The Nepali Math. Sc. Report
Vol. 20 No. 1&2, 2002

Cox's Regression Model for the Growth of a Tumour Followed by Death from that Tumour

GURPRIT GROVER AND NEETA MAKHIJA

Abstract: In this paper we introduce Cox's Regression Model to study the growth process of a tumour and death from that tumour. The hazard rate is taken to be a function of the explanatory variables and unknown regression coefficients multiplied by an arbitrary and unknown function of time. Formulas for the density function, the distribution function and the expectation of lifetime are outlined. An equivalent reformulation of the model is also given in terms of the intensities of counting processes. Estimation procedures in the model are discussed and large sample properties of the estimators are outlined.

1. Introduction

A tumour is a swelling of any kind in any par of the body. Tumours occur when the cells of a tissue or organ multiply in an uncontrolled fashion unrelated to the biological requirements of the body and not to meet the needs of repair or of normal replacement. The growth of tumour in man is supposed to depend upon an interplay between factors in the environment and the genetic component of body cells. When an individual is exposed to one or more environmental factors (or carcinogenic agents) such as different occupations, lifestyles, exposure to injurious chemical agents, drugs, ultraviolet light, certain tumour inducing viruses etc., the body's metabolic and other biological reactions render most of the absorbed molecules inactive and are excreted from the body system, but a few metabolites remain and become carcinogenic which are responsible for the growth of tumour in man. The potential carcinogenic metabolites bind with DNA in cells to cause mutation of DNA and result in tumour. The formation of tumour is a manifestation of this continuous biological process that depends on the amount of toxic material retained in the body.

The appearance of tumour cannot always be detected, either because the tumour has not yet reached the detectable size, or because it is hidden in the body. The retained amount of toxic material in the body becomes carcinogenic only after a specified period of time which may be months or years. The shortest period of time is known as the Latent Period. Suppose this period is denoted by $[0, t_0]$, where t_0 is the length of this time.

For every $t > t_0$, the amount of toxic material absorbed during the interval $[0, t - t_0]$, will be carcinogenic at time t and the amount absorbed during the interval $[t - t_0, t]$ will be potentially carcinogenic.

The growth process of tumour can be explained by using Cox's Regression Model. Cox'x Regression model, based on the method of 'Partial Likelihood', for analysing censored survival data allowing for covariates, is beautifully adapted to the kind of data obtained in clinical cancer trials. By incorporating time-varying random covariates, it becomes a highly flexible tool for model building. The hazard rate at any point of time does not depend on time only but also on a host of explanatory variables or covariates, some of which may not be expressed in quantitative form.

2. Development of the Model

Suppose $T_i, i = 1,2, \ldots, n,$ are independent continuously distributed positive random variables representing the times of death of n individuals suffering from tumour, each of whom is observed for a fixed time interval $[0, c_i]$ for certain censoring times $c_i, i = 1,2, \ldots, n$. Suppose that individual i has hazard rate.

(2.1) $$\lambda_i(t) = \lim_{dt \to 0} \frac{1}{dt} P\left[T_i \le t + dt \mid T_i \ge t\right]$$;

which can be written as

(2.2) $$\lambda_i(t) = \lambda_0(t)\, e^{\beta' z_i(t)}$$

where

$\lambda_i(t) =$ hazard rate for the i^{th} individual at time t which is a function of overall hazard rate $\lambda_0(t)$, a function of time t only irrespective of other covariates, and hazard rate $e^{\beta' z_i(t)}$ with respect to p factors having corresponding time independent intensities $(\beta_1, \beta_2, \ldots, \beta_p)$ with which the i^{th} individual will be affected at time t. β' is the transpose of the column vector β of p unknown coefficients and $\underset{\sim}{z_i}$ is a column vector of p time-dependent covariates.

