A spatial model for rare binary events

May 22, 2014

1 Introduction

2 New Model?

Let $Y_i \in \{0, 1\}$ be the binary response at spatial location $\mathbf{s}_i \in \mathcal{D}$, and $\mathbf{X}_i$ be the associated p-vector of covariates with first element equal to one for the intercept. We relate the covariates with the response using the link function g so that $\mathrm{P}(Y_i = 1) = p_i = g(\mathbf{X}_i \boldsymbol{\beta})$, where $\boldsymbol{\beta}$ is the p-vector of regression coefficients. For example, Wang and Day (2010) propose the GEV link function $p_i = 1 - \exp\left[(1 - \xi \mathbf{X}_i \boldsymbol{\beta})^{-1/\xi}\right]$ for rare binary data. We will also consider logit and probit links.

– Not quite sure why the article uses this. I think we should use

$$p_i = 1 - \exp\left[-\left(1 + \xi \mathbf{X}_i \boldsymbol{\beta}\right)^{-1/\xi}\right] \tag{1}$$

We propose a copula (Nelsen, 1999) to account for spatial dependence while preserving the marginal event probabilities. Let $Y_i = I(Z_i > z_i)$, where Z_i is a continuous latent variable and z_i is the appropriate threshold so that $P(Y_i = 1) = p_i$. The latent Z_i is modeled using spatial extreme value analysis methods to capture dependence between rare events. We assume Z follows the max-stable spatial process of Reich and Shaby (2012). Under this model, the marginal distribution of each Z_i is GEV(1,1,1) with $\mathrm{P}(Z_i > c) = 1 - \exp(-1/c)$. Therefore, we must set $z_i = -1/\log(1 - p_i)$ so that $P(Y_i = 1) = p_i$.

Spatial dependence is determined by the joint distribution of $\mathbf{Z} = (Z_1, \ldots, Z_n)$,

$$G(\mathbf{z}) = \mathrm{P}[Z_1 < z_1, \ldots, Z_n < z_n] = \exp\left\{-\sum_{l=1}^{L}\left[\sum_{i=1}^{n}\left(\frac{w_l(\mathbf{s}_i)}{z_i}\right)^{1/\alpha}\right]^{\alpha}\right\}, \tag{2}$$

where $\mathbf{z} = (z_1, \ldots, z_n)$. This is a special case of the multivariate GEV distribution with asymmetric Laplace dependence function (Tawn, 1990). The parameter $\alpha \in (0, 1)$ determines the strength of dependence, with α near zero giving strong dependence and $\alpha = 1$ giving joint independence. The weights $w_{li} > 0$ determine the spatial dependence structure, and are discussed in detail in Section 3. Many weight functions are possible, but the weights must be constrained so that $\sum_{l=1}^{L} w_l(\mathbf{s}_i) = 1$ for all $i = 1, \ldots, n$ to preserve the marginal GEV distribution.

3 Spatial dependence

The weights $w_l(\mathbf{s}_i)$ in (2) should vary smoothly across space to induce spatial dependence. For example, Reich and Shaby (2012) take the weights to be scaled Gaussian kernels with knots $\mathbf{v}_l$, that is

$$w_l(\mathbf{s}_i) = \frac{\exp\left[-0.5\left(||\mathbf{s}_i - \mathbf{v}_l||/\rho\right)^2\right]}{\sum_{j=1}^{L}\exp\left[-0.5\left(||\mathbf{s}_i - \mathbf{v}_j||/\rho\right)^2\right]}. \tag{3}$$

To kernel bandwidth $\rho > 0$ determines the spatial range of the dependence, with large ρ giving long-range dependence and vice versa.

Then in a bivariate setting, the probability of observing a joint exceedances as a function of α is

$$
\begin{aligned}
\mathrm{P}(Y_i = 1, Y_j = 1) &= 1 - \exp\left\{-\frac{1}{z_i}\right\} - \exp\left\{-\frac{1}{z_j}\right\} + \exp\left\{-\sum_{l=1}^{L}\left[\left(\frac{w_l(\mathbf{s}_i)}{z_i}\right)^{1/\alpha} + \left(\frac{w_l(\mathbf{s}_i)}{z_j}\right)^{1/\alpha}\right]^{\alpha}\right\} \\
&= p_i + p_j - \left(1 - \exp\left\{-\sum_{l=1}^{L}\left(\left[-\log(1-p_i)w_l(\mathbf{s}_i)\right]^{1/\alpha} + \left[-\log(1-p_j)w_l(\mathbf{s}_j)\right]^{1/\alpha}\right)^{\alpha}\right\}\right).
\end{aligned}
\tag{4}
$$

To describe the tail dependence, we use the χ statistic of Coles et al. (1999). Assume that Y_i and Y_j have the same marginal distributions, then $p_i = p_j = p$ for all i, j. As shown in Appendix A.2,

$$
\chi = 2 - \vartheta(\mathbf{s}_i, \mathbf{s}_j). \tag{5}
$$

where $\vartheta(\mathbf{s}_i, \mathbf{s}_j) = \sum_{l=1}^{L}\left[w_l(\mathbf{s}_i)^{1/\alpha} + w_l(\mathbf{s}_j)^{1/\alpha}\right]^{\alpha}$ is the pairwise extremal coefficient given by Reich and Shaby (2012). In the case of complete dependence, $\chi = 1$, and in the case of complete independence, $\chi = 0$. This is relatively easy to show for $\alpha = 1$, but I don't know of a way to prove $\lim_{\alpha \to 0} \chi = 1$. Any thoughts?

