# thesis note

October 14, 2020

## 1 Kernel estimation of Hazard function

### 1.1 rank kernel

Suppose $T_1$,...,$T_n$ i.i.d. with cdf $F_T$, independent of $C_1$,...,$C_n$ i.i.d. with cdf $F_C$. $X_i$ = min($T_i$,$C_i$) with indicator $\delta_i$=$I_{T_i \leq C_i}$. Denote $R_j$ the rank of $X_j$.(muhaz in R)

$$\hat{h}_T(t) = \sum_{j=1}^{n}(n - R_i + 1)^{-1}\delta_i K_h(t - X_i) \quad (1)$$

Here K is symmetric non-negative kernel functions, K(t) = o($t^{-1}$) as t $\to \infty$, $\int K(t)dt = 1, K_h(y) = h^{-1}K(y/h)$ where h is tending to zero at appropriate rates.

### 1.2 some commonly used kernel

Normal kernel: k(X) = gaussian(x) Uniform kernel: K(x) = 1/2 for $-1 \leq x \leq 1$
Epanechnikov kernel: K(x)=0.75$(1 - x^2)$ for $-1 \leq x \leq 1$
Biweight kernel: K(x) = 15/16$((1 - x^2)^2)$ for $-1 \leq x \leq 1$
Gaussian kernel:K(x) = $\frac{1}{\sqrt{2\pi}}exp(-x^2/2)$

### 1.3 bandwidth selection

The bandwidth function can be defined as the distance between the point of interest and its $Kth$ nearest neighbour among the remaining deaths. The resulting estimator has a bandwidth which adapts to the configuration of the observation.

To be more specific, an optimal local bandwidth sequence, b(t), can be consistently estimated by minimizing an estimate of local

$$MISE(h) = E[\int(\hat{f}_h(x) - f(x))^2 dx] \quad (2)$$

of the hazard rate estimate with respect to the bandwidth.

Bandwidth estimated by Silverman's Rule of Thumb(Gaussian kernel):

$$h = (\frac{4\hat{\sigma}^5}{3n})^{1/5} \approx 1.06\hat{\sigma}n^{-1/5} \quad (3)$$

# 2 Non-linear regression

## 2.1 polynomial regression

R code example: model = lm(y   poly(x, 5, raw = TRUE))

## 2.2 Spline regression

R code example: library(splines)
knots = quantile(x, p = c(0.25, 0.5, 0.75)) degree = 3
model = lm (y   bs(x, knots = knots))

## 2.3 Generalized additive models

R code example: library(mgcv)
model = gam(y   s(x))
The term s(x) tells the gam() function to find the "best" knots for a spline term.

## 2.4 kernel regression

Given i.i.d. samples $(x_i, y_i), i = 1, ..., n$ from the model

$$y_i = r(x_i) + \epsilon_i, i = 1, ..., n \tag{4}$$

$$\hat{r}(x) = \sum_{i=1}^{n} w(x, x_i) * y_i \tag{5}$$

$$w(x, x_i) = \frac{K(\frac{x_i - x}{h})}{\sum_{j=1}^{n} K(\frac{x_j - x}{h})} \tag{6}$$

Therefore, the kernel regression estimator is:

$$\hat{r}(x) = \frac{\sum_{i=1}^{n} K(\frac{x_i - x}{h}) * y_i}{\sum_{j=1}^{n} K(\frac{x_j - x}{h})} \tag{7}$$

where K(x) is a kernel function satisfying $\int K(x)dx = 1,\ K(x) = K(-x)$

# 3 Influence function

To investigate the influence of outlying data(contaminated) we consider the $\alpha - influencefunction$ defined as the difference between estimates before and after contamination divided by the contamination rate,$i.e.1/n$:

$$IF(z) = n(\hat{\beta}^z - \hat{\beta}) \tag{8}$$

where $\hat{\beta}^z$ is the estimate based on the contaminated data with outlying observation z, and $\hat{\beta}$ is the estimate based on uncontaminated data.

