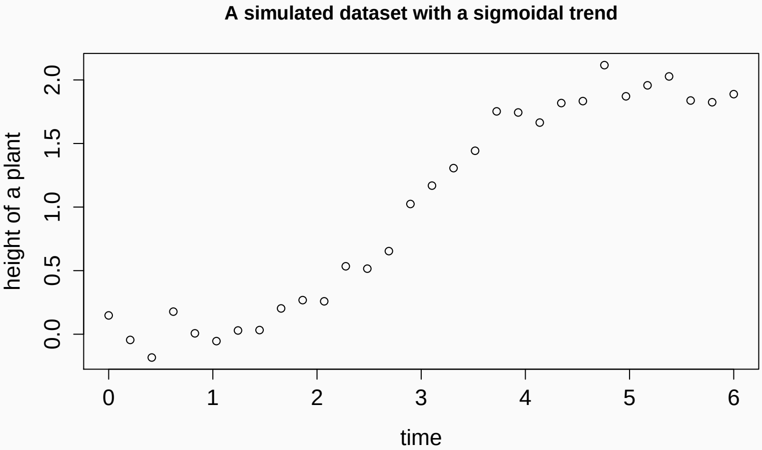
A flexible sequential Monte Carlo algorithm for shape-constrained regression

Kenyon Ng, with Kevin Murray, Berwin A. Turlach

4 September 2018

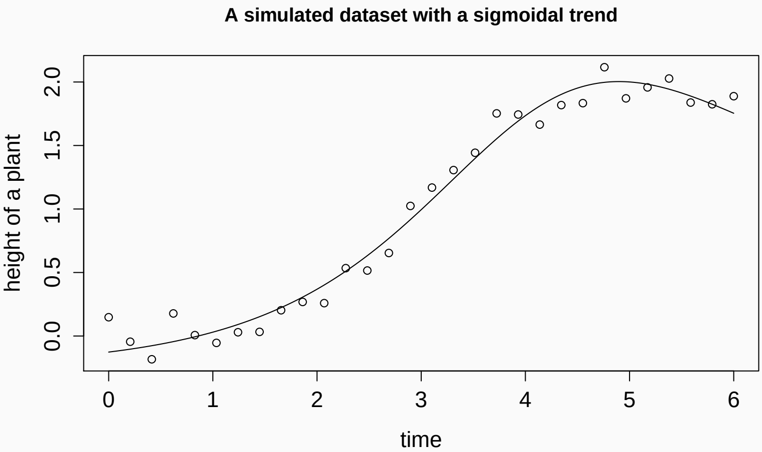
Department of Mathematics and Statistics,
The University of Western Australia.

A rational function model



$$\text{Model: } r(x;\boldsymbol{\beta}) = \frac{\beta_1 + \beta_2 x}{1 + \beta_3 x + \beta_4 x^2}$$

Least squares



Minimise residual sum of squares $\text{RSS}(\beta) = \sum\{y_i - r(x_i; \beta)\}^2$

Issues

In general,

- regression is equivalent to a minimisation problem;
- shape-constrained regression is equivalent to a constrained minimisation problem;
- example: find a least squares $r(x; \boldsymbol{\beta})$ and monotonic increasing when $x \in [0, 6]$, same as minimise $\text{RSS}(\boldsymbol{\beta})$ subject to $\frac{\partial}{\partial x} r(x; \boldsymbol{\beta}) \geq 0$ for all $x \in [0, 6]$;
- infinite inequality constraints to satisfy!

Goals

- A curve fitting algorithm that works with:
  - any loss function $\ell(\boldsymbol{\beta})$, e.g. RSS, Tukey's biweight;
  - a majority of shape constraints.
- Construct $\mathbb{1}_{\mathcal{S}}(\boldsymbol{\beta})$ where $\mathcal{S} := \{\boldsymbol{\beta} | \boldsymbol{\beta} \text{ that satisfies shape constraints}\}$.
- Minimise $\ell(\boldsymbol{\beta})$ subject to $\boldsymbol{\beta} \in \mathcal{S}$.

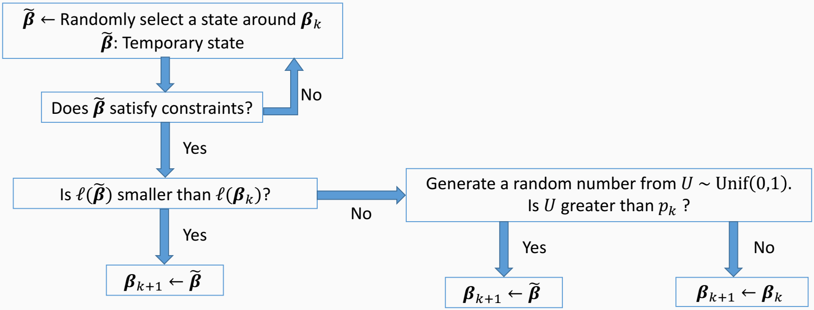
A MCMC approach to minimisation

- Construct a Boltzmann distribution $\pi(\boldsymbol{\beta}) \propto \exp\left\{-\frac{\ell(\boldsymbol{\beta})}{T}\right\}\mathbb{1}_{\mathcal{S}}(\boldsymbol{\beta})$.
- When $T \to 0$, $\pi(\boldsymbol{\beta})$ converges to a degenerate distribution with point mass at $\boldsymbol{\beta}^*$, the global minimum.[1]
- If we can sample from $\pi(\boldsymbol{\beta})$, we are done.
- Use Markov chain Monte Carlo to sample from $\pi(\boldsymbol{\beta})$.

