# A note on testing separability in spatial-temporal marked point processes

Renato Assunção and Alexandra Maia
LESTE - Laboratório de Estatística Espacial
Departmento de Estatística, Universidade Federal de Minas Gerais
31270-901 Belo Horizonte, MG, Brazil

December 25, 2005

**Abstract**

In environmental risk analysis, it is common to assume the stochastic independence (or separability) between the marks associated with the random events of a spatial-temporal point process. Schoenberg (2004, *Biometrics* **60**, 471–481) proposed several test statistics for this hypothesis and used simulated data to evaluate their performance. He found that a Cramér-von Mises-type test is powerful to detect gradual departures from separability although it is not uniformly powerful over a large class of alternative models. We present a semi-parametric approach to model alternative hypotheses to separability and derive a score test statistic. We show that there is a relationship between this score test and some of the test statistics proposed by Schoenberg. Specifically, all are different versions of weighted Cramér-von Mises-type statistics. This gives some insight into the reasons for the similarities and differences between the test statistics performance. We also point out some difficulties controlling the type I error probability in Schoenberg's residual test.
**keywords** spatial-temporal point processes; separability; marked point processes; space-time interaction; Poisson process.

# 1 Introduction

Many risk assessment analyses use point process data which arise from spatial-temporal marked point processes (Brillinger et al, 2003). Modelling the joint distribution of spatial locations, occurrence times, and the marks is challenging, and few models are flexible enough to be applied over a wide large of applications. Because it is difficult to find models for the interactions of marks and random event times that are simple to fit, it is common to assume separability. That is, to assume that marks are independent of the spatial-temporal process. Under separability, data analysis can be decomposed into a purely spatial-temporal analysis and a statistical analysis of the marks data.

In practice, the separability hypothesis is not realistic, and it should at least be tested before assumed. Schoenberg (2004) proposed and compared several non-parametric tests to investigate the separability of multi-dimensional point processes. By simulation, he showed that a Cramér-von Mises-type test is very powerful at detecting gradual departures from separability. For inhibition between marks and time of occurrence of the events, the non-parametric tests had very low power. As a complement to these tests, Schoenberg (SC henceforth) suggested another test, called a residual test, that was based on randomly rescaling of the process and showed good performance against these latter nonseparable alternatives.

In this note, we introduce a non-parametric approach to model alternative hypotheses to separability that yields a closed form score test statistic. Since a score test is a locally most powerful test for non-separability, we expect good power properties from this proposed test.

We show that the best test statistics considered by SC, as well as our score test statistic, can be seen as Cramér-von Mises-type statistics with different weight functions. This gives some insight into the reasons for the similarities and the differences between the test statistics performance. We also point out some problems with the control of the type I error probability of Schoenberg's residual test.

# 2 Separability

During the time period $[0, t_F]$, random events are observed within a finite $d$-dimensional region. Commonly, we have $d = 2$ or 3 as in, for example, modelling the occurrence of wildfires or earthquakes. Additionally to their location in the space and time, the events have associated marks, real-valued measures $m$ of size or other characteristic of interest.

To set notation, let $\mathbf{x}_i = (x_{1i}, \ldots, x_{di})$ be the locations, $t_i$ be the reference times, and $m_i$ be the marks of $n$ random events of the point process $N$ occurring within a specified finite region $A \subset \mathbb{R}^d$, during time period $[0, t_F]$, with marks $m \in \mathbb{R}$, and indexed by $i$, $i = 1, \ldots, n$. Let $\mathcal{X}$ be the set $A \times [0, t_F] \times \mathbb{R}$ and denote the expected number of events in a measurable set $B$ contained in $\mathcal{X}$ by $E(N(B))$.

