Automatic Control of Physiological State and Function

Abdulkabir, et al., Automat Control Physiol State Func 2015, 2:2
DOI: 10.4172/2090-5092.1000108

Research Article Open Access

Gradient Curve of Cox Proportional Harzard and Weibull Models

Muritala Abdulkabir[1*], Ahmadu Abdulaziz Oshioke[1], Udokang Anietie Edem[2] and Raji Surajudeen Tunde[2]

[1]*Postgraduate Student University of Ilorin, Ilorin, Nigeria*

[2]*Mathematics and Statistics, Federal Polytechnic, Offa, Kwara State, Nigeria*

[*]**Corresponding author:** Muritala Abdulkabir, Postgraduate Student University of Ilorin, Ilorin, Nigeria, Tel: 234-031-221691-4; Fax: 031-221937; E-mail: kaybeedydx@gmail.com

Received date: August 6, 2015; **Accepted date:** August 18, 2015; **Published date:** August 21, 2015

Abstract

This study centers on the comparison between the gradient curve of the cox proportional hazard and weibull models. It has two faces, the simulated and the real life data approach. The data used for this research work is on tuberculosis diseases for the year 2011. It was observed from the real life data that the shape parameter of the weibull model does not depend or have effect on the performance of the Cox proportional hazard model. It was also observed that both models perform similarly when the distributional assumptions are not met except when sample size is small and the weibull model out-perform the Cox model when the distributional assumption are met and the shape parameter known.

Keywords: Tuberculosis; Survival; Cox proportional hazard model; Weibull model; Parameter

Introduction

Tuberculosis (short for tubercle bacillus), in the past also called phthisis, is a widespread, and in many cases fatal, infectious disease caused by various strains of mycobacterium. It is usually called *Mycobacterium tuberculosis*. Tuberculosis typically attacks the lungs, but can also affect other parts of the body. It is spread through the air when people who have an active TB infection cough, sneeze, or otherwise transmit respiratory fluids through the air. Most infections do not have symptoms, known as latent tuberculosis. About one in ten latent infections eventually progresses to active disease which, if left untreated, kills more than 50% of those so infected.

The classic symptoms of active TB infection are a chronic cough with blood-tinged sputum, fever, night sweats, and weight loss (the latter giving rise to the formerly common term consumption). Infection of other organs causes a wide range of symptoms. Diagnosis of active TB relies on radiology (commonly chest X-rays), as well as microscopic examination and microbiological culture of body fluids. Diagnosis of latent TB relies on the tuberculin skin test (TST) and blood tests. Treatment is difficult and requires administration of multiple antibiotics over a long period of time. Social contacts are also screened and treated if necessary. Antibiotic resistance is a growing problem in multiple drug-resistant tuberculosis (MDR-TB) infections. Prevention relies on screening programs and vaccination with the bacillus Calmette-Guérin vaccine.

In this paper, the main focus will be on the Cox proportional hazard model that depend on the shape parameter of the Weibull model and investigate if there exist an advantage of using a parametric form of the survival distribution (Weibull distribution) instead of the semi parametric Cox proportional hazard model when the parametric form of the model is known.

Simulation studies

The data used for this research is a secondary data collected from university of Ibadan teaching hospital in Ibadan Oyo state. It covers the survival rate of tuberculosis infection in Nigeria for the year 2011. Various sample sizes where considered and the MSE of each sample sizes where replicated 1000 times. For all the simulation works and data analysis R statistical packages was employed.

Exponential and Weibull Distributions

Statisticians chose the exponential distribution to model life data because the statistical methods for it were fairly simple [1]. The exponential density function is

$$f(t)=\lambda \exp\{-\lambda t\},\ for\ \lambda>0\ and\ t>0$$

It has a constant hazard function h(t)=λ and its survival function is S(t)=exp{-λt}

Thus, a large λ implies a high risk and a short survival. Conversely, a small λ indicates a low risk and a long survival. This distribution has the memory less property meaning that how long an individual has survived does not affect its future survival. It is used with ordered data, that is, the first individual to fail is the weakest, the second to fail is the second weakest, and so on [2]. The exponential distribution is limited in applicability because it has only one parameter, the scale parameter λ. By adding a shape parameter the distribution becomes more flexible and can fit more kinds of data. The generalization of the exponential distribution to include the shape parameter is the Weibull distribution. The cumulative distribution function of the Weibull distribution is