# 4 Partial(conditional) likelihood

In Cox's (1972) proportional hazard model, it is assumed that the conditional hazard function $\lambda\left(t|x_i\right)$ for individual $i$ with a vector explanatory variables $x_i = (x_{i1}, \ldots, x_{ip})$ at time $t$ is exp $\left(x_i\beta\right)\lambda_0(t)$, where $\beta$ is an unknown $p$ dimensional parameter vector and $h_0(t)$ is the baseline hazard function i.e. the hazard rate at $x = 0$.
Let $t_1 < \ldots < t_k$ represent $k$ distinct observed times among a sample of $n$ possibly censored observed times, and denote the individuals who are free of the event of interest and uncensored just prior to $t_i$, the risk set, by $R\left(t_i\right).$ Then the conditional probability that an individual with covariate vector $x_i$ will meet the event of interest at $t_i$ is defined as follow:

$$\exp\left(x_i\beta\right) / \sum_{l\in R(t_i)} \exp\left(x_i\beta\right) \tag{9}$$

Multiplying these probabilities over each of the $k$ lifetimes gives the 'partial likelihood' (Cox, 1975 ), now known as conditional likelihood for estimating $\beta$ :

$$L(\beta) = \prod_{i=1}^{k}\left[\exp\left(x_i\beta\right) \mid \sum_{l\in R(t_i)} \exp\left(x_l\beta\right)\right] \tag{10}$$

Taking tied events into account, the former equation is modified (Breslow, 1974) as:

$$L(\beta) = \prod_{i=1}^{k}\left\{\left[\exp\left(S_i\beta\right) \mid \sum_{l\in R(t_i)} \exp\left(x_l\beta\right)\right]^{d_i}\right\} \tag{11}$$

where $d_i$ is the number of observed times equal to $t_i$ and $S_i$ is the sum of the covariate vectors for these individuals.
Then following the idea of maximum likelihood estimate(MLE), the parameter vector $\beta$ can be estimated by finding the vector B that maximize the likelihood $L(\beta)$, or equivalently finding the vector B that maximize the log-likelihood arise from $L(\beta)$:

$$\log L(\beta) = \sum_{i=1}^{k} S_i\beta - d_i \log\left[\sum_{l\in R(t_i)} \exp\left(x_l\beta\right)\right] \tag{12}$$

which can be done by setting the first derivatives of log-L to zero and solve the equation numerically.

# 5 Hellinger-distance based hazard estimation(Ying)

$$\begin{array}{ll} N_n(x) = \sum_{i=1}^{n} I_{\{Z_i\le x, \delta_l=1\}}, & \bar{N}_n(x) = N_n(x)/n \\ Y_n(x) = \sum_{i=1}^{n} I_{\{Z_i\ge x\}}, & \bar{Y}_n(x) = Y_n(x)/n \end{array} \tag{13}$$

We define our minimum Hellinger-type distance estimator, restricted to $[a, b]$, by $\hat{\theta}_n(a, b) = \xi\left(\hat{\lambda}_n, \bar{Y}_n, a, b\right)$. In other words, $\hat{\theta}_n(a, b)$ can be found as :

$$\hat{\theta}_n(a, b) = arg \min_{\theta \in \Theta} \int_a^b \left[\lambda_\theta^{1/2}(x) - \hat{\lambda}_n^{1/2}(x)\right]^2 \bar{Y}_n(x) dx \tag{14}$$

where $\hat{\lambda}_n = \frac{1}{d_n} \int_{-\infty}^{\infty} K\left(\frac{x-u}{d_n}\right) d\hat{\Lambda}_n(u)$ with $\hat{\Lambda}_n(x) = \int_{-\infty}^{x} \frac{dN_n(u)}{Y_n(u)}$ is the kernel estimator of the underlying hazard rate function $\lambda$, $\lambda_\theta$ is the corresponding hazard rate function estimator from a parametric family $\Theta \subset R^p$. Note that $dN_n(x)/Y_n(x) \approx \hat{\lambda}_n(x)dx$.

