

A spatial model for rare binary events

September 9, 2015

1 Introduction

2 Binary regression

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. The goal in binary regression is to relate a set of covariates with the response using the link function g so that $P(Y_i = 1) = \pi_i = g(\mathbf{X}_i\boldsymbol{\beta})$, where \mathbf{X}_i is the vector of covariates for observation i , and $\boldsymbol{\beta}$ is the p -vector of regression coefficients. Two very commonly used types of binary regression include logistic regression with $\pi_i = \frac{\exp \mathbf{X}_i\boldsymbol{\beta}}{1 + \exp \mathbf{X}_i\boldsymbol{\beta}}$ and probit regression with $\pi_i = \Phi(\mathbf{X}_i\boldsymbol{\beta})$ where $\Phi(\cdot)$ represents the standard normal distribution function. One limitation to these link functions is that they assume the data are symmetric which may not always be the case. More recently, Wang and Dey (2010) introduced the generalized extreme value (GEV) link function for rare binary data where

$$\pi_i = 1 - \exp \left[- (1 - \xi \mathbf{X}_i\boldsymbol{\beta})^{-1/\xi} \right]. \quad (1)$$

In the case that $\xi = 0$, this is the complementary log-log (cloglog) link function. The GEV link function is attractive because it allows for asymmetry thereby providing a more flexible model for the data.

3 Spatial dependence

In this section we extend the GEV link function to allow for spatial correlation. First, to simplify the notation, we let $z_i = (1 - \xi \mathbf{X}_i \beta)^{1/\xi}$ from (1). To incorporate spatial dependence into the model, we consider the hierarchical max-stable process of Reich and Shaby (2012). The spatial dependence is determined by the joint distribution of $\mathbf{Z} = (Z_1, \dots, Z_n)$,

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

where $\mathbf{z} = (z_1, \dots, z_n)$, $w_l(\mathbf{s}_i)$ are a set of weights that determine the spatial dependence structure and are discussed further in Section 3.1, and $\alpha \in (0, 1)$ determines the strength of dependence, with α near zero giving strong dependence and $\alpha = 1$ giving joint independence. This is a special case of the multivariate GEV distribution with asymmetric Laplace dependence function (Tawn, 1990). One nice feature to this hierarchical model is that the lower-dimensional marginal distributions also follow a multivariate extreme value distribution. More importantly, at a single site i , the marginal distribution gives $P(Y_i = 1) = 1 - \exp \left\{ -\frac{1}{z_i} \right\}$ which is the same as the marginal distributions given by Wang and Dey (2010).

3.1 Weight functions

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 $i = 1, \dots, n$ to preserve the marginal GEV distribution. 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

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

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

35 **4 Multivariate distribution**

36 As shown in Appendix A.1, the joint probability mass function of $\mathbf{Y} = (Y_1, \dots, Y_n)$ has a convenient form
 37 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
 38 generality the data are ordered so that the $Y_1 = \dots = Y_K = 1$. Then

$$P(Y_1 = y_1, \dots, Y_n = y_n) = \begin{cases} G(\mathbf{z}) & K = 0 \\ G(\mathbf{z}_{(1)}) - G(\mathbf{z}) & K = 1 \\ G(\mathbf{z}_{(12)}) - G(\mathbf{z}_{(1)}) - G(\mathbf{z}_{(2)}) + G(\mathbf{z}) & K = 2 \end{cases} \quad (4)$$

39 where $G(\mathbf{z}_{(1)}) = P(Z_2 < z_2, \dots, Z_n < z_n)$, $G(\mathbf{z}_{(2)}) = P(Z_1 < z_1, Z_3 < z_3, \dots, Z_n < z_n)$, and
 40 $G(\mathbf{z}_{(12)}) = P(Z_3 < z_3, \dots, Z_n < z_n)$. Similar expressions can be derived for all K , but become cumber-
 41 some for large K .

42 **4.1 Bivariate distribution**

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

$$P(Y_i = 1, Y_j = 1) = 1 - \exp\left\{-\frac{1}{z_i}\right\} - \exp\left\{-\frac{1}{z_j}\right\} + \exp\left\{-\sum_{l=1}^L \left[\left(\frac{w_l(\mathbf{s}_i)}{z_i}\right)^{1/\alpha} + \left(\frac{w_l(\mathbf{s}_j)}{z_j}\right)^{1/\alpha}\right]^\alpha\right\} \quad (5)$$

44 In the literature on extremes, one common metric to describe the bivariate dependence is the χ statistic of
 45 Coles et al. (1999). The χ statistic between two observations z_1 and z_2 is given by

$$\chi(\mathbf{s}_1, \mathbf{s}_2) = \lim_{c \rightarrow \infty} = P(Z_1 > c | Z_2 > c). \quad (6)$$

