Research Update - Simulation Study

Sam Morris

7 May 2015

Numerical stability in the simulation study

Part of the instability from the simulation study arises from the restrictions on the model parameters. For example, in the GEV link we need to ensure that $\xi \mathbf{X}\beta < 1$. So, the bounds on the parameter space for ξ depend on the current iteration of $\mathbf{X}\beta$ and vice versa. We also encountered moves in α that made the likelihood evaluate as `NaN`. Part of the problem was in how I had coded up the likelihood function. The original function was `lly = (1 - y) * log[P(Y = 0)] + y * log[P(Y = 1)]` which caused problems when $y = 0$ and $P(Y = 1) \approx 0$ due to the fact that $0 * -\infty$ is not a number. This has been addressed by setting `lly = log[P(Y = 0)]` when $y = 0$, and `lly = log[P(Y = 1)]` when $y = 1$.

Alternatives to estimating α and ρ in the MCMC

Given that it is challenging to estimate α and ρ (the bandwidth for the kernel) for the max-stable process when we actually observe the latent variable z, estimating these terms when z is not observed is even more challenging. One idea is to try to estimate them from the data using some numerical technique, and then running the MCMC using the empirically estimated value. The rest of this document is meant to explore how empirical estimates of $\vartheta(h)$ change with α and ρ.

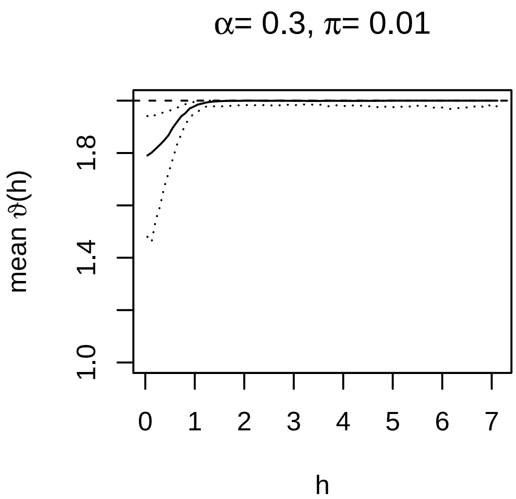
Exploring $\vartheta(h)$

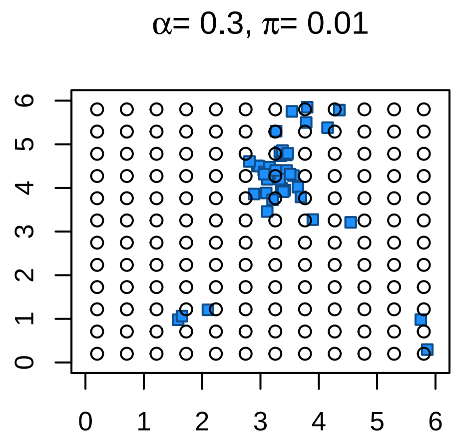
To get a sense of how $\vartheta(h)$ changes with different simulated data settings, I tried looking at two different levels of rareness, $\pi = 0.01, 0.05$, for $\alpha = 0.3, 0.7$ and $\rho = 0.5, 3$. In the following sections, we generate 200 datasets with 4000 observations and obtain an empirical estimate for set i of $\chi(h)_i = P[Y(\mathbf{s}_1 = 1 | Y(\mathbf{s}_2 = 1)$ for each dataset. This is related to the $\chi(h)$ statistic from the extreme value literature in the sense that $\lim_{u \to \infty} \chi_u(h) = \chi(h)$ where u is the threshold for z at which $y = 1$. The maximum distance is $h = 8.44$, so we split the interval (0, 8.44) into 99 evenly sized bins. Using the relationship for bivariate max-stable distributions, $\vartheta(h)_i = 2 - \chi(h)_i$, we can then estimate the extremal coefficient for each set. The plots show the median$[\vartheta(h)]$ for the 200 datasets using a solid line, and $q(0.025)$ and $q(0.975)$ using the dotted lines. Next to each $\vartheta(h)$ plot, we also give a sample simulated dataset.

