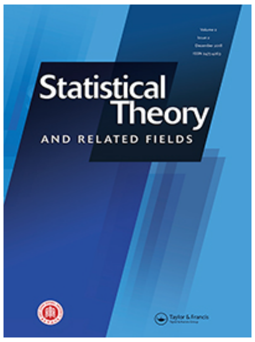

Statistical Theory and Related Fields

ISSN: 2475-4269 (Print) 2475-4277 (Online) Journal homepage: https://www.tandfonline.com/loi/tstf20

Some issues on longitudinal data with nonignorable dropout, a discussion of "Statistical Inference for Nonignorable Missing-Data Problems: A Selective Review" by Niansheng Tang and Yuanyuan Ju

Lei Wang

STATISTICAL THEORY AND RELATED FIELDS
2018, VOL. 2, NO. 2, 137–139
https://doi.org/10.1080/24754269.2018.1522575

Taylor & Francis
Taylor & Francis Group

SHORT COMMUNICATION

Check for updates

Some issues on longitudinal data with nonignorable dropout, a discussion of "Statistical Inference for Nonignorable Missing-Data Problems: A Selective Review" by Niansheng Tang and Yuanyuan Ju

Lei Wang

School of Statistics and Data Science & LPMC, Nankai University, Tianjin, People's Republic of China

ARTICLE HISTORY Received 7 September 2018; Accepted 10 September 2018

We thank Tang and Ju for their review on statistical inference for univariate response data with nonignorable missing. In this paper, we mainly discuss some issues on longitudinal data with nonignorable dropout. In research areas such as medicine, population health, economics, social sciences and sample surveys, data are often collected from every sampled subject at T time points, which are referred to as longitudinal data. Let $Y=(y_1,\dots,y_T)$ be a T dimensional vector of the study variable with distribution denoted by $p(Y)$, and X be a q-dimensional time-independent continuous covariate associated with Y. Our interest is to estimate parameters in $p(Y)$ such as the mean vector $E(Y)$. We consider the situation where X is always observed, but subjects may drop out prior to the end of the study, which results in incomplete Y data. Let $R=(r_1,\dots,r_T)$ be the vector of response indicators, where $r_t=1$ if y_t is observed and $r_t=0$ if $y_t,\dots,y_T$ are not observed. Dropout is ignorable if the propensity $p(R\mid Y,X)$ is a function of the observed values (Little & Rubin, 2002), where $p(\cdot\mid\cdot)$ is a generic notation for conditional distribution or density. Otherwise, dropout is nonignorable.

When missing data are ignorable dropout, Little (1995) presented some well-established methods. However, in practice dropout is often nonignorable (Troxel, Harrington, & Lipsitz, 1998). In this case, for identifiability Wang, Qi, and Shao (2018) assumed that $X=(U,Z)$ with an instrument Z satisfying $p(R\mid Y,X)=p(R\mid Y,U)$ and that $p(Y\mid U,Z)$ depends on Z. Furthermore, a parametric dropout propensity model is also imposed as follows,

$$\begin{aligned}\Pr(r_t=1\mid r_{t-1}=1,U,y_1,\dots,y_t)\\ =\Psi(\alpha_t+\beta_t y_t+\gamma_t\mathcal{O}_t^{\mathrm{T}}),\quad t=1,\dots,T,\end{aligned}\tag{1}$$

where $\mathcal{O}_t=(U,y_1,\dots,y_{t-1})$, $\theta_t=(\alpha_t,\beta_t,\gamma_t)$ is a row vector of unknown parameters, Ψ is a known monotone function defined on $[0,1]$, r_0 is defined to be 1 and $\mathcal{O}_1$ is defined to be U. For $t=1,\dots,T$, define the following $L=t+q$ estimating equations

$$g_t(Y,X,R,\theta_t)=r_{t-1}\{r_t\omega(\theta_t)-1\}S_t,$$

where $S_t=(1,Z,\mathcal{O}_t)^{\mathrm{T}}$ and $\omega(\theta_t)=\{\Psi(\alpha_t+\beta_t y_t+\gamma_t\mathcal{O}_t^{\mathrm{T}})\}^{-1}$. If θ_t^0 is the true value of θ_t, it can be verified that

$$E\{g_t(Y,X,R,\theta_t^0)\}=0.$$

Let $\{(X_i,Y_i,R_i):i=1,\dots,n\}$ be an independent and identically distributed sample from (X,Y,R), y_{ti} and r_{ti} be the tth components of Y_i and R_i for $t\le T$, respectively, where the covariate vector X_i is fully observed and the response y_{ti} is observed if and only if $r_{ti}=1$. A sample version of estimating equation is $G_t(\theta_t)=n^{-1}\sum_{i=1}^n g_t(Y_i,X_i,R_i,\theta_t)=0$. Wang et al. (2018) applied the two-step generalised method of moments (GMM) to estimate the unknown parameter vector θ_t as follows:

