A flexible sequential Monte Carlo algorithm for shape-constrained regression

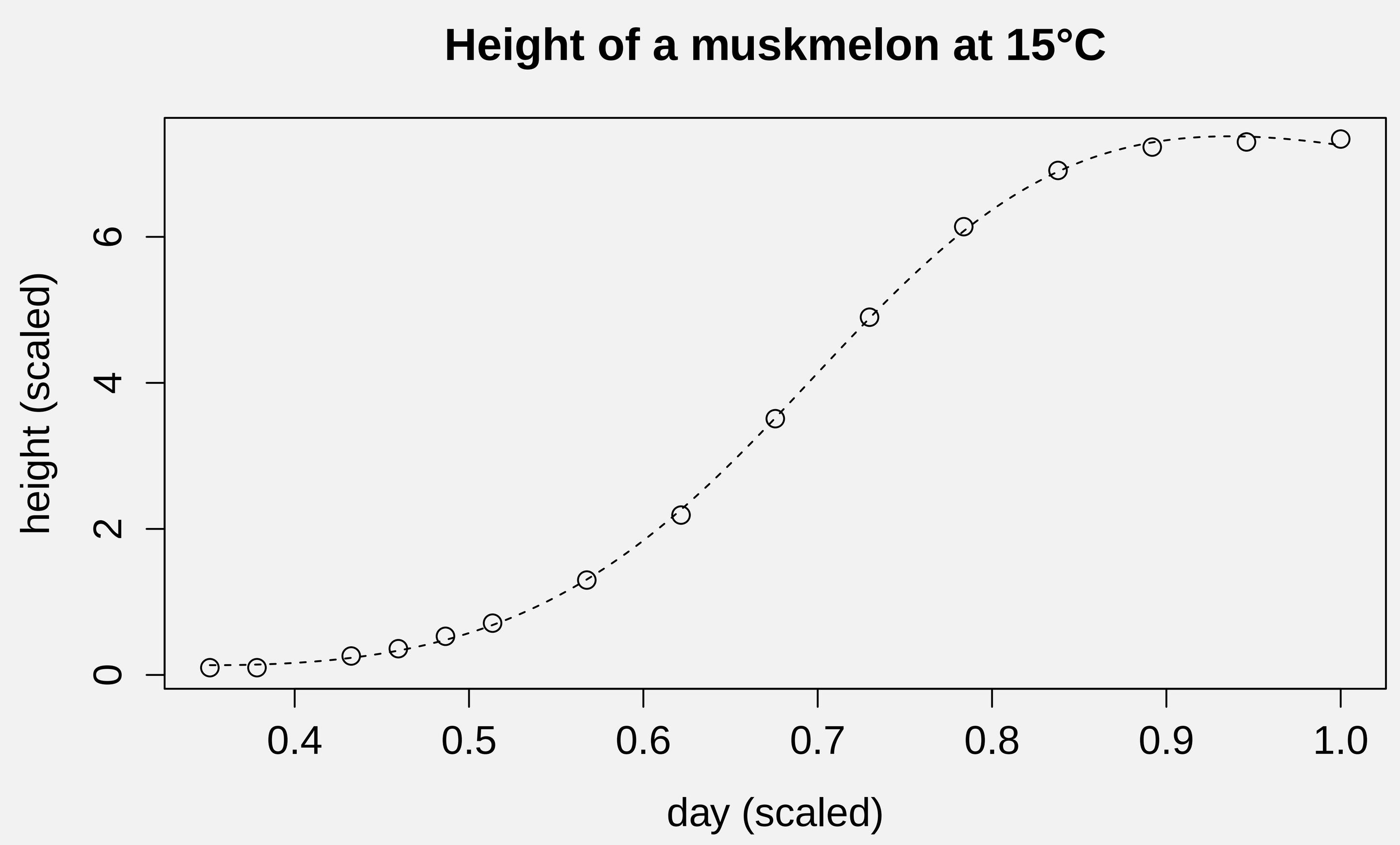
Kenyon Ng[1], Kevin Murray[2], Berwin A. Turlach[1]

[1] Department of Mathematics and Statistics, The University of Western Australia.