Further, to ensure consistency and efficiency, we add an weight term $w_n$(may be regarded by $\hat{Y}_n$)into Hellinger-distance and $\hat{\theta}_n$ can be found as:

$$\hat{\theta}_n = arg \min_{\theta \in \Theta} \int_{-\infty}^{\infty} \left[\lambda_\theta^{1/2}(x) - \hat{\lambda}_n^{1/2}(x)\right]^2 \bar{Y}_n(x) w_n(x) dx \tag{15}$$

Note that, $\hat{\theta}_n$ is also approximating the maximum likelihood estimator of $\theta$ if $f \in \{f_\theta : \theta \in \Theta\}$.

# 6 another hazard estimation(bshazard in R)

Consider a sample of subjects that are observed from a common origin (e.g. start of exposure,time of diagnosis) to the occurrence of the event of interest. Observation can be right censored and/or left truncated.

By partitioning the time axis into very small intervals, the number of events in each interval is approximately Poisson with mean $\mu(t) = h(t)P(t)$, where $P(t)$ is the total person-time in the interval (in the simplest case without censoring $P(t)$ will be the product of the number of individuals at risk at the beginning of the interval and the length of the interval). The time at risk of each subject can thus be split into n bins or time intervals (common to all subjects).

For a data set with one record per individual including entry time, exit time and censoring indicator, this splitting of time can be implemented easily by the *splitLexis* function in the *Epi* package. Using (common) break points supplied by the user,the function divides each data record into disjoint follow-up intervals each with an entry time, exit time and censoring indicator, stacking these as separate ‘observations’ in the new data set. The bins should be small enough so that $h(t)$ can be considered approximately constant within each interval.

Then we use t to denote the vector of the midpoints of all bins, with $t_i$ representing the midpoint for the $ith$ bin ($i = 1, ..., n$). The vectors $y(t)$ and $P(t)$ represent the total number of events observed and the total person-time in each interval. Using the Poisson likelihood to approximate the general likelihood for survival data (Lee et al., 2006; Lambert and Eilers, 2005; Whitehead, 1980), the hazard can be estimated by modeling the expected number of events, $\mu(t)$, in each interval as a Poisson variable by using $P(t)$ as an offset term.

$$log[\mu(t)] = f(t) + log[P(t)], \tag{16}$$

where f(t)denotes the logarithm of the hazard. In this context it is straightforward to account for possible covariates in a proportional hazards scheme. After the splitting, data can be aggregated according to the bin (time) and also to the covariate values for each subject, so that the final data is organised with one record for each bin and covariate combination. Denoting by $X$ the design matrix which contains the covariate values (fixed in time) and by $\beta$ the corresponding vector of coefficients,the model becomes:

$$log[\mu(t)] = X\beta + f(t) + log[P(t)] \tag{17}$$

Further, we can use B-splines, which provides a numerically efficient choice of basis functions (De Boor, 1978), to estimate $f(t) = log[h(t)]$ . B-splines consist of polynomial pieces of degree m, joined at a number of positions, called knots, along the time axis. The total number of knots (k) and their positions are arbitrarily chosen and are quite crucial for the final estimate, since the function can have an inflection at these locations. By using B-splines to estimate f(t), the expected number of events above can be rewritten as

$$log[\mu(t)] = X\beta + Zv + log[P(t)] \tag{18}$$

where $Z$ is the matrix whose q columns are the B-splines, i.e. the values of the basis functions at the midpoints of the time bins (that will be repeated for each covariate combination) and $v$ is a vector of length q of coefficients whose magnitude determines the amount of inflection at the knot locations.The number of basis functions $q = k+m1$, where $k$ is the total number of knots, including minimum and maximum, and $m$ is the degree of the polynomial splines.Thus the problem of estimating the hazard function reduces to the estimation of coefficients in a Generalised Linear Model framework, which can be done through iterative algorithm.