46 However, in this latent variable approach, $\lim_{c \rightarrow \infty}$ may not be the most reasonable metric because the
 47 observed data are a series of zeros and ones. Therefore, we chose the κ statistic of Cohen (1960) defined by

$$\kappa = \frac{P(A) - P(E)}{1 - P(E)} \quad (7)$$

48 where $P(A)$ is the joint probability of agreement and $P(E)$ is the joint probability of agreement under an
 49 assumption of independence. We believe this measure of dependence to be reasonable because,

$$\lim_{\beta_0 \rightarrow \infty} \kappa(h) = \chi(h) = 2 - \vartheta(\mathbf{s}_i, \mathbf{s}_j) \quad (8)$$

50 where β_0 is the intercept from $\mathbf{X}^T \boldsymbol{\beta}$ and $\vartheta(\mathbf{s}_i, \mathbf{s}_j) = \sum_{l=1}^L [w_l(\mathbf{s}_i)^{1/\alpha} + w_l(\mathbf{s}_j)^{1/\alpha}]^\alpha$ is the pairwise extremal
 51 coefficient given by Reich and Shaby (2012) (see Appendix A.2). In the case of complete dependence,
 52 $\kappa = 1$, and in the case of complete independence, $\kappa = 0$.

5 Computation

For small K we can evaluate the likelihood directly. In the random effects model, the expression for the joint density conditional on θ is

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

(9)

6 Simulation study

For our simulation study, we generate $n_m = 100$ datasets under 4 different settings to explore the impact of rareness of observations, sample size, and knot spacing. We consider two degrees of rareness $\pi = 0.01, 0.05$ and two sample sizes $n_s = 1000, 2000$. For the different knot spacings, we use knots in $[0, 1] \times [0, 1]$ on a 21×21 grid and 31×31 grid. For each dataset, we fit the model using three different methods, spatial logistic regression, spatial probit regression, and the proposed spatial GEV method. In each case, we fit the model using Bayesian methods with proper, but fairly uninformative priors. For each method, we fit the model using 75% of the observations as a training set, and the remaining observations are used as a validation set to assess the model's predictive power.

7 Data analysis

For the data analysis, we consider data from the eBirds dataset, a citizen-based observation network of bird sightings in the United States (Sullivan et al., 2009). The data are publicly available from <http://ebird.org>. We use data from 2002, and focus specifically on cattle egrets and vesper sparrows.

69 8 Conclusions

70 Acknowledgments

71 A Appendices

72 A.1 Derivation of the likelihood

73 We use the hierarchical max-stable spatial model given by Reich and Shaby (2012). If at each margin,
 74 $Z_i \sim \text{GEV}(1, 1, 1)$, then $Z_i | \theta_i \stackrel{\text{indep}}{\sim} \text{GEV}(\theta, \alpha\theta, \alpha)$. As defined in section 5, we reorder the data such that
 75 $Y_1 = \dots = Y_K = 1$, and $Y_{K+1} = \dots = Y_n = 0$. Then the joint likelihood conditional on the random effect
 76 θ is

$$\begin{aligned}
 P(Y_1 = y_1, \dots, Y_n = y_n) &= \prod_{i \leq K} \left\{ 1 - \exp \left[- \left(\frac{\theta_i}{z_i} \right)^{1/\alpha} \right] \right\} \prod_{i > K} \exp \left[- \left(\frac{\theta_i}{z_i} \right)^{1/\alpha} \right] \\
 &= \exp \left[- \sum_{i=K+1}^n \left(\frac{\theta_i}{z_i} \right)^{1/\alpha} \right] - \exp \left[- \sum_{i=K+1}^n \left(\frac{\theta_i}{z_i} \right)^{1/\alpha} \right] \sum_{i=1}^K \exp \left[- \left(\frac{\theta_i}{z_i} \right)^{1/\alpha} \right] \\
 &\quad + \exp \left[- \sum_{i=K+1}^n \left(\frac{\theta_i}{z_i} \right)^{1/\alpha} \right] \sum_{1 < i < j \leq K} \left\{ \exp \left[- \left(\frac{\theta_i}{z_i} \right)^{1/\alpha} - \left(\frac{\theta_j}{z_j} \right)^{1/\alpha} \right] \right\} \\
 &\quad + \dots + (-1)^K \exp \left[- \sum_{i=1}^n \left(\frac{\theta_i}{z_i} \right)^{1/\alpha} \right]
 \end{aligned} \tag{10}$$

77 Finally marginalizing over the random effect, we obtain

$$\begin{aligned}
P(Y_1 = y_1, \dots, Y_n = y_n) &= \int G(\mathbf{z}|\mathbf{A})p(\mathbf{A}|\alpha)d\mathbf{A}. \\
&= \int \exp \left[- \sum_{i=K+1}^n \left(\frac{\theta_i}{z_i} \right)^{1/\alpha} \right] - \exp \left[- \sum_{i=K+1}^n \left(\frac{\theta_i}{z_i} \right)^{1/\alpha} \right] \sum_{i=1}^K \exp \left[- \left(\frac{\theta_i}{z_i} \right)^{1/\alpha} \right] \\
&\quad + \exp \left[- \sum_{i=K+1}^n \left(\frac{\theta_i}{z_i} \right)^{1/\alpha} \right] \sum_{1 \leq i < j \leq K} \left\{ \exp \left[- \left(\frac{\theta_i}{z_i} \right)^{1/\alpha} - \left(\frac{\theta_j}{z_j} \right)^{1/\alpha} \right] \right\} \\
&\quad + \dots + (-1)^K \exp \left[- \sum_{i=1}^n \left(\frac{\theta_i}{z_i} \right)^{1/\alpha} \right] p(\mathbf{A}|\alpha)d\mathbf{A}. \tag{11}
\end{aligned}$$