[2] School of Population and Global Health, The University of Western Australia.

THE UNIVERSITY OF WESTERN AUSTRALIA

Motivation



- This dataset records the height of a muskmelon (y) over a period of time (x).
- We fit a rational function model

$$r(x;\beta) = \frac{\beta_1 + \beta_2 x + \beta_3 x^2}{1 + \beta_4 x + \beta_5 x^2}.$$

- The β is obtained by minimising a loss function

$$\ell(\beta) = \sum \{y_i - r(x_i;\beta)\}^2.$$

- Muskmelon do not usually grow shorter over time! Should not be decreasing when $x > 0.9$.
- $r(x;\beta)$ should be monotonically increasing when $x \in [0.35, 1]$.

A constrained minimisation problem

- The β needs to satisfy:
  - $\frac{\partial}{\partial x} r(x;\beta) \geq 0$ for all $x \in [0.35, 1]$;
  - $r(x;\beta) < \infty$ for all $x \in [0.35, 1]$.
- This is challenging! There are infinite number of inequality constraints.
- Fortunately, we can build an indicator function

$$\mathbb{1}(\beta) = \begin{cases} 1 & \text{if } \beta \text{ satisfies the shape constraints;} \\ 0 & \text{otherwise.} \end{cases}$$

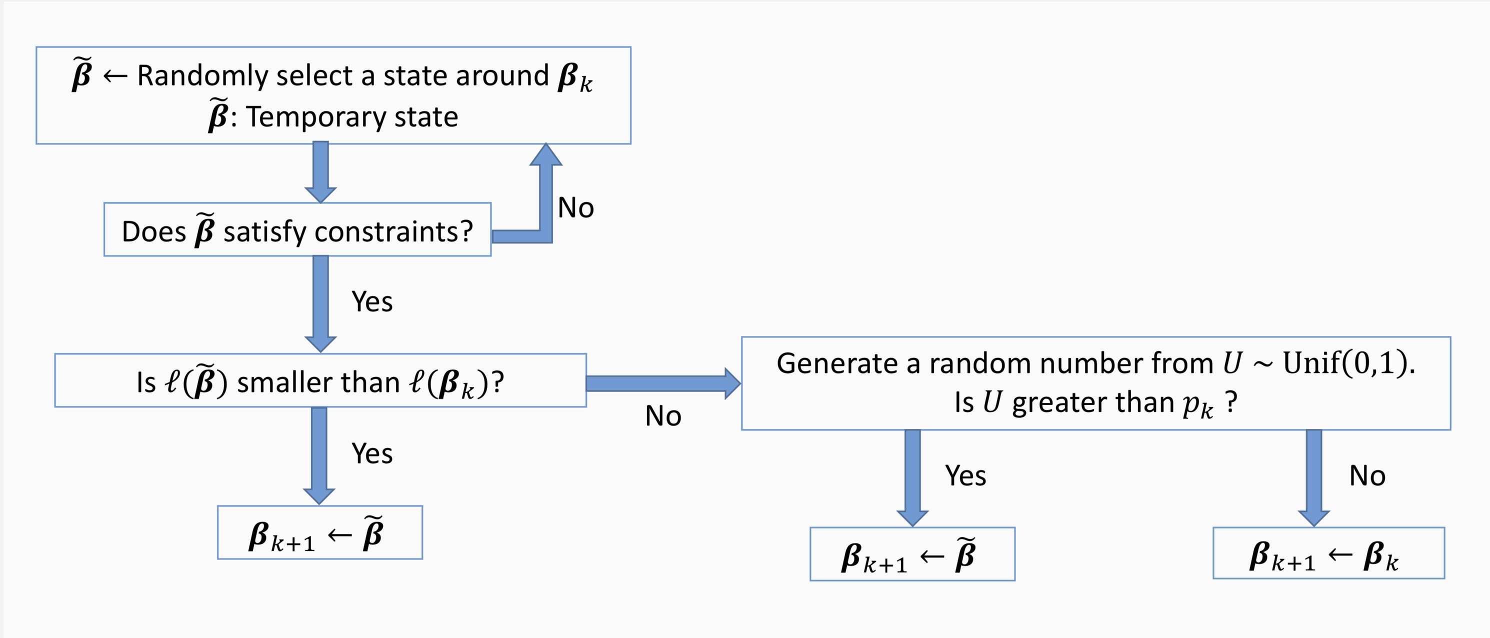
- Now we have a constrained minimisation problem. Find the global minimiser β^* of $\ell(\beta)$ subject to $\mathbb{1}(\beta^*) = 1$.
- [1] shows that

