Shrinkage estimators of the spatial relative risk function

Martin Hazelton[1]

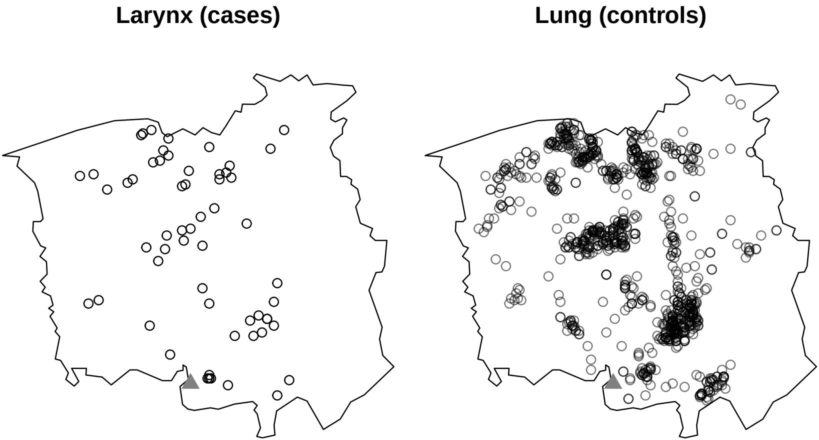
Department of Mathematics & Statistics
University of Otago

4-6 December 2023

[1] Email: martin.hazelton@otago.ac.nz

A Motivational Example

Locations of larynx cancer in Chorley Ribble region of UK



Spatial Relative Risk

- How to assess geographical variation in risk from case-control data?
- Observations lie in compact spatial region W.
- f is density function for spatial coordinates of cases; g for controls
- For $\boldsymbol{x} = (x_1, x_2)^{\mathsf{T}} \in W$, the **relative risk function** (Bithell, 1990) is

$$r(\boldsymbol{x}) = \frac{f(\boldsymbol{x})}{g(\boldsymbol{x})}.$$

- Describes only relative spatial differences, not overall intensity.
- Typical to work with: $\rho(\boldsymbol{x}) = \log(r(\boldsymbol{x})) = \log(f(\boldsymbol{x})) - \log(g(\boldsymbol{x}))$.
- $r(\boldsymbol{x}) = 1 \Leftrightarrow \rho(\boldsymbol{x}) = 0$ is null; $\rho(\boldsymbol{x}) > 0$ for elevated risk at $\boldsymbol{x}$.

Bithell, J.F. (1990). *Statistics in Medicine* **9**, 691–701.

Kernel Smoothing

- Data: marked point pattern $\{(\boldsymbol{x}_1, y_1), \ldots, (\boldsymbol{x}_n, y_n)\}$ on W.
  - $y = 1$ if case; $y = 0$ if control.
- Number cases $n_1 = \sum_{i=1}^{n} y_i$, number controls $n_2 = \sum_{i=1}^{n}(1 - y_i)$.
- Case and control densities estimated by kernel density estimation:

$$\hat{f}(\boldsymbol{x} \mid h) = \frac{1}{n_1} \sum_{i=1}^{n} y_i K_h(\boldsymbol{x} - \boldsymbol{x}_i)$$

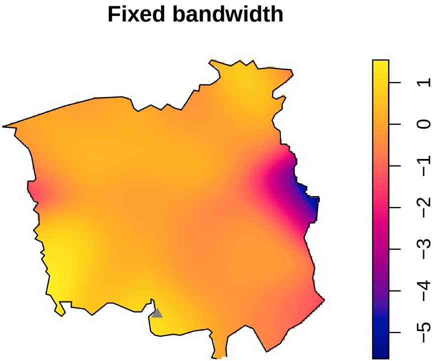
$$\hat{g}(\boldsymbol{x} \mid h) = \frac{1}{n_2} \sum_{i=1}^{n} (1 - y_i) K_h(\boldsymbol{x} - \boldsymbol{x}_i).$$

