

Notes on Permutation Tests for Spatial Cluster Detection with Censored Outcome Data

Andrea J. Cook

Group Health Center for Health Studies, Seattle, Washington 98101, U.S.A.

email: cook.aj@ghc.org

and

Yi Li

Department of Biostatistics, Harvard University and the Dana Farber Cancer Institute, Boston, Massachusetts 02115, U.S.A.

SUMMARY. This short note evaluates the assumptions required for a permutation test to approximate the null distribution of the spatial scan statistic for censored outcomes proposed in Cook, Gold, and Li (2007, *Biometrics* **63**, 540–549). In particular, we study the exchangeability conditions required for such a test under survival models. A simulation study is further performed to assess the impact on the type I error when the global exchangeability assumption is violated and to determine whether the permutation test still well approximates the null distribution.

1. Introduction

Commenges and Liqueur (2007) provided an alternative approach to assessing the distribution of one spatial cluster detection method proposed by Cook, Gold, and Li (2007), under the null hypothesis of no spatial clustering. They elaborated on a potential limitation of the permutation method approach to approximate the asymptotic distribution of the proposed spatial scan statistic, which was briefly mentioned in the original article of Cook et al. (2007). In particular, they argued that the exchangeability condition required for a permutation test could be violated in Cook et al. (2007) as the distribution of the censoring might depend on geographic location. An asymptotic approximation was further proposed by Commenges and Liqueur (2007).

Though we agree that Commenges and Liqueur (2007) did make a valid point, we would like to point out that the spatial scan method developed by Cook et al. (2007) is indeed valid for the data application considered in their original article. The data set arose from a study investigating potential environmental exposures and their relationship to childhood asthma and other respiratory outcomes in the greater Boston area. Details of the study design have been previously published by Celedon et al. (1999). In Cook et al. (2007), the censored outcomes included time to asthma, time to eczema, and time to allergic rhinitis/hayfever, while most of the censoring ($335/370 = 90.5\%$) is due to administrative reasons, and not due to drop-out. The main reasons for administrative censoring were due to end of study duration and family relocation, both of which were irrelevant to the disease status (Celedon et al., 1999). Under such a design, it seems unlikely that there was any influential association between location and censoring. This type of study design is indeed common in

epidemiological studies, particularly with prospective cohorts with long term follow-up (Mangano et al., 1992; Zhang et al., 2006) and therefore the assumption of independence between censoring and geographic location may be reasonable.

Even though it is reasonable to make such an independent assumption in Cook et al. (2007) we do feel that is of interest to investigate the performance of the permutation test when the exchangeability condition is violated. We have conducted simulations in this article to evaluate the type I error when the distribution of censoring does depend on locations. Moreover, as detailed in the next section, adjusting for the impact of covariates may pose additional difficulty for the permutation test. In particular, in the context of survival analysis, the global exchangeability of (martingale) residuals will no longer hold in the presence of covariates, while the first-moment exchangeability might hold (under the null hypothesis). We will explore this issue further and assess, via simulations, the performance of the test when only the first-moment exchangeability holds while the global exchangeability may be violated.

2. Score Spatial Scan Statistic

Cook et al. (2007) proposed a score spatial scan statistic for detecting spatial clustering as follows

$$\sup_k LR_k = \sup_k \frac{U(k)}{\text{var}(U(k))^{1/2}}, \quad (1)$$

where $U(k) = \sum_{i=1}^n Z_i(k) \hat{M}_i$, \hat{M}_i is the martingale residual from the proportional hazards model under the null hypothesis (i.e., without clustering), k denotes the k th potential cluster region,

$$\lambda_N(t | \mathbf{X}_i) = \lambda_0(t) \exp[\beta \mathbf{X}_i], \quad (2)$$

and

$$\begin{aligned} & \text{var}(U(k)) \\ &= \sum_{j=1}^n D_j \left[\frac{\sum_{\{l:Y_l \geq Y_j\}} Z_l(k) e^{\hat{\beta} \mathbf{X}_l}}{\sum_{\{l:Y_l \geq Y_j\}} e^{\hat{\beta} \mathbf{X}_l}} - \left(\frac{\sum_{\{l:Y_l \geq Y_j\}} Z_l(k) e^{\hat{\beta} \mathbf{X}_l}}{\sum_{\{l:Y_l \geq Y_j\}} e^{\hat{\beta} \mathbf{X}_l}} \right)^2 \right]. \end{aligned} \quad (3)$$

