Duration Analysis (and Cox Model)

Contents

Basic Idea

Sometimes, the data we are faced with are the duration of a status.A typical example is studying the length of time between the time you graduate and the time you find a job, i.e., the duration of job search. Suppose we are standing at 2020/12/31.The survey data, which is cross-sectional, typically looks like - Individual 1: graduated at 2020/01/01, find a job at 2020/06/01. Therefore, the duration of his job search is 5 months. - Individual 2: graduated at 2020/03/01, but still looking for job at the time of the survey, 2020/12/31. Therefore, we do know the exact real job search duration for this individual, but we know the duaration is longer than 10 months.

The main message from this data set is that: - first, when dealing with such data, we may have censored data, in the sense we cannot observe the real value but rather its lower bound. - Second, the end of the duration means jumping out from this status to another status, and some specific stochastic distribution is needed to depict such a process.

Denote the real job search duration as T_i^*, whether the individual has already found a job or not as $w_i(w_i = 1$ means 'already found a job', 0 denotes 'still searching'), and the observed job search duratio (standing at the time point of survey) as t_i. Following the general framework in the section 'Missing or censored data', the likelihood function of an individual is therefore

$$f(t_i|x_i, \theta)^{w_i} S(t_i|x_i, \theta)^{1-w_i}$$

in whicht $S(t_i|x_i, \theta) = 1 - F(t_i|x_i, \theta)$. f and F are the density and accumulative probability function of the real duration T^*. This likelihood function is intuitive: at the time of survey, if i has already found a job, then we know the exact $t_i = T_i^*$. If i is till finding a job and his job search has lasted for t_i, all we know is that $T_i^* > t_i$

This is the starting point of transition analysis framework. The next task is to discuss the sepcification of F and f. In most textbooks, however, we often discuss the specification of $\lambda(t, x) \equiv \frac{f(t|x)}{S(t_i|x_i)}$. This expression implies that

$$f(t) = \lambda(t)exp\left[-\int_0^t \lambda(u)du\right], S(t) = exp\left[-\int_0^t \lambda(u)du\right] \tag{1}$$

That is, when we determine the expression of $\lambda(.)$, we know f and F immediately. The $\lambda(t)$ is called 'harzard function', which measures the 'instant 'possibility of switching out of the status given the individual has been in this status for t period of time. The following sections give a detailed discussion on how to specify $\lambda(t,x)$

Proportional Harzard (PH)

A very intuitive and simple way is to assume that

$$\lambda(t,x) = \lambda_0(t)h(x)$$

In which $\lambda_0(t)$ is called 'baseline hazard', which does not depend on the x. The next step is to specify the $\lambda_0(t)$ and $h(x)$. Of course, one principle of the specification is that it should be easy to interprete. Another principal is to make (1) easy to calculate. Often we specify $h(x) \equiv e^{x'\beta}$(why?). Now we discuss the several possible specification of $\lambda_0(t)$

Exponetial regression

If

$$\lambda_0(t) = e^a$$

In which a is a parameter to estimate, then we get the exponential regression.

Weibull Regression

If

$$\lambda_0(t) = pt^{p-1}e^a$$

in which a and p are paramter to estimate, then we get weibull regression.

Gompertz Regression

If

$$\lambda_0(t) = e^{a+\gamma t}$$

, then we get Gompertz regression. By specifying $\gamma > 0$ or $\gamma < 0$, we can make the baseline harzard increase or decline with t. This regression specification is often used in demographics.

Piecewise Constant Harzard

A common disadvantage for the above regression, of course, is the monotonicity: the baseline either increases, or decreases, or keeps constant as t increase. This may not be the realistic case. For example, infants and elder people have high death harzard, while young age people have low death harzard. To depict such a baseline harzard pattern, a natural way is to use piece-wise expression of baseline harzard.