Let

$$R(t) = \{i : T_i \ge t, c_i \ge t\}$$

denote the risk set at time t, i.e., the set of individuals i under observation at time t. Given $R(t)$ and that at time t one individual in $R(t)$ is observed to die, the probability that it is individual i or the probability of death of the i^{th} individual at time t when his censoring time is c_i can be calculated as:

$$\frac{\lambda_0(t)e^{\beta' z(t)}}{\sum_{i\in R(t)} \lambda_0(t)e^{\beta' z(t)}}$$

Thus, the likelihood function is given by

$$L(\beta) = \prod_{T_i \le c_i} \prod_{i=1}^{n} \frac{\exp[\beta' z_i(t)]}{\sum_{j\in R(t)} \exp[\beta' z_i(t)]}$$

This likelihood function, according to Cox (1975) was an example of Partial Likelihood

3. Distribution of Time to Death

Since T_i is a random variable representing the lifetime of the i^{th} individual, who is observed for a fixed time interval $[0, c_i]$, therefore, for every t, $0 \le t \le c_i$, the hazard rate at time t is given by (2.2), i.e.,

$$\lambda_i(t) = \lambda_0(t)e^{\beta' z_i(t)}$$

The cumulative hazard function is given by

$$\int_0^t \lambda_i(x)\,dx = \int_0^t \lambda_0(x)\,e^{\beta z_i(x)}dx \;;\; 0 \le t \le c_i, \tag{3.1}$$

The survival function is given by

$$S(t) = P\text{ (of survival of } i^{\text{th}} \text{ individual atleast time } t)$$

$$= \exp\left\{-\int_0^t \lambda_i(x)\,dx\right\} \;;\; 0 \le t \le c_i \tag{3.2}$$

$$\Rightarrow \; S(t) = \left\{-\int_0^t \lambda_0(x)\,e^{\beta' z_i(x)}\,dx\right\} \;;\; 0 \le t \le c_i \tag{3.3}$$

The corresponding distribution function and the density function of the lifetime T_i of the i^{th} individual can be obtained from the following expressions:

$$F(t) = 1 - S(t)$$

$$\Rightarrow \; F(t) = 1 - \exp\left\{-\int_0^t \lambda_0(x)\,e^{\beta z_i(x)}\,dx\right\} \;;\; 0 \le t \le c_i \tag{3.5}$$

and

$$f(t) = -\frac{d}{dt}\left[\exp\left(-\int_0^t \lambda_0(x)\, e^{\beta z_i(x)}\, dx\right)\right] \; ; \; 0 \le t \le c_i \tag{3.4}$$

The expectation of lifetime of i^{th} individual, T_i for $T_i \le c_i$, can be obtained by using the following expression

$$E\,[T_i \mid T_i \le c_i] = \left(\int_0^{c_i} t\, f(t)\, dt\right)[F(c_i)]^{-1} \tag{3.6}$$

4. Formulation of Cox-Regression Model Based on Counting Process Theory

Cox's Regression Model can be reformulated as a model for the random intensity of a multivariate counting process. The original hazard rate definition of Cox's model can be interpreted as specifying the stochastic intensity of multivariate counting process (counting occurrences of the event "death" for each of the individuals under observation).

The observation of the i^{th} individual may be considered as the observation of a counting process.

$$N_i = \{N_i(t)\,;\, t \ge 0\}$$

where N_i counts 1 if death is observed in the i^{th} individual, otherwise zero, i.e.,

$$N_i = \begin{cases} 1, & \text{if } T_i \le t,\ T_i \le c_i \\ 0, & \text{otherwise} \end{cases} \tag{4.1}$$

The counting process N_i has a random intensity process.

$$\Delta_i = \{\Delta_i(t)\,;\, t \ge 0\}$$

defined by

$\Delta_i(t)dt = P$ {death is observed in the i^{th} individual in a time interval of length dt around time $t \mid F_t$.}

$$\Delta_i(t) = \lim_{dt \to 0} \frac{1}{dt}\, P\{N_i(t+dt) - N_i(dt) = 1 \mid F_t\} \tag{4.2}$$

where

F_t denotes the past upto the beginning of the small interval dt, i.e., everything that has happened until just before time t.

Now, given what has happened before the time interval dt, if the i^{th} individual dies at the observed time $T_i \le t$ and $T_i \le c_i$, or if the i^{th} individual was censored at time $c_i < t$ then

$$P(\text{of death of the } i^{th} \text{ individual in the interval } dt) = 0$$

But, if the i^{th} individual is still alive and uncensored, $T_i \in dt$ or $T_i \geq t$, then by (2.1)

$$P(\text{of death of the } i^{th} \text{ individual in the interval } dt) = \lambda_i(t)dt$$

Define

$$(4.3) \qquad Y_i(t) = \begin{cases} 1, & \text{if individual } i \text{ is under observation just before time } t \\ 0, & \text{otherwise} \end{cases}$$

From (2.2) and (4.2) we get

$$\Delta_i(t)dt = Y_i(t)\,\lambda_i(t)\,dt$$

$$(4.4) \qquad \Rightarrow \quad \Delta_i(t)\,dt = Y_i(t)\,\lambda_0(t)\,e^{\beta' z_i(t)}\,dt$$

Given the past upto (but not including) time t, $Y_i(t)$ & $\Delta_i(t)$ are fixed or non random. Thus Y_i and Δ_i are predictable.

