Improving power in post-selection inference for changepoints by conditioning on less information

Rachel Carrington, Paul Fearnhead

Lancaster University

StatScale ECR Meeting
16 December 2022

Motivation

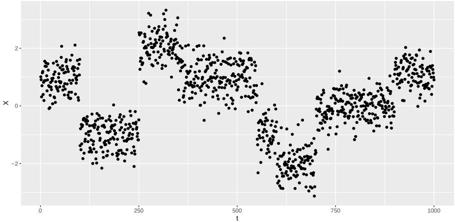
$$X_t = \mu_t + \epsilon_t \ (t = 1, ..., T), \qquad \epsilon_t \sim_{iid} N(0, \sigma^2)$$

where μ_t is piecewise constant, $\mu_t \neq \mu_{t+1}$ only at K changepoints $\{\tau_1, ..., \tau_K\}$.

Motivation

$$X_t = \mu_t + \epsilon_t \ (t = 1, ..., T), \qquad \epsilon_t \sim_{iid} N(0, \sigma^2)$$

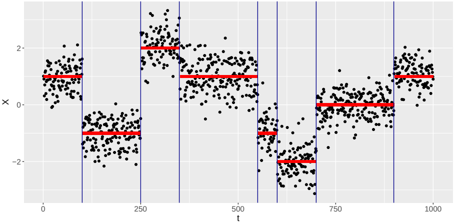
where μ_t is piecewise constant, $\mu_t \neq \mu_{t+1}$ only at K changepoints $\{\tau_1, ..., \tau_K\}$.



Motivation

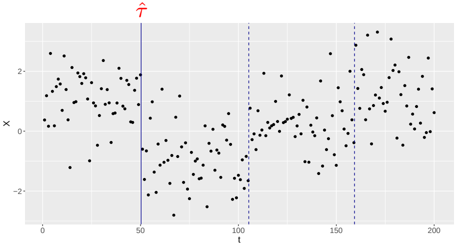
$$X_t = \mu_t + \epsilon_t \ (t = 1, ..., T), \qquad \epsilon_t \sim_{iid} N(0, \sigma^2)$$

where μ_t is piecewise constant, $\mu_t \neq \mu_{t+1}$ only at K changepoints $\{\tau_1, ..., \tau_K\}$.

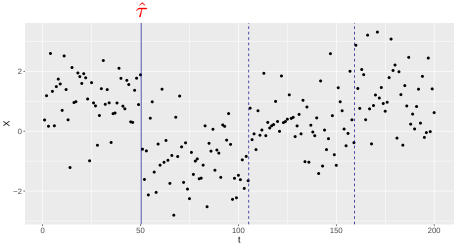


How confident can we be that the changepoints we detect are real?

Set up test



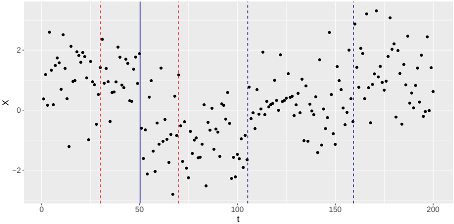
Set up test



H_0: No changepoint in $(\hat{\tau} - h + 1, \hat{\tau} + h)$.

H_1: At least one changepoint in this region

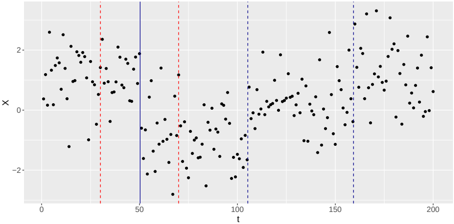
Set up test



H_0: No changepoint in $(\hat{\tau} - h + 1, \hat{\tau} + h)$.

H_1: At least one changepoint in this region

Set up test



H_0: No changepoint in $(\hat{\tau} - h + 1, \hat{\tau} + h)$.