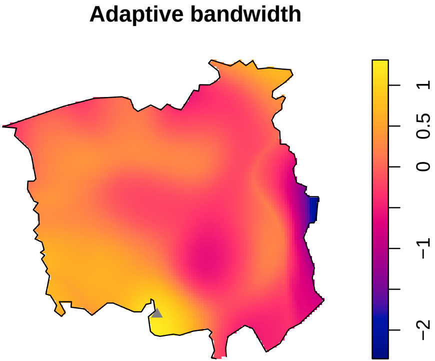
- Kernel K is isotropic density; scaled kernel $K_h(\boldsymbol{x}) = h^{-2} K(\boldsymbol{x}/h)$.
- Bandwidth h controls degree of smoothing.
- Relative risk estimate $\hat{r}(\boldsymbol{x}) = \hat{f}(\boldsymbol{x} \mid h) \Big/ \hat{g}(\boldsymbol{z} \mid h)$.

UNIVERSITY of OTAGO
Te Whare Wānanga o Otāgo
NEW ZEALAND

Application to Chorley Ribble

Estimates of log-relative risk function $\hat{\rho}(\boldsymbol{x})$





Shrinkage Estimation

Bithell's Method

- Null value is $r(\boldsymbol{x}) = 1 \Leftrightarrow \rho(\boldsymbol{x}) = 0$.
- Idea: shrink estimate towards null value in areas of sparse data.
  - Insufficient evidence there to warrant non-null estimate.
- Bithell's estimator:

$$\hat{r}_B(\boldsymbol{x} \mid h, \lambda) = \frac{\lambda k_0/n_1 + \hat{f}(\boldsymbol{x} \mid h)}{\lambda k_0/n_1 + \hat{g}(\boldsymbol{x} \mid h)}$$

- $k_0 = K_h(\mathbf{0}) = K(\mathbf{0})/h^2$.
- λ an interpretable tuning parameter, controlling degree of shrinkage.

UNIVERSITY of OTAGO
Te Whare Wānanga o Otāgo
NEW ZEALAND

Lasso Shrinkage Estimation

Local likelihood

- Consider estimation at $\boldsymbol{x} \in W$.
- Local constant estimator is $\rho(\boldsymbol{z}) = b$ for $\boldsymbol{z}$ in neighbourhood of $\boldsymbol{x}$.
- $\mathrm{P}(Y = 1 \mid \boldsymbol{z}, n_1, n_2) = n_1 e^b/(n_2 + n_1 e^b)$.
- Local log-likelihood [Tibshirani & Hastie (1987)]

$$
\begin{aligned}
L(b) &= \sum_{i=1}^{n} \log\left(\mathrm{P}(Y_i = y_i \mid \boldsymbol{x})\right) K_h(\boldsymbol{x} - \boldsymbol{x}_i) \\
&= b n_1 \hat{f}(\boldsymbol{x}) - [n_1 \hat{f}(\boldsymbol{x}) + n_2 \hat{g}(\boldsymbol{x})] \log\left(1 + \frac{n_1}{n_2} e^b\right) + c
\end{aligned}
$$

- $L(b)$ maximized by standard kernel estimator $\hat{\rho}(\boldsymbol{x}) = \hat{b}$.

Tibshirani R & Hastie T. (1987) *JASA* **82**, 559–567.

Lasso Shrinkage Estimation

Penalized Local likelihood

- Penalize negative local likelihood by L_1 (lasso) penalty:

$$Q(b) = -L(b) + \lambda k_0 |b|$$

- Minimize $Q(b)$ for lasso estimator $\hat{\rho}(\boldsymbol{x}) = \hat{b}$.
- Rationale: lasso estimators will shrink $\hat{\rho}(\boldsymbol{x})$ to exactly zero for sufficiently large λ.
- Bonus: $\hat{\rho}(\boldsymbol{x})$ available in closed form:

$$e^{\hat{b}} = \begin{cases} \frac{\hat{f}(\boldsymbol{x}) - \lambda k_0/n_1}{\hat{g}(\boldsymbol{x}) + \lambda k_0/n_2} & 1 < \frac{\hat{f}(\boldsymbol{x}) - \lambda k_0/n_1}{\hat{g}(\boldsymbol{x}) + \lambda k_0/n_2} \\ \frac{\hat{f}(\boldsymbol{x}) + \lambda k_0/n_1}{\hat{g}(\boldsymbol{x}) - \lambda k_0/n_2} & 0 < \frac{\hat{f}(\boldsymbol{x}) + \lambda k_0/n_1}{\hat{g}(\boldsymbol{x}) - \lambda k_0/n_2} < 1 \\ 1 & \text{otherwise.} \end{cases}$$