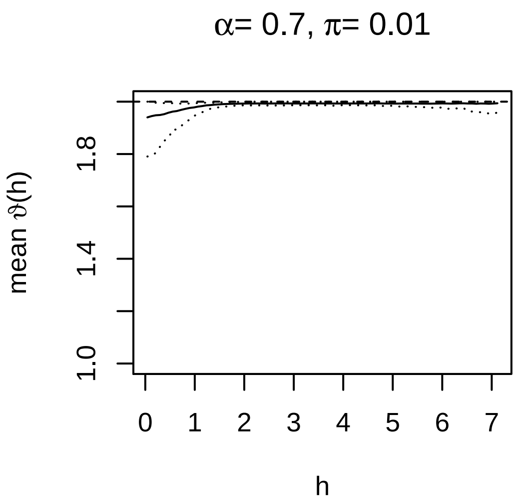
Looking at the following plots, the proportion over the threshold plays a role in the limiting $\vartheta(h)$ as $h \to \infty$. It would also appear that as ρ becomes large relative to knot spacing, that we see $\vartheta(h)$ demonstrate behavior more close to independence. This seems reasonable because as ρ increases, each site will be impacted by more random effects than in the case of small bandwidths. I think the role of ρ is to essentially determine the distance at which $\vartheta(h)$ plateaus. Finally, as usual α controls the strength of the dependence.

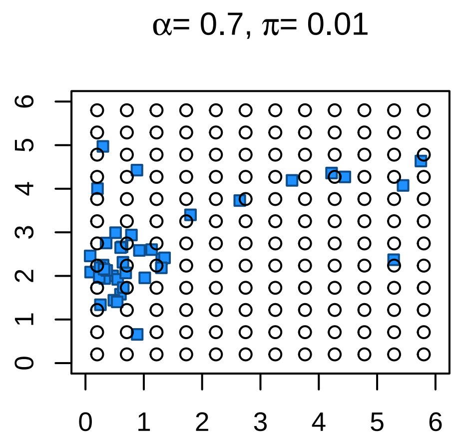
Varying α

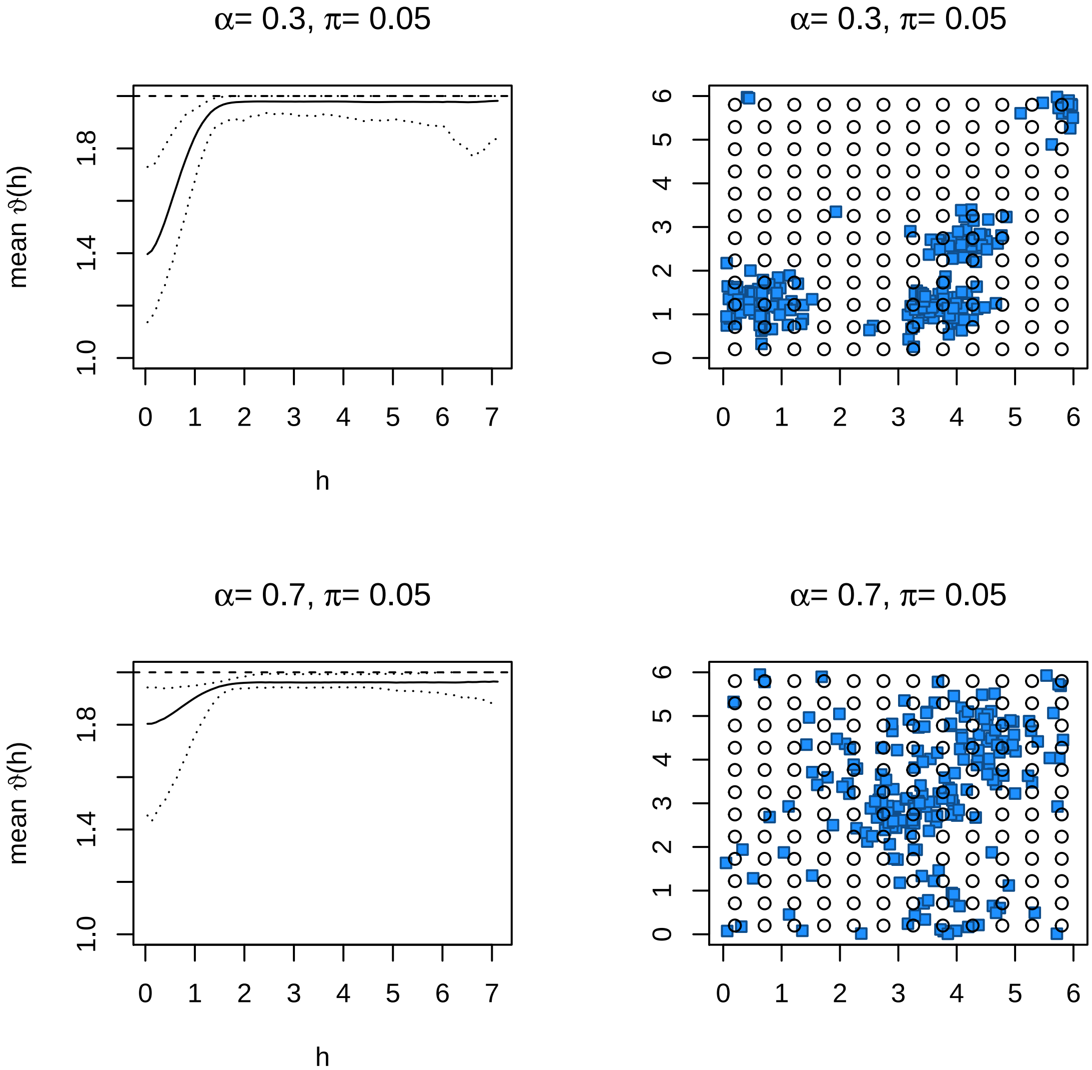
The following plots show median[$\vartheta(h)$] for $\alpha = 0.3, 0.7$ with rareness of $\pi = 0.01, 0.05$ when the true value of ρ is 0.55. I selected ρ to be close to the actual horizontal/vertical knot spacing which is 0.51.







