Air Pollution Studies

Bingyu Wang

April 12, 2017

1 Case Crossover

1.1 General Framework

Let X_{it}^l be the exposure for person i belonging to location l and in interval t, $t = 1, \ldots, T$ and let Y_{it}^l indicates whether person i has the event at location l in interval t(1 - yes, 0 - no). Assume that the outcome Y_{it}^l is rare and that the probability that subject i fails in interval t at location l is given by the relative risk model:

$$\lambda_i(t, X_{it}^l) = \lambda_{it} \exp(\beta X_{it}^l) = \lambda_i \exp(\beta X_{it}^l + \gamma_{it}) \tag{1}$$

Each person is assumed to have his own baseline risk λ_{it} at time t consisting of two parts:

1. λ_i is a constant frailty for person i;
2. $\exp(\gamma_{it})$ is the effect of unmeasured time-varying factors on his risk.

1.2 Case-crossover design

In the case-crossover approach, the exposure of cases in interval t_i is compared to the exposures from a set of references periods, where t_i is event interval and $W(t_i)$ is a set of references periods. For example, $t_i = 8$ indicates the event was on the 8th day and $W(8) = \{7, 8, 9\}$ means the day before and the day after, including itself as the reference periods. The only assumption of a case-crossover design is that the time-varying effect γ_{it} is constant for all $t \in W(t_i)$.

Conditional on an individual being a case within a pre-specified reference windo $W(t_i)$, the probability $p_{it_i}^l$ that subject i belonging to l location and fails at time t_i is

$$\begin{aligned}
p_{it_i}^l &= P(T_i = t_i | X, W(t_i), \sum_{m=1}^{T} Y_{im}^l = 1, L_i = l) && (2)\\
&= \frac{P(T_i = t_i, \sum_{m=1}^{T} Y_{im}^l = 1, L_i = l | X, W(t_i))}{\sum_{j \in W(t_i)} P(T_i = j, \sum_{m=1}^{T} Y_{im}^l = 1, L_i = l | X, W(t_i))} && (3)\\
&= \frac{\lambda_i \exp\{\beta X_{it_i}^l + \gamma_{it_i}\}}{\sum_{j \in W(t_i)} \lambda_i \exp\{\beta X_{ij}^l + \gamma_{ij}\}} && (4)\\
&= \frac{\exp\{\beta X_{it_i}^l\}}{\sum_{j \in W(t_i)} \exp\{\beta X_{ij}^l\}} && (5)
\end{aligned}$$

which is free of terms λ_i and γ_{it_i}.

1.3 Derivation

The likelihood is defined as following, assuming subjects are independent.

$$\mathcal{L}(\beta) = \prod_{i=1}^{n} p_{it_i}^{l} = \prod_{i=1}^{n} \left(\frac{\exp\{\beta X_{it_i}^{l}\}}{\sum_{j \in W(t_i)} \exp\{\beta X_{ij}^{l}\}} \right) \tag{6}$$

Log-likelihood:

$$\ell(\beta) = \sum_{i=1}^{n} \log p_{it_i}^{l} = \sum_{i=1}^{n} \left(\beta X_{it_i}^{l} - \log \sum_{j \in W(t_i)} \exp\{\beta X_{ij}^{l}\} \right) \tag{7}$$

1.3.1 Derivation I

Take derivation directly:

$$\frac{\partial \ell(\beta)}{\partial \beta} = \sum_{i=1}^{n} \left(X_{it_i}^{l} - \frac{1}{\sum_{j \in W(t_i)} \exp\{\beta X_{ij}^{l}\}} \sum_{m \in W(t_i)} \exp\{\beta X_{im}^{l}\} X_{im}^{l} \right) \tag{8}$$

$$= \sum_{i=1}^{n} \left(X_{it_i}^{l} - \sum_{m \in W(t_i)} X_{im}^{l} \frac{\exp\{\beta X_{im}^{l}\}}{\sum_{j \in W(t_i)} \exp\{\beta X_{ij}^{l}\}} \right) \tag{9}$$

This can be used as the updating rule for β. The complexity of this updating rule is correlated to number of subjects(persons) and the size of window, which can be written as $O(n|W|)$

1.3.2 Derivation II

This derivation will be based on group information. Denote the observed number of events Y_t^l at location l in interval time t is $Y_t^l = \sum_{i \in \mathcal{I}} Y_{it}^l$, where $\mathcal{I}$ the subjects satisify the same time t, same location l and same exposures X.

