). This paper deals with the gamma frailty model to evaluate the amount of heterogeneity of 700 tuberculosis patients. A clinical study of these right censored data has been analyzed and the results are discussed.

Introduction

The notion of frailty provides a convenient way to introduce random effects, association and unobserved heterogeneity into models for survival data. In its simplest form, a frailty is an unobserved random proportionality factor that modifies the hazard function of an individual, or of related individuals. In essence, the frailty concept goes back to work of Greenwood and Yule (1920) on "accident proneness". The term frailty itself was introduced by Vaupel et al. (1979) in univariate survival models and the model was substantially promoted by its application to multivariate survival data in a seminal paper by Clayton (1978) (without using the notion "frailty") on chronic disease incidence in families.

Frailty models are extensions of the proportional hazards model which is best known as the Cox model (Cox, 1972), the most popular model in survival analysis. Normally, in most clinical applications, survival analysis implicitly assumes a homogenous population to be studied. This means that all individuals sampled into that study are subject in principle under the same risk (e.g., risk of death, risk of disease recurrence). In many applications, the study population cannot be assumed to be homogeneous but must be considered as a heterogeneous sample, i.e. a mixture of individuals with different hazards. For example, in many cases it is impossible to measure all relevant covariates related to the disease of interest, sometimes because of economical reasons, sometimes the importance of some covariates is still unknown. The frailty approach is a statistical modeling concept which aims to account for heterogeneity, caused by unmeasured covariates. In statistical terms, a frailty model is a random effect model for time-to-event data, where the random effect (the frailty) has a multiplicative effect on the baseline hazard function. This random effect explains the dependence in the sense that had we known the frailty, the events would be independent. In other words, the life times are conditional independent, given the frailty. This approach can be used for survival times of related individuals like family members or recurrent observations on the same person.

The topic of heterogeneity in the analysis of duration times has received much attention in a wide variety of fields, such as demography (vaupel(1990)) econometrics Heckman and Singer (1984) and statistics[clayton and cuzick (1985), Aalen (1988), Nielsem,Gill, Andersen and Sorensen (1992)]. The heterogeneity may because by an unobservable covariates as explained by Heckman and singes or due to block effects as in the Danish twin subject effect when recording serial events per subject [oakes (1991)]

A simple model for heterogeneity which is reminiscent of the proportional hazard model is via the concept of frailty as proposed by vaupel, menton, and stallard (1979), the frailty is an unobserved random factor applied to the base line hazard function of the duration time, so if the hazard function of a subject with a frailty value of 1 is given by $\alpha(t)$, the hazard function of a subject with a frailty value of Z is given by $Z\,\alpha(t)$. In this paper the frailty is assumed to follow gamma distribution with mean 1 and unknown variance, mathematically this is convenient as the gamma distribution is a conjugate prior. In some settings there is some justification for the assumption that the frailty follows a distribution skewed to the right, such as a lognormal or gamma distribution [Aalen(1988)]. At any rate, as discussed by Manton, stallard and vaupel (1986) often the fit of a model is more sensitive to assumptions on the form of $\alpha(t)$ than on the form of the frailty α distribution Heckman and singer(1984) give a functional form to α and leave the form of the frailty distribution unspecified. Other settings allow the specification of both the form of α and the frailty distribution.

Survival function

The term "survival analysis" pertains to a statistical approach designed to take into account the amount of time an experimental unit contributed to a study .ie, it is the study of time between entry into observation and a subsequent event. Application of survival analysis is very vast; now include time until onset of disease, time until stock market crush, time until equipment failure, and time until earth quake and so on.

Different types of probability tools are usually encountered in survival analysis and their properties are described here. Let T be the time variable, considered as a positive real valued variable, having a continuous distribution with finite expectation. For applications, this variable represents the time being in a given state or the time between events. Several functions characterized the distribution of T:

F(t), t≥0 is the probability density of T;

$$S(t) = p(T > t) = \int_{t}^{\alpha} f(x)\,dx = 1 - F(t), \text{ is the}$$

survival function, which is the probability of an individual surviving beyond time t. (F (t) is the cumulative distribution.)

Hazard function

Let T represent survival time. We regard T as a random variable with cumulative distribution function P (t) =Pr (T≤ t) and probability density function P (t) = dP(t)/dt. The more optimistic survival function S (t) is the complement of the distribution function, S (t) = Pr (T> t) = 1-P (t). This is the representation of the distribution of survival times is the hazard function, which demise at time t, conditional on survival to that time:

$$\lim_{\Delta t \to 0} \frac{\Pr\left[(t \le T < t + \Delta t) | T \ge t\right]}{\Delta t} = \frac{f(t)}{s(t)}$$

Modeling of survival data usually employs the hazard function. For example, assuming a constant hazard, h(t) = v, implies an exponential distribution of survival times.

