Generalised-Pareto distribution with Predictive Processes – MCMC updates

Andrew C. Parnell

April 12, 2018

Algorithm details

(Example below – to be filled in):

Algorithm 1: BART Markov chain Monte Carlo

Data: Target variables y (length n; standardised), feature matrix X (n rows and p columns

Result: Posterior list of trees T, values of τ, fitted values $\hat{y}$

Initialisation;

Hyper-parameter values of α, β, τ_μ, ν, λ;

Number of trees M;

Number of iterations N;

Initial value $\tau = 1$;

Set trees T_j; $j = 1, \ldots, M$ to stumps;

Set values of μ to 0;

for *iterations i from 1 to N* **do**

  for *trees j from 1 to M* **do**

    Compute partial residuals from y minus predictions of all trees except tree j;

    Grow a new tree T_j^{new} based on grow/prune/change/swap;

    Set l_{new} = log full conditional of new tree T_j^{new} based on Equation 1 plus Equation 4;

    Set l_{old} = log full conditional of old tree T_j based on same equations;

    Set $a = \exp(l_{new} - l_{old})$;

    Generate $u \sim U(0, 1)$;

    if $a > u$ **then**

      Set $T_j = T_j^{new}$;

    end

    Simulate μ values using Equation 3;

  end

  Get predictions $\hat{y}$ from all trees;

  Update τ using Equation 4;

end

1 DAG

To be filled in

2 Notation

- x_i Multivariate (bivariate or tri-variate) location values for location i, $i = 1, \ldots, n$. Write the matrix of all locations as just x
- y_{ij} Excess j for observation i, $j = 1, \ldots, n_i$ where n_i is the number of excesses at location i
- $\sigma(x_i)$ Scale parameter for location x_i
- $\xi(x_i)$ Shape parameter for observation x_i
- z_k Sub-grid locations $k = 1, \ldots, m$. Together written as z
- A, B Projection matrices of dimension $n \times m$
- $\tilde{\sigma}, \tilde{\xi}$ Gaussian processes defined on sub-grid z
- $\mu(z), \kappa(z)$ Means for the Gaussian processes
- Σ, Ψ Auto-covariance matrices for the Gaussian processes
- $\omega_\sigma, \omega_\xi$ Nugget standard deviation parameters
- $\alpha_\sigma, \alpha_\xi, \beta_\sigma, \beta_\xi$ intercept and slope parameters for Gaussian process means
- ν_σ, ν_ξ Matern smoothness parameters (perhaps fixed?)
- ρ_σ, ρ_ξ Matern scale parameters given discrete uniform prior for simplicity
- τ_σ, τ_ρ Variance parameters for Gaussian process

3 Model outline

In hierarchical notation:

$$
\begin{aligned}
y_{ij} &\sim GPD(\sigma(x_i), \xi(x_i)) \\
\log(\sigma(x)) &= A(x,z)^T \Sigma^{-1}(z,z)\tilde{\sigma}(z) \\
\log(\xi(x)) &= B(x,z)^T \Psi^{-1}(z,z)\tilde{\xi}(z) \\
A(x_i,z_k) &= \tau_\sigma^2 \frac{2^{1-\nu_\sigma}}{\Gamma(\nu_\sigma)} \left(\sqrt{2\nu_\sigma} \frac{\|x_i - z_k\|}{\rho_\sigma} \right)^{\nu_\sigma} K_{\nu_\sigma} \left(\sqrt{2\nu_\sigma} \frac{\|x_i - z_k\|}{\rho_\sigma} \right) \\
B(x_i,z_k) &= \tau_\xi^2 \frac{2^{1-\nu_\xi}}{\Gamma(\nu_\xi)} \left(\sqrt{2\nu_\xi} \frac{\|x_i - z_k\|}{\rho_\xi} \right)^{\nu_\xi} K_{\nu_\xi} \left(\sqrt{2\nu_\xi} \frac{\|x_i - z_k\|}{\rho_\xi} \right) \\
\tilde{\sigma} &\sim MVN_m(\mu(z), \omega_\sigma^2 I_m + \Sigma(z,z)) \\
\tilde{\xi} &\sim MVN_m(\kappa(z), \omega_\xi^2 I_m + \Psi(z,z)) \\
\mu(z) &= \alpha_\sigma + \beta_\sigma^T z \\
\kappa(z) &= \alpha_\xi + \beta_\xi^T z \\
\Sigma(z_k,z_l) &= \tau_\sigma^2 \frac{2^{1-\nu_\sigma}}{\Gamma(\nu_\sigma)} \left(\sqrt{2\nu_\sigma} \frac{\|z_k - z_l\|}{\rho_\sigma} \right)^{\nu_\sigma} K_{\nu_\sigma} \left(\sqrt{2\nu_\sigma} \frac{\|z_k - z_l\|}{\rho_\sigma} \right) \\
\Psi(z_k,z_l) &= \tau_\xi^2 \frac{2^{1-\nu_\xi}}{\Gamma(\nu_\xi)} \left(\sqrt{2\nu_\xi} \frac{\|z_k - z_l\|}{\rho_\xi} \right)^{\nu_\xi} K_{\nu_\xi} \left(\sqrt{2\nu_\xi} \frac{\|z_k - z_l\|}{\rho_\xi} \right)
\end{aligned}
$$

Hyper-parameter prior distributions (subject to change):

$$
\begin{aligned}
\tau_\sigma, \tau_\xi, \omega_\sigma, \omega_\xi &\sim U(0,100) \\
\rho_\sigma, \rho_\xi &\sim DU(0.001, 0.01, 0.1, 1, 10, 100, 1000) \\
\alpha_\sigma, \alpha_\xi, \beta_\sigma, \beta_\xi &\sim N(0,100)
\end{aligned}
$$