Poisson Multi-Changepoint Model

MLE + Laplace Approximation

$y_t \ \forall \ t = 1, 2 \ldots T$

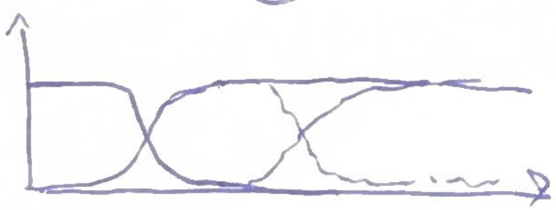
K changepoints ∴ K+1 states.

Changepoints @ $\tau_1, \tau_2 \ldots \tau_{K+1}$ positions. w/ $\tau_0 = 0$ and $\tau_{K+1} = T$

Emmissions := $\lambda_j \ \forall \ j = 0, 1, 2 \ldots K$

$y_t \sim \text{Poisson}(\lambda_j)$

To allow for use of gradients over τ, we define τ as continuous with states being marked using sigmoids.



For each state, we can have a filter: $A_j(t) = \sigma_j(t)\left(1 - \sigma_{j+1}(t)\right)$

Using this, we can define the rates: $\lambda(t) = \sum^{K} \lambda_j \cdot A_j(t)$

where $\cancel{\sigma(t) = \frac{1}{1+e^{-\frac{(t-\tau)}{s}}}}$ $\sigma(t) = \frac{1}{1+e^{-s(t-\tau)}}$, Since s controls the timescale of the sigmoid, we can make sure the effect of s is always negligible to remove it as a free param.

e^{-st} means $1 \to 0$ takes @ $t = e$.
If were working in time indices and not strictly time, we can set $s = 100$ (or something similarly large).

∴ Assuming timepoints are independent, given the parameters:

$$\mathcal{L}(\vec{\lambda}, \vec{\tau}) = \prod^{T} f(y_t \mid \lambda_t) \quad \text{for} \quad f(y_t \mid \lambda(t)) = \lambda(t)^{y_t} e^{-\lambda(t)} / y_t!$$

However, we have to make sure changepoints are sequential, that is $\tau_1 < \tau_2 < .. \tau_k$.

So instead of sampling changepoints directly, we can sample them as steps / deltas

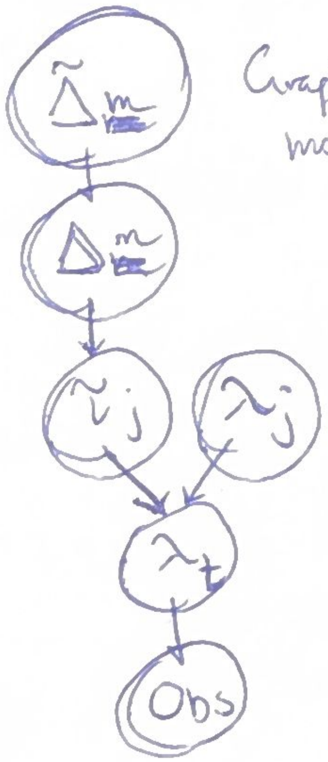
that is $\tau_j = \sum_{m=1}^{j} \Delta_m$.

To allow ~~for~~ unconstrained optimization, we can draw $\Delta_m = e^{\tilde{\Delta}_m}$

Getting back to the log-likelihood: $\ell \propto \sum_t \left[y_t \ln \lambda(t) - \lambda(t) \right]$ (from single changepoint MLE pg 1)

Start MLE:

$$\cancel{\frac{\partial \ell}{\partial \tau} = \frac{\partial \ell}{\partial \lambda(t)} \cdot \frac{\partial \lambda(t)}{\partial A} \cdot \frac{\partial A}{\partial \sigma} \cdot \frac{\partial \sigma}{\partial \tau}} \; ; \quad \tilde{\Delta}_m \xrightarrow{\exp} \Delta_m \xrightarrow{\Sigma} \tau_j \longrightarrow \sigma_j(t) \longrightarrow A_j(t) \longrightarrow \lambda_t \longrightarrow \ell$$