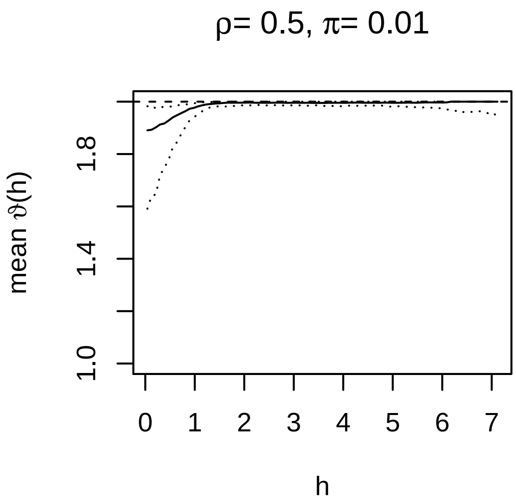


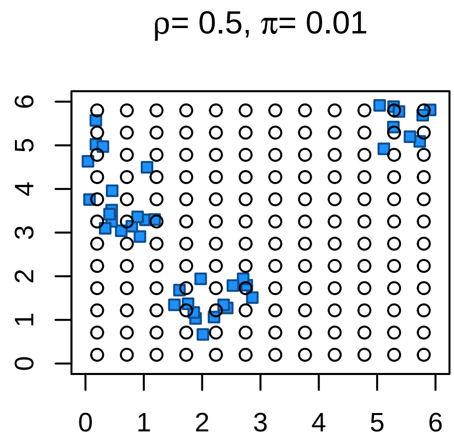


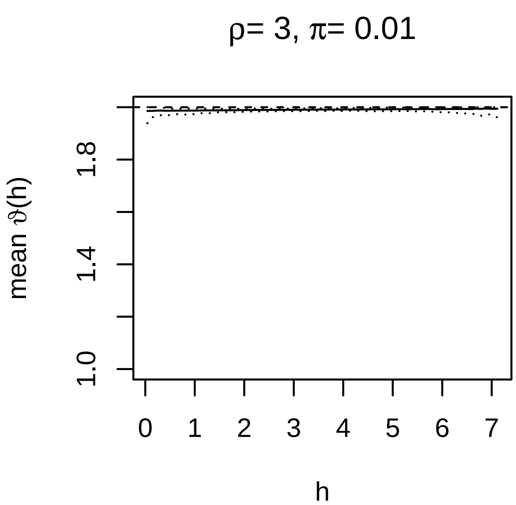
α= 0.3, π= 0.05
mean ϑ(h)
h
α= 0.3, π= 0.05
α= 0.7, π= 0.05
mean ϑ(h)
h
α= 0.7, π= 0.05