Accelarated Failure Time Model

Proportional harzard model assumes that the baseline harzard does not vary with x. Taking Gompertz regression and the job search as example. The gompertz regression specification implies that the two individuals have different harzard of finding a job for a given job search duration t. The 'change' of the harzard of finding job with time, however, is the same for these two individuals, since they have the same expression of baseline hazard. In reality, however, it is totally possible that one individual have a faster change of harzard with time than the other. To depict such a potential difference in the 'change' of the harzard with time, we want to make the baseline hazard also a function of x. There we can write the hazard function

$$\lambda(t,x) = \lambda_0(e^{-x'\beta}t)x'\beta$$

Issues of unobservable heterogeneity

What if there is unobersable heterogeneity in the $\lambda(x_i,t)$? Say,

$$\lambda_i(x_i,t,v_i) = \lambda_0(t)e^{x'_i\beta}v_i$$

in which v_i are something that cannot be observed. Therefore we have

$$S(t,x_i,v_i) = exp\left[-\int_0^t \lambda(u,x_i,v_i)du\right]$$

Of course, since there is unobservable v_i, we need to take integral over v_i before we write down the likelihood function. To take the integrals, we naturally need to assume a distribution of v_i.

$$\lambda_i(x_i,t) = \int \lambda_0(t)e^{x'_i\beta}v_ig(v_i)dv_i = \lambda_0(t)e^{x'_i\beta}\int v_ig(v_i)dv_i$$

$$S(x_i,t) = \int\left[exp\left(-e^{x'_i\beta}\int_0^t \lambda_0(t)du\right)\right]^{v_i} g(v_i)dv_i$$

It is better to choose $g()$ such that the above expressions have analytical expression. Having the above expression we are ready to apply MLE. But what is bad about such unobserved hetergeneity?

A very natural logic is as follows. Suppose we do not consider the heterogeity when writing the likelihood. Suppose that sample, there are many people who find job quickly, i.e., they have small t. What will this affect the estimation? This will in general makes the estimation on $\lambda_0(t)$, the basic harzard, high. But can we be sure that many people find quickly right because that the baseline harzard is high? One alternative reason is that, for these part of people, they have some unbservable characteritics (not in x),say, very high ability of job search ability, that makes them find jobs quickly, while the rest spend a long time on job search since they have low job search ability. If most inviduals in the sample are high ability people, then the average job find time may be low, although the baseline harzard may even be low. A MLE without considering such composition effect brought by unobservable characteristics, therefore, leads to inconsistent estimation on the baseline harzard.

Cox Model

A common disadvantage of the PH or AFT models is that they all assume a specific functional form of the baseline harzard λ_0. However, avoiding a specification of $\lambda_0(t)$ may be a good idea. First, it reduces the potential risk of model misspecification. Second, in fact we may not care too much about the baseline hazard. What we care most is, 'why did this guy spend less time on job search on that guy did?',i.e., we are interested in the relative difference of the harzard, and how such difference can be explained by the independent variables. To better illustrate this , let's go back to PH model, and see that

$$\frac{\lambda(x_i,t)}{\lambda(x_j,t)} = e^{x'_i-x'_j\beta}$$

That is to say, the ratio of hazard between i and j only depends on x and β. This provides a new idea for writing likelihood function.

A Working Example

Consider the following situation

individual	t	x
1	2	4
2	3	1
3	6	3
4	12	2

Here we have four individuals. Indivual 1 finds a job 2 months after he graduates, $t_1 = 2$.Individual 2 finds a job 3 months after he graduates, $t_2 = 3$.. In the basic model we write down the following likelihood function

$$f(t_1 = 2|x_1)f(t_2 = 3|x_2)f(t_3 = 6|x_3)f(t_4 = 12|x_4) \tag{2}$$

which requires we estimate the $\lambda_0(t)$.

The above table, however, provides some other information on the 'order' of finding a job.Denote event: 1 finds job fastest, 2 the second , the 3 the third, and the 4 the slowest. denote the probability of this Event as p. What is the expression of p? - For $t = 2$, i.e, two months after each person's graduation, we know that among 1,2,3,4, There is one person that finds a job. Which one? It is 1.Denote $p_1 :=$ the probality that 1 finds a job given that there is one person in 1,2,3,4 that finds a job - For $t = 3$, i.e, 3 months after each person's graduation, we know that among 2,3,4 (i.e., given that 2,3,4 are searching for jobs),one finds a job. which one? It is 2. Denote $p_2 :=$ the probability that 2 finds a job given that there is one person in 2,3,4 that finds a job - For $t = 6$, i.e.,6 months after each person's graduation, we know that among 3,4 (i.e.,given that 3,4, are still searching for jobs), one finds a job. Which one? It is 3. Denote $p_3 :=$ the probability that 3 finds a job given that there is one person in 3,4 that finds a job - For $t = 12$, i.e., 12 months after each person's graduation, 4 finds a job.Denote $p_4 :=$ the probability that 4 finds a job given that there is one person in 4 that finds a job. Of course $p_4 = 1$.