Like SC, we define the conditional intensity $\lambda(\mathbf{x}, t, m|\mathcal{F}_t)$ giving the mean rate of occurrence of events per space-time-mark volume conditional on the past. More specifically, let $\{\mathcal{F}_t : t \geq 0\}$ be a filtration $\mathcal{F}_t$ representing the history of the point process $N$ up to and including time $t$. Let $dV = (d\mathbf{x}, dt, dm)$ be a small $d + 2$ dimensional region in $\mathcal{X}$ and $N(dV)$ be the number of events in $dV$. Then

$$\lambda(\mathbf{x}, t, m|\mathcal{F}_t) = \lim_{dV \to 0} \frac{P(N(dV) = 1|\mathcal{F}_t)}{|dV|} \tag{1}$$

where $|dV|$ is the volume of $dV$. The log-likelihood of the observed events can be written as

$$\sum_{i=1}^{n} \log\left(\lambda(\mathbf{x}_i, t_i, m_i|\mathcal{F}_t)\right) - \int_{\mathcal{X}} \lambda(\mathbf{x}, t, m|\mathcal{F}_t)\, d\mathbf{x}\, dt\, dm \tag{2}$$

([]Fish76).

We can obtain a non-parametric conditional intensity estimate under a non-separable model based on a kernel estimate:

$$\widehat{\lambda}(\mathbf{x}, t, m|\mathcal{F}_t) = \int_{\mathcal{X}} K(\mathbf{x} - \mathbf{y}, t - u, m - m^*)\, dN(\mathbf{y}, u, m^*)\ . \tag{3}$$

where $K$ is a spatial-temporal-marked $d+2$-dimensional kernel function.

Let $\mu = E\left(N(\mathcal{X})\right)$ be the expected total number of events and define the nonnegative predictable process

$$\lambda_{ST}(\mathbf{x}, t|\mathcal{F}_t) = \mu^{-1} \int_{\mathbb{R}} \lambda(\mathbf{x}, t, m|\mathcal{F}_t)\, dm \tag{4}$$

and the nonnegative marginal random function

$$\lambda_M(m|\mathcal{F}_t) = \mu^{-1} \int_0^{t_F} \int_A \lambda(\mathbf{x}, t, m|\mathcal{F}_t)\, d\mathbf{x}\, dt\ . \tag{5}$$

Under the separability hypothesis, $\lambda_M(m|\mathcal{F}_t) = \lambda_M(m)$ and

$$\lambda(\mathbf{x}, t, m|\mathcal{F}_t) = \mu\, \lambda_{ST}(\mathbf{x}, t|\mathcal{F}_t)\, \lambda_M(m)\ . \tag{6}$$

Define the method of moments estimator $\hat{\mu} = N(\mathcal{X})$. Consider also the non-parametric kernel-based estimates

$$\widetilde{\lambda_{ST}}(\mathbf{x}, t|\mathcal{F}_t) = \hat{\mu}^{-1} \int_{\mathcal{X}} K_{ST}(\mathbf{x} - \mathbf{y}, t - u)\, dN(\mathbf{y}, u, m)\ , \tag{7}$$

and

$$\widetilde{\lambda_M}(m) = \hat{\mu}^{-1} \int_{\mathcal{X}} K_M(m - m^*)\, dN(\mathbf{x}, t, m^*) \;, \tag{8}$$

where $K_{ST}$ and $K_M$ are spatial-temporal and marked kernel densities, respectively. Under the null hypothesis (6), an estimate of $\lambda(\mathbf{x}, t, m|\mathcal{F}_t)$ is given by the product

$$\widetilde{\lambda}(\mathbf{x}, t, m|\mathcal{F}_t) = \hat{\mu}\; \widetilde{\lambda_{ST}}(\mathbf{x}, t|\mathcal{F}_t)\; \widetilde{\lambda_M}(m) \tag{9}$$

To simplify the notation, from here on we drop the dependence of the conditional intensity estimates on $\mathcal{F}_t$.

