Relative Risk Surfaces

Kernel density could using at epidemiology or social economy

- The demonstration script is **`KernelDensityLogRR.r`**.
- In relative risk surfaces the density $\lambda^{+}(\mathbf{s})$ of a case pattern is compared against the density of a baseline control pattern $\lambda^{0}(\mathbf{s})$ by calculating the relative risk ratio $rr(\mathbf{s}) = \frac{\lambda^{+}(\mathbf{s})}{\lambda^{0}(\mathbf{s})}$.
- Note that the relative risk is undefined if its control density $\lambda^{0}(\mathbf{s})$ is virtually zero. Therefore, it is advisable to over-smooth the control density to some degree by selecting a larger bandwidth. The bandwidth of kernel density(study extent) need to do data-driven
- The relative risk is interpreted as

Case vs Control

$$\begin{cases} rr(\mathbf{s}) > 1 & \text{case density is greater than control density at } \mathbf{s} \\ rr(\mathbf{s}) \sim 1 & \text{case density is equal to control density at } \mathbf{s} \\ rr(\mathbf{s}) < 1 & \text{case density is less than control density at } \mathbf{s} \end{cases}$$

- The distribution of the relative risk is highly positively skewed because its theoretical lower bound is zero whereas its theoretical upper bound is plus infinity.
- Therefore, frequently the log relative risk is used with $\log(rr(\mathbf{s})) \in [-\infty, \infty]$. The ***neutral value*** of equal densities now becomes ***zero*** and negative values indicate less case density, whereas positive value measure higher case density.
- The package **`smacpod`** calculates and evaluates log relative risk $\log(rr(\mathbf{s}))$ surfaces.
- Whether at a particular location the log relative risk is significantly larger than zero or significantly less than zero can be only evaluated by simulating alternative random pattern:

- Given the case locations $\{\mathbf{s}_1^+, \mathbf{s}_2^+, \dots, \mathbf{s}_m^+\}$ and the control locations $\{\mathbf{s}_1^o, \mathbf{s}_2^o, \dots, \mathbf{s}_n^o\}$ randomly relocate the marks for case and controls to the given locations $\mathbf{s}$ while maintaining the number of cases and controls m and n, respectively.
- This random mixing will break any case or control clusters up and the expected log relative risk at any location will be approximately zero.
- Repeat the random mixing K times and record the K log relative risk $\log\big(rr(\mathbf{s})\big)_k$ surfaces with $k = 1,2, \dots, K$ at all possible locations $\mathbf{s}$.
- Generate a sorted list of all $\log\big(rr(\mathbf{s})\big)_k$ and place the observed value $\log\big(rr(\mathbf{s})\big)_{obs}$ into this sorted list: $\log\big(rr(\mathbf{s})\big)_{[1]} \le \log\big(rr(\mathbf{s})\big)_{[2]} \le \cdots \le \log\big(rr(\mathbf{s})\big)_{obs} \le \cdots \le \log\big(rr(\mathbf{s})\big)_{[K-1]} \le \log\big(rr(\mathbf{s})\big)_{[K]}$. simulate distribution is very common in pattern analysis
  This sorted sequence of simulated log relative risks provides the distribution under the null hypothesis and allows to evaluate whether $\log\big(rr(\mathbf{s})\big)_{obs}$ falls into the tails of the simulate distribution.
- Base on the rank of $\log\big(rr(\mathbf{s})\big)_{obs}$ its significance level can be evaluated:
  - If $\log\big(rr(\mathbf{s})\big)_{obs}$ has a low tail rank a significantly lower case density is observed at the location $\mathbf{s}$.
  - Whereas if the tail rank of $\log\big(rr(\mathbf{s})\big)_{obs}$ is high a significant cluster of cases relative to the controls is present.

Kernel Densities on Networks

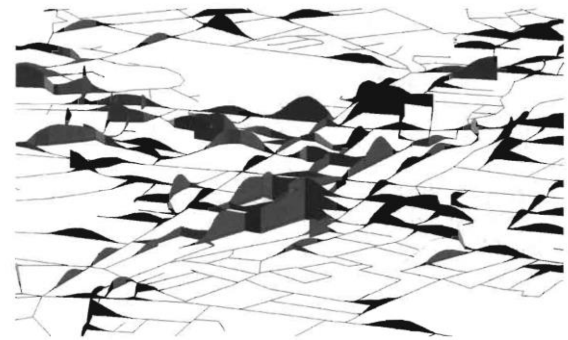

Figure 12.20 The density of beauty parlors around Omotesando subway station in Shibuya ward, Tokyo estimated using the equal-split continuous kernel density function (the distribution of beauty parlors is shown in Figure 12.7.