Here Y_i is the observed survival time, D_i is the censoring indicator, $Z_i(k)$ is an indicator if individual i is in potential cluster k , and \mathbf{X}_i is the associated covariate vector. Cook et al. (2007) approximated the asymptotic distribution of the test statistic by fixing observed outcome and covariate data (Y_i, D_i, \mathbf{X}_i) , while permuting the location of individual (s_i, r_i) ($i = 1, \dots, n$). The test statistic was recalculated on each permuted data set, and inference was drawn based on the empirical distribution of a series of recalculated test statistics.

2.1 Exchangeability

Focusing on the numerator of the test statistic, $U(k) = \sum_{i=1}^n Z_i(k) \hat{M}_i$, as in Commenges and Liquet (2007), we study the exchangeability properties of the martingale residuals, \hat{M}_i , required for the permutation test. It can be easily shown that the martingale residuals are first-moment exchangeable in the presence or in the absence of covariates, \mathbf{X}_i , even without assuming independence between the censoring distribution and location. We define the vector of observations \mathbf{M} as being first-moment [or second-moment] exchangeable when the $E(\mathbf{M}) = E(\mathbf{PM})$ [or $\text{var}(\mathbf{M}) = \text{var}(\mathbf{PM})$] for any permutation vector \mathbf{P} . Specifically for our situation, $E(\hat{M}_i) = 0$ for all i under the null hypothesis of no spatial clustering and given equation (2) is correctly specified. We now consider the exchangeability of these martingale residuals. To illustrate, consider the trivial case presented in Commenges and Liquet (2007) as $\hat{\mathbf{M}} = (\hat{M}_1, \hat{M}_2)$. Obviously, $\hat{\mathbf{M}}$ has expectation $(0,0)$, and so does the permuted vector $\mathbf{P}\hat{\mathbf{M}} = (\hat{M}_2, \hat{M}_1)$. Therefore the martingale residuals are first-moment exchangeable.

Furthermore, in the absence of covariates, the global exchangeability is held, except for when the censoring distribution is dependent on the location. On the other hand, in the presence of covariates, as $\text{Var}(\hat{M}_1) \neq \text{Var}(\hat{M}_2)$ in general, the second-moment exchangeability does not hold, and hence, the global exchangeability fails. But when the covariates are independent of location, the martingale residuals would be independent of location under the null. In this case, the global exchangeability is held when viewing covariates as random variables, as in Section 3.5 of Commenges (2003).

Therefore, the assumptions needed for the permutation test are actually held in certain situations, and are also partially held for other situations. It has been demonstrated that under the first-moment exchangeability, the permutation test still performs reasonably well for residuals under generalized linear models (Jacqmin-Gadda et al., 1997). As no analytical results were available under survival models we ran a simulation study to investigate how well our proposed permutation

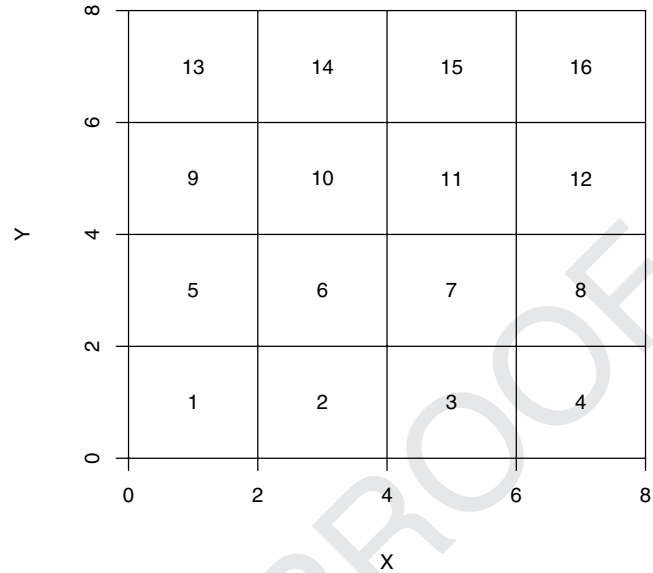


Figure 1. Grid study area created for the simulation to evaluate type I error rates.