4 Computation

As shown in Appendix A.1, the joint probability mass function of $\mathbf{Y} = (Y_1, \ldots, Y_n)$ has a convenient form when the number of events is small. Let $K = \sum_{i=1}^{n} Y_i$ be the number of events, and assume without loss of generality the data are ordered so that the $Y_1 = \ldots = Y_K = 1$. Then

$$
P(Y_1 = y_1, \ldots, Y_n = y_n) = \begin{cases} G(\mathbf{z}) & K = 0 \\ G(\mathbf{z}_{(1)}) - G(\mathbf{z}) & K = 1 \\ G(\mathbf{z}_{(12)}) - G(\mathbf{z}_{(1)}) - G(\mathbf{z}_{(2)}) + G(\mathbf{z}) & K = 2 \end{cases} \tag{6}
$$

where $G(\mathbf{z}_{(1)}) = P(Z_2 < z_2, \ldots, Z_n < z_n)$, $G(\mathbf{z}_{(2)}) = P(Z_1 < z_1, Z_3 < z_3, \ldots, Z_n < z_n)$, and $G(\mathbf{z}_{(12)}) = P(Z_3 < z_3, \ldots, Z_n < z_n)$. Similar expressions can be derived for all K, but become cumbersome for large K. Therefore, for small K we can evaluate the likelihood directly. Most days in our dataset have $K < 4$, so we use this expression for those days. However for days with many events, we must use the latent variable scheme described below (unless you can think of a better way!). I think it should be more computationally efficient to use (6) for any K. At most, we have to calculate the $\left(\frac{w_l(\mathbf{s}_i)}{z_i}\right)^{1/\alpha}$, for all i, l. In the random effects model, the expression for the joint density conditional on θ is

$$
P(Y_1 = y_1, \ldots, Y_n = y_n) = \prod_{i=1}^{n}\left[\exp\left\{\sum_{l=1}^{L} A_l \left(\frac{w_l(\mathbf{s}_i)}{z_i}\right)^{1/\alpha}\right\}\right]^{1-Y_i}\left[1 - \exp\left\{\sum_{l=1}^{L} A_l \left(\frac{w_l(\mathbf{s}_i)}{z_i}\right)^{1/\alpha}\right\}\right]^{Y_i}. \tag{7}
$$

So we still need to compute $\left(\frac{w_l(\mathbf{s}_i)}{z_i}\right)^{1/\alpha}$, but we also need to do the sampling for all the A_l terms as well.

5 Simulation study

6 Data analysis

7 Conclusions

Acknowledgments

Appendix A.1: Derivation of the likelihood

We use the hierarchical max-stable spatial model given by Reich and Shaby (2012). If at each margin, $Z_i \sim \text{GEV}(1,1,1)$, then $Z_i|\theta_i \overset{indep}{\sim} \text{GEV}(\theta, \alpha\theta, \alpha)$. As defined in section 4, we reorder the data such that $Y_1 = \ldots = Y_K = 1$, and $Y_{K+1} = \ldots = Y_n = 0$. Then the joint likelihood conditional on the random effect θ is

$$
\begin{aligned}
P(Y_1 = y_1, \ldots, Y_n = y_n) &= \prod_{i \leq K} \left\{ 1 - \exp\left[-\left(\frac{\theta_i}{z_i}\right)^{1/\alpha} \right] \right\} \prod_{i > K} \exp\left[-\left(\frac{\theta_i}{z_i}\right)^{1/\alpha} \right] \\
&= \exp\left[-\sum_{i=K+1}^{n} \left(\frac{\theta_i}{z_i}\right)^{1/\alpha} \right] - \exp\left[-\sum_{i=K+1}^{n} \left(\frac{\theta_i}{z_i}\right)^{1/\alpha} \right] \sum_{i=1}^{K} \exp\left[-\left(\frac{\theta_i}{z_i}\right)^{1/\alpha} \right] \\
&+ \exp\left[-\sum_{i=K+1}^{n} \left(\frac{\theta_i}{z_i}\right)^{1/\alpha} \right] \sum_{1 < i < j \leq K} \left\{ \exp\left[-\left(\frac{\theta_i}{z_i}\right)^{1/\alpha} - \left(\frac{\theta_j}{z_j}\right)^{1/\alpha} \right] \right\} \\
&+ \cdots + (-1)^K \exp\left[-\sum_{i=1}^{n} \left(\frac{\theta_i}{z_i}\right)^{1/\alpha} \right] \qquad (8)
\end{aligned}
$$