$F(t)=1-\exp\left\{-\theta t^{\gamma}\right\},\ t>0$ Where is the shape parameter and is the scale parameter, and the probability density function of the Weibull distribution is $f(t)=\gamma\theta^{\gamma-1}\exp\left\{-\theta t^{\gamma}\right\},\ t>0$

The survival function and hazard function of the Weibull distribution are

$S(t)=\exp\{-\theta t^{\gamma}\}$ *and* $h(t)=\gamma\theta t^{\gamma-1}$ respectively.

It is easy to see just how flexible the Weibull distribution can be. When γ=1, the Weibull distribution becomes the exponential distribution with θ = λ and the hazard rate remains constant as time increases, and when γ=2 it is the Rayleigh distribution. For 3 ≤ γ ≤ 4, it is close to the normal distribution and when γ is large, say γ ≥ 10 it is close to the smallest extreme value distribution [3]. When γ > 1 the hazard rate increases as time increases, and for γ<1 the hazard rate decreases.

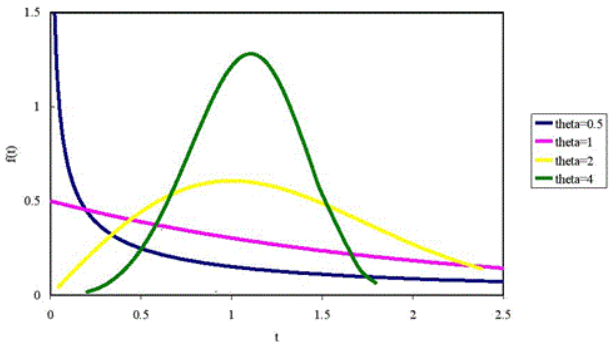

Figure 1: Weibull Distribution for Different Shape Parameters.

Because of the Weibull distribution's flexibility, it is used for many applications including product life and strength/reliability testing. It models the rate of failure as time increases [3]. It can be shown that the mean and standard deviation are

$$E(T)=\left[\frac{1}{\theta}\right]^{\frac{1}{\gamma}}\Gamma\left[1+\frac{1}{\gamma}\right]$$

$$SD=\left[\frac{1}{\theta}\right]^{\frac{1}{\gamma}}\left[\Gamma\left(1+\frac{2}{\gamma}\right)-\Gamma^{2}\left(1+\frac{1}{\gamma}\right)\right]^{\frac{1}{2}}$$

The exponential and weibull distribution are also a models in survival analysis or lifetime analysis.

Cox proportional hazards model

Cox D introduced a model for survival time that allows for covariates but does not impose a parametric form for the distribution of survival times. Specifically he assumed that the survival distribution satisfies the condition [4].

$h(t|x)=h_0(t)\exp\{\beta x\},t>0$

Where x is a covariate, but he made no assumption about the form of h0 (t) which is called the baseline hazard function because it is the value of the hazard function when x = 0.When using a covariate of the form

$$\theta=\exp\{\beta_0+\beta_1 x\}$$

β is incorporated into the baseline hazard function When x is changed, the conditional hazard functions change proportionally with one another. Hazard functions for any pair of different covariate values i and j can be compared using a hazard ratio:

$$HR=\frac{h0(t)\exp\{\beta x_i\}}{h0(t)\exp\{\beta x_j\}}=\exp\left\{\beta(x_i-x_j)\right\}\ for\ i\neq j$$

Hence, the hazard ratio is a constant proportion and the Cox's model is, indeed, a proportional hazards model. This model is used when the covariates have a multiplicative effect on the hazard function and can be extended for multiple regression situations by allowing [5].

$h(t|x)=h_0(t)\exp\{\beta x\}$

where β and x to be vectors. It is mostly used in biostatistics.

Methods and Materials

Data simulation

A simulation study was done to compare the mean square errors of the Weibull maximum likelihood estimate and the Cox proportional hazards model estimate of β=PH-slope when data come from a Weibull distribution.