Graphical model.

$$\frac{\partial \ell}{\partial \tilde{\Delta}_m} = \sum_{j=1}^{k} \frac{\partial \ell}{\partial \tau_j} \cdot \frac{\partial \tau_j}{\partial \tilde{\Delta}_m}$$ since all taus are affected by a single $\tilde{\Delta}_m$

$$\Rightarrow \frac{\partial \ell}{\partial \tau_j} = \sum_{t=1}^{T} \frac{\partial \ell}{\partial \lambda_t} \cdot \frac{\partial \lambda_t}{\partial \tau_j}$$ since each τ_j alters the complete emission.

Combining: $$\frac{\partial \ell}{\partial \tilde{\Delta}_m} = \sum_{j=1}^{k} \left[\sum_{t=1}^{T} \frac{\partial \ell}{\partial \lambda_t} \cdot \frac{\partial \lambda_t}{\partial \tau_j} \right] \frac{\partial \tau_j}{\partial \tilde{\Delta}_m}$$

☆ Note: λ_t denotes rate at all times while λ_j denotes a single value for the rate of a given state

$$\boxed{\frac{\partial \ell}{\partial \lambda_t} = \frac{y_t}{\lambda_t} - 1}$$ from single changepoint MLE pg. 3

$$\frac{\partial \lambda_t}{\partial \tau_j} = \frac{\partial}{\partial \tau_j} \left[\sum^{k} \lambda_k \cdot A_k(t) \right] = \sum^{k} \lambda_k \cdot \frac{\partial}{\partial \tau_j} A_k(t) = \sum^{k} \lambda_k \frac{\partial A_k(t)}{\partial \sigma_j} \cdot \frac{\partial \sigma_j}{\partial \tau_j}$$

③

Note, any given τ_j will only impact the sigmoid ~~its~~ it is associated with $\therefore$ it will only impact states (A_k) that use that sigmoid.

$$A_k(t) = \sigma_k(t)(1 - \sigma_{k+1}(t))$$

for $k = j-1$: $A_{j-1}(t) = \sigma_{j-1}(t)[1 - \sigma_j(t)]$

$k = j$: $A_j(t) = \sigma_j(t)[1 - \sigma_{j+1}(t)]$

for all other A_k, $\frac{\partial A_k(t)}{\partial \tau_j} = 0$

$$\therefore \frac{\partial \lambda_t}{\partial \tau_j} = \lambda_{j-1} \frac{\partial A_{j-1}(t)}{\partial \tau_j} + \lambda_j \frac{\partial A_j(t)}{\partial \tau_j}$$

Need to find $\frac{\partial \sigma_j}{\partial \tau_j} = \frac{\partial}{\partial \tau_j} \frac{1}{1 + e^{-s(t-\tau_j)}} = -s\,\sigma_j(1-\sigma_j)$ [from single chp. MLE pg 3]

$$\frac{\partial A_{j-1}(t)}{\partial \tau_j} = \frac{\partial}{\partial \tau_j}\left[\sigma_{j-1}(t)[1 - \sigma_j(t)]\right] = \sigma_{j-1}[1 - \cancel{\sigma_j(t)}]^{0} \frac{\partial}{\partial \tau_j}(1 - \sigma_j(t)) = -\sigma_{j-1}[-s\,\sigma_j(1-\sigma_j)]$$

$$\Rightarrow +s\,\sigma_{j-1}\sigma_j(1-\sigma_j)$$

$$\frac{\partial A_j(t)}{\partial \tau_j} = \frac{\partial}{\partial \tau_j}\left[\sigma_j(t)[1 - \sigma_{j+1}(t)]\right] = (1 - \sigma_{j+1}(t)) \cdot \frac{\partial \sigma_j(t)}{\partial \tau_j} = [1 - \sigma_{j+1}(t)][-s\,\sigma_j(1-\sigma_j)]$$

