Introduction to cox regression

Notes of class

Iván Andrés Trujillo Abella

Facultad de ingeniería

0.1 Exponential Distribution

Exponential distribution dont have memory equall to the geometric distribution. with one parameter λ the time between until a poisson process occur. Assume that $y \sim G(p)$ then

$$P(Y \leq z)$$

The exponential distribution have the following Probability Distribution Function:

$$f(x) = \lambda e^{-\lambda x}$$

Therefore we said, $X \sim exp(\lambda)$ that x follow a exponential distribution with one only parameter in this case lambda.

Notes exponential distribution it is related with **gamma distribution**.

0.1.1 Survivor and hazard function

we have a variable y that means time of survivor $f(y)$ it is its PDF. the cdf will be $F(y) = \int_0^y F(t)dt$ and survivor function is defined by

$$S(y) = P(Y > y) = 1 - F(y)$$

that is the probability of survive beyond T.

in this order of ideas the *Hazard function* is defined as

$$h(y) = \frac{f(y)}{S(y)}$$

then this will be undesertand as a risk. but

$$f(y) = \lim_{\Delta \to 0} \frac{F(y + \Delta y) - F(y)}{(\Delta y)}$$

$$\frac{f(y)}{S(y)} = \frac{\lim_{\Delta \to 0} \frac{F(y+\Delta y) - F(y)}{(\Delta y)}}{1 - F(y)}$$

if we translate the above expression in probability terms we have *the probality of failure in a small time period of change* due the survial period.

$$h(y) = -\frac{d}{dy}ln[1 - F(y)] \tag{1}$$
$$\int_0^y h(y) = -ln(S(y)) \tag{2}$$
$$S(y) = e^{\int_0^y -h(t)dt} \tag{3}$$

Note here that survival function and hazard are related.

0.1.2 Hazard in exponential

we need get the **CDF** this is

$$f(x) = \lambda e^{-\lambda x}$$
$$\int_0^x f(t)dt = 1 - e^{\lambda x}$$
$$S(x) = e^{-\lambda x}$$
$$h(x) = \lambda$$

thus the hazard it is a contant when failure time it is exponential.

Another formulation:

$$h(y) = \lim_{\Delta \to 0} \frac{P(y < Y \leq y + \Delta y \mid Y > y)}{\Delta y} \tag{4}$$

Therefore we can see that the denominator is a conditional probability that we could rewrite in terms of joint distribution, and how the one event is a subset of another by intersection we have:

$$h(y) = \lim_{\Delta \to 0} \frac{P(y < Y \leq y + \Delta y)}{P(Y > y)\Delta y} \tag{5}$$

Note that $P(Y > y)$ is the survival function.

$$h(y) = \lim_{\Delta \to 0} \frac{P(Y \leq y + \Delta y) - P(Y < y)}{S(y)\Delta y} \tag{6}$$

note that is equal to the result that we get previously given that; $P(Y \leq y + \Delta y) = F(y + \Delta y)$.

Hint: *You could read a hazard (λ) as the risk of occur the event at t period of time*

1 Cox hazard

the main parameter is *hazard rate* the expected number of event by period time.

hazard ratio could be compated with odds ratio.

observed to expected this is very important this concept due it is have a notion about ramdon.

$_0(t)$ risk can change over time according to covariates. the proportional hazard condition establish that covariates are multiplicative related to hazard.

the measure of effect is the hazard rate, not is probability (the expected number of events per unit of time). Sometimes we are interested, in compare rates among groups.

$$\lambda(t|z) = \lambda_0(t)e^{\sum_{i=1}^{n} \beta_i x_i}$$

in this case there are not β_0 due if $z = 0$ then $\lambda_0(t)$ it is the baseline hazard.

The cox regression give us the probability that occur the event of interest by a t period, given a vector of inputs.

```
import pandas
```

1.1 Wald

1.2 Ratio Likelihood

1.3 Coursera

hazard : the risk of death at a given moment in time.

Hazard ratio is as odds ratio.

1.4 Missing values

2 Residuals

The difference among $t_i - \hat{t}_i$.

2.1 Schoendfeld residual

2.2 Martingale Residual

2.3 Deviance Test

2.4 What means proportional hazard ratios?

It is simply that the hazard of the event according to a category is parallel to the another, for instance, the hazard of art student is twice of the hazard of economic student.

3 Hazard ratio

$$\frac{\lambda_i}{\lambda_j} \tag{7}$$

4 Interpret the model

Each binary variable is a estimator of hazard ratio. Measure the risk of presence regarding lacking the feature.

4.1 continuous variables

$$e^{\beta_i} \tag{8}$$

The hazard ratio to increase at one unit the variable i.

5 Prediction

5.1 ROC_t curve

5.1.1 Concordance

If j and i underwent the event of interest then, and suppose that $t_j > t_i$ and the predicted survival of j is $\hat{S(j)}$, therefore if $\hat{S}(j) > \hat{S}(i)$ occur a match.

$$Adjusment = \frac{\sum_{i \neq j}^{n} match(i,j)}{C(i,j) \in n} \tag{9}$$

What happens if compare with a censored data?