test for the score spatial scan statistic performs when only first-moment exchangeability is held.

3. Simulation Study

We conducted simulations to evaluate the type I error probability for the spatial scan statistic using the permutation test described in Cook et al. (2007). For computational efficiency, we allowed a finite range of radii of 0.5 to 2 sequenced by 0.5 and created 1000 data sets per simulation.

A simulated data set was generated by dividing the area into 16 equally sized squares of size 2×2 as depicted in Figure 1. The study population size ranged from 50 to 300 and each participant's location (r_i, s_i) was randomly assigned with a uniform distribution over the study area.

To assess the magnitude of impact of dependence between censoring distribution and location on the permutation test, we made the censoring distribution within grids #6 and #10 different from the rest of the study area. Therefore, we first independently generated location coordinates r_i and s_i ($i = 1, \dots, n$) each with uniform distribution over $[0, 8]$. Then we generated random variables C_i and F_i ($i = 1, \dots, n$) from the exponential distributions with constant hazards λ_{ci} and λ_{fi} , respectively. If participant i , has location (r_i, s_i) within grids #6 or #10, then $\lambda_{ci} = \lambda_{c1}$ otherwise $\lambda_{ci} = \lambda_{c0}$. We set $\lambda_{fi} = \lambda_f$ for all i . Given F_i and C_i , define $D_i = I(F_i \leq C_i)$ and $Y_i = \min(C_i, F_i)$ to complete the randomly generated failure time data set. If $\lambda_{c1} > \lambda_{c0}$, censoring is more likely to occur in grids #6 and #10 compared to the rest of the study area; if $\lambda_{c1} < \lambda_{c0}$, it is vice versa.

The results are displayed in Table 1, which reveals that the type I error is being held for all simulations. This is by no means a proof that there is no effect of dependence between censoring distribution and spatial location, but rather a simple display of limited effect in one simple case when the censoring is dependent on residence location within a single

Table 1

Type I error rate for different sample sizes, percentage of failure events, and censoring location dependence

| | | Number of observations | | | | | |
|--------------------------------|------------|------------------------|-------|-------|-------------------|-------|-------|
| | | $\lambda_f = 1/8$ | | | $\lambda_f = 1/3$ | | |
| | | 50 | 100 | 300 | 50 | 100 | 300 |
| Censoring | (1/2, 1/4) | 0.034 | 0.056 | 0.064 | 0.054 | 0.044 | 0.050 |
| $(\lambda_{c1}, \lambda_{c0})$ | (1/2, 1/2) | 0.032 | 0.055 | 0.049 | 0.062 | 0.047 | 0.038 |
| | (1/4, 1/2) | 0.046 | 0.046 | 0.042 | 0.054 | 0.062 | 0.056 |

Table 2

Type I error rate assessing the effect of covariate adjustment for differing magnitude of covariate and location dependence

| | | $P(X = 1 \mid \text{grid}) : (p_1, p_0)$ | |
|---|--------------|--|------------|
| (λ_0, λ_x) | | (2/3, 1/4) | (2/3, 1/2) |
| $\lambda_{f_i} = \lambda_0 + \lambda_x X_i$ | (1/8, 1/12) | 0.058 | 0.036 |
| | (1/8, 1/8) | 0.050 | 0.056 |
| | (1/8, -1/12) | 0.046 | 0.036 |
| | (1/3, 1/12) | 0.046 | 0.046 |
| | (1/3, 1/8) | 0.066 | 0.052 |
| | (1/3, -1/12) | 0.069 | 0.050 |

rectangular grid. However, this was an important scenario to evaluate, as the proposed spatial scan statistic is to detect a single, or a small number, of spatial clusters. Further, we assessed, via varying λ_{c1} and λ_{c0} , the scenarios with larger differences in censoring distributions. These scenarios are unlikely to exist in most data sets, but make us more confident that the dependence of the censoring distribution and location does not strongly affect the results of the spatial scan statistic.

