Small Scale Sim

Morris, S.

July 1, 2015

Small-scale simulation results

Data settings

I generated data at 2000 sites using the following settings:

- $\alpha = 0.3$
- $\rho = 0.1$
- $\xi = 0$

This is 1500 sites for training, and 500 sites for cross validation. The MCMC ran for 45000 iterations with 35000 burnin, and convergence isn't great, but it's more stable than before.

Methods

There were 5 methods used to fit the datasets.

1a: Fix α and ρ in the MCMC to be the estimates from the pairwise composite likelihood. Only fit β and random effects in the MCMC.

1b: Fix *alpha* in the MCMC to be the estimate from the pairwise composite likelihood, and fix *rho* in the MCMC to be the knot spacing. Only fit β and random effects in the MCMC.

1c: Fit α, ρ, *beta*, and random effect in the MCMC.

2: Logit

3: Probit

Methods 1a, 1b, and 1c are all variations on rare binary. Method 1a uses the pairwise composite likelihood to estimate α and ρ. Method 1b uses the pairwise composite likelihood to estimate α with ρ taken as the knot spacing. Both 1a and 1b take β to be the estimate for β marginally for the sites when estimating α and ρ.

Results

Here are the results for 10 datasets:

Table 1: Simulation results (x 100) for 10 datasets

	1a	1b	1c	2	3
1	2.739	2.356	2.331	5.477	2.317
2	2.071	1.881	1.813	1.66	1.752
3	3.932	3.799	3.816	3.745	3.763
4	2.201	1.938	1.929	1.862	2.076
5	1.33	0.9368	0.8578	1.019	0.9547
6	0.6124	0.2163	NA	0.3369	0.3013
7	2.017	1.767	1.757	1.789	1.834
8	4.6	0.7708	0.8166	0.6557	0.6397
9	2.605	2.531	2.564	2.866	2.596
10	1.528	1.147	1.036	1.162	1.153
Mean	2.364	1.734	1.88	2.057	1.739

So, some combination of 1b and 1c tend to help the performance of our rare binary method the most. I also encountered a little bit of problem with dataset 6 when trying to fit all the parameters in the MCMC. When I didn't fix α and ρ, as the MCMC was getting started, eventually the random effects would stop moving in the MCMC. I don't know why they stopped.