The data were simulated from a Weibull distribution with survival function

$S(t)=(e^{-t}2)e^{x}$

That is, the model is Weibull with β=1 for the slope of the covariate x, shape parameter γ=2, and baseline survival function $h0(t)=e^{-t}2$. The values of the covariate are x assumed to normally distributed as X~N (N,0,1). The total sample sizes are 15, 45, 90 and 180 with 5, 15, 30 or 60 observations for each value of x. The data were simulated using the fact that the random variable U=F(T) has a uniform distribution where T is a Weibull random variable with cumulative distribution function F(t). For this study, a value of T was obtained at $T=(-\ln(U)\beta e^{-X\beta})^{\frac{1}{\alpha}}$ here U~U(0,1), a=2 is the Weibull shape parameter and β=1 is the Weibull scale parameter.

The uniform random variable was generated using the R random number generator. Data were simulated without censoring and with ten percent random censoring. With random censoring a uniform variable U* was generated independently of U and an observation was denoted as censored if U* ≤ 0

The maximum likelihood estimate of PH-slope using the parametric Weibull model was obtained from SURVREG as $\hat{\beta}=-\hat{\gamma}\hat{\delta}_1$ where $\hat{\gamma}$ is the estimate of the shape parameter and $\hat{\delta}_1$ is the estimate of the slope of the Weibull model as parameterized in SURVREG. Since the shape parameter is known to be 2, an estimate of PH-slope that takes advantage of this fact, $-2\hat{\delta}_1$, was also obtained. The estimate of PH-slope from the Cox proportional hazards model was computed using COXPH.

One-thousand replications of each sample size were run and the mean square estimated as

$$\Sigma_{i=1}^{1000}\frac{(\hat{\beta}_i-\beta)^2}{1000}$$

where β=1. The standard error of the mean square error was computed as the standard deviation of the squared deviations $\left(\hat{\beta}_i-\beta\right)^2$,i=1,2...1000,divided by the square root of 1000. The

distributions of from the maximum likelihood estimates of the Weibull parameters and from the Cox proportional hazards model do not depend on the value of the shape parameter γ. Thus, the mean square errors apply to all Weibull shape parameters.

R Implementation

The parameters in the Weibull model may be estimated in R with the SURVREG procedure which uses the maximum likelihood estimates. The parameters in the Cox proportional hazards model may be estimated with the COXPH procedure which uses a form of a partial likelihood function proposed by Breslow N as the default option [6]. When calculating parameter estimates, it is important to understand that SURVREG and COXPH use different parameterizations. The coefficients that are estimated by the two procedures are not the same, but they are related. COXPH uses the model.

$h(t)=h_0(t)\exp\{\beta x\}$Where h(t) is the hazard function and is the baseline hazard function. Survreg uses the model $G=G^{*}e^{\delta_0+\delta_1 x}$ where G is the survival time and G* is a random variable that has the Weibull survival function $S^{*}(t)=\exp\{-t^{\gamma}\}$

In terms of the survival function, the parameterization of the Weibull model for G is

$$SURVREG: S\left(te^{-\delta 0+\delta 1 x}\right)=e^{-\left(te^{-\delta 0+\delta 1 x}\right)^{\gamma}}=\left(e^{-t^{\gamma}e^{-\gamma\delta 0}}\right)^{e^{-\gamma\delta 1 x}}$$

On the other hand, the parameterization for COXPH gives the following form of the survival function,

$$COXPH: S(t)=\left(e^{-t^{\gamma}e^{-\gamma\delta 0}}\right)^{e}\beta x$$

It follows that the relationship between the parameterizations of the Weibull model for these SURVREG and COXPH is $-\gamma\delta_1=\beta$

If are estimates of the slope and shape parameters from SURVREG and β is the estimate of the slope from COXPH, it follows that are estimates of the same parameter which we call "PH-slope". This chapter shows numerical examples of estimates of PH-slope using real data and compares the mean square errors of estimates of this parameter when estimated by the maximum likelihood method for the Weibull model and the [6] method for the semi-parametric Cox proportional hazards model. Computations are done using SURVREG and COXPH.