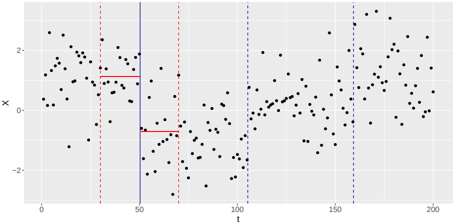
H_1: At least one changepoint in this region

Test statistic: $\phi = \boldsymbol{\nu}^T \boldsymbol{X}$, where

$$\nu_t = \begin{cases} 0 & \text{if } t \leq \hat{\tau} - h \text{ or } t \geq \hat{\tau} + h + 1 \\ \frac{1}{h} & \text{if } \hat{\tau} - h + 1 \leq t \leq \hat{\tau} \\ -\frac{1}{h} & \text{if } \hat{\tau} + 1 \leq t \leq \hat{\tau} + h. \end{cases}$$

Under H_0, $\phi \sim N(0, \frac{2}{h}\sigma^2)$.

Set up test



H_0: No changepoint in $(\hat{\tau}-h+1, \hat{\tau}+h)$.

H_1: At least one changepoint in this region

Test statistic: $\phi = \boldsymbol{\nu}^T \boldsymbol{X}$, where

$$\nu_t = \begin{cases} 0 & \text{if } t \leq \hat{\tau}-h \text{ or } t \geq \hat{\tau}+h+1 \\ \frac{1}{h} & \text{if } \hat{\tau}-h+1 \leq t \leq \hat{\tau} \\ -\frac{1}{h} & \text{if } \hat{\tau}+1 \leq t \leq \hat{\tau}+h. \end{cases}$$

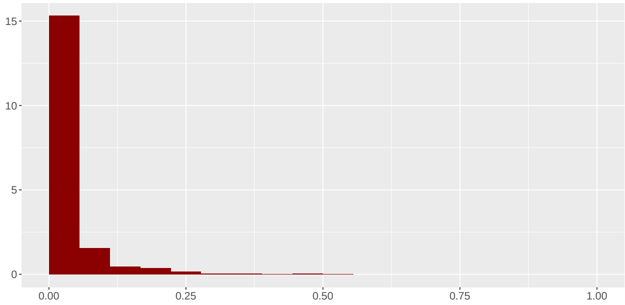
Under H_0, $\phi \sim N(0, \frac{2}{h}\sigma^2)$.

Naive p-value

$$p = \Pr\left(|\phi| \geq |\boldsymbol{\nu}^T \boldsymbol{X}_{obs}|\right)$$

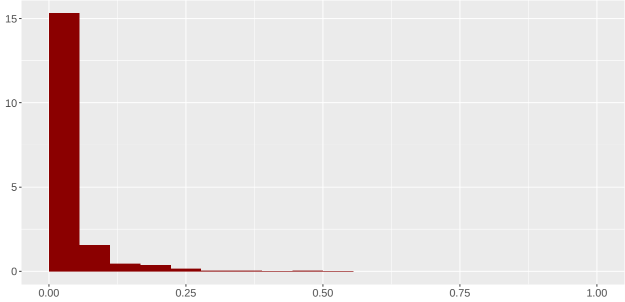
Naive p-value

$$p = \Pr\left(|\phi| \geq |\boldsymbol{\nu}^T \boldsymbol{X}_{obs}|\right)$$



Naive p-value

$$p = \Pr\left(|\phi| \geq |\boldsymbol{\nu}^T \boldsymbol{X}_{obs}|\right)$$



We need to condition on the information used to choose the test.

Selective inference

Selective p-value:

$$p = \Pr\left(|\phi| \geq |\boldsymbol{\nu}^T \boldsymbol{X}_{obs}| | \hat{\tau} \in \mathcal{M}(\boldsymbol{X})\right).$$

Selective inference

Selective p-value:

$$p = \Pr\left(|\phi| \geq |\boldsymbol{\nu}^T \boldsymbol{X}_{obs}| | \hat{\tau} \in \mathcal{M}(\boldsymbol{X})\right).$$

How to calculate this?