$$\hat\theta_t=\mathrm{argmin}_{\theta_t\in\Theta_t}G_t(\theta_t)^{\mathrm{T}}\hat W_t^{-1}G_t(\theta_t),\tag{2}$$

where Θ_t is the parameter space for θ_t, $\hat W_t^{-1}$ is the inverse of the $L\times L$ matrix $n^{-1}\sum_{i=1}^n g_t(Y_i,X_i,R_i,\hat\theta_t^{(1)})$ $g_t(Y_i,X_i,R_i,\hat\theta_t^{(1)})^{\mathrm{T}}$ and $\hat\theta_t^{(1)}=\mathrm{argmin}_{\theta_t\in\Theta_t}G_t(\theta_t)^{\mathrm{T}}$ $G_t(\theta_t)$. Then, the marginal mean of y_t, $\mu_t=E(y_t)$, can be estimated by

$$\hat\mu_t^{\mathrm{ipw1}}=\frac{1}{n}\sum_{i=1}^n\frac{r_{ti}y_{ti}}{\hat\pi_{ti}}\quad\text{or}$$

$$\hat\mu_t^{\mathrm{ipw2}}=\sum_{i=1}^n\frac{r_{ti}y_{ti}}{\hat\pi_{ti}}\Bigg/\sum_{i=1}^n\frac{r_{ti}}{\hat\pi_{ti}},\tag{3}$$

where $\hat\pi_{ti}=\prod_{s=1}^t\Psi(\hat\alpha_s+\hat\beta_s y_{si}+\hat\gamma_s\mathcal{O}_{si}^{\mathrm{T}})$ and $\hat\theta_t=(\hat\alpha_t,\hat\beta_t,\hat\gamma_t)$ is a consistent estimator of θ_t. In theory, the method proposed by Wang et al. (2018) can handle the

CONTACT Lei Wang lwangstat@nankai.edu.cn School of Statistics andData Science & LPMC, Nankai University, Tianjin, People's Republic of China

longitudinal data with nonignorable dropout. However, its performance may be hindered by the following three problems.

Optimal estimation The efficiency of proposed GMM estimators $\hat{\theta}_t$ in (2) depends on the choice of S_t, which may not be optimal if we only use the first order moments of data. Other moments or characteristics of Z and $\mathcal{O}_t$ may provide more information and, hence, result in more efficient GMM estimators. When Y is univariate, Morikawa and Kim (2016) investigated the efficient estimation of the parameters in the propensity and derived the corresponding semiparametric efficiency bound. In addition, Morikawa, Kim, and Kano (2017) proposed to improve the efficiency based on the semiparametric maximum likelihood approach. Recently, Ai, Linton, and Zhang (2018) proposed a simple and efficient estimation method based on the GMM. If the number of moments increases appropriately, they showed that the GMM estimator can achieve the semiparametric efficiency bound derived in Morikawa and Kim (2016), but under weaker regularity conditions. Motivated by Morikawa and Kim (2016) or Morikawa et al. (2017), we may propose two adaptive estimators for θ_t with a parametric working model or a nonparametric estimator for $p(y_t \mid Z, \mathcal{O}_t, r_t = 1)$ by estimating the efficient score functions, or extend the semiparametric maximum likelihood method by assuming a parametric form for $p(r_t \mid y_t, Z, \mathcal{O}_t)$ only. On the other hand, as in Ai et al. (2018), we can consider a set of known basis functions based on $(Z, \mathcal{O}_t)$ and then obtain the efficient GMM estimator.

Variable selection For longitudinal data, the dimension of $\mathcal{O}_t$ is not low when t increases. If the unknown parameter vector γ_t is sparse, the unregularized two-step GMM estimator of θ_t may lose some efficiency. To obtain more efficient estimator, we can apply the penalised-GMM as follows

$$\hat{\theta}_t^p = \operatorname{argmin}_{\theta\in\Theta_t} \times \left\{ G_t(\theta)^{\mathrm{T}} \hat{W}_t G_t(\theta) + \lambda_t \sum_{j=1} P_{tn}(|\theta_{tj}|) \right\}, \quad (4)$$

where $P_{tn}(\cdot)$ is a penalty function, λ_t is a tuning parameter and θ_{tj} is the jth element of θ_t. The penalty function is a nonnegative, nondecreasing and differentiable function on $(0, \infty)$ (Fan & Li, 2001; Zou, 2006). The tuning parameter λ_t determines the amount of shrinkage These properties ensure that the estimates of $\hat{\theta}_t^p$ in (4) can shrunk to zero if they are small. The corresponding covariates of the estimates that are zero are the insignificant predictors whereas the estimates that are not zero correspond to those $\mathcal{O}_t$ which are statistically significant predictors.