# 3 Schoenberg Test Statistics

Under the separability hypothesis, the intensity estimates $\widehat{\lambda}(\mathbf{x}, t, m)$ and $\widetilde{\lambda}(\mathbf{x}, t, m)$ are similar and reasonable test statistics can be constructed by comparing them. SC considered several non-parametric test statistics. Large values of these test statistics are evidence against the separability hypothesis:

$$S_1 = \sup \left[ |\widehat{\lambda}(\mathbf{x}, t, m) - \widetilde{\lambda}(\mathbf{x}, t, m)| / \sqrt{\widetilde{\lambda}(\mathbf{x}, t, m)}\, ; (\mathbf{x}, t, m) \in \mathcal{X} \right]$$

$$S_2 = \inf \left[ |\widehat{\lambda}(\mathbf{x}, t, m) - \widetilde{\lambda}(\mathbf{x}, t, m)| / \sqrt{\widetilde{\lambda}(\mathbf{x}, t, m)}\, ; (\mathbf{x}, t, m) \in \mathcal{X} \right]$$

$$S_3 = \int_{\mathcal{X}} \left[ \widehat{\lambda}(\mathbf{x}, t, m) - \widetilde{\lambda}(\mathbf{x}, t, m) \right]^2 dt\, d\mathbf{x}$$

$$S_4 = \int_{\mathcal{X}} \left[ \log \widehat{\lambda}(\mathbf{x}, t, m) - \log \widetilde{\lambda}(\mathbf{x}, t, m) \right] N(dt, d\mathbf{x}) - \int_{\mathcal{X}} \left[ \widehat{\lambda}(\mathbf{x}, t, m) - \widetilde{\lambda}(\mathbf{x}, t, m) \right] d\mathbf{x}\, dt$$

$$S_5 = n^{-1} \sum_i \left( \widehat{\lambda}(\mathbf{x}_i, t_i, m_i) - \widetilde{\lambda}(\mathbf{x}_i, t_i, m_i) \right)^2$$

SC used simulated data to evaluate the performance of the test statistics and found that $S_3$ was the most powerful against two alternatives of gradual departures from separability. Test $S_5$ ranked either as second or third best while test $S_4$ ranked either as third or fourth in these comparisons. However, test $S_3$ is not uniformly most powerful among the tests considered. In all four time-mark clustering models of his Table 1, the type II error probability of $S_5$ was considerably smaller than the second best and third best, $S_4$ and $S_3$, respectively. In the uniform additive process, test $S_5$ and, to a lesser extent, test $S_4$ had reasonable performance (see his Figure 2b).

Ignoring the nonseparability models in the form of inhibition of the marks, tests $S_3$, $S_4$, $S_5$ ranked among the best test statistics considered, none of them uniformly. Many times, the power of some of these tests were approximately equal. In Section (5), we give some insight into these results. For temporal and marked inhibition models the test statistics power is very low. SC suggested to complement the above tests with the use of one additional residual test based on randomly rescaling the point process. We discuss this test in Section 6.

# 4 A score test statistic

We derive a score test statistic using a likelihood approach. For this derivation, we need to assume a specific distribution for the point process under the nonseparable hypothesis. If the process is non-separable, there exists a constant $\epsilon$ and a certain predictable function $g(\mathbf{x}, t, m|\mathcal{F}_t)$ such that, almost surely, we have

$$\lambda(\mathbf{x}, t, m|\mathcal{F}_t) = \mu\,\lambda_{ST}(\mathbf{x}, t|\mathcal{F}_t)\,\lambda_M(m|\mathcal{F}_t)\,(1 + \epsilon\, g(\mathbf{x}, t, m|\mathcal{F}_t)) \ . \qquad (10)$$

To see this, define $\epsilon\, g(\mathbf{x}, t, m|\mathcal{F}_t)$ as the relative difference between $\lambda(\mathbf{x}, t, m|\mathcal{F}_t)$ and $\mu\, \lambda_{ST}(\mathbf{x}, t|\mathcal{F}_t)\, \lambda_M(m|\mathcal{F}_t)$. Under the separability hypothesis, we have $\lambda_M(m|\mathcal{F}_t) = \lambda_M(m)$ and $g(\mathbf{x}, t, m|\mathcal{F}_t) \equiv 0$, which implies the identity (6).

