Spatio – Temporal Methods in Environmental Epidemiology: Supplementary Material for Chapter 7

Gavin Shaddick and James V Zidek

Measurement error in exposures

In this section we consider measurement error in the exposure $\mathbf{z}_{ik}$ rather than ecological bias, whereas the next section will consider both jointly. For simplicity so that ecological bias is not a problem we assume that each individual within area k has the same exposures, namely $\mathbf{z}_{ik} = \mathbf{z}_k$. However $\mathbf{z}_k$ is unknown and only M mis-measured estimates $\mathbf{w}_k(\mathbf{y^{(2)}}^1_k, \ldots, \mathbf{y^{(2)}}_{k^M})$ are available. Then adopting a classical measurement error model we obtain the decomposition

$$f(y_k^{(1)}, \mathbf{z}_k, \mathbf{y^{(2)}}_k | \mathbf{x}_k) = f(y_k^{(1)} | \mathbf{z}_k, \mathbf{y^{(2)}}_k, \mathbf{x}_k) f(\mathbf{y^{(2)}}_k | \mathbf{z}_k, \mathbf{x}_k) f(\mathbf{z}_k | \mathbf{x}_k)$$

where the first element on the right-hand side is the disease model, the second is the measurement error model (classical in this case) and the third is the exposure model. As $(\mathbf{z}_k)$ is constant across individuals the individual level Bernoulli risk model can be aggregated to a Binomial model, meaning that a measurement error model with log link is given by:

$$\begin{aligned}
y_k^{(1)} | \mathbf{z}_k &\sim \text{Binomial}(n_k, p_k) \\
\ln(p_k) &= \beta_0 + \mathbf{z}_k^T \beta_z + \mathbf{x}_k^T \beta_x, \\
\mathbf{y^{(2)}}^i_k &\sim \text{N}(\mathbf{z}_k, \Sigma_w) \\
\mathbf{z}_k &\sim \text{N}(\mu_k, \Sigma_z).
\end{aligned}$$

Using Bayes theorem the conditional distribution $\mathbf{z}_k | \bar{^{(2)}}_k$ can be calculated where $\bar{^{(2)}}_k$ is the mean of the samples. It is given by $\mathbf{z}_k | \bar{^{(2)}}_k \sim \text{N}(\mathbf{m}_k, V_k)$, where

$$\begin{aligned}
\mathbf{m}_k &= (I - Q)\mu_k + Q\bar{\mathbf{w}}_k \\
V_k &= (I - Q)\Sigma_z
\end{aligned}$$

where $Q = \Sigma_z(\Sigma_z + \Sigma_{y^{(2)}})^{-1}$. Then as $\mathbf{z}_k$ is unknown we require a distribution for $y_k^{(1)} | \mathbf{w}_k$ rather than for $y_k^{(1)} | \mathbf{z}_k$. The former is still a Binomial model as the risk function for each individual within area k is the same (exposure is constant). Therefore $y_k^{(1)} | \mathbf{y^{(2)}}_k \sim \text{Binomial}(n_k, p_k^*)$, where

$$\begin{aligned}
\mathbb{E}[y_k^{(1)} | \mathbf{y^{(2)}}_k] &= \mathbb{E}[\mathbb{E}[y_k^{(1)} | \mathbf{y^{(2)}}_k, \mathbf{z}_k]] \\
&= \mathbb{E}[n_k p_k | \mathbf{y^{(2)}}_k] \\
&= \mathbb{E}[n_k \exp(\beta_0 + \mathbf{z}_k^T \beta_z + \mathbf{x}_k^T \beta_x)] \\
&= n_k \exp(\beta_0 + \mathbf{z}_k^T \beta_z) \mathbb{E}[\mathbf{x}_k^T \beta_x | \mathbf{y^{(2)}}_k] \\
&= n_k \exp(\beta_0 + \mathbf{x}_k^T \beta_x) \exp((I - Q)\mu_k \beta_z + Q\bar{\mathbf{y^{(2)}}}_k \beta_z + \beta_z^T (I - Q)\Sigma_z \beta_z)
\end{aligned}$$

Here the term $(I-Q)\mu_k\beta_z$ has no dependence on $\mathbf{y^{(2)}}_k$ and can be absorbed into the intercept term.

References