Understanding the Cox Regression Models with Time-Change Covariates

Mai Zhou
University of Kentucky

The Cox regression model is a cornerstone of modern survival analysis and is widely used in many other fields as well. But the Cox models with time-change covariates are not easy to understand or visualize.

We therefore offer a simple and easy-to-understand interpretation of the (arbitrary) baseline hazard and time-change covariate. This interpretation also provides a way to simulate variables that follow a Cox model with arbitrary baseline hazard and time-change covariate. Splus/R codes to generate/fit various Cox models are included. Frailty model is also included.

KEY WORDS: Arbitrary baseline hazard; Splus code; Simulations.

The Cox regression model is invariably difficult for students to grasp, partly because it is so different from the classical linear regression models. The added concept of time-change covariates further increases the difficulty.

After several years of teaching a master's level survival analysis course, we have settled on a teaching approach that uses exponential distributions in conjunction with a transformation to the Cox model. This approach is not found in the current text books and has proven to be the most beneficial in enabling the students to grasp the ideas quickly. For students that are increasingly computer-savvy, the codes we included in this paper helps them understood what are the various Cox models; including Cox model with time-change covariate, stratified Cox model and frailty model.

1. Cox regression model with a fixed covariate

We approach this topic by asking "How can one simulate variables that follow a Cox model?" We start with an easy-to-understand special case: the parametric exponential regression model.

In the exponential regression model, every outcome is exponentially distributed except the rates are different, i.e. for $i = 1, 2, \cdots, n$ the survival time of the ith subject, Y_i, follows an exponential distribution with a (subject-specific) parameter $\lambda_i = \exp(\beta z_i)$:

$$P(Y_i \le t) = 1 - \exp(-\lambda_i t) \quad \text{with} \quad \lambda_i = \exp(\beta z_i), \tag{1.1}$$

where z_i is the covariate and β is the parameter.

The linear form βz_i is borrowed from the ordinary linear regression model and the $\exp(\cdot)$ function is introduced to assure the parameter values stay positive (since the exponential distribution must have a positive parameter λ_i). This exponential regression model is relatively easy to understand and serves as our starting point. The Splus code to generate those random variables is appended at the end of this paper.

Next we introduce an **arbitrary** monotone increasing transformation $g(\cdot)$ with

$$g(0) = 0, \quad \text{and} \quad g(t) \nearrow \quad \text{for} \quad t > 0. \tag{1.2}$$

From the above we notice that $g^{-1}(t)$ is also well defined and $\nearrow$.

Now we **claim:** The Cox model with an arbitrary baseline hazard function can be obtained by replacing the $Y_i's$ above with $g(Y_i)'s$.

A few remarks are in order.

First, since $g(\cdot)$ is monotone increasing, the ranks of the $Y_i's$ are the same as those of the $g(Y_i)'s$. Therefore any statistical procedure that uses only the ranks of the $Y_i's$ can still operate on the $g(Y_i)'s$ as if no transformation had taken place. The Cox partial likelihood function is such a function that depends only on the ranks of the $Y_i's$.

Second, if Y is an exponential random variable then $g(Y)$ no longer has the memoryless property (unless $g(t) = c\cdot t$). It could have either positive or negative memory. If g is a power function, then $g(Y)$ has a Weibull distribution. We argue that the distributions obtained by $g(Y)$ are more flexible and more suitable for modeling a wider variety of survival times.

Before we prove the claim, we shall recast the exponential regression model (1.1) in terms of hazards and define a Cox model.

For any distribution function $F(t)$ with density $f(t)$, the corresponding hazard function $h(t)$ and cumulative hazard function $H(t)$ are defined as

$$h(t) = \frac{f(t)}{1 - F(t)}\ , \qquad H(t) = \int_{-\infty}^{t} h(s) ds.$$

One useful formula is

$$1 - F(t) = \exp(-H(t)). \tag{1.3}$$

It is readily verified that $Y_i \sim \exp(\lambda_i)$ if and only if the hazard function of Y_i is equal to λ_i. Therefore an equivalent way to formulate model (1.1) is

$$h_{Y_i}(t) = \text{the hazard function of } Y_i = \lambda_i = \exp(\beta z_i) \ ,$$

or

$$H_{Y_i}(t) = \text{the cumulative hazard function of } Y_i = t \times \lambda_i = t \times \exp(\beta z_i) \ .$$

The Cox regression model can now be defined as

$$h_{Y_i}(t) = \text{the hazard function of } Y_i = h_0(t) \times \lambda_i = h_0(t) \times \exp(\beta z_i) \ ,$$

or

$$H_{Y_i}(t) = \text{the cumulative hazard function of } Y_i = H_0(t) \times \lambda_i = H_0(t) \times \exp(\beta z_i) \ ,$$

where $h_0(t)$ ($H_0(t)$) is an arbitrary, unspecified baseline (cumulative) hazard function.