Parameter	Estimate	Std error	z-value	Pr(>|z|)
Intercept	3.3837	0.03838	8.82	<0.0001
Age	-0.0166	0.0053	-3.13	0.0017
Sex	0.4249	0.2007	2.12	0.0343
Log(scale)	-0.1968	0.0942	-2.09	0.0367
Scale=0.821 chisq=13.06 df=2 p-value=0.0015				

Table 1: R results from SURVREG.

Covariate	Estimate	Std. error	z-value	Pr(>|z|)
Age	0.019008	0.006524	2.914	0.00357
Sex	-0.410526	0.250585	-1.638	0.10137
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' '				

Table 2: R result COXPH.

Age is a continuous covariateand Censor indicates censoring where Censor=1 is a censored observation. The estimate of the slope and shape parameters in SURVREG are -0.0166, 0.4249 and .821, respectively. Using the relationship above, the estimate of PH-slope from SURVREG is -(0.821)×(-0.0166)=0.0136286, (0.4249)*(-0.821)=-0.3488429. This compares to 0.019008, -0.410526 with standard error 0.006524, 0.006524, which is the estimate of PH-slope from the COXPH procedure [7-9].

The transformed survival time

Parameter	Estimate	Std error	z-value	Pr(>|z|)
Intercept	1.69185	0.019192	8.82	<0.0001
Age	-0.0331	0.00265	-3.13	0.0017
Sex	0.81245	0.10037	2.12	0.0343
Log(scale)	-0.88994	0.09420	-9.45	<0.0001

Scale=0.411 chisq=13.06 df=2 p-value=0.0015

Table3: R results survreg.

Covariate	Estimate	Std. error	z-value	Pr(>|z|)
Age	0.019008	0.006524	2.914	0.00357
Sex	-0.410526	0.250585	-1.638	0.10137
Signif. Codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1				

Table 4: R result coxph.

For illustrative purposes, Table 2 shows the results for SURVREG when analyzing the square root of survival times. If the original data are Weibull, the transformed data will also be Weibull but with different slope and scale parameters. The estimate of the PH-slope using COXPH will not change because the estimate using the Breslow partial likelihood depends only on the order of the observations and the pattern of censoring, not the actual survival times. Therefore the estimate of PH-slope with respect to AGE using the results of SURVREG is - (0.411) × (-0.0331)=0.0136041and that of SEX is - (0.411)*(0.81245)= -0.33391695 which except for rounding is the same as obtained by the analysis of the original survival data, showing that the semi-parametric model does not depend on the shape parameter.

Simulated data

This section displays simulation of both censored and uncensored data (tuberculosis).

Models	N=15	N=45	N=90	N=180
Weibull MLE shape known	0.168(0.028224)	0.288(0.082944)	0.516(0.266256)	0.65(0.4225)
Weibull MLE shape unknown	0.830354(0.689)	0.54374(0.2957)	0.60922(0.3711)	0.7897(0.6237)
Cox-ph slope estimate	0.5804(0.33686)	0.5309(0.28623)	0.6082(0.37021)	0.7816(0.61920)

Table 5: MSEs and Standard Errors for Complete samples.

Table 5 has the means square errors and the standard error for the complete sample case. Here it can be seen that when the shape parameter is unknown, the estimates of the Cox proportion hazards model and the maximum likelihood estimates of the Weibull model perform almost similarly, but when the shape parameter is knowni.e-2δ, the estimate far out-performs the Cox proportional hazards model. From this, we can advise that when the distributional assumptions are not known, or are not met, the Cox proportional hazards model should be considered keeping in mind that the weibull model when the distributional assumptions are not met stand a good chance as well.

NOTE: As the sample size increases from N=180 to N=450 the MSE's for maximum likelihood estimate of the Weibull is approximately the same as that of the Cox proportional hazard models.

N=450	
Weibull MLE shape known	0.568(0.322624)
Weibull MLE shape unknown	0.96459(0.93043)
Cox-pH slope estimate	0.95967(0.91423)

Table 6: MSE and Standard Error of sample size N=450.