Now, N_i is a simple multivariate counting process, each component of which jumps utmost once, with intensity process Δ_i, satisfying

$$(4.5) \qquad \Delta_i(t)\,dt = Y_i(t)\,\lambda_0(t)\,e^{\beta' z_i(t)}\,dt$$

where

The fixed covariate $z_i(t)$ is replaced by a random covariate $Z_i(t)$. N_i, Y_i and Z_i are processes that can be observed and Y_i and Z_i are predictible since $Y_i(t)$ and $Z_i(t)$ are fixed given what has happened before time t. Thus, the Cox's Model is reformulated by

$$(4.6) \qquad L(\beta) = \prod_{t \geq 0} \prod_{i=1}^{n} \left(\frac{Y_i(t) \exp(\beta' Z_i(t))}{\sum_{j=1}^{n} Y_j(t) \exp\{\beta' Z_j(t)\}} \right)^{dN_i(t)}$$

where $dN_i(t)$ is the increment of N_i over a small interval dt around the time t and the product over t is a product over disjoint intervals. So (4.6) reduces to a finite product over all i and t for which N_i jumps at time t, i.e., death is observed in the ith individual at time $t(dN_i(t) = 1)$; elsewhere $dN_i(t) = 0$. Now, (4.6) is the original likelihood function for β_0 which is the true value β. Let $\hat{\beta}$ be the value of β maximising $L(\beta)$. Define $L(\beta,u)$ as the likelihood function for β based on the observations on the time interval $[0,u]$, in which the product over $t \geq 0$ in (4.6) is replaced by a product $0 \leq t \leq u$. From (4.6) we have

$$(4.7) \qquad L(\beta_0, u) = L(\beta_0) = \prod_{0 \leq t \leq u} \prod_{i=1}^{n} \left(\frac{Y_i(t) \exp(\beta_0' Z_i(t))}{\sum_{j=1}^{n} Y_j(t) \exp\{\beta_0' Z_j'(t)\}} \right)^{dN_i(t)}$$

5. Large-Sample Properties of $\hat{\beta}$

Taking logarithm of the Cox's likelihood (4.7) we get

$$\log L(\beta_0,u)=\sum_{i=1}^{n}\sum_{t\le u} dN_i(t)\log\left(\frac{Y_i(t)\exp(\beta_0' Z_i(t))}{\sum_{j=1}^{n} Y_j(t)\exp\{\beta_0' Z_j(t)\}}\right) \tag{5.1}$$

$$\Rightarrow \log L(\beta_0,u)=\sum_{i=1}^{n}\sum_{t\le u} dN_i(t)\left[\log\left(Y_i(t)\exp(\beta_0' Z_j(t))\right)-\log\left(\sum_{j=1}^{n} Y_j(t)\exp\{\beta_0' Z_j'(t)\}\right)\right]$$

Now,

$$D\log L(\beta_0,u)=\sum_{i=1}^{n}\sum_{t\le u} dN_i(t)\left[\frac{Y_i(t)Z_i(t)\,e^{\beta_0' Z_i(t)}}{Y_i(t)\,e^{\beta_0' Z_i(t)}}-\frac{\sum_{j=1}^{n} Y_j(t)Z_j(t)\exp\{\beta_0' Z_j(t)\}}{\sum_{j=1}^{n} Y_j(t)\exp\{\beta_0' Z_j(t)\}}\right] \tag{5.2}$$

where Dlog $L(\beta)$ denotes the vector of partial derivatives $\frac{\partial}{\partial\beta}\log L(\beta)$ evaluated at β.