6 Considerations

7 Cox and survival

By the equation, (3) we have that

$$\lambda(t) = \lambda_0(t) e^{\sum \beta_i x_i} \tag{10}$$

$$S(t) = e^{\int_0^t -\lambda_0(t)e^{\beta \mathbf{x}}dt} \tag{11}$$

With this equation we can get the probability that the individual, survive until t time according to input features $\beta\mathbf{x}$.

hints over cox regression: linear model of hazard ratio.

$$log\left(\frac{\lambda_i(t)}{\lambda_0(t)}\right) = \sum_j = \beta_j x_j i \tag{12}$$

the baseline hazard function:

8 Statistics and properties

A 'estadístico' is a function that

$$F : \mathbb{R}^n \to \mathbb{R}$$

for instance the median, variance we can call this parameter as θ and define a probability density function for this parameter. for instance the mean is a random variable, whose distribution it is ready known.

therefore the distribution of mean have its own first moment or mean.

a statistics $\hat{\theta}$ is used to estimate a parameter then will be a estimator. for instance $\bar{x} = \frac{1}{n-1}\sum x_i$ it is a estimator of μ.

now $E(\hat{\theta}) = \varsigma(\theta)$ we search that $\varsigma(\theta) = \theta$ then we said that $\hat{\theta}$ is a unbiased estimator.

Another wished characteristic is that $MSE(\theta) = E((\hat{\theta} - \theta)^2)$.

i can rewrite as:

$$MSE(\theta) = var(\hat{\theta}) + (\xi(\theta))^2 \tag{13}$$

where $\xi(\theta)$ is the bias, the two term we can read as accuracy. the is wished $\min_\theta MSE(\theta)$

the Rao-Cramer: establish a cote over the variance of a estimator.

assume that $X_i \sim f_\theta$ and the variables are iid, and the join distribution is $f(\vec{x}) = \prod f_\theta(x_i)$.

$$var(\hat{\theta}) \geq \frac{(\varsigma'(\theta))^2}{nE\left(\frac{\partial}{\partial\Theta}log(f_\theta(X)\right)} \tag{14}$$

note that the numerator of the last equation is equal to one, when the estimator is unbiased.

therefore we said that a estimator is efficient if have the min possible variance and it is unbiased.

rao-cramer allow us determine if a estimator is or not efficient.

suppose that $X \sim poisson(\lambda)$ now used $\hat{\lambda} = \bar{X}$.Now $E(\bar{X}) = \lambda$.

by properties: $var(\bar{X}) = \frac{1}{n^2}\sum var(x_i) = \frac{\lambda}{n}$.

now to establish the bound, then reduce: $var(\hat{\lambda}) \geq \frac{1}{nE\left((\frac{\partial}{\partial\lambda}log\frac{e^{-\lambda}\lambda^x}{x!})^2\right)} = \frac{\lambda}{n}$.

consistency we said that $\hat{\theta}$ is consistency estimator of θ if $\forall\epsilon > 0 \lim_{n\to\infty} P(|\hat{\theta} - \theta| \leq \epsilon) = 1$ note that n is the sample size.

the weak law of big number:

remember that asymptoticaly unbiased implies that n tend to infinity.

what is the difference among consistency and unbiased estimator.

Thus a estimator could be biased but consistency thus, the estimated parameter and the real could converge.

9 Partial likelihood

10 wald test

$$\begin{aligned} H_0 :&\theta = \theta_0 \\ H_1 :&\theta \neq \theta_0 \end{aligned} \tag{15}$$

based in the distance about the θ of H_0

11 likelihood ratio test

12 Information matrix

12.1 Gauss Markov Theorem

12.2 Matrix semi positive

12.3 Lab excersices

12.4 Exponential Distribution

Useful times of occurrence, is derived from gamma distribution. This is related with Poisson distribution.

$$pdf = f(x) = \lambda e^{-\lambda x} \tag{16}$$

```
def dexp(x, parameter, mode='pdf'):
    if mode=='pdf':
        pdf = parameter * np.exp(-parameter*x)
        return pdf
    if mode=='cdf':
        cdf = 1 - np.exp(np.exp(-parameter*x))
        return cdf
```

visually this is :

```
x = np.linspace(1,10,100)
lmbd = 0.90
plt.plot(x, dexp(x,lmbd))
```

12.5 Real life example

13 hazard

event occurred over the number at risk in a certain period,if we think in terms of probability $f(x)$ is based in the relative frequency of events in the period, and the cumulative over that in the are in risk.

14 Hazard ratio

15 Empirical to take decisions

We can estimate a $S_i(t)$ by the i individual and its feature vector $X_i$compare with a mean in sample, thus we can take decisions over borrow or not. Also allow us consider health indicator of risk management.

16 Nelson-Aalaen estimator

proceso no homogeneo de poisson.

17 Neyman person lemma

remember the likelihood function

18 likelihood ratio test

18.1 Cumulative hazard function

$$H(t) = \int_0^t h(u)du \tag{17}$$

19 Coursera

all this are insights about cox proportional hazard, given that we define the $h(t)$ hazard in function of time, risk set is the number of individuals that are susceptible of underwent the event.

20 sickit-survival

installing in linux machine

```
sudo pip install scikit-survival
```

exist a baseline hazard function, and covariates change the risk proportionally. the ratio of risk among two patients remain constant in all period.

we need transform the variables in dummy.

21 tied data

efron and breslow method.

22 Breslow estimator

maximum likelihood estimator for cumulative baseline hazard function.

22.1 no parametric maximum likelihood estimation

stochastic process stochastic process are as functions of random variables. this is interested in study the evolution of this variables in time.

22.2 probability space

$F - \sigma - algebra$

23 Counting process

23.1 convolution

23.2 Concordance

Suppose two models of machine learning we compare what individuals classified equally, however in this procedure could be exist a random assignment.

23.2.1 Kappa cohen

This affected by prevalence.