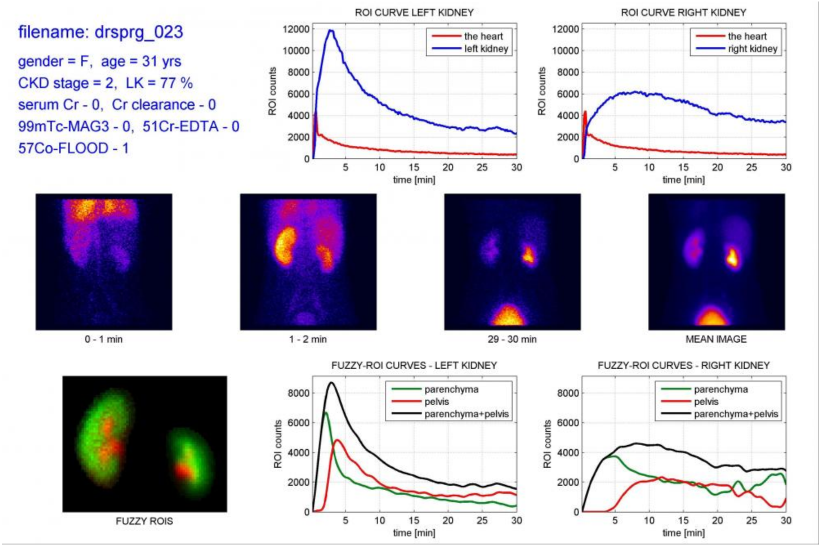
Mixture Model Challenge

Václav Šmídl

March 31, 2020

Patlak Rutland plot

Sequence of scintigraphic images of kidneys.

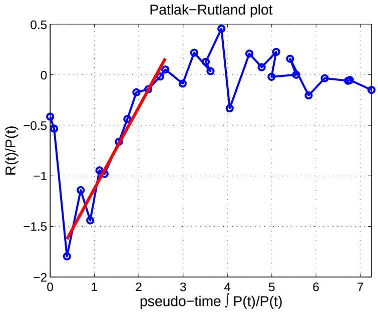


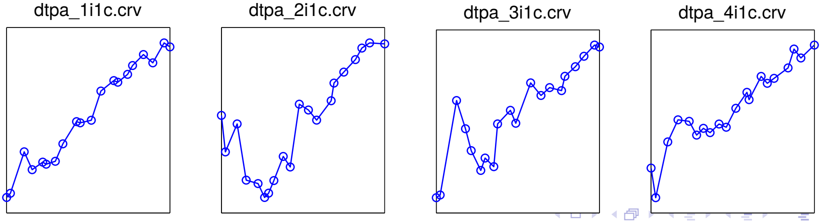
Challenge: Patlak Ruland plot

Patlak Rutland plot is a ratio of the parenchyma curve over the cumulative sum of the heart curve.

Diagnostically important is the **slope** of the linear part, which:

- typically starts around 1min
- typically ends around 3min
- with outliers
- manual fit is the norm





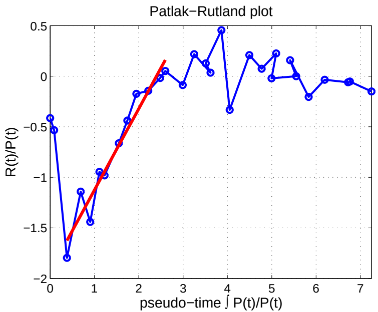
Models?

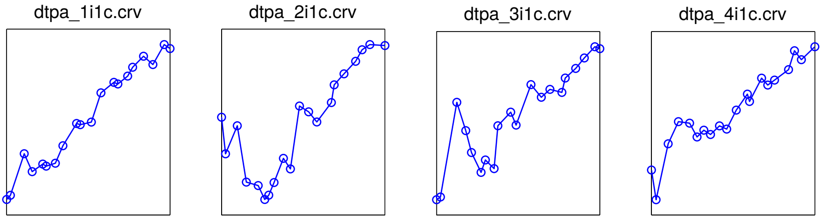
- no heuristics
- probabilistic model providing either:
  - estimate of $\hat{a}$
  - posterior distribution $p(a)$

  where a is the slope of:

$$y_i = ax_i + b + e_i$$

for points belonging to it.



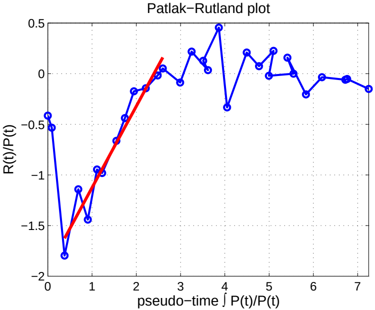


Model 1: Robust regression

Model

$$y_i = ax_i + b + e_i$$

is valid for all $x_i \in [1, 3]$ with heavy-tailed noise $p(e_i)$.