Theorem 1 If the random variables Y_i follow exponential distributions with parameters $\lambda_i = \exp(\beta z_i), i = 1, 2, \cdots, n$ as in (1.1), then with $g(\cdot)$ as defined in (1.2), the $g(Y_i)$'s follow a Cox proportional hazards model with baseline cumulative hazard function $g^{-1}(t)$.

PROOF: Since $P(Y_i > t) = \exp(-\lambda_i \times t)$, the survival function of $g(Y_i)$ is

$$P(g(Y_i) > t) = P(Y_i > g^{-1}(t)) = \exp(-\lambda_i \times g^{-1}(t)).$$

Therefore the cumulative hazard function of $g(Y_i)$ is, according to (1.3),

$$\lambda_i \times g^{-1}(t) = \exp(\beta z_i) \times g^{-1}(t).$$

This is a Cox model with $H_0(t) = g^{-1}(t)$. ♢

The crazy clock interpretation: The above discussion also offers another way of interpreting the Cox model using a variable time clock: different subjects originally are following different exponential distributions (that is, with different $\lambda_i's$). However, the common time clock (the transformation $g^{-1}(t)$) is not a "linear clock" (one with constant speed) but only a monotone clock with changing speed. The clock sometimes goes faster and sometimes slower, resulting in the Cox model.

As an example we consider the case of a Weibull regression model. The Weibull regression model is $Y_i^* \sim$ Weibull with shape parameter α and scale parameter $\lambda_i = \exp(\beta z_i)$, i.e.

$$h_{Y_i^*}(t) = \alpha t^{\alpha-1} \exp(\beta z_i) \ .$$

This is a special case of the Cox model with $h_0(t) = \alpha t^{\alpha-1}$. We can obtain Y_i^* by first generating exponential variables Y_i that follow (1.1), and then letting $Y_i^* = (Y_i)^{1/\alpha}$, i.e. using $g(t) = (t)^{1/\alpha}$ as in Theorem 1. It is readily verified that $d/dt[g^{-1}(t)] = h_0(t) = \alpha t^{\alpha-1}$. The advantage of a general Cox model is, of course, its flexibility in that the transformation $g(\cdot)$ need not be specified except to satisfy (1.2).

2. Cox model with a time-change covariate

We discuss only Cox models with covariates that change in time as step functions. This simplifies the model but is general enough according to Therneau (1999, p.18). In practice the measurement of the covariate is usually taken at intervals, hence the step function.

Similar to the previous section, here a regression model of piecewise exponential distributions becomes a Cox model with time-change covariates, after a monotone transformation $g(\cdot)$.

We illustrate the idea using only one possible time-change point t_{i0} for each covariate. Several time-change points can be dealt with similarly but the notation becomes tedious.

Piecewise exponential regression model: two pieces

Suppose we generate a random variable Y_i that follows a 2-piece exponential distribution, i.e. (see eg. Barlow and Proschan 1981, p.87)

$$\text{hazard of } Y_i = \begin{cases} C_{i1} & \text{for } t \le g^{-1}(t_{i0}); \\ C_{i2} & \text{for } t > g^{-1}(t_{i0}). \end{cases} \tag{2.1}$$

(The cut-off point is so chosen to make it look simpler after transformation.) See below for an Splus code to generate random numbers that follow 2-piece exponential distributions.

Introducing a time-fixed covariate z_{i1} and parameters β_1, β_2, we rewrite the positive constants C_{i1}, C_{i2} into a regression model using exponentials (as in (1.1))

$$C_{i1} = \exp[\beta_1 z_{i1}],$$

$$C_{i2} = \exp[\beta_1 z_{i1} + \beta_2], \tag{2.2}$$

where z_{i1} is the (fixed) covariate for Y_i.