Selective inference

Selective p-value:

$$p = \Pr\left(|\phi| \geq |\boldsymbol{\nu}^T \boldsymbol{X}_{obs}| | \hat{\tau} \in \mathcal{M}(\boldsymbol{X})\right).$$

How to calculate this?

Condition on $\boldsymbol{\Pi}_\nu \boldsymbol{X} = \boldsymbol{\Pi}_\nu \boldsymbol{X}_{obs}$.

Reduces to 1-dimensional problem:

$$\boldsymbol{X}'(\phi) = \boldsymbol{X}_{obs} - \frac{1}{||\boldsymbol{\nu}||_2^2}\boldsymbol{\nu}\boldsymbol{\nu}^T \boldsymbol{X}_{obs} + \frac{1}{||\boldsymbol{\nu}||_2^2}\boldsymbol{\nu}\phi.$$

How to find S?

Let

$$S = \{\phi : \hat{\tau} \in \mathcal{M}(\boldsymbol{X}'(\phi))\}$$

Calculating this is algorithm dependent.

[1] Sean Jewell, Paul Fearnhead, and Daniela Witten. "Testing for a Change in Mean After Changepoint Detection". In: *To appear in Journal of the Royal Statistical Society, Series B (Statistical Methodology)* ().

How to find S?

Let

$$S = \{\phi : \hat{\tau} \in \mathcal{M}(\boldsymbol{X}'(\phi))\}$$

Calculating this is algorithm dependent.

(Jewell et. al)[1]: divide real line into intervals giving different output

[1] Sean Jewell, Paul Fearnhead, and Daniela Witten. "Testing for a Change in Mean After Changepoint Detection". In: *To appear in Journal of the Royal Statistical Society, Series B (Statistical Methodology)* ().

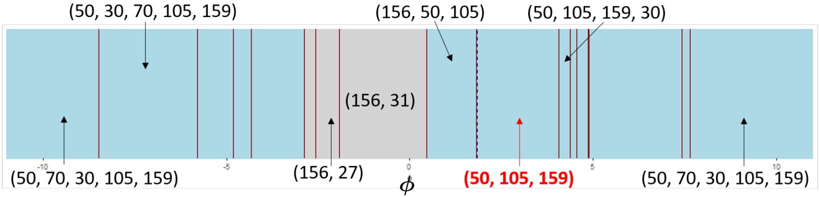
How to find S?

Let

$$S = \{\phi : \hat{\tau} \in \mathcal{M}(\boldsymbol{X}'(\phi))\}$$

Calculating this is algorithm dependent.

(Jewell et. al)[1]: divide real line into intervals giving different output



[1] Sean Jewell, Paul Fearnhead, and Daniela Witten. "Testing for a Change in Mean After Changepoint Detection". In: *To appear in Journal of the Royal Statistical Society, Series B (Statistical Methodology)* ().

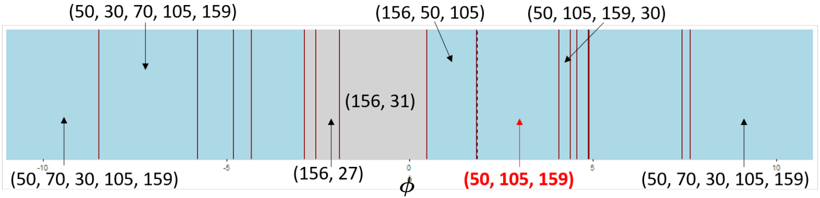
How to find S?

Let

$$S = \{\phi : \hat{\tau} \in \mathcal{M}(\boldsymbol{X}'(\phi))\}$$

Calculating this is algorithm dependent.

(Jewell et. al)[1]: divide real line into intervals giving different output



$$\hat{p} = \Pr\left(|\phi| \geq |\boldsymbol{\nu}^T \boldsymbol{X}_{obs}| \mid \phi \in S\right)$$

[1] Sean Jewell, Paul Fearnhead, and Daniela Witten. "Testing for a Change in Mean After Changepoint Detection". In: *To appear in Journal of the Royal Statistical Society, Series B (Statistical Methodology)* ().