[1] H. E. Romeijn and R. L. Smith. "Simulated annealing for constrained global optimization". In: *Journal of Global Optimization* 5.2 (1994), pp. 101–126.

Metropolis-Hastings

At k^{th} iteration of the Markov chain



Metropolis-Hastings

- The *acceptance probability* p_k is given by

$$p_k = \min\left\{1, \exp\left\{-\frac{\ell(\tilde{\boldsymbol{\beta}}) - \ell(\boldsymbol{\beta}_k)}{T}\right\}\right\}$$

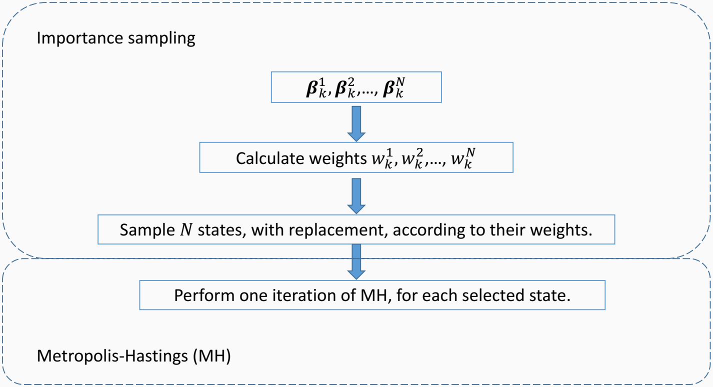
- When the *temperature* T is low, p_k is always 1 or almost 0.
- Especially true in the beginning, as $\ell(\tilde{\boldsymbol{\beta}}) - \ell(\boldsymbol{\beta}_k)$ is usually large.
- Likely to converge to a local minimum.
- In practice:
  - start T with a high value and decrease after each iteration;
  - simulate multiple chains to speed up the process.

Sequential Monte Carlo

- Effectively sampling from a sequence $\pi_k(\boldsymbol{\beta}) \propto \exp\left\{-\frac{\ell(\boldsymbol{\beta})}{T_k}\right\}\mathbb{1}_{\mathcal{S}}$.
- Sampling from $\pi_0(\boldsymbol{\beta})$ is easy, $\pi_\infty(\boldsymbol{\beta})$ is difficult.
- Sequential Monte Carlo sampler to the rescue.
- Designed to sample this type of distribution sequence.

Sequential Monte Carlo

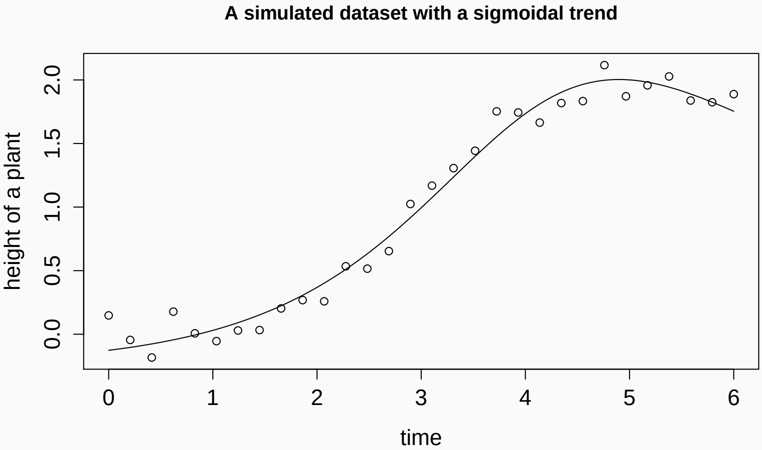
At k^{th} iteration



A note on the temperature

- The only difference between $\pi_k(\beta)$ and $\pi_{k+1}(\beta)$ is temperature.
- $\pi_k(\beta)$ and $\pi_{k+1}(\beta)$ are too close,
  - computational time wasted.
- $\pi_k(\beta)$ and $\pi_{k+1}(\beta)$ are too far,
  - will not converge.
- Convergence conditions:
  - $\left|\frac{1}{T_k} - \frac{1}{T_{k+1}}\right|$ is monotonically decreasing;
  - $T_k \to 0$.