We next ran a simulation to evaluate the effect of covariate adjustment and the holding of type I error when there is dependence between the covariate and location. This scenario would frequently happen, as, for example, a person's age, sex, race, and socioeconomic status may be strongly related to both residence location and the outcome of interest. To assess the effect of this type of association, we ran a simple simulation with a binary covariate only. We assumed that $P(X_i = 1)$ within grids #6 and #10 (Figure 1) and the rest of the study area was different. Specifically, we first independently generated location coordinates r_i and s_i ($i = 1, \dots, n$) each with uniform distribution over $[0, 8]$. Then generated X_i as follows, if participant i has location within grids #6 and #10, then $P(X_i = 1) = p_1$, else $P(X_i = 1) = p_0$. Next we generated random variables C_i and F_i ($i = 1, \dots, n$) from the exponential distributions with constant hazards $\lambda_{c_i} = 1/2$ and $\lambda_{f_i} = \lambda_0 + \lambda_x X_i$, respectively. Given F_i and C_i , define $D_i = I(F_i \leq C_i)$ and $Y_i = \min(C_i, F_i)$ to complete the randomly generated failure time data set.

Table 2 displays the type I error simulation to evaluate the impact of covariate adjustment. It appears that even when the covariate depends upon a single cluster location, the type I error is still roughly held at the nominal level for the scenarios considered.

4. Conclusion

The proposed use of the permutation test to approximate the distribution of the spatial scan statistic using the score test in certain scenarios has good properties even when global exchangeability is not met. Based on our numerical experience, we feel that the permutation testing approach may still be a valid and preferred method, particularly for smaller data sets, when the asymptotic results proposed by Commenges and Liquet (2007) may not hold.

Finally, we would like to convey that the spatial cumulative residual test, the main product of Cook et al. (2007), deviates much from the spatial scan test considered here and in Commenges and Liquet (2007), and does not require such exchangeability conditions. It indeed provides a testing procedure that is valid under broader circumstances.

The authors would like to thank the editor and an associate editor for their helpful comments. This work was supported by NIH grant R01 CA95747.

REFERENCES

- Celedon, J., Litonjua, A., Weiss, S., and Gold, D. (1999). Day care attendance in the first year of life and illnesses of the upper and lower respiratory tract in children with a familial history of atopy. *Pediatrics* **104**, 495–500.
- Commenges, D. (2003). Transformations which preserved exchangeability and applications to permutation tests. *Nonparametric Statistics* **15**, 171–185.
- Commenges, D. and Liquet, B. (2007). Asymptotic distribution of score statistics for spatial cluster detection with censored data. *Biometrics Response Letter*.
- Cook, A., Gold, D., and Li, Y. (2007). Spatial cluster detection for censored outcome data. *Biometrics* **63**, 540–549.
- Jacqmin-Gadda, H., Commenges, D., Nejari, C., and Dartigues, J. (1997). Test of geographical correlation with adjustment for explanatory variables: Application to dyspnoea in the elderly. *Statistics in Medicine* **16**, 1283–1297.
- Mangano, D., Browner, W., Hollenberg, M., Li, J., and Tateo, I. (1992). Long-term cardiac prognosis following noncardiac surgery. the study of perioperative ischemia research group. *JAMA* **268**, 233–239.
- Zhang, H., Thijs, L., Kuznetsova, T., Fagard, R., Li, X., and Staessen, J. (2006). Progression to hypertension in the non-hypertensive participants in the flemish study on environment, genes and health outcomes. *Journal of Hypertension* **24**, 1719–1727.