Semiparametric model Note that the estimators in (3) are not consistent unless the parametric model (1) is correct. Since the parametric approach is sensitive to failure of the assumed models, we may consider a semiparametric propensity model (Kim & Yu, 2011) as follows:

$$\Pr(r_t = 1 \mid r_{t-1} = 1, U, y_1, \ldots, y_t) = [1 + g_t(\mathcal{O}_t) h_t(y_t, \beta_t)]^{-1}, \quad t = 1, \ldots, T, \quad (5)$$

where $h_t(y_t, \beta_t)$ is a known function of y_t with an unknown q-dimensional parameter vector β_t and $g_t(\mathcal{O}_t)$ is a completely unspecified function of $\mathcal{O}_t$, $t = 1, \ldots, T$. Semiparametric model (5) encompasses a large class of dropout propensity models. As in Shao and Wang (2016), we have

$$E\{\delta_t \omega_t(\theta_t) - 1 \mid \delta_{t-1} = 1, \mathcal{O}_t\} = 0, \quad (6)$$

which is equivalent to

$$g_t(\mathcal{O}_t) = \frac{E\{1 - \delta_t \mid \delta_{t-1} = 1, \mathcal{O}_t\}}{E\{\delta h_t(y_t, \beta_t) \mid \delta_{t-1} = 1, \mathcal{O}_t\}}.$$

Then, we can obtain the following kernel regression estimator for $g_t(\mathcal{O}_t)$,

$$\hat{g}_t(\mathcal{O}_t, \beta_t) = \frac{\sum_{\delta_{(t-1)i}=1} (1 - \delta_{ti}) K_h(\mathcal{O}_t - \mathcal{O}_{ti})}{\sum_{\delta_{(t-1)i}=1} \delta_{ti} h_t(y_{ti}, \beta_t) K_h(\mathcal{O}_t - \mathcal{O}_{ti})}, \quad (7)$$

where $K_h(\cdot) = h^{-1} K(\cdot/h)$, $K(\cdot)$ is a symmetric kernel function and h is a bandwidth. After $g_t(\mathcal{O}_t)$ is profiled by (7), $\hat{\beta}_t$ in (5) can be obtained by two-step GMM. Therefore, we have a consistent estimator $\hat{g}_t(\mathcal{O}_t, \hat{\beta}_t)$ of $g_t(\mathcal{O}_t)$, and the estimators of unknown quantities in $p(Y \mid X)$ or the marginal of Y can be obtained using the inverse propensity weighting with the estimated propensity as the weight function.

Disclosure statement

No potential conflict of interest was reported by the author.

Funding

This research was also supported by the National Natural Science Foundation of China (11501208, 11871287) and Fundamental Research Funds for the Central Universities .

Notes on contributor

Dr Lei Wang holds a PhD in statistics from East China Normal University. He is an assistant professor of statistics at Nankai University. His research interests include empirical likelihood and missing data problems.

References

Ai, C., Linton, O., & Zhang, Z. (2018). A simple and efficient estimation method for models with nonignorable missing data. ArXiv preprint arXiv:1801.04202.

Fan, J., & Li, R. (2001). Variable selection via nonconcave penalized likelihood and its oracle properties. *Journal of the American statistical Association*, *96*, 1348–1360.

Kim, J. K., & Yu, C. L. (2011). A semiparametric estimation of mean functionals with nonignorable missing data. *Journal of the American Statistical Association*, *106*, 157–165.

Little, R. J. A. (1995). Modeling the drop-out mechanism in repeated-measures studies. *Journal of the American Statistical Association*, *90*, 1112–1121.

Little, R. J. A., & Rubin, D. B. (2002). *Statistical Analysis with Missing Data*. 2nd ed. New York: Wiley.

Morikawa, K., & Kim, J. K. (2016). Semiparametric adaptive estimation with nonignorable nonresponse data. ArXiv preprint arXiv:1612.09207 .

Morikawa, K., Kim, J. K., & Kano, Y. (2017). Semiparametric maximum likelihood estimation with data missing not at random. *Canadian Journal of Statistics*, *45*, 393–409.

Shao, J., & Wang, L. (2016). Semiparametric inverse propensity weighting for nonignorable missing data. *Biometrika*, *103*, 175–187.

Troxel, A. B., Harrington, D. P., & Lipsitz, S. R. (1998). Analysis of longitudinal data with non-ignorable non-monotone missing values. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, *47*, 425–438.

Wang, L., Qi, C., & Shao, J. (2018). Model-assisted regression estimators for longitudinal data with nonignorable dropout. *International Statistical Review*, to appear.

Zou, H. (2006). The adaptive lasso and its oracle properties. *Journal of the American statistical association*, *101*, 1418–1429.