Proportionality hazard ratio

The hazard Ratio of a survival function is $HR = \frac{h_0(t)\exp(X_1\beta)}{h_0(t)\exp(X_0\beta)}$

$$= \exp((X_1 - X_0)\beta)$$

$$= \exp((Z_0)\beta)$$

From a proportionality hazard perspective it is easy to see how α corresponds to an omitted covariates.

$$h\left(t_i/\alpha_i\right) = \alpha_i h(t_i)$$

$$= \alpha_i h_0(t_i) e(x_i^t \beta)$$

The conditional survival function is

$$S\left(\frac{t}{\alpha}\right) = e\left\{-\int_0^t h(u/\alpha)du\right\} = \{s(t)\}^{\alpha}$$

$$h(t_i) = h_0(t_i)\exp(x_i^t\beta)$$

for some baseline hazard $h_0(t)$

The proportional hazard frailty model assumes that for given frailty variable Z_i and covariates Xi individual has hazard

$$h_i(t) = z_i \alpha(t) e^{\beta^t x_i}$$

At time t, the model is then completed by assuming a parametric distribution for zi. If the dependence of α(t) on t is specified by finitely many parameters M.L.E is in principle a routine matter However based on the prominent place in survival analysis.

Frailty modeling

Frailty models without observed covariates, such models are used when only survival data are available for the analysis, or when additional information is of no interest. These models are described by the stochastic hazards (Z, x)=$z\mu_0$(x), or

$$\mu(Z,x) = Z\mu_{0(x)} + \tilde{\mu}(x)$$

Here the non-negative random variable Z is called frailty, μ0(x) is an underlying hazard and is a $\tilde{\mu}$(x) background hazard. This model is non-identifiable from survival data, i.e., different combinations of μ0$\tilde{\mu}(x)$, and frailty distributions may produce the dame marginal hazard rate. The model becomes identifiable when the parametric structures of μ0(x) and $\tilde{\mu}(x)$ are fixed and Z assumed to belong to some parametric distribution family, e.g. gamma.

The density of gamma distributions with parameters α and β is

$$g(\theta) = \frac{\theta^{\alpha-1}e^{-\beta\theta}\beta^{\alpha}}{\Gamma\alpha} \quad \text{for } \alpha, \beta > 0$$

The mean and variance are $E(\theta) = \frac{\alpha}{\beta}$ and $var(\theta) = \frac{\alpha}{\beta^2}$

So the coefficient of variation is $\frac{1}{\sqrt{\alpha}}$

It is often convenient to take E (θ) =1 so $\alpha=\beta=1/\sigma^2$ this entails no loss of generality because the average level of frailty can always absorbed into the baseline hazard.

If the hazard is constant for each individual but people are heterogeneous and frailty has a gamma distribution then the population hazard is

$$\lambda(t) = \frac{\lambda}{1 + \sigma^2\lambda t}$$

Where λ is the average individual hazard and σ^2 is the variance of frailty.

The Laplace transformation of g is given by $\bar{g}(\epsilon) = \left(\frac{\beta}{\beta+\epsilon}\right)^{\alpha}$ distributed as gamma with unit mean and variance $1/\sigma^2$

The likelihood ratio of subject observed from $(t_{0ij}, t_{ij}]$ is

$$L\left(t_{ij}/t_{0ij}, \alpha_i\right) = \left\{\frac{s(t_{ij})}{s(t_{0ij})}\right\}^{\alpha_i} \{\alpha_i h(t_{ij})\}^{d_{ij}}$$

Gamma frailty model

A convenient assumption used by many authors is that θ has a gamma distribution. This distribution has the appropriate rang (0, ∞) and is mathematically traceable.

The density of gamma distributions with parameters α and β is

$g(\theta) = \frac{\theta^{\alpha-1}e^{-\beta\theta}\beta^{\alpha}}{\Gamma\alpha}$ for α, β>0

The mean and variance are and

So the coefficient of variation is

It is often convenient to take E (θ) =1 so $\alpha=\beta=1/\sigma^2$ this entails no

loss of generality because the average level of frailty can always absorbed into the baseline hazard.

If the hazard is constant for each individual but people are heterogeneous and frailty has a gam ma distribution then the population hazard is

Where λ is the average individual hazard and σ^2 is the variance of frailty.

The Laplace transformation of g is given by distributed as gamma with unit mean and variance

The likelihood ratio of subject observed from is

The gamma frailty model is a natural extension of the estimation Cox proportional hazards model in survival analysis. Because the frailties are unobserved, an E-M approach is often used for. Such an approach is shown to lead to finite sample underestimation of the frailty variance, with the corresponding regression parameters also being underestimated as a result. For the univariate case, we investigate the source of the bias with simulation studies and a complete enumeration. The rank – based E-M approach, we note, only identifies frailty through the order in which failures occur; additional frailty which is evident in the survival times is ignored, and as a result the frailty variance is underestimated. An adapt ion of the standard E-M approach is suggested, whereby the non parametric Breslow estimate is replaced by a local likelihood formulation for the baseline hazard which allows the survival times themselves to enter the model. Simulations demonstrate that this approach substantially reduces the bias, even at small sample sizes. The method developed and applied to survival data from the North West Regional Leukemia Register.

