Modèle à hasards non proportionnels et survie marginale

Roxane Duroux [1], Cécile Chauvel [2] & John O'Quigley [3]

[1] *Laboratoire de Statistique Théorique et Appliquée, Université Pierre et Marie Curie - Paris VI, Paris, France, roxane.duroux@upmc.fr*
[2] *Laboratoire Jean Kuntzmann, Université Joseph Fourier - Grenoble, France, cecile.chauvel@imag.fr*
[3] *Laboratoire de Statistique Théorique et Appliquée, Université Pierre et Marie Curie - Paris VI, Paris, France, john.oquigley@upmc.fr*

Résumé. Nous nous plaçons dans le cadre de l'analyse de survie, c'est-à-dire de l'analyse de données censurées. On notera T la variable aléatoire modélisant le temps de décès et Z un vecteur de covariables. Nous souhaitons modéliser la loi de T sachant Z en prenant en compte dans l'analyse la fonction de survie marginale de T. Nous proposerons dans ce cadre un nouvel estimateur convergent vers un paramètre β^* proche de $E[\beta(T)]$.

Mots-clés. Modèle à hasards proportionnels de Cox, Equation estimatrice, Effets non constants, Score pondéré.

Abstract. We place ourself in the context of survival analysis, *i.e.* analysis of censored data. We denote by T the failure time random variable and Z a vector of covariates. We would like to include in our analysis the marginal survival function of T. In this context, we propose a new consistent estimator of a parameter β^* close to $E[\beta(T)]$.

Keywords. Cox's proportional hazards model, Estimating equation, Time-varying effects, Weighted score.

1 Présentation de l'étude

Nous nous plaçons dans le cadre de l'analyse de survie, c'est-à-dire de l'analyse de données censurées. On notera T la variable aléatoire de fonction de répartition F modélisant le temps de décès, $Z(.) \in \mathbb{R}^d$ un vecteur de covariables et C le mécanisme de censure que l'on choisit indépendant de T conditionnellement à $Z(.)$.

Nous souhaitons modéliser la loi de T sachant Z sans imposer une structure trop forte et aimerions prendre en compte dans l'analyse la fonction de survie marginale, S, de T. En effet, avant d'étudier la régression de T par rapport à Z, on peut posséder à l'avance de l'information sur la loi marginale de T. C'est le cas dans le cadre de la survie relative lorsqu'on utilise des données de registre par exemple. Il est alors possible d'avoir un estimateur précis de cette loi marginale. On peut aussi supposer que la variable aléatoire

T que l'on étudie est générée de la même manière que dans une autre étude qui a déjà été menée. Enfin, on peut choisir un modèle pour la loi marginale de T, en se basant sur des connaissances annexes par exemple. C'est ce à quoi on s'intéressera ici. Une fois la loi marginale connue, ou supposée connue, il suffit d'en estimer les paramètres et de les considérer comme fixés. La question naturelle qui vient alors est : est-ce que la prise en compte de cette information est utile ? Si oui, à quel point ?

On se placera dans le cadre d'un modèle à hasards non proportionnels pour le risque relatif, qu'on écrira sous la forme

$$\lambda(t|Z) = \lambda_0(t)\exp\{\beta(t)Z\}, \tag{1}$$

où λ_0 est la fonction de hasard de base qui eut être interprétée comme le hasard $\lambda(t|Z=0)$. On remarque que le modèle à hasards proportionnels de Cox (Cox (1972)) est un cas particulier de ce modèle : le cas où la fonction $\beta(.)$ est constante. Pour distinguer ces modèles, nous noterons le coefficient de régression $\beta(t)$ dans le cas d'un modèle à risques non proportionnels (Murphy et Sen (1991)) et β dans le cas du modèle de Cox. Le modèle (1) a été étudié entre autres par Moreau, O'Quigley et Mesbah (1985), O'Quigley et Pessione (1989), O'Quigley et Pessione (1991), Liang, Self et Liu (1990), Zucker et Karr (1990), Murphy et Sen (1991), Gray (1992), Hastie et Tibshirani (1993), Verweij et Houwelingen (1995), Lausen et Schumacher (1996), Marzec et Marzec (1997). Leur principal objectif était l'estimation du coefficient de régression $\beta(t)$ comme une fonction de t. Mais on pourrait aussi penser à essayer d'estimer plutôt un effet moyen, toujours lorsque cet effet varie avec le temps. En effet, on peut supposer que cette estimation sera plus simple que d'estimer totalement la fonction $\beta(t)$. De plus, on peut espérer que ce résultat donnera un premier résumé de l'information sur le coefficient de régression avant d'aller plus en avant dans l'analyse des données.