We want to contrast the null separability hypothesis (6) with the hypothesis (10) where space-time and marks interact in a non-parametric way. Of particular interest are alternatives that are not extreme departures from the null hypothesis. In the extreme cases, it is likely that no test would be required since the data would clearly point out to the inadequacy of the null hypothesis. For the less extreme cases, we can consider $\epsilon$ to be small enough such that (10) can be seen as a local departure from (6).

Assume for now that $\lambda_{ST}(\mathbf{x}, t|\mathcal{F}_t)$, $\lambda_M(m|\mathcal{F}_t)$ and $g(\mathbf{x}, t, m|\mathcal{F}_t)$ are known functions. Omitting the filtration $\mathcal{F}_t$ to simplify the notation, under the alternative hypothesis, the log-likelihood of the single parameter $\epsilon$ is given by

$$\begin{aligned} l(\epsilon) &= \sum_i \left(\log \mu\lambda_{ST}(\mathbf{x}_i, t_i)\lambda_M(m_i) + \log\left(1 + \epsilon g(\mathbf{x}_i, t_i, m_i)\right)\right) \\ &\quad - \int_{\mathcal{X}} \left(1 + \epsilon\, g(\mathbf{x}, t, m)\right) \mu\lambda_{ST}(\mathbf{x}, t)\, \lambda_M(m)\, d\mathbf{x}\, dt\, dm \end{aligned}$$

and therefore the score statistic is given by

$$\frac{\partial l}{\partial \epsilon} = \sum_{i=1}^{n} \frac{g(\mathbf{x}_i, t_i, m_i)}{1 + \epsilon g(\mathbf{x}_i, t_i, m_i)} - \int_{\mathcal{X}} \mu\lambda_{ST}(\mathbf{x}, t)\, \lambda_M(m)\, g(\mathbf{x}, t, m)\, d\mathbf{x}\, dt\, dm \quad (11)$$

When evaluated at $\epsilon = 0$, it is the score test statistic

$$\frac{\partial l}{\partial \epsilon}\Big|_{\epsilon=0} = \sum_{i=1}^{n} g(\mathbf{x}_i, t_i, m_i) - \int_{\mathcal{X}} \mu\lambda_{ST}(\mathbf{x}, t)\, \lambda_M(m)\, g(\mathbf{x}, t, m)\, d\mathbf{x}\, dt\, dm\ . \quad (12)$$

A test based on the score statistic is the locally most powerful test in the sense that it maximizes the derivative of the power function at $\epsilon = 0$ ([]CoxHin, page 113). However, this test is not feasible because, usually, the functions $\lambda_{ST}$, $\lambda_M$ and $g$ are not known.

To circumvent this problem note that

$$\epsilon\, g(\mathbf{x}, t, m) = \frac{\lambda(\mathbf{x}, t, m)}{\mu \lambda_{ST}(\mathbf{x}, t) \lambda_M(m)} - 1\,.$$

Then, an estimate of $\epsilon\, g(\mathbf{x}, t, m)$ is given by $\widehat{\lambda}(\mathbf{x}, t, m)/\widetilde{\lambda}(\mathbf{x}, t, m) - 1$. We replace $g$ (up to a proportionality constant) in (12) with this empirical estimate to obtain a new non-parametric test statistic

$$\begin{aligned} T &= \sum_{i=1}^{n} \left( \frac{\widehat{\lambda}(\mathbf{x}_i, t_i, m_i)}{\widetilde{\lambda}(\mathbf{x}_i, t_i, m_i)} - 1 \right) - \int_{\mathcal{X}} \widetilde{\lambda}(\mathbf{x}, t, m) \left( \frac{\widehat{\lambda}(\mathbf{x}, t, m)}{\widetilde{\lambda}(\mathbf{x}, t, m)} - 1 \right) d\mathbf{x}\, dt\, dm \\ &= \sum_{i=1}^{n} \frac{\widehat{\lambda}}{\widetilde{\lambda}}(t_i, \mathbf{x}_i, m_i) - \int_{\mathcal{X}} \left[ \widehat{\lambda}(\mathbf{x}, t, m) - \widetilde{\lambda}(\mathbf{x}, t, m) \right] d\mathbf{x}\, dt\, dm - n\,. \end{aligned} \tag{13}$$

The sampling distribution of $T$ defined in (13) is intractable. Its null hypothesis distribution can be obtained by a Monte Carlo procedure conditionally on the realizations of the process spatial and temporal components as explained by SC.