$$\pi(\beta) \propto \exp\left\{-\frac{\ell(\beta)}{T}\right\}\mathbb{1}(\beta)$$

will converge to a Dirac delta function $\delta(\beta - \beta^*)$ when $T \to 0$.
- If we can sample from $\pi(\beta)$ with small T, we will get β^*.
- We have converted a regression problem to a sampling problem.

Metropolis-Hastings algorithm [2, 3]

- Use Markov chain Monte Carlo (MCMC) to sample from $\pi(\beta)$.
- Idea: simulate a Markov chain with an invariant distribution of $\pi(\beta)$ for a long enough period.
- First attempt: Metropolis-Hastings (a type of MCMC) to simulate this Markov chain. The flow chart at k^{th} iteration is illustrated:



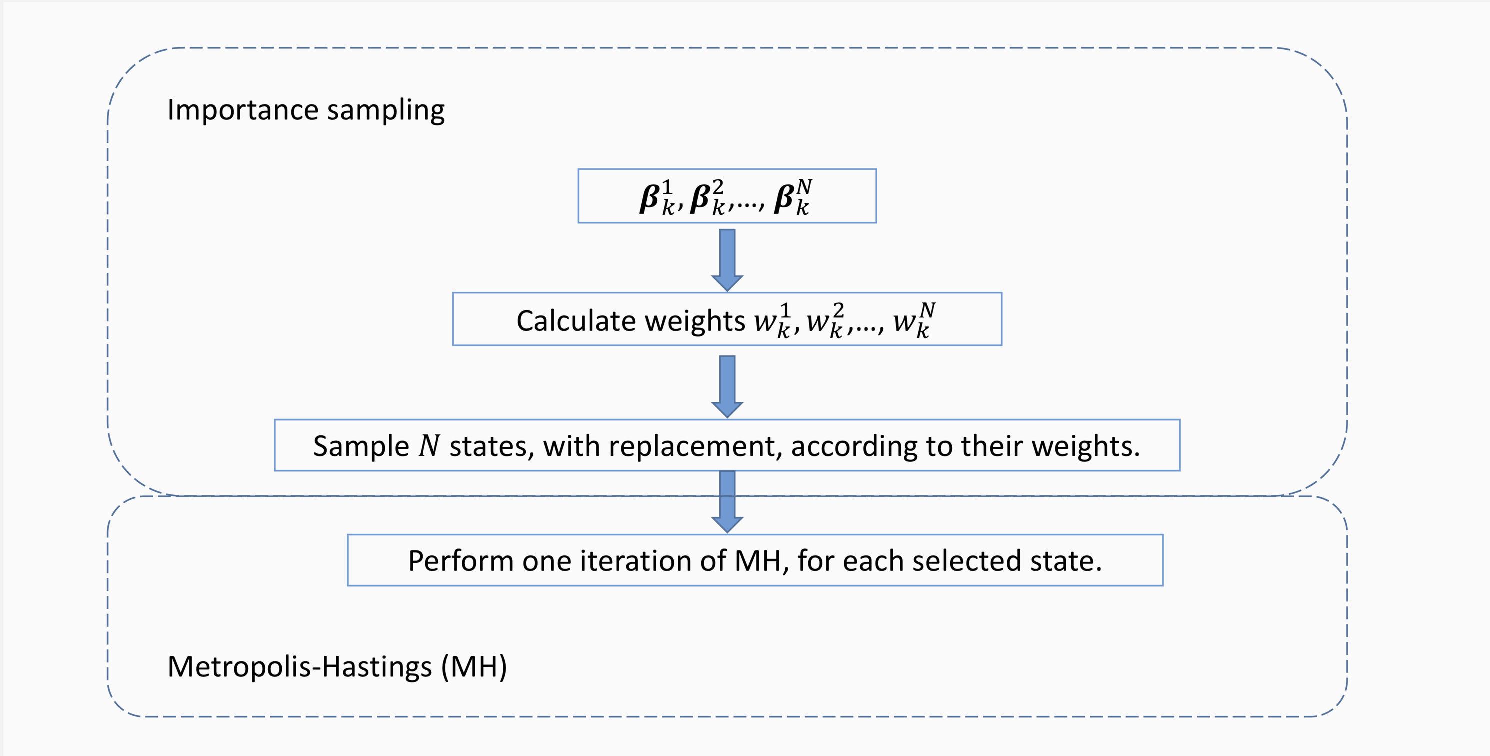
- Issue: p_k is small when T is small. The chain will barely move (i.e. $\beta_{k+1} \leftarrow \beta_k$ always happens).
- We can start with a large T and decrease it after each iteration.

