Comparison of methods to assess divergent areas in disease mapping

Annibale Biggeri and Dolores Catelan
Department of Statistics "G. Parenti" University of Florence (IT)
Biostatistics Unit, ISPO Cancer Prevention and Research Institute, Florence

The first entry of the small-area variations issue in the epidemiological literature was on health service research (McPherson et al. 1982). The hierarchical model was clearly stated and the inference on area-specific relative risks was based on posterior summaries. League tables were built using posterior relative risk estimates or using posterior ranks (Goldstein and Spiegelhalter, 1996). Spatial dependence was not addressed.
Disease mapping focused on relative risk surface estimation. This explain why great emphasis was put on spatial patterns. Since the seminal paper of Clayton and Kaldor (1987) spatially-structured priors were considered in all the proposed models in the literature. The Besag, York and Mollié model (1991) is popular because it is more flexible. It combines spatially structured and un-structured random effects. Inference on area-specific relative risks received little attention. Stern and Cressie (1999, 2000) used cross-validation posterior predictive distribution to explore model fitting. Interestingly they noted that this approach resembles "a traditional significance-testing approach in the sense that a specific alternative model is not specified" (2000, page 2385). Let see in more detail this approach.
The posterior predictive distribution is:

$$P\left(\mathbf{Y}^{rep}|\mathbf{Y}\right)=\int P\left(\mathbf{Y}^{rep}|\,\boldsymbol{\theta}\,,\mathbf{Y}\right)P\left(\boldsymbol{\theta}\,|\,\mathbf{Y}\right)d\boldsymbol{\theta}=\int P\left(\mathbf{Y}^{rep}|\,\boldsymbol{\theta}\right)P\left(\boldsymbol{\theta}\,|\,\mathbf{Y}\right)d\boldsymbol{\theta}$$

Note that in deriving the predictive distribution we assume conditional independence of Y^{rep} and Y given the parameters.
Now define a discrepancy measure, for example the deviance, and compare its posterior distribution to its posterior predictive distribution. A posterior predictive p-value is defined and a summary measure over the joint posterior distribution or a vector of expected posterior predictive p-values can be derived (Gelman et al 1996):

$$p=E_{\boldsymbol{\theta}|\mathbf{Y}}\left[\Pr\left(D(\mathbf{Y}^{rep},\boldsymbol{\theta})\geq D(\mathbf{Y},\boldsymbol{\theta})\right)\right]$$

If we choose as a discrepancy measure a test statistic, which does not depend on model parameters, the posterior predictive p-values are given directly by comparing to the observed data. However, they are conservative and not uniformly distributed under the null. We can also say that we are actually using the data twice, for deriving posteriors and for obtaining replicates (Marshall and Spiegelhalter 2007). This reflects criticisms on posterior predictive criteria for model checking (Plummer 2008).
The usual solution to control for excess in optimism (Efron 1983) is cross-validation. The posterior predictive distribution is replaced by the cross-validation (leave-one-out) posterior predictive distributions:

$$P\left(Y_i^{rep}|\,\mathbf{Y}_{-i}\right)=\int P\left(Y_i^{rep}|\,\theta_i\right)P\left(\theta_i\,|\,\mathbf{Y}_{-i}\right)d\theta$$

the probability density calculated at Y_i was proposed by Geisser (1993) as a diagnostic check under the name *conditional predictive ordinate*. Notice that in disease mapping a log-linear random effects model is specified for the relative risk parameter. The model is highly parameterized with at least one random effect per area. The cross-validation posterior for θ_i is:

$$P(\theta_i \mid \mathbf{Y}_{-i}) = \int P(\theta_i \mid \boldsymbol{\theta}_{-i}, \mathbf{a})\, P(\boldsymbol{\theta}_{-i} \mid \mathbf{a}, \mathbf{Y}_{-i}) P(\mathbf{a} \mid \mathbf{Y}_{-i})\, \mathrm{d}\boldsymbol{\theta}_{-i}\, \mathrm{d}\mathbf{a} \ .$$

Stern and Cressie did not fully exploited the potentiality of this approach: they stay on model checking because the reference distribution was obtained under the alternative hypothesis and any discrepancy detected was interpreted as a symptom of lack of fit.
Marshall and Spiegelhalter (2003) noted this point: "… There are essentially two reasons why regions may be divergent. First, the statistical assumptions underlying the model may be incorrect. … [second], these regions could represent genuine 'hot-spots' of disease requiring further investigation." They developed further this second point in a series of papers on provider profiling, where they specified a hierarchical modelling of the null (Marshall and Spiegelhalter 2007, Ohlssen et al. 2007). Thus, we are back to health service research and the emphasis on profiling / league tables raise the question of multiple comparisons.
Again Marshall and Spiegelhalter (2007): "… As mentioned previously, it may at first appear strange that a Bayesian modelling procedure should lead to considerations of multiple comparisons and adjustment of p-values and so on. … it has been suggested that a more appropriate Bayesian procedure would be to model an alternative hypothesis and hence produce posterior probabilities of the null and alternative hypothesis, rather than p-values (Efron et al. 2001)…" They did not explain in details.

Here, we aim to show that there is a full range of different ways to model divergence. Instead of modelling the most plausible alternative and checking lack-of-fit or modelling the most plausible null and detecting sensible outlying observations, we consider a non parametric mixture prior. The advantage of this approach is that multiple comparison is inherently accounted for and a whole range informative priors can be easily incorporated.

McPherson et al. N Engl J Med. 1982 Nov 18;307(21):1310-4.

Clayton, D. and Kaldor, J. Biometrics 1987, 43:671-681.

Besag, J., York J. and Mollié A. Annals of the Institute of Statistical Mathematics, 1991; 43:1-59.

Goldstein, H, Spiegelhalter, DJ. Journal of the Royal Statistical Society A, 1996, 159:385-443

Ohlssen, D.I., Sharples, L.D. and Spiegelhalter, D.J. JRSS A, 2007, 170: 865-890.

Stern HS and Cressie N. in Lawson et al (eds) Wiley 1999

Stern HS and Cressie N. Statist. Med. 2000; 19:2377-2397.

Plummer, M. Biostatistics. 9, 523-539 (2008)

Efron B. Journal of the American Statistical Association 1983, 78, 316–331.

Geisser S. Chapman and Hall, 1993.

Marshall EC and Spiegelhalter DJ. Bayesian Analysis (2007) 2, Number 2, pp. 409-444

Gelman A, Meng X-L, and Stern, HS. Statistica Sinica, 1996, 6: 733-760.