# 5 The relationship between the test statistics

It is possible to show that test statistic $S_4$ and $T$ are approximately equal. Using the first-order Taylor expansion, we have $\log\left(\widehat{\lambda}/\widetilde{\lambda}\right) \approx \left(\widehat{\lambda}/\widetilde{\lambda} - 1\right)$. After substituting into $S_4$ we have:

$$\begin{aligned} S_4 &\approx \sum_{i=1}^{n} \left( \frac{\widehat{\lambda}}{\widetilde{\lambda}} - 1 \right) - \int_{\mathcal{X}} \left[ \widehat{\lambda}(\mathbf{x}, t, m) - \widetilde{\lambda}(\mathbf{x}, t, m) \right] d\mathbf{x}\, dt \\ &= \sum_{i=1}^{n} \frac{\widehat{\lambda}}{\widetilde{\lambda}}(t_i, \mathbf{x}_i) - \int_{\mathcal{X}} \left[ \widehat{\lambda}(\mathbf{x}, t, m) - \widetilde{\lambda}(\mathbf{x}, t, m) \right] d\mathbf{x}\, dt - n \\ &= T\,. \end{aligned}$$

Therefore, under the separability hypothesis, the test statistics $S_4$ and $T$ should be approximately equal. Under local alternatives to the null hypothesis, we can expect they are also approximately equal and hence their properties, such as their power, should be similar. Although this result is known from general statistical inference theory, it is reassuring to find that it holds for a specific semi-parametric modelling approach that produce closed form expressions.

We can also find a relationship between $S_4$ and $S_3$ that give some insight into their differences and similarities. For a non-negative, predictable function $f$, we have

$$E\left[\sum_i f(\mathbf{x}_i, t_i, m_i)\right] = E\left[\int_{\mathcal{X}} f(\mathbf{x}, t, m)\lambda(\mathbf{x}, t, m)\, d\mathbf{x}\, dt\, dm\right] \qquad (14)$$

and hence the observed sum $\sum_i f(\mathbf{x}_i, t_i, m_i)$ is a moment estimate of (14). Holding estimates $\hat{\lambda}$ and $\tilde{\lambda}$ *fixed*, we can see the term $\sum_i \log(\hat{\lambda}/\tilde{\lambda})(\mathbf{x}_i, t_i, m_i)$ in $S_4$ as an approximation to the integral

$$\begin{aligned} E\left[\sum_i \log\frac{\hat{\lambda}}{\tilde{\lambda}}(\mathbf{x}_i, t_i, m_i)\right] &= E\left[\int_{\mathcal{X}} \left(\log\frac{\hat{\lambda}}{\tilde{\lambda}}\right) \lambda(\mathbf{x}, t, m)\, d\mathbf{x}\, dt\, dm\right] \\ &\approx \int_{\mathcal{X}} \left(\log\frac{\hat{\lambda}}{\tilde{\lambda}}\right) \lambda(\mathbf{x}, t, m)\, d\mathbf{x}\, dt\, dm \end{aligned}$$

where $\lambda(\mathbf{x}, t, m)$ is the true unknown conditional intensity of $N$.