- Objective: Model the density of events, which are located on a network.
- The density follows the network and not the underlying surface. Thus, the network space is not homogenous.
  ⇨ Different locations are surrounded by different configurations of lines.
- Networks can be directed and acyclic (i.e., one cannot return to the origin). An example is a river network.
- Technically, the density associated with just one location $\mathbf{s}_i$ on a network segment still needs to integrate to one. This causes problems at segment forks (vertices), when the density is split at these forks:

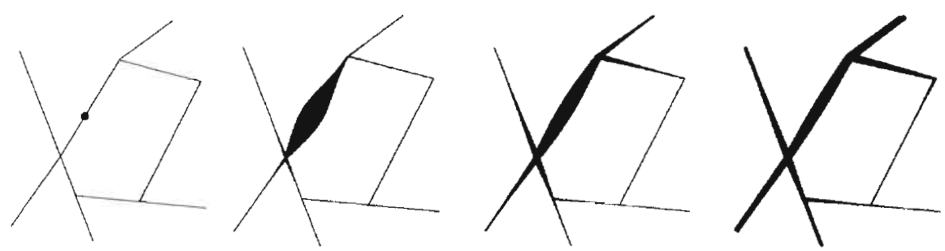

Figure 17.6. *Smoothing kernel on a network.* Left: *simple example of linear network and one data point u* (•). Other panels: *Line-thickness display of the smoothing kernel* $\kappa(u,v)$ *as a function of* v, *with bandwidth* $\sigma = 0.1, 0.2,$ *and* 0.3 *(left to right).*

- At the vertices the transition of the density curve can be continuous or discontinuous.

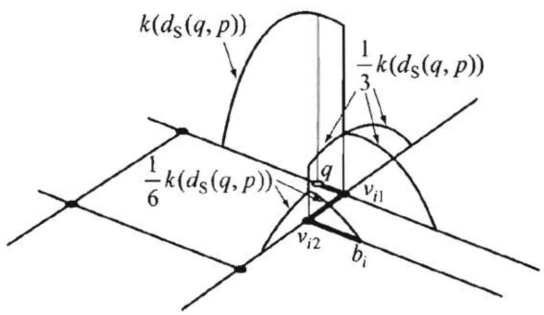

Figure 9.6 An equal-split discontinuous kernel density function centered at q, where $k(d_S(q,p))$ is a base kernel density function (the shortest path from q to b_i, $s(q,b_1)$, is indicated by the bold line segments).

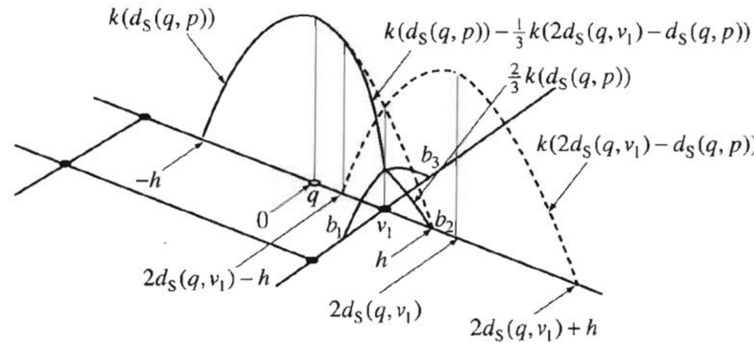

Figure 9.7 An equal-split continuous kernel density function centered at q, where $k(d_S(q,p))$ is a base kernel density function.

- The wider the bandwidth becomes and if a continuous transition is selected, the more computationally extensive the calculation of the network kernel density becomes.
- Okabe & Sugihara (2012). *Spatial Analysis along Networks. Statistical and Computational Methods.* Wiley, focus on the analysis of event distributions on a network. Their software **SANET**, however, becomes outdated.
- The package **`spatstat`** has function and specialized object classes for point patterns on a network.
- In **`spatstat`** points are snapped (relocated) onto their closest line segment on the network.
- See -script **`NetworkKernel.R`**.