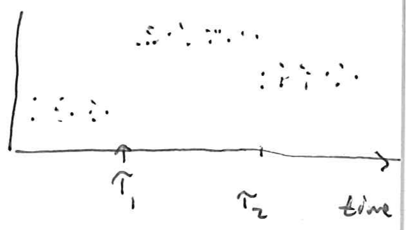
Changepoint analysis

* We study one or more changes in distr of time series
* online vs offline
* Some examples : : detect engine failure, fraud detection, etc.



Today: Fearnhead (2006)

Paul Fernhead
- prof at Lancaster
- worked with Martin Tretan
- changepoints, seq MC, etc

→ Contribution of paper:
It gives a method for exact simulation from posterior in $O(n^2)$ time for a class of Bayesian multiple changepoints models.

For Bayesian multiple changepoints models, reversible jump MCMC is usually needed to simulate from the posterior.
Problems:
- computationally expensive
- not exact
- may not get convergence/mixing

Model

We observe responses $y_1, \ldots y_n$. Notation: $y_{1:t} = (y_1, \ldots y_t)$.

We assume there are m changepoints (a random quantity) at times $0 < \tau_1 < \ldots < \tau_m < n$.

Conditional on # of changepoints and their positions, we assume each segment $(y_{\tau_{j-1}+1}, \ldots, y_{\tau_j})$ is associated a parameter θ_j s.t. conditional on θ_j, $y_i \overset{ind}{\sim} f(y; \theta_j)$. Each θ_j is independently drawn from the same prior, so $\theta_j \overset{iid}{\sim} \pi(\theta_j)$.

Comment: Independence between θ_j, θ_{j+1} (eg mean in two consecutive segments) is kind of restrictive.

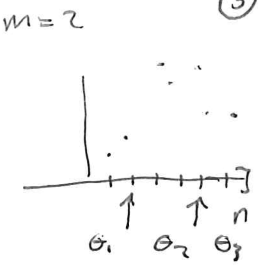
The number of changepoints is $m \sim \pi(m)$,

and cond on m, the m-th changepoint is

$\tau_m \sim \pi_m(\tau_m)$. while conditioned on m, τ_m, $\tau_{m-1} \sim \pi_m(\tau_{m-1} | \tau_m)$,

and so on. $\rightarrow$ $\tau_j | m, \tau_{j-1} \sim \pi_m(\tau_j | \tau_{j-1})$.

Rmk: Fearnhead also considers another prior on the changepoints, where we take the changepoints as the points of a point process.
I belive this "point process prior" is less general and less interesting, so we omit this in the interest of time.



A heuristic way to look at the DAP:

1) Draw the number of changepoints, $m \sim \pi(m)$
2) Draw the last changepoint $\tau_m \sim \pi_m(\tau_m)$
3) Draw $\tau_j \sim \pi_m(\tau_j \mid \tau_{j+1})$ for $j = m-1, \ldots, 1$
4) Draw $\theta_j \sim \pi(\theta_j)$ indep for $j = 1, \ldots, m+1$, ie $\forall$ segments

   (eg θ_j is mean of segment j)

5) For $i = 1, \ldots, n$, draw $y_i \overset{ind}{\sim} f(y_i \mid \theta_j)$, where j is the segment in which y_i lies

<u>Road ahead</u>

1. We will define a set of probabilities that will allow us to simulate $\tau_1 \ldots \tau_m$ directly from the posterior, conditional on m (and unconditional)

2. We prove a thm that gives us a method for computing the marginal probabilities recursively in $O(n^2)$ time.

3. (If time) A (simple) application to data.

We would like to simulate $\tau_1, \ldots \tau_m$ from the posterior conditional on m. That is, to draw samples

$$\tau_1, \ldots \tau_m \sim \Pr(\tau_1, \ldots \tau_m \mid y_{1:n}, m).$$

<u>Definition</u>

For $s \geq t$, define

$$P(t,s) = Pr(y_{t:s} \mid t,s \text{ in same segment}) \qquad \text{(conditional density)}$$

$$= \int \prod_{i=t}^{s} f(y_i|\theta)\,\pi(\theta)\,d\theta.$$ In practice we will need conjugate prior here.

For $j=0,\dots,m$, and $t=j+1,\dots,n-m+j$, define

$$Q_j^{(m)}(t) = Pr(y_{t:n}, \tau_j = t-1 \mid m \text{ changepoints}).$$ (Typo in paper)

How do these help us? Well, if $P(t,s)$ and $Q_j^{(m)}(t)$ are known,

$$\begin{aligned}
Pr(\tau_1 \mid y_{1:n}, m) &= Pr(\tau_1, y_{1:n} \mid m) / Pr(y_{1:n} \mid m) \\
&= Pr(\tau_1, y_{1:n} \mid m) / Q_0^{(m)}(1) \\
&= Pr(\tau_1, y_{\tau_1+1:n} \mid m)\, Pr(y_{1:\tau_1} \mid \tau_1, y_{\tau_1+1:n}, m) / Q_0^{(m)}(1) \\
&= Q_1^{(m)}(\tau_1)\, Pr(y_{1:\tau_1} \mid \tau_1, m) / Q_0^{(m)}(1) \qquad \text{(independence)} \\
&= Q_1^{(m)}(\tau_1)\, P(1,\tau_1) / Q_0^{(m)}(1).
\end{aligned}$$