Finally marginalizing over the random effect, we obtain

$$
\begin{aligned}
P(Y_1 = y_1, \ldots, Y_n = y_n) &= \int G(\mathbf{z}|\mathbf{A}) p(\mathbf{A}|\alpha) d\mathbf{A}. \\
&= \int \exp\left[-\sum_{i=K+1}^{n} \left(\frac{\theta_i}{z_i}\right)^{1/\alpha} \right] - \exp\left[-\sum_{i=K+1}^{n} \left(\frac{\theta_i}{z_i}\right)^{1/\alpha} \right] \sum_{i=1}^{K} \exp\left[-\left(\frac{\theta_i}{z_i}\right)^{1/\alpha} \right] \\
&+ \exp\left[-\sum_{i=K+1}^{n} \left(\frac{\theta_i}{z_i}\right)^{1/\alpha} \right] \sum_{1 < i < j \leq K} \left\{ \exp\left[-\left(\frac{\theta_i}{z_i}\right)^{1/\alpha} - \left(\frac{\theta_j}{z_j}\right)^{1/\alpha} \right] \right\} \\
&+ \cdots + (-1)^K \exp\left[-\sum_{i=1}^{n} \left(\frac{\theta_i}{z_i}\right)^{1/\alpha} \right] p(\mathbf{A}|\alpha) d\mathbf{A}. \qquad (9)
\end{aligned}
$$

Consider the first term in the summation,

$$\int \exp\left\{-\sum_{i=K+1}^{n}\left(\frac{\theta_i}{z_i}\right)^{1/\alpha}\right\} p(\mathbf{A}|\alpha)d\mathbf{A} = \int \exp\left\{-\sum_{i=K+1}^{n}\left(\frac{\left[\left(\sum_{l=1}^{L} A_l w_l(\mathbf{s}_i)^{1/\alpha}\right)^{\alpha}\right]}{z_i}\right]^{1/\alpha}\right\} p(\mathbf{A}|\alpha)d\mathbf{A}$$
$$= \int \exp\left\{-\sum_{i=K+1}^{n}\sum_{l=1}^{L} A_l \left(\frac{w_l(\mathbf{s}_i)}{z_i}\right)^{1/\alpha}\right\} p(\mathbf{A}|\alpha)d\mathbf{A}$$
$$= \exp\left\{-\sum_{l=1}^{L}\left[\sum_{i=K+1}^{n}\left(\frac{w_l(\mathbf{s}_i)}{z_i}\right)^{1/\alpha}\right]^{\alpha}\right\}. \tag{10}$$

The remaining terms in equation (9) are straightforward to obtain, and after integrating out the random effect, the joint density is the density given in (6).

Appendix A.2: Derivation of the χ statistic

$$\chi = \lim_{p\to 0} \mathrm{P}(Y_i = 1 | Y_j = 1)$$
$$= \lim_{p\to\infty} \frac{p + p - \left(1 - \exp\left\{-\sum_{l=1}^{L}\left[(-\log(1-p)w_l(\mathbf{s}_i))^{1/\alpha} + (-\log(1-p)w_l(\mathbf{s}_j))^{1/\alpha}\right]^{\alpha}\right\}\right)}{p}$$
$$= \lim_{p\to 0} \frac{2p - \left(1 - \exp\left\{\log(1-p)\sum_{l=1}^{L}\left[w_l(\mathbf{s}_i)^{1/\alpha} + w_l(\mathbf{s}_j)^{1/\alpha}\right]^{\alpha}\right\}\right)}{p}$$
$$= \lim_{p\to 0} \frac{2p - \left(1 - (1-p)^{\sum_{l=1}^{L}\left[(w_l(\mathbf{s}_i))^{1/\alpha} + (w_l(\mathbf{s}_j))^{1/\alpha}\right]^{\alpha}}\right)}{p}$$
$$= \lim_{p\to 0} 2 - \sum_{l=1}^{L}\left[w_l(\mathbf{s}_i)^{1/\alpha} + w_l(\mathbf{s}_j)^{1/\alpha}\right]^{\alpha}(1-p)^{-1+\sum_{l=1}^{L}\left[w_l(\mathbf{s}_i)^{1/\alpha} + w_l(\mathbf{s}_j)^{1/\alpha}\right]^{\alpha}}$$
$$= 2 - \sum_{l=1}^{L}\left[w_l(\mathbf{s}_i)^{1/\alpha} + w_l(\mathbf{s}_j)^{1/\alpha}\right]^{\alpha}. \tag{11}$$

References

Coles, S., Heffernan, J. and Tawan, J. (1999) Dependence measures for extreme value analyses. *Extremes*, **2**, 339–365.

Nelsen, R. B. (1999) *An introduction to copulas*. New York: Springer-Verlag.

Reich, B. and Shaby, B. (2012) A hierarchical max-stable spatial model from extreme precipitation. *The Annals of Applied Statistics*, **6**, 1430–1451.

Tawn, J. A. (1990) Modelling multivariate extreme value distributions. *Biometrika*, **77**, 245–253.

Wang, X. and Day, D. K. (2010) Generalized extreme value regression for binary response data: An application to b2b electronic payments system adoption. *The Annals of Applied Statistics*, **64**, 2000–2023.