# 7 Conditional hazard estimation

Note that the estimation of T is possibly not complete due to random censoring. Here we make an assumption that $T \mid Z$ and $C \mid Z$ are independent. Given $Z = z$, the variables T and C have conditional survival functions $S(. \mid z)$ and $G(. \mid z)$ and conditional densities $f(. \mid z)$ and $g(. \mid z)$, respectively. Furthermore, $H(. \mid z)$ denotes the conditional survival function of T given Z = z and $h(.,. \mid z)$ denotes the conditional density of $(T, \delta)$ then $r(t \mid z) = h(t, 1 \mid z)$.(not censored)

Followed by the independence of T and C, we have

$$H(. \mid z) = S(. \mid z)G(. \mid z). \tag{19}$$

consequently,

$$r(. \mid z) = f(. \mid z)G(. \mid z) \tag{20}$$

The definition of conditional hazard rate function

$$\lambda(t \mid z) = \lim_{\Delta t \to 0} P(X \leq t + \Delta t; \delta = 1 \mid X > t, Z = z)/\Delta t \tag{21}$$

leads to the local linear hazard rate estimator with the presence of censoring:

$$\hat{\lambda}(t \mid z) = f(. \mid z)/S(. \mid z) = \hat{r}(t \mid z)/\hat{H}(t \mid z) \tag{22}$$

To give the estimation of $\hat{r}(t \mid z)$ and $\hat{H}(t \mid z)$, we define

$$w_{h,i}(z) = w((z - Z_i)/h)/h, \;\; k_b(t) = k(t/b)/b \tag{23}$$

where w(.) and k(.) are kernel functions, h and b are bandwidth.

The density estimator $\hat{r}(t \mid z)$ is defined as

$$\hat{r}(t \mid z) = \frac{\sum_i^n w^*_{h,i}(z) k_b(t - X_i)\delta_i}{\sum_i^n w^*_{h,i}(z)} \tag{24}$$

where

$$w^*_{h,i}(z) = w_{h,i}(z)\left(s_{n,2}(z) - (z - Z_i)\, s_{n,1}(z)\right) \tag{25}$$

$$s_{n,\ell}(z) = \sum_{i=1}^{n} w_{h,i}(z)\,(z - Z_i)^{\ell}\,, \quad \ell = 1, 2 \tag{26}$$

Let $K(t) = \int_{-\infty}^{t} k(u)du, \quad K_b(t) = K(t/b)$, then the estimator is written as

$$\hat{H}(t \mid z) = \frac{\sum_i^n w^*_{h,i}(z) K_b(X_i - t)}{\sum_i^n w^*_{h,i}(z)} \tag{27}$$

Note that this local linear hazard rate hazard estimator is not necessarily a true hazard rate, since the numerator may be negative. To deal with this, in practice, we simply truncate the local hazard rate estimator to zero if necessary.

# 8 Notation

## 8.1 basic setting

Suppose the observed data is $(X_i, \delta_i, Z_i)$ i=1,2,...,N where $T_i^{\prime}s$ are event time i.i.d. with cdf $F_T$ , $C_i^{\prime}s$ are censoring time i.i.d. with cdf $F_C$ . $X_i = \min(T_i, C_i)$ with indicator $\delta_i = I_{T_i \leq C_i}$. $Z_i$ is a vector(could be of dimension 1) of covariates for ith individual .

## 8.2 modified notation

Note that in clinical trails these n individuals can usually be grouped into M groups corresponding to M different sets of covariates.Therefore we adjust the notation as $(X_{ij}, \delta_{ij}, Z_i)$ for jth individual in ith group, $i = 1, 2, ..., M, j = 1, 2, ..., n_i$, $n_i$ is the number of individuals in ith group that $\sum_{i=1}^{M} n_i = N$.