$$n^{-1/2}\,D\log L(\beta_0)=n^{-1/2}\sum_{i=1}^{n}\sum_{t\le u} dN_i(t)\left[Z_i(t)-\frac{\sum_{j=1}^{n} Y_j(t)Z_j(t)\,e^{\beta_0 Z_j(t)}}{\sum_{j=1}^{n} Y_j(t)\,e^{\beta_0 Z_i(t)}}\right]$$

$$=n^{-1/2}\sum_{i=1}^{n}\sum_{t\le u}[Z_i(t)-E_0(t)]\,dN_i(t)$$

$$\Rightarrow n^{-1/2}\,D\log L(\beta_0)=n^{-1/2}\sum_{i=1}^{n}\int_{t=0}^{u}[Z_i(t)-E_0(t)]\,dN_i(t) \tag{5.3}$$

where,

(5.4)
$$E_0(t) = \frac{\sum_{j=1}^{n} Y_j(t) Z_j(t) \exp[\beta_0' Z_j(t)]}{\sum_{j=1}^{n} Y_j(t) \exp[\beta_0' Z_j(t)]}$$

$$\Rightarrow \quad n^{-1/2}\ D\log L(\beta_0, u) = \sum_{i=1}^{n} \int_{t=0}^{u} n^{-1/2}\,[Z_i(t) - E_0(t)]\, dN_i(t)$$

Now, since $N_i(t)$ is a counting process with the corresponding intensities $\Delta_i(t)$, therefore, $M_i(t)$ defined as

(5.6)
$$M_i(t) = N_i(t) - \int_0^1 \Delta_i(\tau)\, d\tau$$

is a Martingale.

(5.7)
$$\Rightarrow \quad n^{-1/2}\ D\log L(\beta_0, u) = \sum_{i=1}^{n} \int_{t=0}^{u} n^{-1/2}\,[Z_i(t) - E_0(t)]\, dM_i(t)$$

since

$$dM_i(t) = dN_i(t) - \Delta_i(t)\, dt$$

and

$$\begin{aligned}
\sum_{i=1}^{n} [Z_i(t) - E_0(t)]\, \Delta_i(t) &= \sum_{i=1}^{n} Z_i(t)\, Y_i(t) \lambda_0(t) \exp\left(\beta_0' Z_i(t)\right) - E_0(t) \sum_{i=1}^{n} Y_i(t) \lambda_0(t) \exp\left(\beta_0' Z_i(t)\right) \\
&= \sum_{i=1}^{n} Z_i(t)\, Y_i(t) \lambda_0(t) \exp\left(\beta_0' Z_i(t)\right) - \sum_{j=1}^{n} Z_j(t) Y_j(t) \lambda_0(t) \exp\left(\beta_0' Z_j(t)\right) \\
&= 0
\end{aligned}$$

Now, $n^{-1/2}\,[Z_i(t) - E_0(t)]$ is a vector of predictable processes and it only depends on the fixed parameter β_0 and the predictable processes Y_j and Z_j, $j = 1, 2, \ldots, n$. Therefore, by Martingale Transform Theorem, $M^{(n)}(t) = n^{-1/2}\ D\log L(\beta_0, t)$, considered as a stochastic process in t, is the sum of n (vector) martingales, hence also a martingale.

Now, expanding Dlog $L(\beta, u)$ around β_0, using Taylor's expansion, we get

(5.8)
$$D\log L(\beta, u) = D\log L(\beta_0, u) - I(\beta^*, u)\ (\beta - \beta_0)$$

where β^* is on the line segment between β and β_0, and the positive semidefinite matrix $I(\beta, u)$ is minus the second derivative of $\log L(\beta, u)$ with respect to β i.e.,

$$I(\beta, u) = \frac{-\partial^2 \log L(\beta, u)}{\partial \beta^2} = \frac{-\partial}{\partial \beta} D \log L(\beta, u)$$

Then, $I(\beta, u) = \int_0^u \times$

$$\times \left[\frac{\sum_{j=1}^{n} Y_j(t) e^{\beta' Z_j(t)} \sum_{j=1}^{n} Y_j(t) Z_j(t)^{\otimes 2} e^{\beta' Z_j(t)} - \sum_{j=1}^{n} Y_j(t) Z_j(t) e^{\beta' Z_j(t)} \sum_{j=1}^{n} Y_j(t) Z_j(t) e^{\beta' Z_j(t)}}{\left(\sum_{j=1}^{n} Y_j(t)\, e^{\beta' Z_j(t)} \right)^{\otimes 2}} \right]$$

$$\times\, d\overline{N}(t)$$

where for a column vector 'a' the matrix aa' is denoted by $a^{\otimes 2}$ or $a \otimes a$ where $a = (a_1, a_2, \ldots, a_n)'$ is a p-vector and $\overline{N} = \sum_{i=1}^{n} N_i$