$$\Rightarrow -s\,\sigma_j[1 - \sigma_j][1 - \sigma_{j+1}]$$

$$\frac{\partial \lambda_t}{\partial \tau_j} = \lambda_{j-1}\left[\sigma_{j-1}[+s\,\sigma_j(1-\sigma_j)]\right] + \lambda_j[1 - \sigma_{j+1}][-s\,\sigma_j(1-\sigma_j)]$$

$$\boxed{\frac{\partial \lambda_t}{\partial \tau_j} = s\,\sigma_j(1-\sigma_j)\left[\lambda_{j-1}\sigma_{j-1} - \lambda_j(1 - \sigma_{j+1})\right]}$$

~~$\frac{\partial \lambda_t}{\partial \tau_j} = \sum^{k} \lambda_k \frac{\partial}{\partial \tau_j} \Lambda_k(t) = \sum^{k} \lambda_k \Big[$~~

Final piece for $\frac{\partial \ell}{\partial \tilde{\Delta}_m}$ is $\frac{\partial \tau_j}{\partial \tilde{\Delta}_m}$:

$\frac{\partial \tau_j}{\partial \tilde{\Delta}_m} = \frac{\partial}{\partial \tilde{\Delta}_m} \sum^{j} e^{\tilde{\Delta}_j}$, note $\tilde{\Delta}_m$ only has an effect on τ_j if ~~$j \leq m$~~ $m \leq j$
that is if that Δ occurs before the changepoint, and it alters the position of τ_j by itself, that is by $e^{\tilde{\Delta}_m}$.

In other words,

$$\tau_j = \sum_j e^{\tilde{\Delta}_j} = e^{\tilde{\Delta}_1} + e^{\tilde{\Delta}_2} + \dots e^{\tilde{\Delta}_m} \dots + e^{\tilde{\Delta}_j}$$

$$\frac{\partial}{\partial \tilde{\Delta}_m} \tau_j = \frac{\partial}{\partial \tilde{\Delta}_m} \Big[\qquad \text{"} \qquad \Big] = \frac{\partial}{\partial \tilde{\Delta}_m} e^{\tilde{\Delta}_m} = e^{\tilde{\Delta}_m} \text{ if } m \leq j \text{ else } 0.$$

$$\boxed{\frac{\partial \tau_j}{\partial \tilde{\Delta}_m} = e^{\tilde{\Delta}_m} \cdot \mathbb{I}(m \leq j)}$$

selector variable

No need to write out the full form for $\frac{\partial \ell}{\partial \tau_j}$

We can test the most complicated gradient, $\frac{\partial \lambda_t}{\partial \tau_j}$ numerically for accuracy.

Next, $\frac{\partial \ell}{\partial \lambda_j}$ (individual state emissions): $\frac{\partial \ell}{\partial \lambda_j} = \sum_t \frac{\partial \ell}{\partial \lambda_t} \cdot \frac{\partial \lambda_t}{\partial \lambda_j}$

$\frac{\partial \ell}{\partial \lambda_t}$ we already know.

$$\frac{\partial \lambda_t}{\partial \lambda_j} = \frac{\partial}{\partial \lambda_j} \sum^{K} \lambda_k \cdot A_k(t) = \sum^{k} A_k(t) \frac{\partial \lambda_k}{\partial \lambda_j} = \sum^{k} A_k(t) \quad \text{for } k=j$$

$$\frac{\partial \lambda_t}{\partial \lambda_j} = A_j(t)$$

$$\frac{\partial \ell}{\partial \lambda_j} = \sum_t \frac{\partial \ell}{\partial \lambda_t} \cdot \frac{\partial \lambda_t}{\partial \lambda_j} = \boxed{\sum_t \left(\frac{y_t}{\lambda_t} - 1 \right) \cdot A_j(t) = \frac{\partial \ell}{\partial \lambda_j}}$$

☆ Laplace approximation done numerically.