Conditioning on less information

This gives valid p-values (which can be efficiently computed).

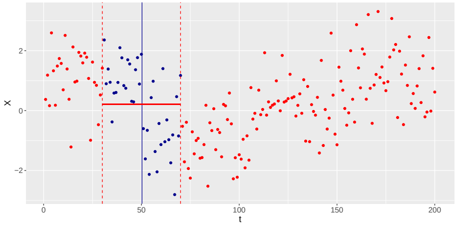
BUT we have to condition on a lot of extra information $\Rightarrow$ less power.

Can we do better?

Conditioning on less information

Condition on:

- The values of $\boldsymbol{X}$ outside of $\{X_{\hat{\tau}-h+1}, ..., X_{\hat{\tau}_h}\}$.
- The mean of $\{X_{\hat{\tau}-h+1}, ..., X_{\hat{\tau}+h}\}$.



Conditioning on less information

We can rewrite $\boldsymbol{X}$:

$$\boldsymbol{X} = \boldsymbol{Z}\boldsymbol{X} + \frac{1}{||\boldsymbol{\nu}||_2^2}\boldsymbol{\nu}\boldsymbol{\nu}^T\boldsymbol{X} + \left(\boldsymbol{B}\boldsymbol{B}^T + \frac{1}{||\boldsymbol{a}||_2^2}\boldsymbol{a}\boldsymbol{a}^T\right)\boldsymbol{X}$$

where $\boldsymbol{Z} = \boldsymbol{I} - \frac{1}{||\boldsymbol{\nu}||_2^2}\boldsymbol{\nu}\boldsymbol{\nu}^T - \frac{1}{||\boldsymbol{a}||_2^2}\boldsymbol{a}\boldsymbol{a}^T - \boldsymbol{B}\boldsymbol{B}^T = \boldsymbol{U}\boldsymbol{U}^T$.

Conditioning on less information

We can rewrite $\boldsymbol{X}$:

$$\boldsymbol{X} = \boldsymbol{Z}\boldsymbol{X} + \frac{1}{||\boldsymbol{\nu}||_2^2}\boldsymbol{\nu}\boldsymbol{\nu}^T\boldsymbol{X} + \left(\boldsymbol{B}\boldsymbol{B}^T + \frac{1}{||\boldsymbol{a}||_2^2}\boldsymbol{a}\boldsymbol{a}^T\right)\boldsymbol{X}$$

where $\boldsymbol{Z} = \boldsymbol{I} - \frac{1}{||\boldsymbol{\nu}||_2^2}\boldsymbol{\nu}\boldsymbol{\nu}^T - \frac{1}{||\boldsymbol{a}||_2^2}\boldsymbol{a}\boldsymbol{a}^T - \boldsymbol{B}\boldsymbol{B}^T = \boldsymbol{U}\boldsymbol{U}^T$.

Let $\phi = \boldsymbol{\nu}^T\boldsymbol{X}$ and $\boldsymbol{\psi} = \boldsymbol{U}^T\boldsymbol{X}$. Then:

Conditioning on less information

We can rewrite $\boldsymbol{X}$:

$$\boldsymbol{X} = \boldsymbol{Z}\boldsymbol{X} + \frac{1}{||\boldsymbol{\nu}||_2^2}\boldsymbol{\nu}\boldsymbol{\nu}^T\boldsymbol{X} + \left(\boldsymbol{B}\boldsymbol{B}^T + \frac{1}{||\boldsymbol{a}||_2^2}\boldsymbol{a}\boldsymbol{a}^T\right)\boldsymbol{X}$$

where $\boldsymbol{Z} = \boldsymbol{I} - \frac{1}{||\boldsymbol{\nu}||_2^2}\boldsymbol{\nu}\boldsymbol{\nu}^T - \frac{1}{||\boldsymbol{a}||_2^2}\boldsymbol{a}\boldsymbol{a}^T - \boldsymbol{B}\boldsymbol{B}^T = \boldsymbol{U}\boldsymbol{U}^T$.