If we introduce a second, time-change covariate

$$z_{i2}(t) = \begin{cases} 0 & \text{for } t \le t_{i0}; \\ 1 & \text{for } t > t_{i0}, \end{cases}$$

then

$$\exp[\beta_1 z_{i1} + \beta_2 z_{i2}(t)] = \begin{cases} C_{i1} & \text{for } t \le t_{i0} \\ C_{i2} & \text{otherwise.} \end{cases} \tag{2.3}$$

Theorem 2 Suppose Y_i, $i = 1, 2, \cdots, n$ are random variables with piecewise exponential distributions as in (2.1)-(2.2). Further suppose $g(\cdot)$ is a monotone increasing function as in (1.2) such that $g^{-1}(t)$ is differentiable. Then $g(Y_i)$ follows a Cox model with a time-change covariate and a baseline hazard $h_0(t) = d/dt[g^{-1}(t)]$.

PROOF: We compute

$$P(g(Y_i) > t) = P(Y_i > g^{-1}(t)) \ .$$

Since Y_i is piecewise exponential, the above probability is

$$P(Y_i > g^{-1}(t)) = \begin{cases} \exp[-C_{i1} g^{-1}(t)] & \text{if } g^{-1}(t) \le g^{-1}(t_{i0}) \ ; \\ \exp[-C_{i1} g^{-1}(t_{i0}) - C_{i2}(g^{-1}(t) - g^{-1}(t_{i0}))] & \text{otherwise.} \end{cases}$$

If we let $h_0(t) = d/dt[g^{-1}(t)]$ then the hazard function of $g(Y_i)$ is seen from above to be

$$\text{hazard of } g(Y_i) = \begin{cases} C_{i1} \times h_0(t) & \text{for } t \le t_{i0} \\ C_{i2} \times h_0(t) & \text{otherwise.} \end{cases}$$

In view of (2.3), we have

$$\text{hazard of } g(Y_i) = h_0(t) \times \exp[\beta_1 z_{i1} + \beta_2 z_{i2}(t)] \ .$$

Therefore $g(Y_i)$ follows a Cox proportional hazards model with baseline hazard $h_0(t)$, one time-change covariate $z_{i2}(t)$, and one time-fixed covariate z_{i1}. $\diamondsuit$

Statistical inference for the piecewise exponential regression model can be simplified by using the memoryless property of the exponential distribution. That is, if an observed failure time Y_i is 308 and there is a rate change at $t_0 = 200$, then this observation is equivalent to *two independent* observations: one with rate λ_1, started at zero, but *censored* at 200; and another with rate λ_2, started at 200, and observed to fail at 308.

This fact can be verified by comparing the two likelihood functions. This also explains the way the data-frame is constructed to feed into the Splus function `coxph`, see Therneau (1999, p.18). Or the new book by Therneau and Grambsch (2000) p. 69.

Remark: For stratified Cox model, there are several baseline hazards, one baseline hazard for each stratum. In the exponential-transformation interpretation, this is equivalent to several (different) transformations, one transformation $g(\cdot)$ for each stratum.

Remark: So far we have not introduced censoring in the model. The ability to handle right-censored data for the inference procedures based on partial likelihood is evident, since one can still write down the partial likelihood function without much extra effort for data that are subject to right-censoring.

Remark: How about frailty model? Well we can use the transformation g as before, The frailty only enters the model by changing the exponential parameters λ_i. The final λ_i^* has an extra multiplicative random effect term $\omega_{j(i)}$, $\lambda_i^* = \lambda_i \times \omega_{j(i)}$.

After this paper was published in *The American Statistician*, I become aware the paper by Leemis, Shih and Reynertson (1990). They apparently also used the idea of transforming the exponential random variables to get proportional hazards random variables. The treatment of time-change covariate is slightly different.

Splus/R codes for generating data that follow Cox models

1. Cox model with fixed covariates.

Assume we have the covariate `zi` (k by n matrix) and parameter `beta0` (1 by k vector) already created in Splus. Otherwise an example of `zi` and `beta0` might be

```
> beta0 <- c(3, -1)
> zi <- matrix(c(1:50/10, runif(50)), nrow=2, ncol=50, byrow=TRUE)
```

The example below took $g(\cdot)$ as the square root function.

```
> lambdai <- exp(as.vector(crossprod(beta0, zi)))
> yi <- rexp(length(lambdai), rate=lambdai)
> gyi <- sqrt(yi) # or other increasing function
```

The responses are `gyi` with covariates `zi`. We are to estimate `beta0` from `gyi` and `zi` without knowing which $g(\dot{)}$ was used.

