Tests du Log-rank ajusté : étude de simulations

Florent Le Borgne [1,2] & Yohann Foucher [1]

[1] *Université de Nantes - EA4275-SPHERE - ITUN*

[2] *IDBC/A2com*

florent.le-borgne@etu.univ-nantes.fr

05/06/2014

Introduction

Contexte : Étude observationnelle en présence de données de survie.

- L'étude de la causalité entre l'exposition et l'événement requiert un ajustement.

  $\Rightarrow$ Estimateur Kaplan-Meier inadéquat

  $\Rightarrow$ Modèle multivarié (Cox) adapté mais perte d'information :

  Résultat souvent résumé en un seul RR, ne présente pas la possible évolution au cours du temps de ce RR (pas de représentation graphique)

Solution

Courbes de survie ajustées utilisant la méthode IPTW (Inverse Probability of Treatment Weighting) basée sur les scores de propension.

Introduction

- Le test du Log-rank est le test standard de comparaison de deux courbes de survie.
- Trois versions de ce test ont été adaptées à l'estimateur de Kaplan-Meier ajusté.
- D'autres méthodes basées sur les scores de propensions existent (stratification, appariement, IPTW).

Objectifs de notre étude de simulations

- Évaluer les performances des tests du log-rank ajustés par rapport au test de référence (modèle de Cox multivarié) en terme d'erreur de 1$^{\text{ère}}$ et 2$^{\text{nde}}$ espèces.

  $\Rightarrow$ Faut-il faire un modèle de Cox multivarié ?
- Choisir le plus performant des trois.

Notations

- On observe un échantillon de taille n.
- Soient T_i les temps de participation $(i = 1, \dots, n)$ et δ_i l'indicateur de censure ($\delta_i = 0$ si T_i est censuré à droite et $\delta_i = 1$ sinon).
- Soit X_i la variable explicative représentant le facteur d'exposition d'intérêt composé de K groupes.
- D_k le nombre de temps d'événements différents observé dans le groupe k, on a alors au temps t_j $(j = 1, \dots, D_k)$:
  - $d_{jk} = \sum_{i:t_i=t_j} \delta_i I(X_i = k)$: nombre de sujets du groupe k subissant l'évènement au temps t_j.
  - $Y_{jk} = \sum_{i:t_i \geq t_j} I(X_i = k)$: nombre de sujets du groupe k à risque au temps t_j.
  - Soient $d_j = \sum_{k=1}^{K} d_{jk}$ et $Y_j = \sum_{k=1}^{K} Y_{jk}$.

La pondération IPTW

- La méthode IPTW propose de corriger la contribution de chaque individu par un poids $w_{ik} = 1/p_{ik}$

  où $p_{ik} = P(X_i = k|Z_i)$

  et Z_i le vecteur des covariables potentiellement facteurs de confusion.

- Les nombres pondérés d'événements et d'individus à risque peuvent alors être obtenus :
  - $d_{jk}^w = \sum_{i:t_i=t_j} w_{ik}\delta_i I(X_i = k)$
  - $Y_{jk}^w = \sum_{i:t_i \geq t_j} w_{ik} I(X_i = k)$
  - $d_j^w = \sum_{k=1}^K d_{jk}^w$ et $Y_j^w = \sum_{k=1}^K Y_{jk}^w$.

- Pour la suite considérons seulement deux groupes, notés $X = 0$ et $X = 1$.

Log-rank ajusté proposé par Xu et al. (2012)

1. Xu et al. proposent un test du log-rank ajusté semblable au test du log-rank standard en remplaçant simplement :
   - Les nombres d'événements par les nombres pondérés d'événements.
   - Les nombres de sujets à risque par les nombres pondérés de sujets à risque.