Model 1: Robust regression

Model

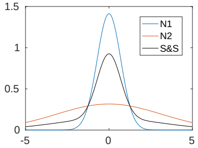
$$y_i = ax_i + b + e_i$$

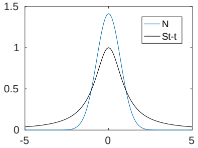
is valid for all $x_i \in [1, 3]$ with heavy-tailed noise $p(e_i)$.
Possibilities:

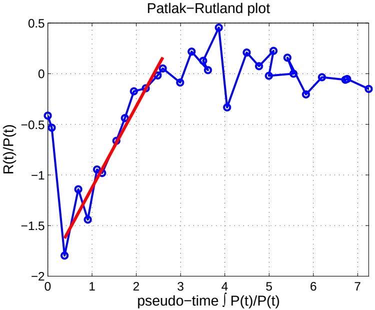
- Spike and slab,

$$p(e_i) = \alpha\mathcal{N}(0, v_1) + (1-\alpha)\mathcal{N}(0, v_2),$$

- Student-t $p(e_i) = \mathcal{S}t(\nu, 0, \sigma)$,







Latent representations

Spike and slab

$$p(e_i|v_1, v_2) = \alpha\mathcal{N}(0, v_1) + (1-\alpha)\mathcal{N}(0, v_2),$$

is the marginal of

$$\begin{aligned} p(e_i, l_i|v_1, v_2) &= p(e_i|l_i, v_1, v_2) \\ p(e_i|l_i = \epsilon_1, v_1) &= \mathcal{N}(0, v_1), \\ p(e_i|l_i = \epsilon_2, v_2) &= \mathcal{N}(0, v_2), \end{aligned}$$

Student-t

$p(e_i) = \mathcal{S}t(\nu, 0, \sigma)$, is the marginal of

$$\begin{aligned} p(e_i) &= \int p(e_i|\beta_i)p(\beta_i) \\ p(e_i|\beta_i) &= \mathcal{N}(0, \beta_i), \\ p(\beta_i) &= i\mathcal{G}(\gamma_0, \delta_0), \end{aligned}$$

Spike&Slab

Spike and slab prior of the noise implies:

$$p(e_i, l_i) = \mathcal{N}(0, v_1)^{l_i=\epsilon_1}\mathcal{N}(0, v_2)^{l_i=\epsilon_2},$$
$$p(y_i, l_i|x_i) = \mathcal{N}(ax_i + b, v_1)^{l_i=\epsilon_1}\mathcal{N}(ax_i + b, v_2)^{l_i=\epsilon_2},$$

Do the point outside of the line follow the $ax_i + b$ curve?

Spike&Slab

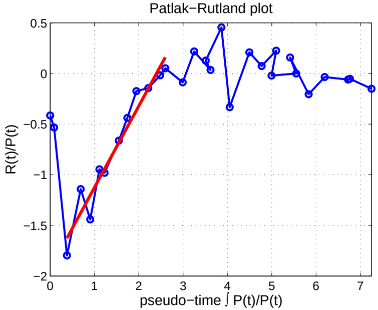
Spike and slab prior of the noise implies:

$$p(e_i, l_i) = \mathcal{N}(0, v_1)^{l_i=\epsilon_1}\mathcal{N}(0, v_2)^{l_i=\epsilon_2},$$
$$p(y_i, l_i|x_i) = \mathcal{N}(ax_i + b, v_1)^{l_i=\epsilon_1}\mathcal{N}(ax_i + b, v_2)^{l_i=\epsilon_2},$$

Do the point outside of the line follow the $ax_i + b$ curve?

What model do they follow?

$$p(y_i, l_i|x_i) = \mathcal{N}(ax_i + b, v_1)^{l_i=\epsilon_1} p(x_i|l_i = \epsilon_2)^{l_i=\epsilon_2},$$



Possibilities:

$$p(x_i|l_i = \epsilon_2) = \mathcal{U}(y_{min}, y_{max}),$$
$$p(x_i|l_i = \epsilon_2) = \mathcal{N}(ax, v(x_i, \psi)),$$
$$\vdots$$

Be creative.