(5.9) $$\Rightarrow I(\beta, u) = \int_0^u \left[\frac{\sum_{j=1}^{n} Y_j(t) Z_j(t)^{\otimes 2}\, e^{\beta' Z_j(t)}}{\sum_{j=1}^{n} Y_j(t)\, e^{\beta' Z_j(t)}} - \left(\frac{\sum_{j=1}^{n} Y_j(t) Z_j(t)\, e^{\beta' Z_j(t)}}{\sum_{j=1}^{n} Y_j(t)\, e^{\beta' Z_j(t)}} \right)^{\otimes 2} \right] d\overline{N}(t)$$

Now, from (5.8)

$$D\log L(\beta, u) - D \log L(\beta_0, u) = -\, I(\beta^*, u)\ (\hat{\beta} - \beta_0).$$

Inserting $\hat{\beta}$ we get

(5.10) $$D \log L(\hat{\beta}, u) - D \log L(\beta_0, u) = -\, I(\beta^*, u)\ L(\hat{\beta}, \beta_0)$$

But $$D \log L(\hat{\beta}, u) = 0$$

since $\hat{\beta}$ is the solution of the likelihood equation $\frac{\partial}{\partial \beta} \log L(B, u) = 0$

$$\Rightarrow \quad D\log L(\beta_0, u) = I(\beta^*, u)\ (\hat{\beta} - \beta_0)$$

$$\Rightarrow \quad n^{-1/2} D \log L(\beta_0, u) = \{n^{-1} I(\beta^*, u)\} n^{1/2} (\hat{\beta} - \beta_0).$$

Now, to prove the asymptotic normality of $n^{1/2}(\hat{\beta}_0, \beta_0)$ we have to prove that the martingale $n^{-1/2}\, D \log L(\beta_0, u)$ weakly converges to a Gaussian Process and $n^{-1} I(\beta^*, u)$ converges in probability to a non-singular matrix. This can be derived by using the Martingale Central Limit Theorem (Rebolledo, 1980) and the counting process approach of Andersen and Gill (1982) with some regularity conditions.

Thus, $\hat{\beta}$ is consistent and asymptotically normally distributed as $n \to \infty$ under mild regularity conditions on the covariate processes and the variance of $\hat{\beta}$ can be estimated consistently from the second derivative $L(\beta)$ evaluate at $\hat{\beta}$.

Consistency and asymptotic normality of the maximum partial likelihood estimates of the regression parameters can also be established by using the weak convergence results as suggested by Tsiatis (1981).

6. Conclusion

The purpose of this paper is to develop a model for the study of growth of a tumour followed by death from that tumour based on partial likelihood method of Cox. It has been shown that how this model can be developed further by using counting process theory. Distribution of time to death has also been outlined. The model also permits the estimation of parameters affecting the growth process of tumour. Consistency and asymptotic normality for the maximum partial likelihood estimate of the regression parameter in Cox's Regression Model can be established by using Martingales techniques suggested by Andersen and Gill (1982) or by using weak convergence results suggested by Tsiatis (1981).

REFERENCE

[1] Andersen, P.K., and Gill, R.D., (1982), *"Cox's Regression Model for Counting Processes: A Large Sample Study"*, Annals of Statistics, 10, 1100–1120.

[2] Biswas, S., and Thomas, M., (1993), *On a Martingale based approach for the estimation of the parameters affecting the growth of tumour*. Sankhya, Series-B, Vol. 55, Part-2, page 219–228.

[3] Chiang, C.L., and Conforti, P.M., (1989), *"A survival model and the estimation of time to tumour. Mathematical Bioscience*, 94, 1–29.

[4] Cox, D.R., (1975), *Partial Likelihood*, Biometrika, 62, 269–276.

[5] Gill, R.D., (1984), *Understanding Cox's Regression Model, A Matingale approach*, Journal of American Statistical Association. Vol. 70, Page 441–447.

[6] Oakes, D., (1981), *Survival Times. aspects of partial likelihood* (with discussion), Int. Statist. Rev. 49, 235–264.

[7] Tsiatis. A.A., (1981a), *A Large Sample Study of Cox's Regression Model Annals of Statistics* 9, 93–108.

GURPRIT GROVER
Department of Statistics
University of Delhi

NEETA MAKHIJA
Department of Statistics
University of Delhi