Let $\phi = \boldsymbol{\nu}^T\boldsymbol{X}$ and $\boldsymbol{\psi} = \boldsymbol{U}^T\boldsymbol{X}$. Then:

$$\boldsymbol{X} = \boldsymbol{X}'(\phi, \boldsymbol{\psi}) = \boldsymbol{U}\boldsymbol{\psi} + \frac{1}{||\boldsymbol{\nu}||_2^2}\boldsymbol{\nu}\phi + \left(\frac{1}{||\boldsymbol{a}||_2^2}\boldsymbol{a}\boldsymbol{a}^T + \boldsymbol{B}\boldsymbol{B}^T\right)\boldsymbol{X}_{obs}$$

Monte Carlo estimation

P-value:

$$p = \Pr\left(|\phi| \geq |\boldsymbol{\nu}^T \boldsymbol{X}_{obs}| \mid \hat{\tau} \in \mathcal{M}(\boldsymbol{X}'(\phi, \boldsymbol{\psi}))\right)$$
$$= \frac{\Pr\left(|\phi| \geq |\boldsymbol{\nu}^T \boldsymbol{X}_{obs}| \cap \phi \in S_\psi\right)}{\Pr\left(\phi \in S_\psi\right)}$$

where

$$S_\psi = \{\phi : \hat{\tau} \in \mathcal{M}(\boldsymbol{X}'(\phi, \boldsymbol{\psi}))\}$$

Monte Carlo estimation

P-value:

$$p = \Pr\left(|\phi| \geq |\boldsymbol{\nu}^T \boldsymbol{X}_{obs}| \mid \hat{\tau} \in \mathcal{M}(\boldsymbol{X}'(\phi, \boldsymbol{\psi}))\right)$$
$$= \frac{\Pr\left(|\phi| \geq |\boldsymbol{\nu}^T \boldsymbol{X}_{obs}| \cap \phi \in S_\psi\right)}{\Pr\left(\phi \in S_\psi\right)}$$

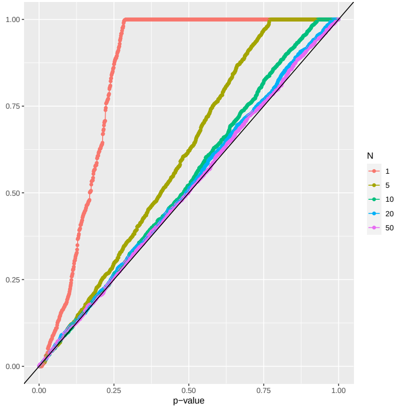
where

$$S_\psi = \{\phi : \hat{\tau} \in \mathcal{M}(\boldsymbol{X}'(\phi, \boldsymbol{\psi}))\}$$

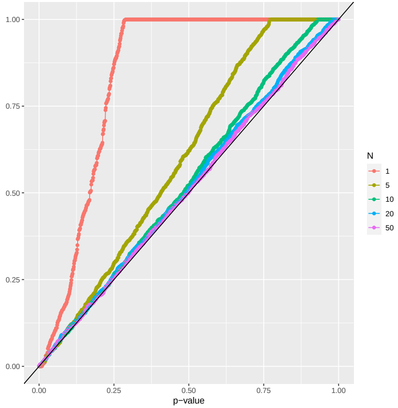
We can use Monte Carlo to estimate the p-value:

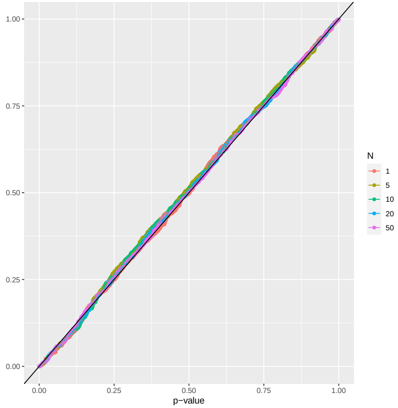
$$p = \frac{\Pr\left(|\phi| \geq |\boldsymbol{\nu}^T \boldsymbol{X}_{obs}| \cap \phi \in S_\psi\right)}{\Pr\left(\phi \in S_\psi\right)}$$
$$\approx \frac{\frac{1}{N}\sum_{j=1}^N \Pr\left(|\phi| \geq |\boldsymbol{\nu}^T \boldsymbol{X}_{obs}|, \phi \in S_{\psi^{(j)}} \mid \boldsymbol{\psi}^{(j)}\right)}{\frac{1}{N}\sum_{j=1}^N \Pr\left(\phi \in S_{\psi^{(j)}} \mid \boldsymbol{\psi}^{(j)}\right)}.$$