Choice of Shrinkage Parameter

Penalized Local likelihood

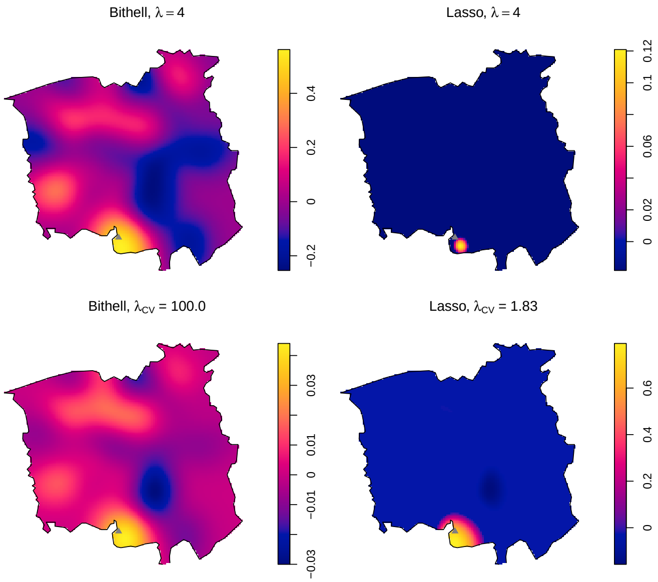
Rule of Thumb

- For lasso method, method effectively changes λ cases at $\boldsymbol{x}$ into controls.
- Suggests $\lambda = 4$.
- $\hat{\rho}$ shrunk to zero except in locations $\boldsymbol{x}$ where we would tend to reject $H_0 : \rho(\boldsymbol{x}) = 0$.

Cross-Validation

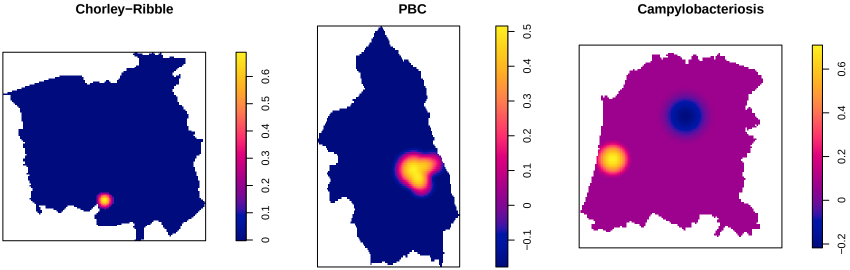
- Employ leave-one-out cross-validation based on Bernoulli log-likelihood.
- Choose left-hand local minimum in cases of multiple extrema.