Tutorial model

Recall matrix notation from linear regression:

$$y_i = \boldsymbol{x}_i^\top \theta, \qquad y_i = \begin{bmatrix} x_i & 1 \end{bmatrix} \begin{bmatrix} a & b \end{bmatrix}^\top,$$

Then

$$p(y_i | x_i, l_i) = \mathcal{N}(\boldsymbol{x}_i^\top \theta, \sigma)^{l_{1,i}} U(-2, 1)^{l_{2,i}},$$

$$p(l_i|\alpha) = \prod_{k=1}^{K} \alpha_k^{l_{k,i}} \qquad p(\alpha) = \prod_k \alpha_k^{z_0 - 1}$$

$$p(\theta) = \mathcal{N}(0, \tau I_2), \quad p(\omega) = \Gamma(\gamma_0, \delta_0)$$

Joint distribution:

$$\mathcal{L} = \sum_{i=1}^{n} \log p(y_i, l_i | x_i, \theta, \omega, \alpha) p(\theta, \omega, \alpha) = \frac{1}{2} \sum_i l_{1,i} \left(\log(\omega/2\pi) - \omega(y_i - \boldsymbol{x}_i^\top \theta)^2\right)$$

$$+ \sum_i l_{2,i} \log \frac{1}{3} + \sum_i \sum_k l_{k,i} \log \alpha_k + (z_0 - 1) \sum_k \log \alpha_k +$$

$$+ \frac{2}{2} \log \tau - \frac{1}{2} \tau \theta^\top \theta + (\gamma - 1) \log \omega - \delta_0 \omega,$$

$$\text{s.t} \sum l_{i,k} = 1, \ \sum \alpha_k = 1,$$

Expectation maximization (EM) algorithm: M-step

Expectation over l: $l_{1,i} \to \hat{l}_{1,i}$, $l_{2,i} \to \hat{l}_{2,i}$

$$
\begin{aligned}
\mathrm{E}_l(\mathcal{L}) = & \frac{1}{2} \sum_i \hat{l}_{1,i} \log \omega - \frac{1}{2} \sum_i \hat{l}_{1,i} \omega (y_i - \boldsymbol{x}_i^\top \theta)^2 \\
& + \sum_i \hat{l}_{2,i} \log \frac{1}{3} + \sum_i \sum_k \hat{l}_{k,i} \log \alpha_k + (z_0 - 1) \sum_k \log \alpha_k + \\
& + \frac{2}{2} \log \tau - \frac{1}{2} \tau \theta^\top \theta + (\gamma_0 - 1) \log \omega - \delta_0 \omega + \text{n.c.}, \qquad \text{s.t. : } \sum_k \alpha_k = 1,
\end{aligned}
$$

In θ (via moment matching of $q(\theta) = \mathcal{N}(\hat{\theta}, \Sigma_\theta)$):

$$
\begin{aligned}
\mathcal{L} &= -\frac{1}{2} \textstyle\sum_i \hat{l}_{1,i} \omega (y_i - \boldsymbol{x}_i^\top \theta)^2 - \frac{1}{2} \tau \theta^\top \theta \\
&= -\frac{1}{2} \textstyle\sum_i \hat{l}_{1,i} \omega (y_i^2 - 2 y_i \boldsymbol{x}_i^\top \theta + \theta^\top \boldsymbol{x}_i \boldsymbol{x}_i^\top \theta) - \frac{1}{2} \tau \theta^\top \theta \\
&= -\frac{1}{2} \left(\omega \textstyle\sum_i \hat{l}_{1,i} y_i^2 - 2 \omega \sum_i \hat{l}_{1,i} y_i \boldsymbol{x}_i^\top \theta + \theta^\top \left(\omega \sum_i \hat{l}_{1,i} \boldsymbol{x}_i \boldsymbol{x}_i^\top + \tau I_2 \right) \theta \right) \\
\hat{\theta} &= \left(\omega \textstyle\sum_i \hat{l}_{1,i} \boldsymbol{x}_i \boldsymbol{x}_i^\top + \tau I_2 \right)^{-1} \left(\omega \sum_i \hat{l}_{1,i} y_i \boldsymbol{x}_i^\top \right)
\end{aligned}
$$

Expectation maximization (EM) algorithm: M-step

Expectation over l: $l_{1,i} \to \hat{l}_{1,i}$, $l_{2,i} \to \hat{l}_{2,i}$

$$\mathcal{L} = \frac{1}{2}\sum_i \hat{l}_{1,i} \log\omega - \frac{1}{2}\sum_i \hat{l}_{1,i}\omega(y_i - \boldsymbol{x}_i^\top\theta)^2$$
$$+ \sum_i \hat{l}_{2,i} \log\frac{1}{3} + \sum_i\sum_k \hat{l}_{k,i}\log\alpha_k + (z_0 - 1)\sum_k \log\alpha_k +$$
$$+ \frac{2}{2}\log\tau - \frac{1}{2}\tau\theta^\top\theta + (\gamma_0 - 1)\log\omega - \delta_0\omega + \text{n.c.}, \qquad \text{s.t.}: \sum_k \alpha_k = 1,$$