Using a second-order Taylor expansion, we have

$$\log\left(\frac{\widetilde{\lambda}}{\widehat{\lambda}}\right) \approx \frac{\widetilde{\lambda} - \widehat{\lambda}}{\widehat{\lambda}} - \frac{1}{2}\left(\frac{\widetilde{\lambda} - \widehat{\lambda}}{\widehat{\lambda}}\right)^2 .$$

Note that we used the inverse ratio of $S_4$ in the above approximation. Therefore

$$
\begin{aligned}
-S_4 &\approx \int_{\mathcal{X}} \left(\log \frac{\tilde{\lambda}}{\hat{\lambda}}\right) \lambda(\mathbf{x},t,m)\, d\mathbf{x}\, dt\, dm - \int_{\mathcal{X}} (\tilde{\lambda} - \hat{\lambda})\, d\mathbf{x}\, dt\, dm \\
&\approx \int_{\mathcal{X}} \frac{\widetilde{\lambda} - \widehat{\lambda}}{\widehat{\lambda}} \lambda(\mathbf{x},t,m)\, d\mathbf{x}\, dt\, dm - \frac{1}{2}\int_{\mathcal{X}} \left(\frac{\widetilde{\lambda} - \widehat{\lambda}}{\widehat{\lambda}}\right)^2 \lambda(\mathbf{x},t,m)\, d\mathbf{x}\, dt\, dm \\
&\quad - \int_{\mathcal{X}} (\tilde{\lambda} - \hat{\lambda})\, d\mathbf{x}\, dt\, dm \qquad (15)
\end{aligned}
$$

Assuming that $\hat{\lambda}$ is a good non-parametric estimate of $\lambda$, the first and third terms of (15) approximately cancel out and the second term can be further simplified to finally produce the approximation

$$S_4 \approx \frac{1}{2}\int_{\mathcal{X}} (\widehat{\lambda} - \widetilde{\lambda})^2 \, \frac{1}{\lambda}\, d\mathbf{x}\, dt\, dm \qquad (16)$$

Like the test statistic $S_3$, the right hand side of (16) is a Cramér-von Mises-type of statistic but weighting more heavily the discrepancies in the low intensity regions.

In summary, when the true point process distribution is close to a separable distribution, we can say that

$$T \approx S_4 \approx \int_{\mathcal{X}} \left(\frac{\hat{\lambda}}{\tilde{\lambda}} - 1\right)^2 \lambda\, d\mathbf{x}\, dt\, dm \approx \frac{1}{2}\int_{\mathcal{X}} \left(\hat{\lambda} - \tilde{\lambda}\right)^2 \frac{1}{\lambda}\, d\mathbf{x}\, dt\, dm \qquad (17)$$

By similar arguments, we have

$$S_5 = \frac{1}{n}\sum_i (\widehat{\lambda}_i - \widetilde{\lambda}_i)^2\, d\mathbf{x}\, dt\, dm \approx \frac{1}{n}\int_{\mathcal{X}} (\widehat{\lambda} - \widetilde{\lambda})^2\, \lambda\, d\mathbf{x}\, dt\, dm \qquad (18)$$

To understand the differences between the test statistics in one specific case, consider the nonseparable exponential additive example of SC where $\mathcal{X} = [0,1]^4$ and the conditional intensity is given by

$$\lambda(\mathbf{x},t,m|\mathcal{F}_t) = \lambda(\mathbf{x},t,m) = e^{3+m} + e^{3(1+t)} \qquad (19)$$

which does not depend on $\mathbf{x}$. We have $\mu = e^3(e^3/3 + e - 4/3)$,

$$\lambda_{ST}(\mathbf{x}, t|\mathcal{F}_t) = \lambda_{ST}(\mathbf{x}, t) = \frac{1}{\mu}\left(e^{3(1+t)} + e^3(e-1)\right)$$