If we assume the group subjects share the same exposure, $X_{it}^l = X_t^l$, the Log-likelihood could be written as

$$
\begin{aligned}
\frac{\partial \ell(\beta)}{\partial \beta} &= \sum_{i=1}^{n} \left(X_{it_i}^l - \sum_{m \in W(t_i)} X_{im}^l \frac{\exp\{\beta X_{im}^l\}}{\sum_{j \in W(t_i)} \exp\{\beta X_{ij}^l\}} \right) && (10)\\
&= \sum_{l \in L} \sum_{i \in l} \left(X_{it_i}^l - \sum_{m \in W(t_i)} X_{im}^l \frac{\exp\{\beta X_{im}^l\}}{\sum_{j \in W(t_i)} \exp\{\beta X_{ij}^l\}} \right) && (11)\\
&= \sum_{l \in L} \left[\sum_{i \in l} \left(X_{t_i}^l - \sum_{m \in W(t_i)} X_{m}^l \frac{\exp\{\beta X_{m}^l\}}{\sum_{j \in W(t_i)} \exp\{\beta X_{j}^l\}} \right) \right] && (12)\\
&= \sum_{l \in L} \left[\sum_{t=1}^{T} Y_t^l \left(X_{t}^l - \sum_{m \in W(t)} X_{m}^l \frac{\exp\{\beta X_{m}^l\}}{\sum_{j \in W(t)} \exp\{\beta X_{j}^l\}} \right) \right] && (13)\\
&= \sum_{l \in L} \left[\sum_{t=1}^{T} Y_t^l X_t^l - \sum_{t=1}^{T} Y_t^l \sum_{m \in W(t)} X_{m}^l \frac{\exp\{\beta X_{m}^l\}}{\sum_{j \in W(t)} \exp\{\beta X_{j}^l\}} \right] && (14)\\
&= \sum_{l \in L} \left[\sum_{t=1}^{T} Y_t^l X_t^l - \sum_{t=1}^{T} \sum_{m=1}^{T} Y_t^l X_{m}^l \frac{I(m \in W(t)) \exp\{\beta X_{m}^l\}}{\sum_{j=1}^{T} I(j \in W(t)) \exp\{\beta X_{j}^l\}} \right] && (15)\\
&= \sum_{l \in L} \left[\sum_{t=1}^{T} Y_t^l X_t^l - \sum_{m=1}^{T} \left(X_{m}^l \sum_{t=1}^{T} Y_t^l \frac{I(m \in W(t)) \exp\{\beta X_{m}^l\}}{\sum_{j=1}^{T} I(j \in W(t)) \exp\{\beta X_{j}^l\}} \right) \right] && (16)\\
&= \sum_{l \in L} \left[\sum_{t=1}^{T} Y_t^l X_t^l - \sum_{t=1}^{T} \left(X_{t}^l \sum_{m=1}^{T} Y_m^l \frac{I(t \in W(m)) \exp\{\beta X_{t}^l\}}{\sum_{j=1}^{T} I(j \in W(m)) \exp\{\beta X_{j}^l\}} \right) \right] && (17)\\
&= \sum_{l \in L} \left[\sum_{t=1}^{T} X_t^l \left(Y_t^l - \sum_{m=1}^{T} Y_m^l \frac{I(t \in W(m)) \exp\{\beta X_{t}^l\}}{\sum_{j=1}^{T} I(j \in W(m)) \exp\{\beta X_{j}^l\}} \right) \right] && (18)\\
&= \sum_{l \in L} \left[\sum_{t=1}^{T} X_t^l \left(Y_t^l - \sum_{m \in R(t)} Y_m^l \frac{\exp\{\beta X_{t}^l\}}{\sum_{j \in W(m)} \exp\{\beta X_{j}^l\}} \right) \right] && (19)\\
&= \sum_{l \in L} \left[\sum_{t=1}^{T} X_t^l \left(Y_t^l - \exp\{\beta X_{t}^l\} \sum_{m \in R(t)} \frac{Y_m^l}{\sum_{j \in W(m)} \exp\{\beta X_{j}^l\}} \right) \right] && (20)
\end{aligned}
$$

Now the updating rule has been transformed into the one related to number of locations, times and references window size. The time complexity is $O(LT|W|)$. The advantage of this updating method is that we shrink lots of duplicated persons row data into much smaller number of groups, which share the same location l and exposures(features) X_t^l at the same time t.