78 Consider the first term in the summation,

$$\begin{aligned}
\int \exp \left\{ - \sum_{i=K+1}^n \left(\frac{\theta_i}{z_i} \right)^{1/\alpha} \right\} p(\mathbf{A}|\alpha)d\mathbf{A} &= \int \exp \left\{ - \sum_{i=K+1}^n \left(\frac{\left[\sum_{l=1}^L A_l w_l(\mathbf{s}_i)^{1/\alpha} \right]^\alpha}{z_i} \right)^{1/\alpha} \right\} p(\mathbf{A}|\alpha)d\mathbf{A} \\
&= \int \exp \left\{ - \sum_{i=K+1}^n \sum_{l=1}^L A_l \left(\frac{w_l(\mathbf{s}_i)}{z_i} \right)^{1/\alpha} \right\} p(\mathbf{A}|\alpha)d\mathbf{A} \\
&= \exp \left\{ - \sum_{l=1}^L \left[\sum_{i=K+1}^n \left(\frac{w_l(\mathbf{s}_i)}{z_i} \right)^{1/\alpha} \right]^\alpha \right\}. \tag{12}
\end{aligned}$$

79 The remaining terms in equation (11) are straightforward to obtain, and after integrating out the random

80 effect, the joint density is the density given in (4).

81 A.2 Derivation of the χ statistic

$$\begin{aligned}
\chi &= \lim_{p \rightarrow 0} P(Y_i = 1 | Y_j = 1) \\
&= \lim_{p \rightarrow \infty} \frac{p + p - \left(1 - \exp \left\{ - \sum_{l=1}^L \left[(-\log(1-p)w_l(\mathbf{s}_i))^{1/\alpha} + (-\log(1-p)w_l(\mathbf{s}_j))^{1/\alpha} \right]^\alpha \right\} \right)}{p} \\
&= \lim_{p \rightarrow 0} \frac{2p - \left(1 - \exp \left\{ \log(1-p) \sum_{l=1}^L \left[w_l(\mathbf{s}_i)^{1/\alpha} + w_l(\mathbf{s}_j)^{1/\alpha} \right]^\alpha \right\} \right)}{p} \\
&= \lim_{p \rightarrow 0} \frac{2p - \left(1 - (1-p)^{\sum_{l=1}^L \left[w_l(\mathbf{s}_i)^{1/\alpha} + w_l(\mathbf{s}_j)^{1/\alpha} \right]^\alpha} \right)}{p} \\
&= \lim_{p \rightarrow 0} 2 - \sum_{l=1}^L \left[w_l(\mathbf{s}_i)^{1/\alpha} + w_l(\mathbf{s}_j)^{1/\alpha} \right]^\alpha (1-p)^{-1 + \sum_{l=1}^L \left[w_l(\mathbf{s}_i)^{1/\alpha} + w_l(\mathbf{s}_j)^{1/\alpha} \right]^\alpha} \\
&= 2 - \sum_{l=1}^L \left[w_l(\mathbf{s}_i)^{1/\alpha} + w_l(\mathbf{s}_j)^{1/\alpha} \right]^\alpha. \tag{13}
\end{aligned}$$

82 References

- 83 Cohen, J. (1960) A Coefficient of Agreement for Nominal Scales. *Educational and Psychological Measure-*
84 *ment*, **20**, 37–46.
- 85 Coles, S., Heffernan, J. and Tawn, J. (1999) Dependence Measures for Extreme Value Analyses. *Extremes*,
86 **2**, 339–365.
- 87 Reich, B. J. and Shaby, B. A. (2012) A hierarchical max-stable spatial model for extreme precipitation. *The*
88 *Annals of Applied Statistics*, **6**, 1430–1451.
- 89 Sullivan, B. L., Wood, C. L., Iliff, M. J., Bonney, R. E., Fink, D. and Kelling, S. (2009) eBird: A citizen-
90 based bird observation network in the biological sciences. *Biological Conservation*, **142**, 2282–2292.
- 91 Tawn, J. A. (1990) Modelling multivariate extreme value distributions. *Biometrika*, **77**, 245–253.
- 92 Wang, X. and Dey, D. K. (2010) Generalized extreme value regression for binary response data: An appli-
93 cation to B2B electronic payments system adoption. *The Annals of Applied Statistics*, **4**, 2000–2023.