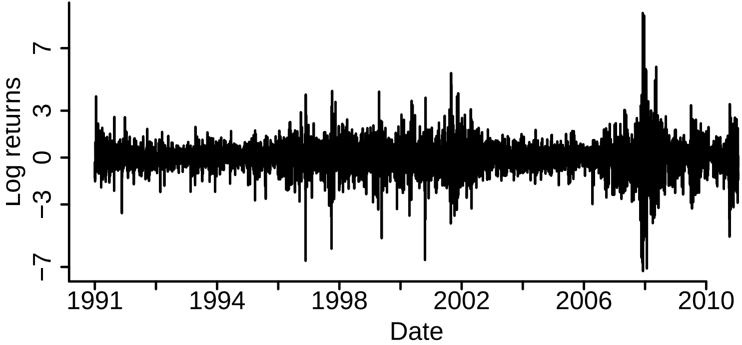
Lecture 6: Change Point Detection

Alex Donov

29 February 2019

- So far we have considered situations where Y_i is independent and identically distributed, $Y_i \sim p(Y|\theta)$
- However, this assumption may not hold in practice.
- Consider the Value-at-Risk problem where we wish to find the probability of losing more than \$D in a single day.
- We start by specifying a probability model, e.g. $N(0, \theta)$, for the returns $Y_1, \ldots, Y_n$, and then use historical data to estimate θ.

Some Real Financial Data



- The pattern we see here with the constantly changing variance occurs in pretty much all real financial data.
- In the case the returns $Y_1, \ldots, Y_n$ are clearly not identically distributed, due to the changing variance.

The Problem

- In cases like this, we cannot just naively estimate the unknown parameters θ as we have been doing.
- Suppose we have 10 years of observed data. If a change in the distribution $p(Y|\theta)$ occurred last year (e.g. if the variance increased), then the observations before that point should not be used to estimate the current value of θ because they will give us a misleading estimate.
- In other words we need to estimate where the most recent change point has occurred and only look at the data after that point. We will then use only this data to compute the quantities we are interested in (e.g. the predictive distribution, and probability of extreme events occurring)

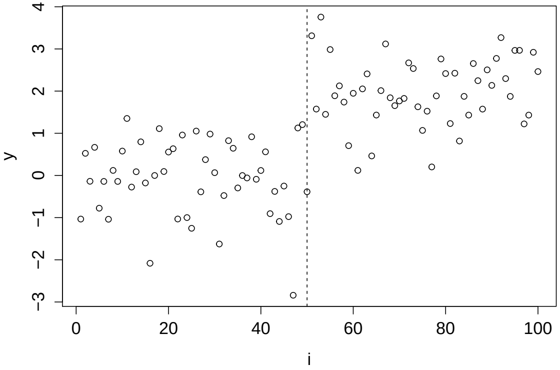
Change Point Detection

- The setting is as follows:
  - we have a sequence of time ordered data $Y_1, \ldots, Y_n$.
  - suppose we believe there is a single change point but we do not know when it occurs. Denote this unknown change point by τ.
  - before the change point, the unknown parameter θ has value θ_1, while after it changes to θ_2.
- The distribution of the data is then:

$$Y_i = \begin{cases} p(Y|\theta_1) & \text{if } i \leq \tau \\ p(Y|\theta_2) & \text{if } i > \tau \end{cases}$$

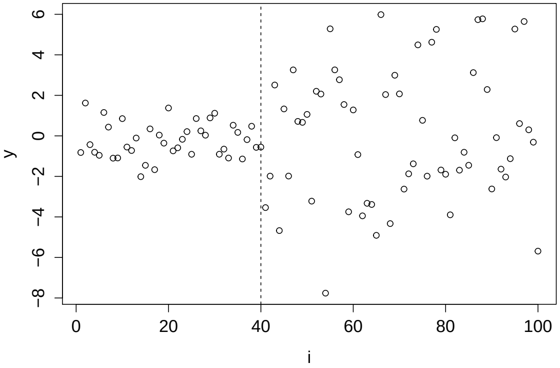
Example 1

Let $Y_1, \ldots, Y_{100} \sim N(\mu, 1)$, where $\mu = 0$ before the change point at $\tau = 50$, and $\mu = 1$ after:



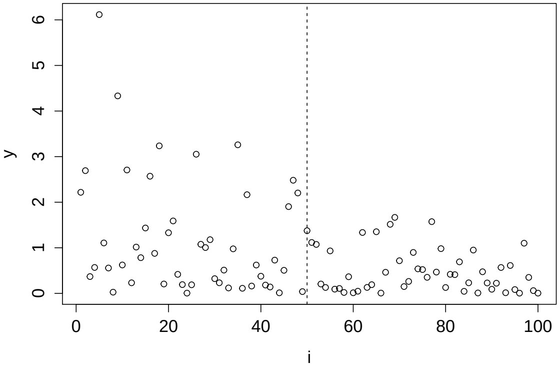
Example 2

Let $Y_1, \dots, Y_{100} \sim N(0, \sigma^2)$ where the change point is at $\tau = 40$ and $\sigma^2 = 1$ before this, and $\sigma^2 = 9$ after:



Example 3

Let $Y_1, \ldots, Y_{100} \sim Exponential(\lambda)$ where $\lambda = 1$ before the change point at $\tau = 50$, and $\lambda = 3$ after:



How to Detect Change Points

- In these examples the location of the change point is fairly obvious to the eye since large changes in parameters have been deliberately chosen. However, this will not be the case when the magnitude of change is smaller.
- We will start with the assumption that there is a **single** change point τ, and we want to estimate its location.
- We will work out the mathematics for the third example above (Exponential sequence) in detail, but identical techniques apply to essentially any other type of distribution (Normal, Binomial, etc)

Our Setting

- So our setting is that we have observed $Y_1, \ldots, Y_n$ where:

$$Y_i = \begin{cases} Exponential(\lambda_1) & \text{if } i \leq \tau \\ Exponential(\lambda_2) & \text{if } i > \tau \end{cases}$$

  where $\lambda_1, \lambda_2, \tau$ are all unknown.

- We seek to estimate τ.

How to Detect Change Points

To estimate τ using Bayesian approach, we proceed as usual:

- start with a prior $p(\tau)$
- specify likelihood $p(\tau|Y)$
- apply the Bayes' Theorem to obtain the posterior distribution $p(\tau|Y)$ which represents our knowledge about the change point after seeing the data.

$$p(\tau|Y) = \frac{p(\tau)p(Y|\tau)}{p(Y)}$$

The Prior

- Often we do not have strong beliefs about where the change point τ occurs.
- In this case we can proceed using a non-informative prior, which treats every possible location as equally likely.
- There are n observations and the last one cannot be a change point since we require at least one observation on both sides of τ.
- As such, there are $n-1$ possible locations, so the non-informative prior is just a discrete uniform distribution on $1, \dots, n-1$:

$$p(\tau = k) = \frac{1}{n-1}, \quad k = 1, \dots, n-1$$

The Likelihood

- Next consider the likelihood $p(Y|\tau)$, which depends on the unknown parameters λ_1 and λ_2.
- Suppose that these quantities are known.
- Then, the likelihood would simply be:

$$p(Y|\tau) = \prod_{i=1}^{\tau} p(Y_i|\lambda_1) \prod_{i=\tau+1}^{n} p(Y_i|\lambda_2) \tag{1}$$

$$= \prod_{i=1}^{\tau} \lambda_1 e^{-\lambda_1 Y_i} \prod_{i=\tau+1}^{n} \lambda_2 e^{-\lambda_2 Y_i} \tag{2}$$