Furthermore, for $j=2,\dots m$,

$$\begin{aligned}
Pr(\tau_j \mid \tau_{j-1}, y_{1:n}, m) &= \frac{Pr(\tau_{j-1}, \tau_j, y_{1:n} \mid m)}{Pr(\tau_{j-1}, y_{1:n} \mid m)} \\
&= \frac{Pr(\tau_{j-1}, \tau_j, y_{1:n} \mid m)}{Pr(\tau_{j-1}, y_{\tau_{j-1}+1:n} \mid m)\, Pr(y_{1:\tau_{j-1}} \mid \tau_{j-1}, m)} \qquad \text{(independence)} \\
&= \frac{Pr(\tau_j, y_{\tau_j+1:n} \mid m)\, Pr(\tau_{j-1} \mid \tau_j, y_{\tau_j+1:n}, m)\, Pr(y_{1:\tau_j} \mid \tau_j, \tau_{j-1}, y_{\tau_j+1:n}, m)}{Pr(\tau_{j-1}, y_{\tau_{j-1}+1:n} \mid m)\, Pr(y_{1:\tau_{j-1}} \mid \tau_{j-1}, m)} \cdots \\
&= \frac{Pr(\tau_j, y_{\tau_j+1:n} \mid m)\, Pr(\tau_{j-1} \mid \tau_j, m)\, Pr(y_{\tau_{j-1}+1:\tau_j} \mid \tau_j, \tau_{j-1})}{Pr(\tau_{j-1}, y_{\tau_{j-1}+1:n} \mid m)}
\end{aligned}$$

$$= Q_j^{(m)}(\tau_j+1)\,\pi_m(\tau_{j-1}\mid\tau_j)\,P(\tau_{j-1}+1,\tau_j)\,/\,Q_0^{(m)}(\tau_{j-1}+1).$$

Thus, we can use $P(t,s)$ and $Q_j^{(m)}(t)$ to simulate $\tau_1,\dots\tau_m$ from the posterior. Fearnhead also gives an algorithm to quickly simulate M samples at once.

We can also use $P(t,s)$ and $Q_j^{(m)}(t)$ to find $P(m\mid y_{1:n})$.

The following theorem states that we can compute $Q_j^{(m)}(t)$ recursively ($O(n^2)$ time).

Thm (Fearnhead, 2006)

1. $Q_m^{(m)}(t) = P(t,n)\pi_m(\tau_m=t-1)$, $t=m+1,\dots,n-1$ (typo in article I think)

2. $$Q_j^{(m)}(t) = \sum_{s=t}^{n-m+j} P(t,s)\,Q_{j+1}^{(m)}(s+1)\,\pi_m(\tau_j=t\mid\tau_{j+1}=s)$$

   for $j=1,\dots,m-1$, and $t=j+1,\dots,n-m-1+j$

3. $$Q_0^{(m)}(1) = Pr(y_{1:n}\mid m)$$

   $$= \sum_{s=1}^{n-m} P(1,s)\,Q_1^{(m)}(s+1).$$

Pf (very similar to proof of thm 2 in Fearnhead.)

We only prove the second statement. We have

$$Q_j^{(m)}(t) = \Pr(y_{t:n}, \tau_j = t-1 \mid m) \qquad (\text{def})$$

$$= \sum_{s=t}^{n-m+j} \Pr(y_{t:n}, \tau_{j+1} = s, \tau_j = t-1 \mid m)$$

$$= \sum_{s=t}^{n-m+j} \Big\{ \Pr(\tau_{j+1} = s, y_{s+1:n} \mid m) \Pr(\tau_j = t-1 \mid \tau_{j+1} = s, y_{s+1:n}, m)$$
$$\Pr(y_{t:s} \mid \tau_j = t, \tau_{j+1} = s, y_{s+1:n}, m) \Big\}$$

$$= \sum_{s=t}^{n-m+j} \Big\{ \Pr(\tau_{j+1} = s, y_{s+1:n} \mid m) \Pr(\tau_j = t-1 \mid \tau_{j+1} = s, m)$$
$$\Pr(y_{t:s} \mid \tau_j = t, \tau_{j+1} = s, m) \Big\}$$

$$= \sum_{s=t}^{n-m+j} Q_{j+1}^{(m)}(s+1)\, \pi_m(\tau_j = t-1 \mid \tau_{j+1} = s)\, P(t,s).$$

$\square$

Simulation based on Fearhead (2006)

Per August Jarval Moen

7/9/2021

Model

We condition on $m = 1$ and let the sample size be $n = 100$. The prior distribution for the changepoint τ_1 is uniform. We assume prior $\theta_j \sim \mathrm{N}(0, 5)$ and $y_i|\theta_j \sim \mathrm{N}(0, 2)$, where j is the segment y_i is in.

To generate an example dataset, we set the realized value of τ_1 to be 50, $\theta_1 = 10$ and $\theta_2 = 15$.

Simulated example