Monte Carlo estimation



Monte Carlo estimation





Monte Carlo estimation

Theorem

Let

$$\hat{p}_N = \frac{\sum_{j=1}^{N} \Pr\left(|\phi| \geq |\boldsymbol{\nu}^T \boldsymbol{X}_{obs}|, \phi \in S_{\psi^{(j)}} \mid \boldsymbol{\psi}^{(j)}\right)}{\sum_{j=1}^{N} \Pr\left(\phi \in S_{\psi^{(j)}} \mid \boldsymbol{\psi}^{(j)}\right)}.$$

Given that there is one $j^* \in \{1, ..., N\}$ such that $\boldsymbol{\psi}^{(j^*)}$ corresponds to the observed data under H_0, and the other $\boldsymbol{\psi}^{(j)}$ are drawn independently from their distribution under H_0, then under H_0, $\hat{p}_N \sim U(0, 1)$.

Proof

Let $\boldsymbol{\psi}^{(1:N)}$ denote the set of $\boldsymbol{\psi}^{(j)}$ values after shuffling, and I the label of $\boldsymbol{\psi}^{(j)}$ that corresponds to the observed data.

Proof

Let $\boldsymbol{\psi}^{(1:N)}$ denote the set of $\boldsymbol{\psi}^{(j)}$ values after shuffling, and I the label of $\boldsymbol{\psi}^{(j)}$ that corresponds to the observed data.

$$f(\phi, \boldsymbol{\psi}^{(1:N)}, I|S) \propto f(\phi) \left(\prod_{j=1}^{N} g(\boldsymbol{\psi}^{(j)}) \right) \mathbb{I}_{\{\phi \in S_{\boldsymbol{\psi}^{(I)}}\}}.$$

Proof

Let $\boldsymbol{\psi}^{(1:N)}$ denote the set of $\boldsymbol{\psi}^{(j)}$ values after shuffling, and I the label of $\boldsymbol{\psi}^{(j)}$ that corresponds to the observed data.

$$f(\phi, \boldsymbol{\psi}^{(1:N)}, I|S) \propto f(\phi) \left(\prod_{j=1}^{N} g(\boldsymbol{\psi}^{(j)}) \right) \mathbb{I}_{\{\phi \in S_{\boldsymbol{\psi}^{(I)}}\}}.$$

If we condition on $\boldsymbol{\psi}^{(1:N)}$, then we get

$$f(\phi, I|\boldsymbol{\psi}^{(1:N)}, S) = \frac{1}{W} f(\phi) \mathbb{I}_{\{\phi \in S_{\boldsymbol{\psi}^{(I)}}\}},$$

where

$$W = \sum_{j=1}^{N} \int_{\phi} f(\phi) \mathbb{I}_{\{\phi \in S_{\boldsymbol{\psi}^{(j)}}\}} d\phi = \sum_{j=1}^{N} \Pr(\phi \in S_{\boldsymbol{\psi}^{(j)}}).$$

Proof

We can now marginalise out I to get

$$\begin{aligned} f(\phi|\boldsymbol{\psi}^{(1:N)}, S) &= \sum_{I=1}^{N} f(\phi, I|\boldsymbol{\psi}^{(1:N)}, S) \\ &= \frac{1}{W} f(\phi) \sum_{I=1}^{N} \Pr(\phi \in S_{\psi^{(I)}}) f(\phi|\phi \in S_{\psi^{(I)}}). \end{aligned}$$