A monotonic rational function model

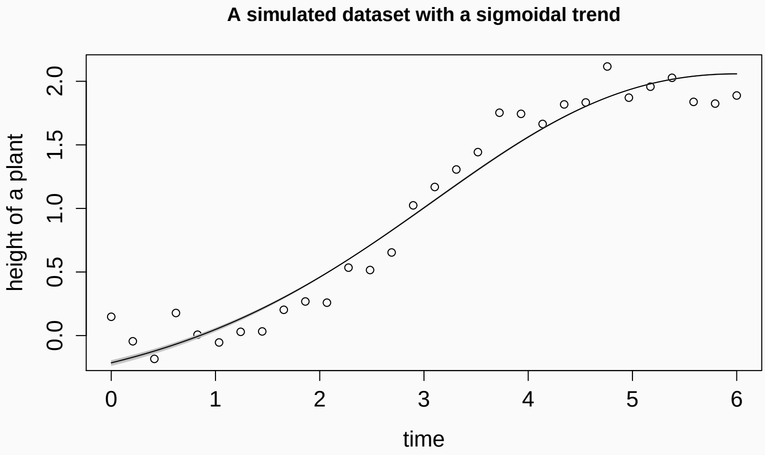


Model: $r(x; \boldsymbol{\beta}) = \dfrac{\beta_1 + \beta_2 x}{1 + \beta_3 x + \beta_4 x^2}$ Shape: Monotonic increasing

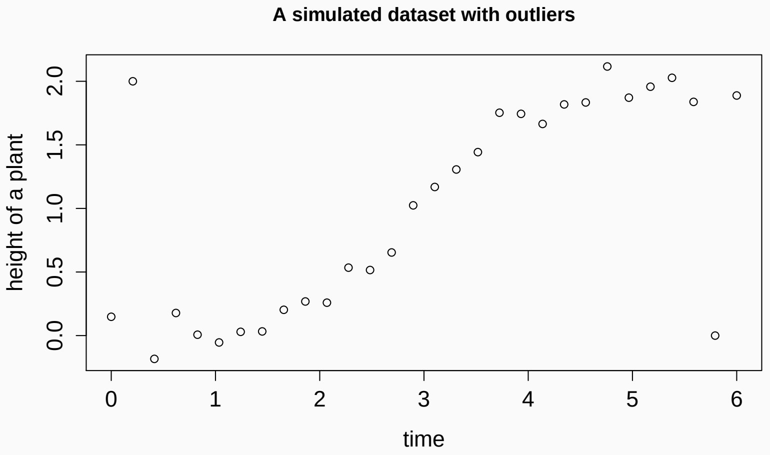
Algorithm specs

- Loss function $\ell(\beta) = \sum\{y_i - r(x_i; \beta)\}^2$
- Cooling schedule $T_k = \dfrac{|\ell(\beta_k^*)|}{1 + 0.5(k-1)^2}$
- Run 100 times.

Monotonic least squares curves



A dataset with outliers



Robust regression

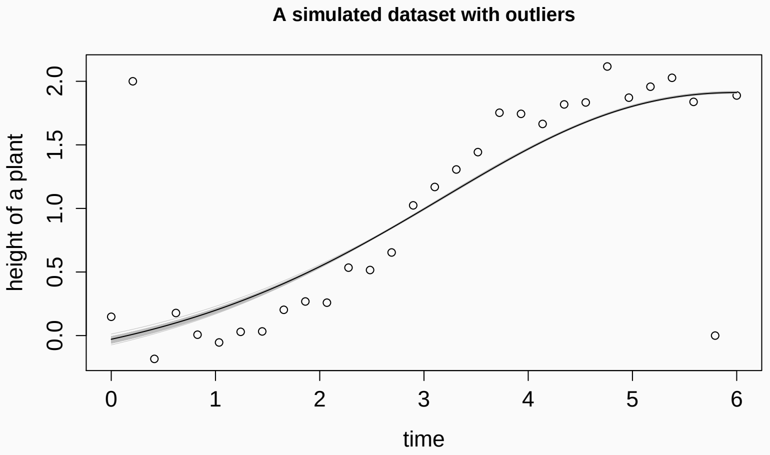
In the presence of outliers, we minimise Tukey's biweight function:

$$\ell(\boldsymbol{\beta}) = \sum \rho\left(y_i - r(x_i; \boldsymbol{\beta})\right)$$

where

$$\rho(u) = \begin{cases} \dfrac{4.685^2}{6}\left(1 - \left(1 - \left(\dfrac{u}{4.685}\right)^2\right)^3\right) & |u| \leq 4.685\,, \\ \dfrac{4.685^2}{6} & |u| > 4.685\,. \end{cases}$$

Tukey's biweight



Conclusion and future work

A generic algorithm for constrained optimisation.

Future works:

- Explore different sampling techniques (e.g. parallel tempering, Hamiltonian Monte Carlo).

Thank you for your attention!