and

$$\lambda_M(m) = \frac{1}{\mu}\left(e^{3+m} + \frac{e^3}{3}(e^3 - 1)\right) .$$

In the top row of Figure 1 we show the surface $\lambda(\mathbf{x}, t, m)$ on the left hand side and $s(\mathbf{x}, t, m) = \mu\lambda_{ST}(\mathbf{x}, t)\lambda_M(m)$ on the right hand side. The $z$-axis range is the same in both plots. The bottom row of plots is associated with the test statistics. The left plot shows the squared differences $(\lambda(\mathbf{x}, t, m) - s(\mathbf{x}, t, m))^2$ and the volume under this surface is estimated by $S_3$. The regions contributing most to the integral $S_3$ are those where both, marks and time, are close to their extreme values. The contributions are considerably larger when the marks (and the conditional intensity) are close to their maximum values. The middle plot shows the surface $(\lambda(\mathbf{x}, t, m) - s(\mathbf{x}, t, m))^2/\lambda(\mathbf{x}, t, m)$, associated with $S_4$ and $T$. Regions where either marks or time are extreme contribute most to the volume. In contrast to the previous plot, this integral is substantially affected by regions of very low intensity and the different regions contribute more equally to the total volume. The rightmost plot shows the surface $(\lambda(\mathbf{x}, t, m) - s(\mathbf{x}, t, m))^2\,\lambda(\mathbf{x}, t, m)$, associated with $S_5$. It is similar to the first one but gives almost no weight to the low intensity regions where marks have low values. Hence, it seems that differences between the test statistics are due to their different treatment of regions in which marks and times have extreme values.

By means of simulation, we investigated the power of our score test statistic $T$, compared to the test statistics $S_1 - S_5$ from SC. As one would expect from our approximations, the tests had similar performance for several models. As in Schoenberg's results, there is not a uniformly best test across different alternative models. The test $T$ had a slightly greater power than the other test only in one case, a process made of the superposition of two different point processes. Test $S_3$ showed the greatest power most of the time.

# 6 Type I error of residual test

One additional residual test based on randomly rescaling the point process was proposed by SC as a complement to the other non-parametric test statistics. A transformed point process with points $(\mathbf{x}, t, m)$ is obtained by shifting the mark coordinate from $m$ to $\int_0^m \lambda(\mathbf{x}, t, n)\, dn$ and this resulting process is a Poisson point process with unit rate (Schoenberg, 1999). With the transformed point process, the test is carried out using the $L(s)$ function, the normalized version of Ripley's K-function, where $s$ is a distance in $\mathbb{R}^{d+2}$. Upper $U(s)$ and lower $M(s)$ 95% confidence bounds are created and the null hypothesis is rejected if the observed $L(s)$ function lies outside this confidence envelope.

This test has a real significance level that can be very different from its nominal 0.05 level, a difficulty that was not mentioned by SC. The confidence bounds $U(s)$ and $M(s)$ are created such that, under the null hypothesis and for each arbitrary and *fixed* $s$, we have $1 - P(M(s) \leq L(s) \leq U(s)) = 0.05$ which is quite different from the type I error probability $1 - P(M(s) \leq L(s) \leq U(s) \forall s) = 0.05$. Therefore, the significance level is not under control and, due to positive correlation between successive values of $L(u)$, it is likely that it is larger than the 0.05 nominal significance level. If this is corrected, the power of the residual test will decrease. There have been some recent attempts to solve this problem (Pallini, 2002).

Unfortunately, SC did not present power estimates for this residual test. Its power could be estimated from the proportion of times the $L(s)$ curve crosses the confidence bounds in several simulated datasets. However, due to the lack of significance level control, it is important to calculate the power under the null hypothesis also to establish the real type I error probability of the test.

# 7 Conclusion

In this note, we showed that a semi-parametric model for nonseparable alternatives can motivate a score test statistic $T$. Test $T$ and several other good non-parametric tests with good power proposed by SC are approximately Cramér-von Mises type statistics with different weight functions. This explains their similar power behavior for several models, and clarifies why tests such as $S_3$ consistently have the best power. In our simulations we found that no test is uniformly best across different alternative models, but $S_3$ is best most of the time and should be recommended. We also pointed out some problems with the control of error type I probability of Schoenberg's residual test.

# Acknowledgements

The first author is grateful for partial support from grants 305567/2004-7 (PQ-1D) and 505608/04-9 (SIMVIDA project) from the Conselho Nacional de Pesquisa Científica e Tecnológica (CNPq). The second author acknowledges partial support from CAPES.