How to Detect Change Points

- However, parameters λ_1 and λ_2 are not known.
- Suppose we use the usual conjugate Gamma(α, β) prior.
- Consider the first segment to the left of the change point, i.e. observations $Y_1, Y_2, \ldots, Y_\tau$ only.
- Since λ_1 is unknown, we can use the Law of Total Probability and integrate out the parameter λ_1:

$$p(Y_1, \ldots, Y_\tau | \tau) = \int \left(\prod_{i=1}^{\tau} p(Y_i | \lambda_1) \right) p(\lambda_1) d\lambda_1$$

How to Detect Change Points

- In the Exponential-Gamma case, we can easily compute this integral:

$$p(Y_1,\ldots,Y_\tau|\tau) = \int \prod_{i=1}^{\tau} p(Y_i|\lambda_1)p(\lambda_1)d\lambda_1 \tag{3}$$

$$= \int \prod_{i=1}^{\tau} \left(\lambda_1 e^{-\lambda_1 Y_i}\right)\left(\frac{\beta^\alpha}{\Gamma(\alpha)}\lambda_1^{\alpha-1}e^{-\beta\lambda_1}\right)d\lambda_1 \tag{4}$$

$$= \frac{\beta^\alpha}{\Gamma(\alpha)}\frac{\Gamma(\alpha+\tau)}{(\beta+S_1)^{\alpha+\tau}} \tag{5}$$

where $S_1 = \sum_{i=1}^{\tau} Y_i$

How to Detect Change Points

- It is easy to show using the same argument that the likelihood for the observations $Y_{\tau+1}, Y_{\tau+2}, \ldots, Y_n$ to the right of the change point is:

$$p(Y_{\tau+1}, \ldots, Y_n|\tau) = \frac{\beta^\alpha}{\Gamma(\alpha)} \frac{\Gamma(\alpha + n - \tau)}{(\beta + S_2)^{\alpha+n-\tau}} \quad \text{where} \quad S_2 = \sum_{i=\tau+1}^{n} Y_i$$

- Combining this result together with the previous expression gives the full likelihood:

$$p(Y|\tau) = \left(\frac{\beta^\alpha}{\Gamma(\alpha)}\right)^2 \frac{\Gamma(\alpha + \tau)}{(\beta + S_1)^{\alpha+\tau}} \frac{\Gamma(\alpha + n - \tau)}{(\beta + S_2)^{\alpha+n-\tau}}$$

How to Detect Change Points

- Finally, combining with the prior $p(\tau)$ gives the posterior distribution for the change point τ:

$$p(\tau|Y) = \frac{p(\tau)p(Y|\tau)}{p(Y)} = \frac{\frac{1}{n-1}\left(\frac{\beta^\alpha}{\Gamma(\alpha)}\right)^2 \frac{\Gamma(\alpha+\tau)}{(\beta+S_1)^{\alpha+\tau}} \frac{\Gamma(\alpha+n-\tau)}{(\beta+S_2)^{\alpha+n-\tau}}}{p(Y)}$$

How to Detect Change Points

What do we do about the $p(Y)$ part on the denominator? It turns out we actually do not need to compute it.

Think about what the expression $p(\tau|Y)$ really means. It means that if we plug in $\tau = k$ into the formula on the previous page $p(\tau = k|Y)$, we get the posterior probability that $\tau = k$

There are only $n-1$ possible locations for the change point τ, i.e. $p(\tau|Y)$ is only non-zero at the values $\tau = 1, 2, \ldots, n-1$

Note that the part on the bottom $p(Y)$ does not depend on τ

How to Detect Change Points

Since in this case there are only a finite number of values for τ, $(1, 2, \ldots, n-1)$, we can instead just evaluate the **numerator** of $p(\tau|Y)$ at these values and ignore the $p(Y)$.

However, the resulting posterior will not sum to 1. But we can rescale it so it sums to 1.