2 Cox Proportional Hazards

2.1 Model

The hazard function for the Cox proportional hazard model has the form:

$$h(y_i|\boldsymbol{\beta}) = h_0(y_i|\boldsymbol{\beta}) \exp(\boldsymbol{\beta}^T \boldsymbol{x}_i)$$

where

- $h_0(y_i|\boldsymbol{\beta})$: the unspecified baseline hazard function.

- $i \in [1, n]$: each individual and n is the total number of individuals. .
- $y_i = \min(t_i, c_i)$: t_i is time-to-event(failure time) and c_i is right-censoring time.
- $\boldsymbol{x}_i = (x_{i1}, x_{i2}, \ldots, x_{ip})^T$: p-vector of features for the individual i.
- $\boldsymbol{\beta} = (\beta_1, \beta_2, \ldots, \beta_p)^T$: p-vector of underlying model parameters.

The n observed data $\boldsymbol{D} = \{(y_i, \delta_i, \boldsymbol{x}_i) : i = 1, \ldots, n\}$, where $\delta_i = I(t_i \leq c_i)$ is an indicator variable such that $\delta_i = 1$ if the observation is not censored and 0 otherwise.

2.2 Partial Likelihood

To estimate the underlying parameters $\boldsymbol{\beta}$, the original likelihood $L(\boldsymbol{\beta}|\boldsymbol{D})$ is hard to maximize. Cox proposed to maximize the partial likelihood:

$$L_p(\boldsymbol{\beta}|\boldsymbol{D}) = \prod_{i=1}^{n} \left(\frac{\exp(\boldsymbol{\beta}^T \boldsymbol{x}_i)}{\sum_{t \in R(y_i)} \exp(\boldsymbol{\beta}^T \boldsymbol{x}_t)} \right)^{\delta_i}$$

where $R(y_i)$ is the risk set of the i-th observation, defined as $R(y_i) = \{t : y_t \geq y_i\}$

2.3 Estimate Parameters

Maximizing partial likelihood is equivalent to maximize log-partial likelihood:

$$l_p(\boldsymbol{\beta}|\boldsymbol{D}) = \sum_{i=1}^{n} \delta_i \left\{ \boldsymbol{\beta}^T \boldsymbol{x}_i - \log \left(\sum_{t \in R(y_i)} \exp\left(\boldsymbol{\beta}^T \boldsymbol{x}_t\right) \right) \right\}$$

The negated log-partial likelihoods are convex, and a wide range of optimization algorithms can be utilized. In our experiments, we apply Limited-memory BFGS algorithm to minimize the negated log-partial likelihoods:

$$\boldsymbol{\beta}^* = \arg\min_{\boldsymbol{\beta}} -l_p(\boldsymbol{\beta}|\boldsymbol{D})$$

To apply L-BFGS, we have to calculate the first derivatives of $-l_p(\boldsymbol{\beta}|\boldsymbol{D})$ with respect to $\boldsymbol{\beta}$:

$$-l'_p(\beta_j) = -\sum_{i=1}^{n} \delta_i \left(x_{ij} - \frac{\sum_{t \in R(y_i)} x_{tj} \exp\left(\boldsymbol{\beta}^T \boldsymbol{x}_t\right)}{\sum_{t \in R(y_i)} \exp\left(\boldsymbol{\beta}^T \boldsymbol{x}_t\right)} \right)$$