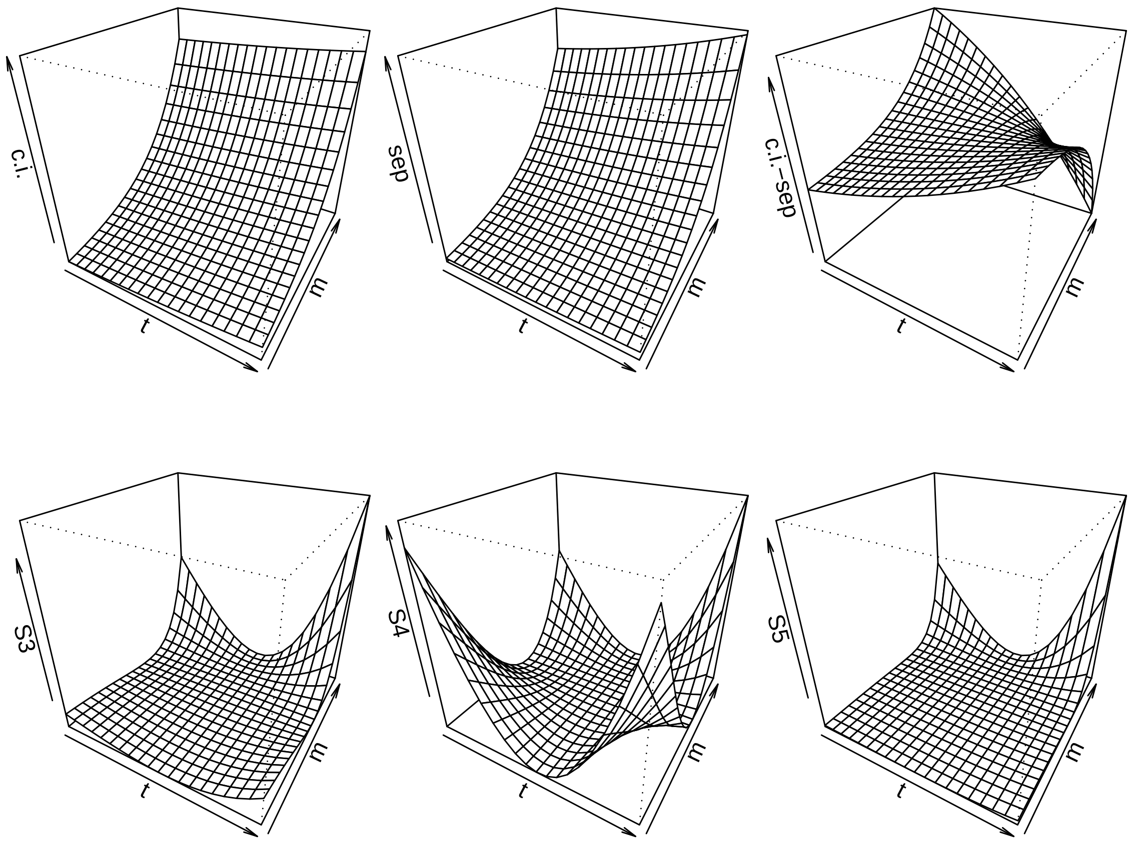


Figure 1: Nonseparable exponential additive model. The top row plots are the surfaces of the conditional intensity $\lambda = \lambda(\mathbf{x}, t, m)$ (left) and the product of the marginal intensities $s = s(\mathbf{x}, t, m)$ (middle). The vertical scales are identical, ranging from 40 to 500 in both plots. The third plot (right) is the surface of the difference between $\lambda(\mathbf{x}, t, m)$ and $s(\mathbf{x}, t, m)$ ranging from $-34$ to 25. From left to right, the bottom row shows the surfaces $(\lambda - s)^2$, $(\lambda - s)^2/\lambda$, and $(\lambda - s)^2\,\lambda$.