Censored data

MSEs and Standard Errors for Complete samples				
	N=15	N=45	N=90	N=180
Weibull MLE shape known	0.094(0.008836)	0.514(0.264196)	0.9118(0.83137)	0.676(0.456976)
Weibull MLE shape unknown	0.752583(0.566)	0.98227(0.9649)	0.9699(0.94071)	0.8819(0.77775)

Cox-pH slope estimate	0.2597(0.06744)	0.7118(0.50666)	0.95943(0.9205)	0.83981(0.7053)

Table 7: Result for censored data.

Results for the censored sample case are shown in Table 7. The patterns are similar to the uncensored sample case. The MSEs are smaller for the maximum likelihood estimates and the proportional hazards model estimates when the shape parameter is known, but much bigger for the maximum likelihood estimates of the Weibull model when the shape parameter is unknown. The MSEs for censored data are larger than uncensored data in most of the scenario, but not appreciably so, except in one notable case. The small sample case, N=15, the Cox PH model occasionally produces unusual estimates, sometimes very large, in both the uncensored and censored cases yielding inconsistent MSE calculations. This problem is exacerbated in the presence of censored data, but is not present in either case for larger sample sizes. Although the Cox model is generally comparable to the Weibull model, perhaps it is not for small sample sizes. As the sample size increases Cox proportional hazard model tends to be better (smaller MSEs) than weibull model when the shape parameter is unknown. These suggest that for censored data, Cox proportional hazard model should be preferred over weibull model when the distributional assumptions are not met.

Discussion

The real life data were classified into original and transformed data. Table 1 showed the estimate of survreg (estimate of weibull model in R) for both Age and Sex likewise the shape parameter of the model. Table 2 gives the estimate of the cox proportional hazard model for age and sex. Then from the PH-slope relationship given in figure 1, where , the slope of the cox proportional hazard model in table4 was compared to the slope multiply by the shape parameter of the parametric weibull model. The same procedure was also performed for the case of the transformed data. It was seen that the PH-slope estimate for both case (original and transformed data) are almost similar except for rounding up.

The simulated data was also classified into censored and uncensored data. Table 5 Showed means square error and the standard error of the censored data for the simulated data of various sample sizes i.e. N=15, N=45, N=90, and N=180.The maximum likelihood estimate of the weibull model (when the shape parameter is known) out-perform the cox proportional hazard model. But when the shape parameter of the maximum likelihood estimate of the weibull is know they perform almost similarly except for small sample i.e. N=15. It was observed that the parametric weibull (shape parameter unknown) model is compared to cox proportional hazard model but not for small sample case and as the sample sizes increases, they both can be used interchangeably. The censored data/sample simulation case in Table9 showed the means square error of both models. The weibull model with shape know perform better (smaller MSEs) than the cox the cox proportional hazard in all the scenarios while when the shape parameter is know the cox proportional hazard model out-perform the weibull model except when the sample sizes is small N=15.

Conclusions

Based on the result of the analysis, the Weibull model is a better option for analyzing lifetime data if the distributional assumptions can be met and the shape parameter is known. The mean square errors are smallest in this case. However, when the shape parameter is unknown for censored data, the Cox proportional hazards model is a good alternative. But for uncensored data when the distributional assumptions are not met and shape parameter unknown, both models can be used interchangeably. It requires fewer assumptions than the parametric Weibull model provides comparable mean square errors of the estimates of PH-slope. There may be a concern for smaller samples with the Cox proportional hazards model depending on the particular data set being analyzed. Thus the shape parameter of the weibull model those not depends or have effect on the performance of the proportional hazard model.

References

1. Lawless JF (2003) Statistical Models and Methods for Lifetime Data. Wiley, New York.
2. Epstein B, Sobel M (1953) Life Testing. Journal of the American Statistical Association 48: 486-502.
3. Nelson W (1982) Applied Life Data Analysis. John Wiley & Sons, Inc. New York.
4. Cox D (1972) Regression Models and Life Tables. Journal of the Royal Statistical Society 34: 187-220.
5. Angella MC (2008) Comparison between Cox Proportional Hazard and Weibull Models. Kansas State University.
6. Breslow N (1974) Covariance Analysis of Censored survival Data. Biometrics 30: 89-99.
7. R Core Team (2014) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.
8. Wikipedia (2014) Tuberculosis Disease.
9. Zhou, Mai (2000) Understanding the Cox Model with Time-Change Covariates. The American Statistician 55: 153-155.