With $n_i$ individuals in the ith group, we can estimate the conditional hazard function $\hat{\lambda}(t|Z_i)$ through kernel method and find $\hat{\lambda}(X_{ij}|Z_i)$ by plugging in the observed time $X's$.

# 9 New approach

## 9.1 simple case($\lambda_0(t)$ known)

First we consider the simplest case with the baseline hazard function $\lambda_0(t)$ in the Cox-PH model known.

To estimate the parameter vector $\beta$, we propose a Hellinger-distance based method where the estimator $\hat{\beta}$ can be found by minimizing the Hellinger distance.

As shown before, we can estimate the hazard rate at each observed time $\hat{\lambda}(X_{ij}|Z_i)$ through rank-kernel method. Then we divide them by corresponding covaratie effect terms $exp(b * Z_i)$ to construct $v_{ij}(b)$, where b is a reasonable estimator of true $\beta$:

$$v_{ij}(b) = \frac{\hat{\lambda}(X_{ij}|Z_i)}{exp(bZ_i)} = \frac{exp(\beta Z_i) * \lambda_0(X_{ij})}{exp(bZ_i)} \tag{28}$$

By definition, these $v_{ij}(b)'s$ can be viewed as estimation of $\lambda_0(X_{ij})'s$ when b get close enough to true $\beta$. So conversely, we can find the estimator $\hat{\beta}$ by minimizing the overall sum of point-wise distance between these pairs:

$$\hat{\beta}_{MHD} = arg\min_{b \in R} \sum_i \sum_j \left\{ \lambda_0^{1/2}(X_{ij}) - v_{ij}^{1/2}(b) \right\}^2 \tag{29}$$

## 9.2 general case($\lambda_0(t)$ unknown)

In most general cases, the baseline hazard function $\lambda_0(t)$ is unknown so we need to adjust the proposed new approach, but the basic idea is quite similar that we find the estimator by minimizing the Hellinger distance.

The initial part is the same, we first derive all hazard estimator $\hat{\lambda}(X_{ij}|Z_i)$ for each set of $(X_{ij}, \delta_{ij}, Z_i)$ through grouped rank-kernel method and construct the corresponding $v_{ij}(b)'s$. With all these paired $(X_{ij}, v_{ij}(b))$ from different groups, we can combine them and find an estimator $\hat{\lambda}_0(t)$, denoted as overall baseline estimator through kernel regression:

$$\hat{\lambda}_0(t) = \frac{\sum_{i=1}^{M} \sum_{j=1}^{n_i} K(\frac{x_{ij}-t}{h}) * v_{ij}(b)}{\sum_{k=1}^{M} \sum_{l=1}^{n_i} K(\frac{x_{kl}-t}{h})} \tag{30}$$

The key part of estimating $\beta$ is to minimize the distance of overall baseline estimator $\hat{\lambda}_0(t)$ and grouped estimator $v_{ij}(b)$ at all observed time.

Notice that both terms are functions of b so we need an iteration algorithm to approach the estimation.

- Step 1: Set an initial $b^{(0)}$ and plug into the formula of overall baseline estimator(30), so the terms are no longer function of b.
- Step 2: With the overall baseline estimator set fixed, we find the $b^{(1)}$ by minimizing:

$$\hat{\beta}_{MHD} = arg \min_{b \in R} \sum_i \sum_j \left\{ \hat{\lambda}_0^{1/2}(X_{ij}) - v_{ij}^{1/2}(b) \right\}^2 \tag{31}$$

- Step 3: Plug in the $b^{(1)}$ we found in previous steps into equation(30) so as to update the overall baseline estimator.
- Step 4: Repeat Steps 2 and 3, find $b^{(k)}$ by minimizing equation (31) where the overall baseline is set fixed with $b^{(k-1)}$, then update the overall baseline estimator with new $b^{(k)}$, iterate until the convergence is met.