Remember in general that if we have any sequence of numbers $Z_1, Z_2, \ldots, Z_m$ then we can rescale these to sum to 1 by dividing each one by their total sum:

$$\tilde{Z}_1 = \frac{Z_1}{\sum_{i=1}^m Z_i}, \quad \tilde{Z}_2 = \frac{Z_2}{\sum_{i=1}^m Z_i}, \ldots, \tilde{Z}_m = \frac{Z_m}{\sum_{i=1}^m Z_i},$$

Then $\sum_{i=1}^m \tilde{Z}_i = 1$.

Example

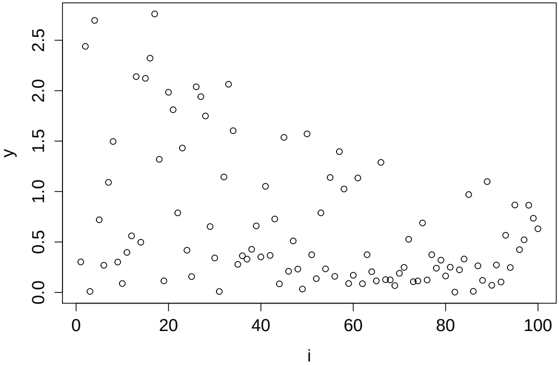
Let's work through an example to make it clearer.

Assume we have 100 observations $Y_1, \ldots, Y_{100}$. The true (and unknown) parameters are $\tau = 50, \lambda_1 = 1, \lambda_2 = 3$.

We use the previous uniform prior on τ, and give λ_1 and λ_2 a Gamma(1,1) prior

Example

The sequence looks like this:



Example

There are 99 possible locations for the unknown τ, corresponding to $\tau \in \{1, 2, \ldots, 99\}$.

We evaluate the numerator of $p(\tau|Y)$ at each of these 99 values of τ (remember we do not need to evaluate $p(Y)$):

$$p(\tau = k|Y) \propto \frac{1}{n-1} \left(\frac{\beta^\alpha}{\Gamma(\alpha)} \right)^2 \frac{\Gamma(\alpha + k)}{(\beta + S_1)^{\alpha+k}} \frac{\Gamma(\alpha + n - k)}{(\beta + S_2)^{\alpha+n-k}}$$

for $k = 1, 2, \ldots, 99$ where

$$S_1 = \sum_{i=1}^{k} Y_i, \quad S_2 = \sum_{i=k+1}^{n} Y_i$$

Example

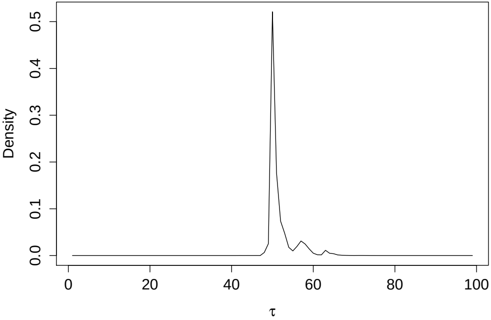
Next, we normalise these values to sum to 1. So, for each k we have:

$$p(\tau = k|Y) = \frac{\tilde{p}(\tau = k|Y)}{\sum_{i=1}^{99} \tilde{p}(\tau = i|Y)}$$

This gives us the true posterior, which is now a properly normalised probability distribution that sums to 1

Example

We can now plot the posterior distribution $p(\tau|Y)$



(note I have smoothed this in the plot to make it clearer but it should really be a histogram: remember that the posterior is only defined on integer values $1, 2, 3, \ldots, n-1$)

Example

So our posterior beliefs are strongly peaked at the change point occurring at around $\tau = 50$. However there is some uncertainty. The actual values are (you will compute these on the exercise sheet)

$$p(\tau = 49|Y) = 0.02$$

$$p(\tau = 50|Y) = 0.52$$

$$p(\tau = 51|Y) = 0.17$$

$$p(\tau = 52|Y) = 0.07$$

And $p(\tau = k|Y) = 0$ for all other values of k.