Proof

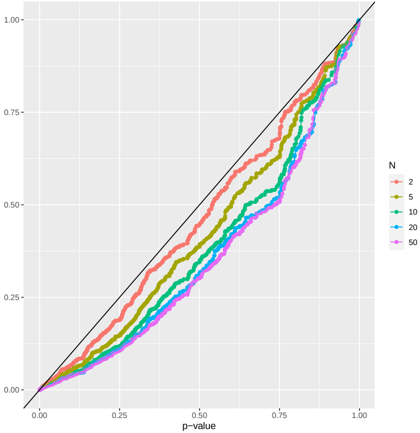
We can now marginalise out I to get

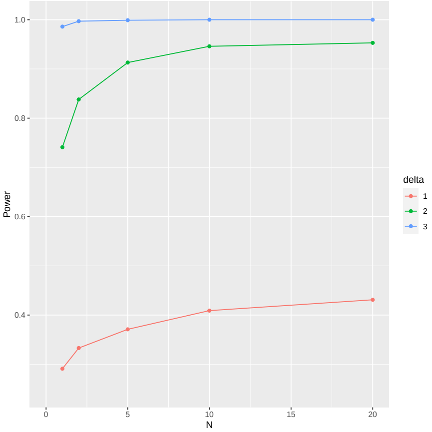
$$f(\phi|\boldsymbol{\psi}^{(1:N)}, S) = \sum_{I=1}^{N} f(\phi, I|\boldsymbol{\psi}^{(1:N)}, S)$$
$$= \frac{1}{W} f(\phi) \sum_{I=1}^{N} \Pr(\phi \in S_{\psi^{(I)}}) f(\phi|\phi \in S_{\psi^{(I)}}).$$

So

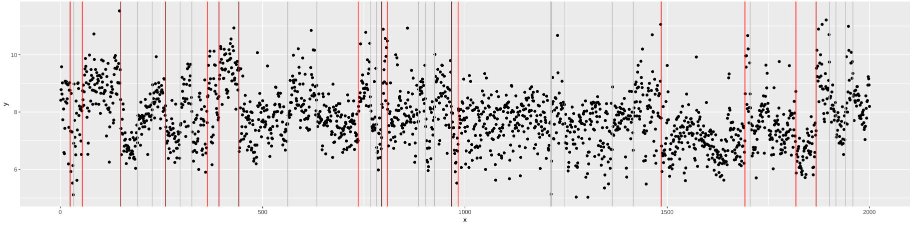
$$\Pr(|\phi| > \alpha|\boldsymbol{\psi}^{(1:N)}, S) = \frac{1}{W} \sum_{I=1}^{N} Pr(\phi \in S_{\psi^{(I)}}) \Pr(|\phi| > \alpha|\boldsymbol{\psi}^{(I)}, S).$$
$$= \frac{\sum_{I=1}^{N} \Pr\left(\phi \in S_{\psi^{(I)}}, |\phi| > \alpha \mid \boldsymbol{\psi}^{(I)}, S\right)}{\sum_{I=1}^{N} \Pr\left(\phi \in S_{\psi^{(I)}} \mid \boldsymbol{\psi}^{(I)}, S\right)}$$

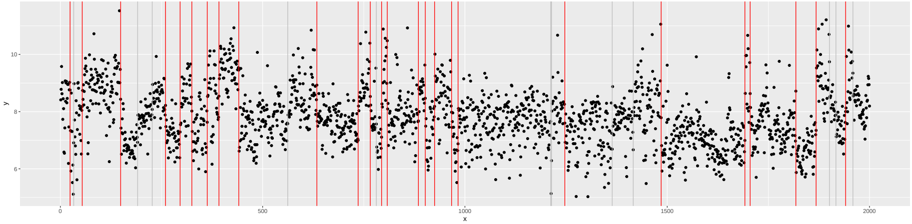
Simulations





Real data





Thanks for listening!

r.j.carrington@lancaster.ac.uk