Frailty models are used to model survival times in the presence of over dispersion or group specific random effects. The latter are distinguished from the former by the term "shared" frailty models. With the release of Stata 7, estimation of parametric non-shared frailty models is now possible, and the new models appear as extensions to the six parametric survival models previously available. The over dispersion in this case is represented by an unobservable multiplicative effect on the hazard, or frailty. For purpose of estimation this frailty is then assumed to either follow a gamma or inverse-Gaussian distribution. Parametric shared frailty models are the next logical step in the development in this area, and will soon be available as an update to Stata 7.

Analysis

The data base consists of 720 Tuberculosis TB cases admitted in a Randomized controlled clinical trial done at tuberculosis research center, ICMR, Chennai. The event of interest is time till the sputum conversion during treatment period.

The covariates under the study are

- Age in years
- Treatment: regiment
- Gender: Male(1) Female(0)
- Drug susceptibility pattern: Res(1) Sen(0)
- Cavity X-ray: Abnormal(1) Normal(0)
- Weight in Kg at the time of admission
- Status: event(1) censoring(0)

The data is analyzed using the statistical tools SPSS and STATA.

- Analysis of data
- Gamma frailty
- No. of subjects = 694
- Number of obs = 694
- No. of failures = 657
- Time at risk = 2058
- LR chi2 (6) = 81.66
- Log likelihood = -699.92992
- Prob > chi2 = 0.0000

Table 41 Analysis of TB Data – Stata output

Covariates	Coef.	Std. Err.	z	P>|z|	[95% CI]	
sexcode	.1109032	.0675308	1.64	0.101	-.0214547	.2432612
senscode	.3730484	.0693049	5.38	0.000	.2372133	.5088834
age	.0065785	.0022897	2.87	0.004	.0020908	.0110661
cavit	.1747142	.0303072	5.76	0.000	.1153131	.2341153
treat	.0061397	.0074198	0.83	0.408	-.0084029	.0206822
wt0	.0020118	.0051698	0.39	0.697	-.0081207	.0121443
theta	.9898061	.2490807			.6044338	1.620883

Likelihood-ratio test of theta=0: chibar2(01) = 36.73
Prob>=chibar2 = 0.000

Result

Using the Gamma frailty (from Table 4.1) we could access that the following covariates are said to be the influencing factor for the time until the sputum conversion are give in the table below.

Table 42 Significant Covariates obtained by Gamma Frailty Model

Significant variables	Haz. Ratio	Standard Error	95% CI	
senscode	.3730484	.0693049	.2372133	.5088834
age	.0065785	.0022897	.0020908	.0110661
cavit	.1747142	.0303072	.1153131	.2341153

The amount of heterogeneity measured using gamma frailty model is 36.73

The deviance measured using gamma frailty is -2LL = (-2*-699.92992) = 1399.86

Conclusion

For the analysis of survival data using gamma frailty model (from Table 4.2) the heterogeneity observed was 36.73.

The better the model, the less unobserved heterogeneity there will be. Zorn argues that the less heterogeneity in the model, the more appropriate it is to interpret any observed duration dependence in substantive terms.

In the model, the more appropriate it is to interpret any observed duration dependence in substantive terms.

Acknowledgement

My sincere acknowledgement to

Dr. P. Venkatesan Former Scientist "E" & Head,

Department of Statistics, National Institute for Research in Tuberculosis, ICMR, Chennai. e-mail: venkaticmr@gmail.com.

REFERENCE

[1] Aalen.O.O (1988) Heterogeneity in survival analysis. Statistics in Medicine | [2] Andreas wienke, Max-Planck- Institute for demographic Forschung, Frailty models MPIDR working paper WP 2003-032 September 2003. | [3] Clayton, D., Cuzick.J (1985) The semi-parametric pareto model for regression analysis of survival times. Preceedings of the centenary session of the International Statistical Institute, Amsterdam. | [4] Hougaard.P(1995) Frailty models for survival data, Springer, Newyork. | [5] Hougaard, P(2000) Analysis of Multivariate survival data, Springer, New york. | [6] John.P-Klein, Melvin .L. Moeschberger Survival Analysis, Techniques for censored and truncated Data. | [7] Vaupel,J.W., Manton,K.G., Stallard, E.(1979), The impact of Heterogeneity in Individual frailty on the Dynamics of Mortality Demography | [8] Mc. Gilchrist, C.A., Aisbett, C.W.(1991) Regression with frailty in survival analysis Biometrics. | [9] Marubini,E. & Valsechi, M.G1995: Analysing Survival data from Clinical trials and observational study | [10] Swinne C, Cornette. P, Schoevaerdts D, et al,(1998) frailty in the medical literature | [11] Tsiatis, A.A(1981) A large sample study of Cox's Regression Model, Annals of Statistics. |