So we can quantify our posterior belief that the change point occurs at any particular location. We believe there is a 52% chance it occurs at $\tau = 50$, a 17% probability it occurs at $\tau = 51$, and so on.

Point Estimate

As always, we can get a point estimate for τ by taking an appropriate summary of the posterior.

Recall that if we seek to minimise the mean square error, we choose the posterior mean. In this case, our point estimate is:

$$\hat{\tau} = \sum_{k=1}^{99} \tau p(\tau|Y)$$

which is $\hat{\tau} = 51.62$ in this case.

Next we will focus on the following three questions:

1. How do we use the information we learned about the change point to reason about the probability of extreme values occurring?
2. How do we decide whether there actually is a change point at all?
3. What do we do if there is more than one change point?

Predicting the Occurrence of Large Values

Suppose we have observed one observation a day for n days $Y_1, \ldots, Y_n$ (e.g. the number of terrorist attacks per day, or the amount of financial loss, etc)

We want to reason about the probability of seeing large values in the future, i.e. we want the predictive distribution $p(\tilde{Y}|Y_1, \ldots, Y_n)$.

As we have seen before, we get this by using the theorem of total probability and integrating over the posterior distribution of any parameters we don't know. i.e. if θ is a vector of unknown parameters then:

$$p(\tilde{Y}|Y_1, \ldots, Y_n) = \int p(\tilde{Y}|\theta)p(\theta|Y_1, \ldots, Y_n)d\theta$$

Predicting the Occurrence of Large Values

Lets refresh our memory about how to do this in the Exponential case where there is **no change point**. We have:

$$Y_i \sim Exponential(\lambda), \quad \text{for all } i$$

$$\lambda \sim Gamma(\alpha, \beta)$$

We learn λ based on $Y_1, \ldots, Y_n$ and our prior knowledge (represented as a Gamma prior) Recall that the posterior distribution $p(\lambda | Y_1, \ldots, Y_n)$ is:

$$p(\lambda | Y_1, \ldots, Y_n) = Gamma(\alpha + n, \beta + \sum_{i=1}^{n} Y_i) = Gamma(\tilde{\alpha}, \tilde{\beta})$$

Predicting the Occurrence of Large Values

$$p(\tilde{Y}|Y_1,\ldots,Y_n) = \int p(\tilde{Y}|\theta)p(\theta|Y_1,\ldots,Y_n)d\theta =$$

$$= \int \lambda e^{-\lambda\tilde{Y}} \frac{\tilde{\beta}^{\tilde{\alpha}}}{\Gamma(\tilde{\alpha})}\lambda^{\tilde{\alpha}-1}e^{-\tilde{\beta}\lambda}d\lambda =$$

$$= \frac{\tilde{\beta}^{\tilde{\alpha}}}{\Gamma(\tilde{\alpha})}\frac{\Gamma(\tilde{\alpha}+1)}{(\tilde{\beta}+\tilde{Y})^{\tilde{\alpha}+1}} = \quad (\text{using } \Gamma(z+1)/\Gamma(z) = z)$$

$$= {\color{red}\tilde{\alpha}}\frac{\tilde{\beta}^{\tilde{\alpha}}}{(\tilde{\beta}+\tilde{Y})^{\tilde{\alpha}+1}} = \frac{\tilde{\alpha}}{\tilde{\beta}}\left(\frac{\tilde{\beta}+\tilde{Y}}{\tilde{\beta}}\right)^{-\tilde{\alpha}-1} =$$

$$= \frac{\tilde{\alpha}}{\tilde{\beta}}\left(1+\frac{\tilde{Y}}{\tilde{\beta}}\right)^{-\tilde{\alpha}-1}$$

Predicting the Occurrence of Large Values

if we want to find $p(\tilde{Y} > D|Y_1, \ldots, Y_n)$ then we do this by:

$$p(\tilde{Y} > D|Y_1, \ldots, Y_n) = \int_D^\infty p(\tilde{Y}|Y_1, \ldots, Y_n) d\tilde{Y}$$

which in the Exponential case is:

$$p(\tilde{Y} > D|Y_1, \ldots, Y_n) = \int_D^\infty \frac{\tilde{\alpha}}{\tilde{\beta}} \left(1 + \frac{\tilde{Y}}{\tilde{\beta}}\right)^{-\tilde{\alpha}-1} d\tilde{Y}$$

In cases where we can't do this integral by hand, numerical integration (e.g. Simpson's Rule or quadratures) can be used.