# 10 experiment

## 10.1 basic setting

All the experiments in this paper are conducted on R. Here we consider the Cox proportional hazards model where $\beta$ is one-dimensional with true value $\beta_0$=2, while $\lambda_0(t)$=0.1+0.2t. Covariate Z can take 0, 1 and 2.
So given Z, hazard function is

$$\lambda(t|Z) = exp(\beta_0 * Z) * (0.1 + 0.2t) \tag{32}$$

Cumulative hazard function is

$$H(t|Z) = exp(\beta_0 * Z) * (0.1t + 0.1t^2) \tag{33}$$

Further,

$$S(t|Z) = exp[-H(t|Z)] = exp[-exp(\beta_0 * Z) * (0.1t + 0.1t^2)] \tag{34}$$

## 10.2 generating data

For group 0(Z=0) we have survival function $S(t|Z=0) = exp[-(t + 0.1t^2)]$ and density function $f(t|Z=0) = -\frac{dS(t|Z=0)}{dt} = (0.1 + 0.2t) * exp[-(0.1t + 0.1t^2)]$ so we can generate survival time through inversion sampler.
Similar when Z=1 and Z=2 so we can generate a data set including survival time of "individuals" from three groups and we know that the baseline hazard function is $\lambda_0(t) = 0.1 + 0.2t$ while true value of $\beta$ is $\beta_0$=2.

### 10.3 censoring data

For group i, we generate a set of censoring data $C_{ij}$ uniformly from an interval $I_i$(varies from group to group). And the final observed time is $min(T_{ij}, C_{ij})$. We adjust the length of intervals to make sure the overall censoring rate is close to what was designed, i.e. 30% and 60%.

### 10.4 hazard estimation

Here we use R function *muhaz* based on rank-kernel smoother to estimate the hazard rate at each event time in each group, i.e. $\hat{\lambda}(X_{ij}|Z_i)$. Note that these are point estimation rather than a continuous function estimator but it is adequate for further calculation since we consider the point-wise distance only.
We construct $v_{ij}(b)$ as proposed and combine all these $v_{ij}(b)$ to estimate the overall baseline estimator $\hat{\lambda}_0(t)$ through kernel regression.

### 10.5 Minimizing the Hellinger distance

Define a function with respect to parameter b, which calculates the Hellinger distance between $\lambda_0(t)$ and $\lambda(t|Z_i)/exp(b * Z_i)$ in each group then we take the sum over all groups.
The target is to minimize equation(31). We use R function "optimize" to find the $\hat{\beta}$ which minimizes the sum of distance and carry out the algorithm through a loop iteration.

## 11 Comparison(to be rearranged)

### 11.1 $\lambda_0(t)$is known

1) Ying's proposal: $\beta$ can be estimated by minimizing the Hellinger distance between kernel hazard estimator and parametric hazard estimator in each group, then take the sum over all groups

$$\sum_i \int \left[(\lambda_0 * exp(\theta * Z_i))^{1/2}(x) - \hat{\lambda}_n^{1/2}(x|Z_i)\right]^2 \bar{Y}_n(x)dx \tag{35}$$

2) New proposal: $\beta$ can be estimated by minimizing the discrete Hellinger distance between constructive kernel estimator of $\hat{\lambda}_0$ and the known baseline hazard function $\lambda_0$

$$\sum_i \sum_j ||\lambda_0(X_{ij}) - \hat{\lambda}(X_{ij}|Z_i)/exp(b * Z_i)|| \tag{36}$$

Note that when the baseline hazard function $\lambda_0(t)$ is known, it is suggested to make full use of this information. Moreover, the second one is generally more efficient since we consider the differences between two hazard functions at the observed time only, thus reduce the error in the smoothing process. The simulation study shows the new approach is competitive comparing with commonly used PMLE method.