Maintenant que nous avons posé le cadre de notre étude, revenons aux questions posées plus haut. Si on se réfère à certains travaux antérieurs sur l'efficacité d'estimateurs obtenus à l'aide de la vraisemblance partielle, nous ne pouvons pas espérer de gains significatifs (Struthers et Kalbfleisch (1986), Lin (1991), Xu et O'Quigley (2000)). Un premier résultat est que, sous un modèle à risques proportionnels (Cox (1972)), si la loi de T est mal spécifiée, l'estimateur obtenu reste consistant pour le paramètre de régression du risque relatif. On constate malgré tout une faible perte d'efficacité. Par contre, dans le cas d'un modèle à risques non proportionnels, on observe un gain d'efficacité. Plus précisément, l'estimateur de la vraisemblance partielle va converger vers une quantité dépendant fortement du mécanisme de censure, même si la censure et les covariables sont indépendantes (Kalbfleisch et Prentice (1980)). Dans le cas d'un modèle à hasards non proportionnels, par exemple dans le cas d'effets diminuant avec le temps, on aimerait être capable d'estimer de manière convergente $E[\beta(T)]$.

A l'aide de la structure particulière des équations estimatrices et la connaissance de la survie marginale de T, nous définissons un nouvel estimateur $\tilde{\beta}$. En l'absence de censure, cet estimateur coïncide avec l'estimateur du maximum de vraisemblance partielle si on pose un modèle à risques proportionnels sur des données générées selon (1); Sous un modèle à risques proportionnels, les deux estimateurs sont consistants pour le vrai paramètre de régression. De plus l'estimateur est simple à calculer en pratique et la connaissance de la survie marginale de T nous apporte un gain en interprétation.

Bibliographie

[1] Cox, D.R. (1972), Regression models and life-tables, *Journal of the Royal Statistical Society, Series B (Methodological)*, 34(2), 187-220.

[2] Gray, R.J. (1992), Flexible methods for analyzing survival data using splines, with applications to breast cancer prognosis, *Journal of the American Statistical Association*, 87(420), 942-951.

[3] Hastie, T. et Tibshirani, R. (1993), Varying-coefficient models, *Journal of the Royal Statistical Society, Series B (Methodological)*, 55(4), 757-796.

[4] Kalbfleisch, J.D. et Prentice, R.L. (1980), The statistical analysis of failure time data, *Wiley seeries in probability and mathematical statistics : Applied probability and statistics.*

[5] Lausen, B. et Schumacher, M. (1996), Evaluating the effect of optimized cutoff values in the assessment of prognostic factors, *Comput. Stat. Data Anal.*, 21(3), 307-326.

[6] Liang, K.Y., Self, S.G. et Liu, X. (1990), The cox proportional hazards model with change point : an epidemiologic application, *Biometrics*, 46(3), 783-793.

[7] Lin, D.Y. (1991), Goodness-of-fit analysis for the cox regression model based on a class of parameter estimators, *Journal of the American Statistical Association*, 86(415), 725-728.

[8] Marzec, L. et Marzec, P. (1997), On fitting cox's regression model with time-dependent coefficients, *Biometrika*, 84(4), 901-908.

[9] Moreau, T., O'Quigley, J. et Mesbah, M. (1985), A global goodness-of-fit statistic for the proportional hazards model, *Journal of the Royal Statistical Society, Series C (Applied Statistics)*, 34(3), 212-218.

[10] Murphy, S.A. et Sen, P.K. (1991), Time-dependent coefficients in a cox-type regression model, *Stochastic Processes and their Applications*, 39(1), 153-180.

[11] O'Quigley, J. et Pessione, F. (1989), Score tests for homogeneity of regression effect in the proportional hazards model, *Biometrics*, 45(1), 135-144.

[12] O'Quigley, J. et Pessione, F. (1991), The problem of a covariate-time qualitative interaction in a survival study, *Biometrics*, 47(1), 101-115.

[13] Struthers, C.A. et Kalbfleisch, J.D. (1986), Misspecified proportional hazard models, *Biometrika*, 73(2), 363-369.

[14] Verweij, P. et Van Houwelingen, H. (1995), Time-dependent effects of fixed covariates in cox regression, *Biometrics*, 51(4), 1550-1556.

[15] Xu, R. et O'Quigley, J. (2000), Estimating average regression effect under non-proportional hazards, *Biostatistics*, 1(4), 423-439.

[16] Zucker, D.M. et Karr, A.F. (1990), Nonparametric survival analysis with time-dependent covariate effects : a penalized partial likelihood approach, *Ann. Statist.*, 18(1), 329-353.