Sequential Monte Carlo sampler [4]

- Decreasing T after each iteration $\implies$ sampling from a sequence of distributions $\pi_k(\beta)$

$$\pi_k(\beta) \propto \exp\left\{-\frac{\ell(\beta)}{T_k}\right\}\mathbb{1}(\beta)$$

- Second attempt: Sequential Monte Carlo (SMC) sampler, a more efficient way of running multiple Metropolis-Hastings algorithms.
- Starting from $\{\beta_0^i\}_{i=1}^N$, the flow chart at k^{th} iteration is illustrated:



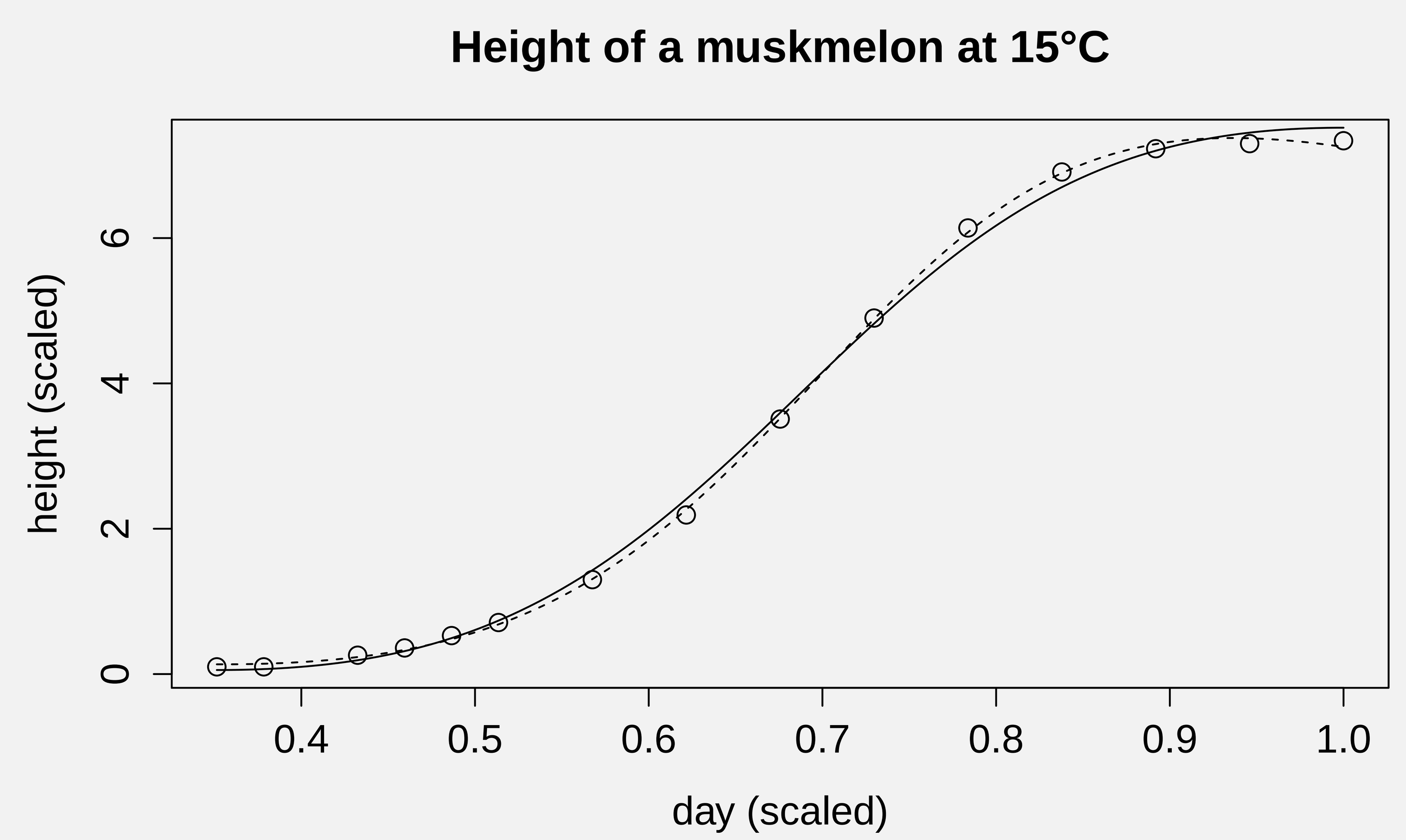
- SMC sampler can be seen as a two step process: importance sampling and a Metropolis-Hasting minimisation step.
- The weight of β_k^i, w_k^i, is inversely related to $\ell(\beta_k^i)$.
- β_k^i with a lower $\ell(\beta_k^i)$ has a better chance of getting minimised.

A note on the temperature

- The only difference between $\pi_k(\beta)$ and $\pi_{k+1}(\beta)$ is T.
- If $\pi_k(\beta)$ and $\pi_{k+1}(\beta)$ are too close, computational resources are wasted.
- If $\pi_k(\beta)$ and $\pi_{k+1}(\beta)$ are too far, the algorithm will not converge to β^*.
- Convergence conditions for the SMC sampler [5]:
  - $\left|\frac{1}{T_k} - \frac{1}{T_{k+1}}\right|$ is monotonically decreasing;