Intuitively , we have

$$p = p_1 p_2 p_3 p_4 \tag{3}$$

That is, instead of (2), here the (3) models the relative speed of finding job. We next explore the expression of p_1,p_2,p_3, which is easy given their definitions.

$$p_1 = \frac{\lambda(2|x_1)}{\lambda(2|x_1) + \lambda(2|x_2) + \lambda(2|x_3) + \lambda(2|x_4)} = \frac{e^{x_1'\beta}}{e^{x_1'\beta} + e^{x_2'\beta} + e^{x_3'\beta} + e^{x_4'\beta}}$$

$$p_2 = \frac{\lambda(3|x_2)}{\lambda(3|x_2) + \lambda(3|x_3) + \lambda(3|x_4)} = \frac{e^{x_2'\beta}}{e^{x_2'\beta} + e^{x_3'\beta} + e^{x_4'\beta}}$$

$$p_3 = \frac{\lambda(6|x_3)}{\lambda(6|x_3) + \lambda(6|x_4)} = \frac{e^{x_3'\beta}}{e^{x_3'\beta} + e^{x_4'\beta}}$$

Since we have Proportional hazard, the baseline hazard disappears from both the nominator and denominator. Given this, we can write down the the expression of p immediately. To summarize, this p captures the probability for the observed 'order' of the job finding, but not the absolute time point. In some sense, such likelihood function only captures part of the information of the data, therefore it is 'partial likelihood function'

Another Working Example

In the above example, different individuals have different t. Now consider the following case

individual	t	x
1	2	4
2	3	1
3	3	3
4	12	2

Now both 2 and 3 find a job 3 months after they graduate. Denote event: 1 finds job fastest, 2 and 3 the second , the 4 the slowest. denote the probability of this Event as p. What is the expression of p? - For $t = 2$, i.e, two months after each person's graduation, we know that among 1,2,3,4, There is one person that finds a job. Which one? It is 1.Denote $p_1 :=$ the probality that 1 finds a job given that there is one person in 1,2,3,4 that finds a job - For $t = 3$, i.e, 3 months after each person's graduation, we know that among 2,3,4 (i.e., given that 2,3,4 are searching for jobs),two people find jobs. which two? It is 2 and 3. Denote $p_2 :=$ the probability that 2 and 3 find a job given that there two people in 2,3,4 that finds a job - For $t = 12$, i.e., 12 months after each person's graduation, 4 finds a job.Denote $p_4 :=$ the probability that 4 finds a job given that there is one person in 4 that finds a job. Of course $p_4 = 1$.

The expression for p_1 is the same as in the first example. The p_{23} is also easy to calculate given its definition.

$$p_2 = \frac{\lambda(3|x_2)\lambda(3|x_3)}{\lambda(3|x_2)\lambda(3|x_3) + \lambda(3|x_3)\lambda(3|x_4) + \lambda(3|x_4)\lambda(3|x_2)} = \frac{e^{x'_2\beta}}{e^{x'_2\beta}e^{x'_3\beta} + e^{x'_3\beta}e^{x'_4\beta} + e^{x'_4\beta}e^{x'_2\beta}}$$

Stratified Cox Model

From the above analysis, we know that one assumption of COX model is that it must the satisfy proportion harzard specfication, that is,

$$\lambda(t, x) = \lambda_0(t)e^{x'\beta}$$

Only under this assumption can we write down the above likelihood functions that are independent of t!. But sometimes, we may not be able satisfy such a condition. For example, gender. It is quite reasonable that male and female have different baseline hazards,meaning that gender variable cannot be sepearated from the baseline harzard, and we can no longer kill the baseline harzard part from the likelihood function.

One method, naturally, is to allow for difference in the baseline harzard for male $\lambda_0^m(t)$ and for female $\lambda_0^f(t)$, and write down the partial likelihood function for male and men respectively. The harzard for male and female are respectively:

$$\lambda(t, x, g) = \begin{cases} \lambda_0^m(t)e^{x'\beta}, g = male \\ \lambda_0^f(t)e^{x'\beta}, g = female \end{cases}$$

To be specific, for male, we consider the relative order of finding a job among the male, and write down the likelihood function. male have common baseline hazard function, there we can get a likelihood function independent of baseline hazard. The same is for female. Intuitively, there are two main disadvantages in this approach.

- We are not able to estimate the effect of gender on the job search hazard.
- Information loss: we only consider the order of job finding inside each gender group, but the information on the order of job finding for all individuals are lost. . . .(?)