### 11.2 $\lambda_0(t)$is unknown

1) If we can find the estimator of baseline hazard function $\hat{\lambda}_0(t)$ with some of the data(with Z=0) first , then we can adapt the new approach as minimizing the Hellinger distance between constructive kernel estimator of $\hat{\lambda}_0$ and the estimated $\hat{\lambda}_0$

$$\sum_i \sum_j ||\hat{\lambda}_0(X_{ij}) - \hat{\lambda}(X_{ij}|Z_i)/exp(b * Z_i)|| \tag{37}$$

2) Instead of using grouped data to estimate $\hat{\lambda}_0(t)$, we first use all data to find the constructive estimator $\hat{\lambda}_0(t)$, then we can find $\hat{\beta}$ by minimizing the weighted distance between overall constructive estimator and grouped constructive estimator, and take the sum over all groups.

$$\sum_i ||[\frac{\hat{\lambda}(t|Z_i)}{exp(bZ_i)}]^{1/2} - (\hat{\lambda}_0(t))^{1/2}]\bar{Y}(t)|| \tag{38}$$

The simulation study shows that both methods are competitive comparing with the commonly used PMLE method.

## 12 advanced experiment

### 12.1 outlier

To check the sensibility against outliers of both methods, we further replace the last observation in group 2 with a random large number from $N(40, 1)$ and carry out the estimation tests.

| | new approach | | PMLE | |
|---|---|---|---|---|
| | Bias | MSE | Bias | MSE |
| N=50 | -0.06233 | 0.014605 | -0.02592 | 0.010347 |
| N=100 | -0.04448 | 0.007121 | -0.01635 | 0.005009 |
| N=300 | -0.02071 | 0.002075 | -0.01024 | 0.00164 |

Compared with the former results where the observations are not contaminated, we can see that the new approach works worse when contamination occurs but it is still competitive comparing with PMLE.

## 12.2 quadratic setting

Here we consider the Cox proportional hazards model where $\beta$ is one-dimensional with true value $\beta_0$=0.5, while $\lambda_0(t) = 0.1 + 0.2t + 0.3t^2$. Covariate Z can take 0, 1 and 2.
So given Z, hazard function is:

$$\lambda(t|Z) = exp(\beta_0 * Z) * (0.1 + 0.2t + 0.3t^2)$$

$\dot{\text{C}}$umulative hazard function is:

$$H(t|Z) = exp(\beta_0 * Z) * (0.1t + 0.1t^2 + 0.1t^3)$$

Further, we have survival function:

$$\begin{aligned}
S(t|Z) &= exp[-H(t|Z)] \\
&= exp[-exp(\beta_0 * Z) * (0.1t + 0.1t^2 + 0.1t^3)] \\
f(t|Z) &= -\frac{dS(t|Z)}{dt} \\
&= (0.1 + 0.2t + 0.3t^2) * exp[-exp(\beta_0 * Z) * (0.1t + 0.1t^2 + 0.1t^3)]
\end{aligned}$$

so we can generate survival time through independent sampler from three groups and we know that the baseline hazard function is $\lambda_0(t) = 0.1+0.2t+0.3t^2$ while true value of $\beta$ is $\beta_0$=0.5.

# 13 summary

## 13.1 method improvement

Minimizing $\sum_{i=1}^{M} ||[\hat{\lambda}(t|Z_i)/exp(bZ_i)]^{1/2} - (\hat{\lambda}_0(t))^{1/2}||$ leads to better result than minimizing $\sum_{i=1}^{M} ||\hat{\lambda}^{1/2}(t|Z_i) - (\hat{\lambda}_0(t) * exp(bZ_i))^{1/2}||$ since the latter one might expand the measurement error.

## 13.2 comparison with PMLE

Overall, estimator maximizing partial(conditional) likelihood performs better than the new approach. However, the new approach is less sensitive when censoring rate is high.

## 13.3 advantages and shortcomes

# References