- La statistique de test obtenue est : $G^w / \sqrt{\mathrm{Var}(G^w)}$ où :
  - D est le nombre de temps différents pour lesquels des événements sont observés quelque soit le groupe.
  - $G^w = \sum_{j=1}^{D} d_{j1}^w - Y_{j1}^w \left(\frac{d_j^w}{Y_j^w} \right)$
  - $\mathrm{Var}(G^w) = \sum_{j=1}^{D} \left\{ \frac{Y_{j0}^w Y_{j1}^w d_j^w (Y_j^w - d_j^w)}{(Y_j^w)^2 (Y_j^w - 1)} \right\}$

Xu S. and al. Extension of kaplan-meier methods in observational studies with time-varying treatment. *Value in Health.* (2012)

Log-rank ajusté proposé par Sugihara (2010)

2. Une deuxième variante du test du log-rank ajusté est proposée par Sugihara.

- Se différencie de la première par la formule de la variance utilisée.

$$G^w = \sum_{j=1}^{D} d_{j1}^w - Y_{j1}^w \left(\frac{d_j^w}{Y_j^w}\right)$$

$$\mathrm{Var}(G^{w'}) = \sum_{j=1}^{D} \left\{ \frac{d_j(Y_j - d_j)}{Y_j(Y_j - 1)} \sum_{i=1}^{Y_j} \left[\left(\frac{Y_{j0}^w}{Y_j^w}\right)^2 {w_i}^2 X_i + \left(\frac{Y_{j1}^w}{Y_j^w}\right)^2 {w_i}^2 (1 - X_i) \right] \right\}$$

Sugihara M. Survival analysis using inverse probability of treatment weighted methods based on the generalized propensity score. *Pharmaceutical statistics.* (2010)

Log-rank ajusté proposé par Xie and Liu (2005)

3. Xie et Liu proposent une autre adaptation du test du log-rank en ajustant les poids des individus au cours du temps.

- Au temps t_j ($j = 1, \dots, D_k$), le poids pour un individu i du groupe k est ré-assigné tel que :

$$w'_{ijk} = w_{ik}.Y_{jk}/Y^{w}_{jk}$$

- Les nombres pondérés d'événements et d'individus à risque deviennent :

$$d^{w'}_{jk} = \sum_{i:t_i=t_j} w'_{ijk}\delta_i I(X_i = k)$$
$$\text{et } Y^{w'}_{jk} = \sum_{i:t_i \geq t_j} w'_{ijk} I(X_i = k)$$

Xie J. and Liu C. Adjusted kaplan-meier estimator and log-rank test with inverse probability of treatment weighting for survival data. *Statistics in medicine.* (2005)

Log-rank ajusté proposé par Xie and Liu (2005)

- D'après les mêmes formules que le test proposé par Sugihara.
- Avec des poids différents.

$$G^{w'} = \sum_{j=1}^{D} d_{j1}^{w'} - Y_{j1}^{w'} \left(\frac{d_j^{w'}}{Y_j^{w'}}\right)$$

$$\mathrm{Var}(G^{w'}) = \sum_{j=1}^{D} \left\{ \frac{d_j(Y_j - d_j)}{Y_j(Y_j - 1)} \sum_{i=1}^{Y_j} \left[\left(\frac{Y_{j0}^{w'}}{Y_j^{w'}} \right)^2 w_{ij}^{'2} X_i + \left(\frac{Y_{j1}^{w'}}{Y_j^{w'}} \right)^2 w_{ij}^{'2} (1 - X_i) \right] \right\}$$

Xie J. and Liu C. Adjusted kaplan-meier estimator and log-rank test with inverse probability of treatment weighting for survival data. *Statistics in medicine.* (2005)

Deux autres modèles basés sur les scores de propension

- Modèle de Cox univarié pondéré proposé par Cole et Hernán (2004).
  - L'exposition : seule variable dans le modèle
  - Pondéré par les poids w_{ik}

Cole S.R. and Hernán M. Adjusted survival curves with inverse probability weights. *Computer Methods and Programs in Biomedicine* (2004)

- Appariement sur le logit du score de propension.
  - Appariement 1 : 1 sans remise avec le plus proche voisin
  - Caliper égal à 0.2 x écart-type
  - Test du log-rank stratifié

Rosenbaum PR. and Rubin DB. Constructing a control group using multivariate matched sampling methods that incorporate the propensity score. *The American Statistician* (1985)

Etude de simulations

- Simulations limitées à 5 variables :
  - 1 exposition binaire
  - 4 facteurs de confusion
- Performances des différents modèles comparées pour différents :
  - Taux de censure (0.30 et 0.68)
  - Tailles d'échantillon (100, 250, 500 et 1500)
  - Pourcentages de sujets exposés (5%, 20% et 40%)
  - Coefficients β_1 associés à la variable d'exposition d'intérêt (0, 0.250, 0.365, 0.500)
- Lorsque $\beta_X = 0$ on calcule le pourcentage de rejet de l'hypothèse nulle (risque de 1$^{\text{ère}}$ espèce).
- Lorsque $\beta_X \neq 0$ on calcule le pourcentage de non rejet de l'hypothèse nulle (risque de 2$^{\text{nde}}$ espèce).