To produce approximate standard errors for the regression coefficients, we need to calculate the second derivatives:

$$-l''_p(\beta_j) = \sum_{i=1}^{n} \delta_i \left\{ \frac{\sum_{t \in R(y_i)} x_{tj}^2 \exp\left(\boldsymbol{\beta}^T \boldsymbol{x}_t\right)}{\sum_{t \in R(y_i)} \exp\left(\boldsymbol{\beta}^T \boldsymbol{x}_t\right)} - \left(\frac{\sum_{t \in R(y_i)} x_{tj} \exp\left(\boldsymbol{\beta}^T \boldsymbol{x}_t\right)}{\sum_{t \in R(y_i)} \exp\left(\boldsymbol{\beta}^T \boldsymbol{x}_t\right)} \right)^2 \right\}$$

3 Proportional Hazard Model with Frailties

3.1 model

For estimation of zip code specific long-term air-pollution mortality risks, we will consider proportional hazard model with multivariate random effects(frailties). For this model, event times from the same group (zip code area) are likely to be correlated. Suppose there are C number of clusters(zip code areas). Then the proportional hazard model with frailties has the form:

$$h(y_{ij}|\boldsymbol{\beta}_i) = h_0(y_{ij}|\boldsymbol{\beta}_i) \exp\left(\boldsymbol{\beta}_i^T \boldsymbol{x}_{ij} + b_i\right)$$

where

- $h_0(y_{ij}|\boldsymbol{\beta}_i)$: the baseline hazard function.
- $i \in [1, C]$: each clusters, C is the total number clusters(zip code areas).
- $j \in [1, n_i]$: each individual from cluster i, and n_i is the total number of individual from ith cluster.
- $y_{ij} = \min(t_{ij}, c_{ij})$: where t_{ij} is the failure time, and c_{ij} is the right-censorinbg time.
- δ_{ij}: is an indicator variable such that $\delta_{ij} = 1$ if the observation is not censored and 0 otherwise.
- $\boldsymbol{x}_{ij}$: p-vector of features for the individual j in cluster i.
- $\boldsymbol{\beta}_i = (\beta_{i1}, \beta_{i2}, \ldots, \beta_{ip})^T$: p-vector of cluster-specific underlying model parameters.
- b_i: the cluster specific random effects.

3.2 Estimate Parameters

To solve the frailty model, serveral methods have been proposed these years. Xue and Ding (1999) used a Gibbs Sampling approach. Ripatti and Palmgren (2000) considered a penalized partial likelihood approach. Vaida and Xu (2000) proposed a nonparametric maximum likelihood estimator, obtained using a Monte Carlo EM algorithm. Cortinas-Abrahantes et al. A comprehensive comparison of these methods can be found in Gamst et al.(2009). We will follow one of these methods mentioned above or other related methods, to solve this problem.

References

[1] Jose Cortinas Abrahantes and Tomasz Burzykowski. A version of the em algorithm for proportional hazard model with random effects. 2005.

[2] Cox R David. Regression models and life tables (with discussion). *Journal of the Royal Statistical Society*, 34:187–220, 1972.

[3] Anthony Gamst, Michael Donohue, and Ronghui Xu. Asymptotic properties and empirical evaluation of the npmle in the proportional hazards mixed-effects model. *Statistica Sinica*, pages 997–1011, 2009.

[4] Kyeong Eun Lee, Yongku Kim, and Ronghui Xu. Bayesian variable selection under the proportional hazards mixed-effects model. *Computational statistics & data analysis*, 75:53–65, 2014.

[5] Tao-Wen Liu. A regularized limited memory bfgs method for nonconvex unconstrained minimization. *Numerical Algorithms*, 65(2):305–323, 2014.

[6] Yun Lu and Scott L Zeger. On the equivalence of case-crossover and time series methods in environmental epidemiology. *Biostatistics*, 8(2):337–344, 2007.

[7] Sushil Mittal, David Madigan, Randall S Burd, and Marc A Suchard. High-dimensional, massive sample-size cox proportional hazards regression for survival analysis. *Biostatistics*, page kxt043, 2013.

[8] Samuli Ripatti and Juni Palmgren. Estimation of multivariate frailty models using penalized partial likelihood. *Biometrics*, 56(4):1016–1022, 2000.

[9] Florin Vaida and Ronghui Xu. Proportional hazards model with random effects. *Statistics in medicine*, 19(24):3309–3324, 2000.