2. Cox model with one time-change covariate and other fixed covariates.

The new code needed here is for the generation of piecewise exponential random variables Y_i. The creation of covariates and the transformation on `yi` is similar to above. We give below a function that transforms standard exponential random variables into a (2-piece) piecewise exponential random variables.

The algorithm can be described as

$$[\text{Generate } Y \sim \exp(1)] \quad \rightarrow \quad \begin{cases} \text{if } [Y \leq t_0] & \text{return } Y/r_1. \\ \text{if } [Y > t_0] & \text{return } t_0/r_1 + (Y - t_0)/r_2. \end{cases}$$

The following Splus code implements this algorithm in a vectorized way. Suppose we want a vector of piecewise exponential random variables with rates `r1` and `r2` before and after `t0`. Here `r1`, `r2` and `t0` are all vectors of length n. The input `y` below should be a sample of n random variables from the standard exponential distribution.

```
> piecewiseexp <- function(r1, r2, t0, y) {
                      val1 <- pmin(y, t0)/r1
                      val1 + pmax((y-t0), 0)/r2 }
```

Once the above function is defined, we can get, for example, 3 independent 2-piece exponential random variables by

```
> piecewiseexp(r1=c(1,2,3), r2=c(5,6,7), t0=c(3.5,6,8.2), y=rexp(3))
```

The first random variable generated will have hazard 1 for $t \leq 3.5$ and hazard 5 for $t > 3.5$, etc.

3. Stratified Cox models with time fixed covariates.

Separate all the `yi` into two groups. We use two different monotone transformations $g_1(\cdot)$ and $g_2(\cdot)$ for each of the groups. This way we get a Cox model with two strata.

4. Frailty models.

Here we need a random term generated from the gamma distribution and this affects the rate of the exponential distribution, `lambdai`.

We first generate the frailty term:

```
litterindex <- c(rep(1,10), rep(3,10), rep(5,10), rep(7,10), rep(9,10))
temp <- rgamma(5, shape=1)
omigai <- temp[ceiling(1:50/10)]
```

We then replace the first line of code in 1) by the following line.

```
lambdai <- omigai * exp(as.vector(crossprod(beta0, zi)))
```

So the first 10 `lambdai` share a frailty, `temp[1]`, the second 10 `lambdai` share a frailty, `temp[2]` etc. Once we got the `lambdai`, the generation of `yi`, `gyi` and `cengyi` are the same as before. The fit of the frailty model can be accomplished by Splus (for censored responses):

```
coxph( Surv(cengyi, delta) ~ t(zi) + frailty(litterindex) )
```

5. Right censoring.

A big advantage of the Cox model over ordinary regression model is that the inference procedures for Cox model can easily handle right-censored responses Y_i. Given `gyi`'s, the following code generates censored `gyi`'s, in the form of `cengyi` ($= \min(Y_i, C_i)$) and `delta` (indicator of censoring).

```
> cen <- rexp(length(gyi), rate=0.1)  # or other censoring variables
> cengyi <- pmin(gyi, cen)
> delta <- as.numeric(cen>=gyi)
```

The above code also works in R, which is a free software similar to Splus.

References

Therneau, T. (1999), *A Package for Survival Analysis in S* http://www.mayo.edu/hsr/people/therneau/survival.ps

Therneau, T. and Grambsch, P. (2000) Modeling Survival Data: extending the Cox model. Springer.

Gentleman, R. and Ihaka, R. (1996). R: A Language for data analysis and graphics. *J. of Computational and Graphical Statistics*, **5**, 299-314.

Barlow, R. and Proschan, F. (1981). *Statistical Theory of Reliability and Life Testing.* TO BEGIN WITH, Silver Spring MD.

Leemis, L.M., Shih, L. and Reynertson, K. (1990). Variate generation for accelerated life and proportional hazards models with time dependent covariates. *Statist. & Prob. Letters*, **10**, 335-339.

Department of Statistics
University of Kentucky
Lexington, KY 40506-0027