Résultats des simulations

β_X	n	Censoring rate $\approx$ 0.68 Cox multivarié	Xie	Sugihara	Xu	Cox pondéré	Apparié
(a) 0.000	100	5.69	6.12	6.94	19.53	10.92	4.70
	250	5.31	6.27	6.58	23.42	8.80	4.81
	500	4.59	5.46	5.73	25.39	7.10	5.29
	1500	4.85	5.43	5.58	27.58	6.38	4.93
(b) 0.250	100	90.89	89.80	89.04	74.44	84.04	94.75
	250	86.05	86.37	86.20	65.78	82.57	92.28
	500	77.57	82.54	82.59	56.41	79.11	89.64
	1500	43.78	64.74	65.14	31.99	60.86	75.99
(b) 0.365	100	86.90	85.87	85.35	70.03	79.50	93.10
	250	74.66	79.30	79.20	55.24	73.37	88.20
	500	56.71	69.26	69.68	40.35	64.21	81.67
	1500	13.30	40.37	40.95	13.79	36.97	54.16
(b) 0.500	100	80.13	79.79	79.06	61.30	71.88	90.89
	250	58.80	67.87	68.06	41.57	60.14	81.77
	500	32.03	54.21	54.94	24.40	47.34	69.80
	1500	0.99	16.15	16.81	2.95	14.76	27.00

TABLE 1: Taux d'erreur obtenus à partir de données avec 40% de sujets exposés et 68% de censure. (a) Risque de 1$^{\text{ère}}$ espèce en pourcentage. (b) Risque de 2$^{\text{nde}}$ espèce en pourcentage. 10 000 échantillons simulés pour chaque scénario.

Conclusion étude de simulations

- Meilleures performances obtenues par le modèle de Cox multivarié.
- Modèle apparié : perte de puissance statistique.
- Modèle de Cox univarié pondéré : un risque de 1ère espèce plus important.
- Parmi les 3 tests du log-rank ajustés :
  - Celui proposé par Xu et al. ne respecte pas le risque de 1ère espèce.
  - Les deux autres montre des risques de 1ère et 2nde espèces très légèrement supérieurs à ceux du modèle de Cox multivarié.
  - Un risque de 1ère espèce légèrement meilleur pour celui proposé par Xie et Liu.

Discussion(1)

- Deux limites apparaissent à notre étude :
  - Nous avons uniquement considéré le cas où la variable d'exposition est binaire.
    - Les courbes de survie ajustées sont facilement généralisables pour plus de 2 groupes (régression logistique multinomiale)
    - Le test du log-rank ajusté nécessite plus de développements
  - Seul le contexte où les données respectent l'hypothèse de proportionnalité des risques a été simulé.

Discussion(2)

- Pour évaluer la causalité entre une exposition et un événement dans une étude observationnelle nous retenons 2 bonnes méthodes.
  - Le modèle de Cox multivarié
  - Les courbes de survie ajustées avec le test du log-rank proposé par Xie et Liu
- Le modèle de Cox multivarié : le plus efficace, requiert la vérification d'hypothèse, résume le résultat en un RR.
- Les courbes de survie ajustées : illustrent plus précisément les différences de survie entre les groupes, moins bonnes performances du test du log-rank ajusté associé.

Références

- Xu S. and al. Extension of kaplan-meier methods in observational studies with time-varying treatment. *Value in Health.* (2012)
- Sugihara M. Survival analysis using inverse probability of treatment weighted methods based on the generalized propensity score. *Pharmaceutical statistics.* (2010)
- Xie J. and Liu C. Adjusted kaplan-meier estimator and log-rank test with inverse probability of treatment weighting for survival data. *Statistics in medicine.* (2005)
- Cole S.R. and Hernán M. Adjusted survival curves with inverse probability weights. *Computer Methods and Programs in Biomedicine* (2004)
- Rosenbaum PR. and Rubin DB. Constructing a control group using multivariate matched sampling methods that incorporate the propensity score. *The American Statistician*