  - $T_k \to 0$.
- In practice, these conditions can be relaxed.

Results

- Run SMC sampler 40 times.
- The temperature follows $T_k = \frac{|\ell(\beta_k^*)|}{1 + 0.95(k-1)^2}$, β_k^* is the best β observed up until k^{th} iteration.



- Dotted line: the unconstrained curve; solid black: the best constrained curve out of the 40 estimates.

References

[1] H. Edwin Romeijn and Robert L. Smith. Simulated annealing for constrained global optimization. *Journal of Global Optimization*, 5(2):101–126, Sep 1994.

[2] Nicholas Metropolis, Arianna W. Rosenbluth, Marshall N. Rosenbluth, Augusta H. Teller, and Edward Teller. Equation of state calculations by fast computing machines. *The Journal of Chemical Physics*, 21(6):1087–1092, 1953.

[3] Marco Locatelli. Simulated annealing algorithms for continuous global optimization. In *Handbook of global optimization*, pages 179–229. Springer, 2002.

[4] Pierre Del Moral, Arnaud Doucet, and Ajay Jasra. Sequential Monte Carlo samplers. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 68(3):411–436, 2006.

[5] Enlu Zhou and Xi Chen. Sequential Monte Carlo simulated annealing. *Journal of Global Optimization*, 55(1):101–124, 2013.