Application to Chorley-Ribble Dataset



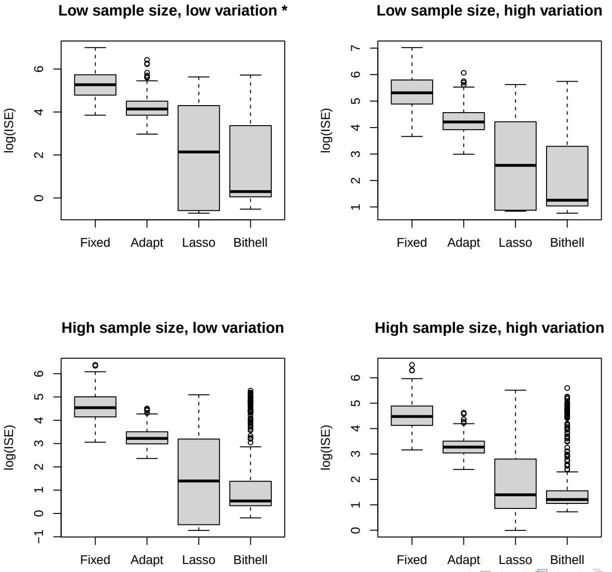
Some Numerical Results

Test Problems



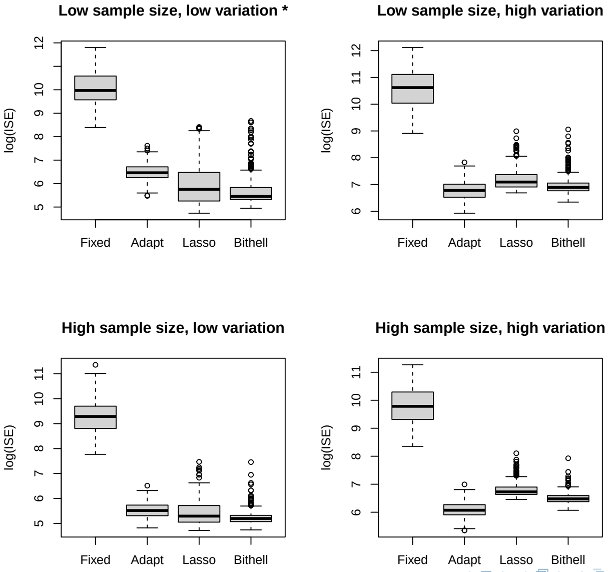
Some Numerical Results

Results for Problem 1



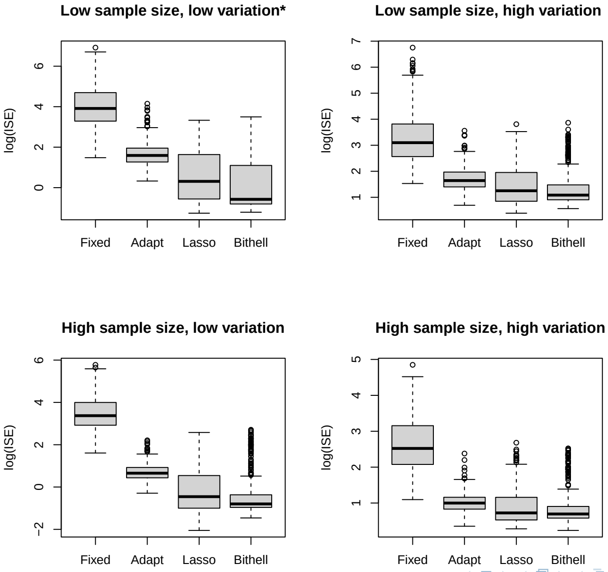
Some Numerical Results

Results for Problem 2



Some Numerical Results

Results for Problem 3



To Learn More ...

Journal Article

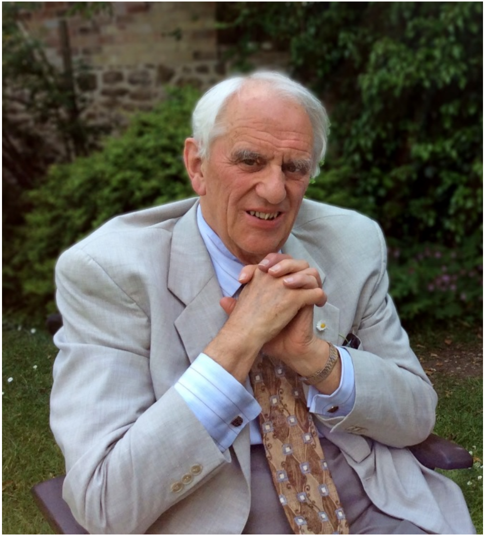
Hazelton, M.L. (2023). Shrinkage estimates of the spatial relative risk function. *Statistics in Medicine*, **42**, 4556-4569. `10.1002/sim.9875`.

Software

Implemented in function `risk` (argument `shrink=T`) in `sparr` package for R.

In Memoriam

Dr John Francis Bithell (1939–2020)



UNIVERSITY of OTAGO
Te Whare Wānanga o Otāgo
NEW ZEALAND