In ω, (via moment matching of $q(\omega) = \mathcal{G}(\gamma, \delta)$):

$$\mathcal{L} = \gamma\log\omega - \delta\omega$$
$$\gamma = \frac{1}{2}\sum_i \hat{l}_{1,i} + 1 + \gamma_0 - 1$$
$$\delta = \frac{1}{2}\textstyle\sum_i \hat{l}_{1,i}(y_i - \boldsymbol{x}_i^\top\hat{\theta})^2 + \delta_0$$
$$\hat{\omega} = \gamma/\delta$$

Expectation maximization (EM) algorithm: M-step

Expectation over l: $l_{1,i} \to \hat{l}_{1,i}$, $l_{2,i} \to \hat{l}_{2,i}$

$$\mathcal{L} = \frac{1}{2}\sum_i \hat{l}_{1,i} \log \omega - \frac{1}{2}\sum_i \hat{l}_{1,i}\omega(y_i - \boldsymbol{x}_i^\top \theta)^2$$

$$+ \sum_i \hat{l}_{2,i} \log \frac{1}{3} + \sum_i \sum_k \hat{l}_{k,i} \log \alpha_k + (z_0 - 1)\sum_k \log \alpha_k +$$

$$+ \frac{2}{2}\log \tau - \frac{1}{2}\tau\theta^\top\theta + (\gamma_0 - 1)\log\omega - \delta_0\omega + \text{n.c.}, \qquad \text{s.t.}: \sum_k \alpha_k = 1,$$

In α (via moment matching of $q(\alpha) = \mathcal{D}i(z)$):

$$\mathcal{L} = \sum_k \sum_i \hat{l}_{k,i} \log \alpha_k + (z_0 - 1)\sum_k \log \alpha_k, \qquad \text{s.t.}: \sum_k \alpha_k = 1,,$$

$$= \sum_k z_k \log \alpha_k + \lambda(\sum_k \alpha_k - 1)$$

$$z_k = \sum_i \hat{l}_{k,i} + z_0$$

$$\hat{\alpha}_k = z_k / \sum_k z_k$$

Expectation maximization (EM) algorithm: E-step

Probability of l_i given estimates

$$\mathcal{L}(l_i) = \frac{1}{2} l_{1,i} \log\omega - \frac{1}{2} l_{1,i}\omega(y_i - \boldsymbol{x}_i^\top\theta)^2 - l_{1,i}\frac{1}{2}\log(2\pi)$$
$$+ l_{2,i}\log\frac{1}{3} + \sum_k l_{k,i}\log\alpha_k + (z_0 - 1)\sum_k \log\alpha_k + \text{n.c.}, \qquad \text{s.t. : } \sum_k l_k = 1$$

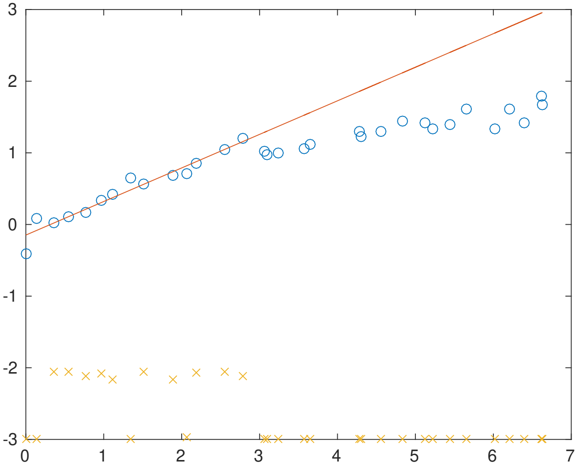
Template:

$$p(l_i|w) \propto \prod_k w_k^{l_{k,i}}, \qquad \sum_k w_k = 1, \qquad \sum_k l_{k,i} = 1 \text{(in our case)}$$

Matching (check with the Bayes rule):

$$\log w_1 = \frac{1}{2}\log\omega - \frac{1}{2}\omega(y_i - \boldsymbol{x}_i^\top\theta)^2 - \frac{1}{2}\log(2\pi) + \log\alpha_1$$
$$\log w_2 = \log\frac{1}{3} + \log\alpha_2$$
$$\hat{l}_1 = w_1/(w_1 + w_2)$$

EM algorithm results



- sensitive to initizalization

Modeling choices

How to encode our knowledge that the linear part is somewhere between 1 and 3min? hard constraints should be avoided.

Modeling choices

How to encode our knowledge that the linear part is somewhere between 1 and 3min? hard constraints should be avoided.

1. Prior on the latent l:

$$p(l|x) = \alpha_0(x),$$

2. Other options? Variance of the noise?

Assignment

Load data Patlak.mat
35 studies with:

- xpr x axis
- ypr y axis
- name name of the study
- int_start ignore
- int_end ignore

Assignment	points
find slope of linear part for all 35 studies	
own code (WLS + uniform)	15
own code (WLS + novel proposition)	25