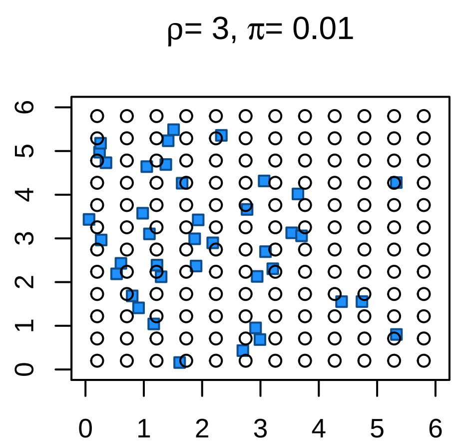
Varying ρ

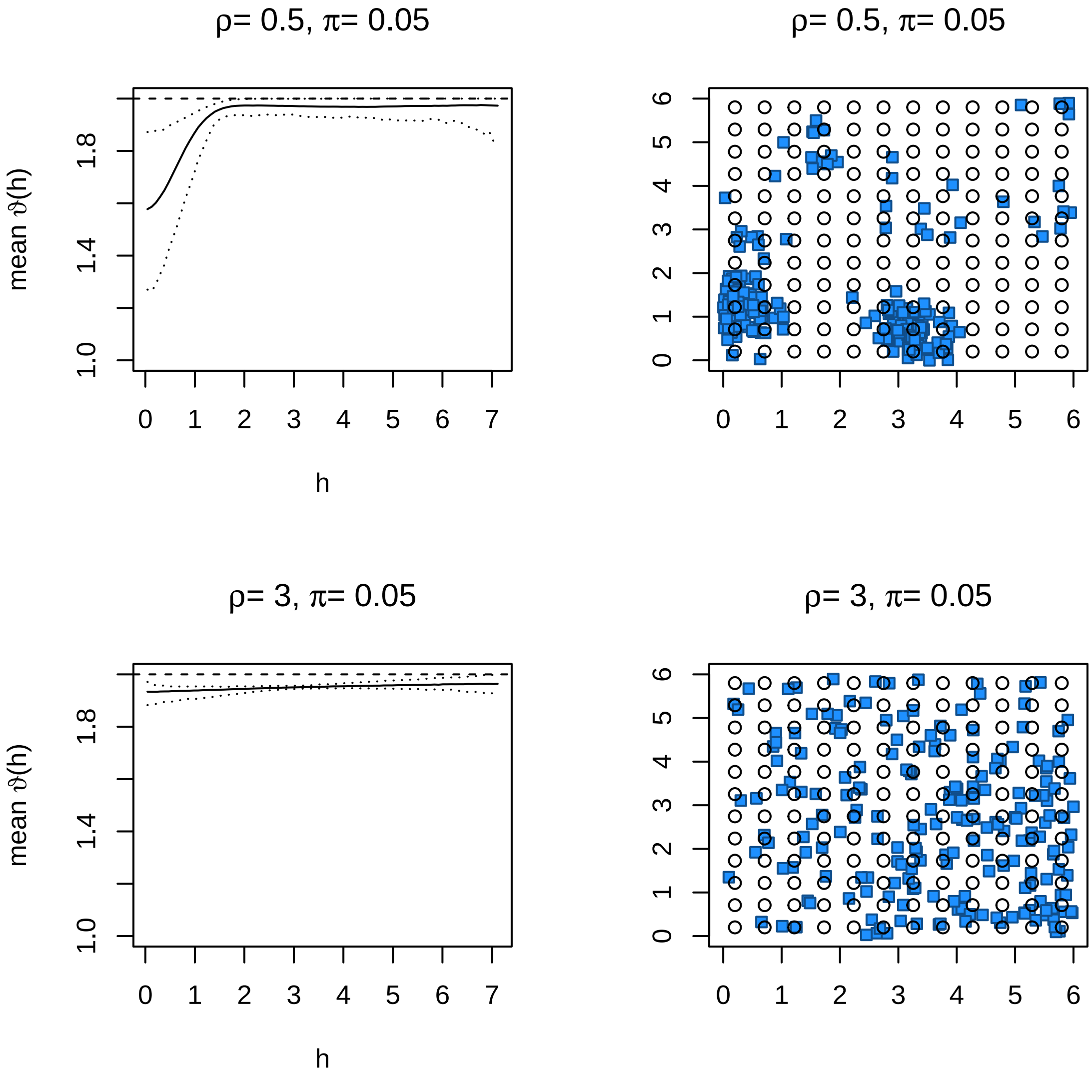
The following plots show median$[\vartheta(h)]$ for $\rho = 0.5, 3$ with rareness of $\pi = 0.01, 0.05$ when the true value of α is 0.5.










ρ= 0.5, π= 0.05
mean ϑ(h)
h
ρ= 0.5, π= 0.05
ρ= 3, π= 0.05
mean ϑ(h)
h
ρ= 3, π= 0.05