Predicting the Occurrence of Large Values

Now lets go back to the case where there is a change point. Suppose we make the following assumptions:

- We know the true value of τ
- No more change points will occur in future

Then:

$$p(\tilde{Y}|Y_1, \ldots, Y_n, \tau) = p(\tilde{Y}|Y_{\tau+1}, \ldots, Y_n)$$

i.e we simply ignore the observations before the change point since they are no longer relevant. – things have changed, so only the observations in the new segment are relevant for predicting the future

Predicting the Occurrence of Large Values

In the Exponential case we hence have:

$$p(\tilde{Y}|Y_1, \ldots, Y_n, \tau) = \frac{\tilde{\alpha}}{\tilde{\beta}}\left(1 + \frac{\tilde{Y}}{\tilde{\beta}}\right)^{-\tilde{\alpha}-1}$$

where:

$$\tilde{\alpha} = \alpha + n - \tau$$

$$\tilde{\beta} = \beta + \sum_{i=\tau+1}^{n} Y_i$$

Predicting the Occurrence of Large Values

Of course, in practice we do not know where τ is. But we saw earlier how to compute its posterior distribution $p(\tau|Y_1, \ldots, Y_n)$.

So, we just do what we always do: average over the posterior:

$$p(\tilde{Y}|Y_1, \ldots, Y_n) = \int p(\tilde{Y}|Y_1, \ldots, Y_n, , \tau)p(\tau|Y_1, \ldots, Y_n)d\tau$$

Since τ can only take finitely many values, this can be written as:

$$p(\tilde{Y}|Y_1, \ldots, Y_n) = \sum_{\tau=1}^{n-1} p(\tilde{Y}|Y_1, \ldots, Y_n, \tau)p(\tau|Y_1, \ldots, Y_n) =$$

$$= \sum_{\tau=1}^{n-1} p(\tilde{Y}|Y_{\tau+1}, \ldots, Y_n)p(\tau|Y_1, \ldots, Y_n)$$

Example

Suppose we have a sequence of 10 observations $Y_1, \ldots, Y_{10}$ from the Exponential distribution, with a single change point

Suppose that when we compute $p(\tau|Y_1, \ldots, Y_{10})$ above we find:

$$p(\tau = 6|Y_1, \ldots, Y_{10}) = 0.1$$

$$p(\tau = 7|Y_1, \ldots, Y_{10}) = 0.7$$

$$p(\tau = 8|Y_1, \ldots, Y_{10}) = 0.2$$

and the probability that $p(\tau = k)|Y_1, \ldots, Y_{10})$ for any other value of k is 0.

If we then want to find the predictive distribution for the next unseen observation (Y_{11}) we have

Example

$$p(\tilde{Y}|Y_1,\ldots,Y_n) = \sum_{\tau=1}^{n-1} p(\tilde{Y}|Y_{\tau+1},\ldots,Y_n,\tau)p(\tau|Y_1,\ldots,Y_n) =$$

$$= 0.1 \times p(\tilde{Y}|Y_7,Y_8,Y_9,Y_{10}) + 0.7 \times p(\tilde{Y}|Y_8,Y_9,Y_{10}) + 0.2 \times p(\tilde{Y}|Y_9,Y_{10})$$

where

$$p(\tilde{Y}|Y_j,\ldots,Y_n) = \frac{\tilde{\alpha}}{\tilde{\beta}}\left(1+\frac{\tilde{Y}}{\tilde{\beta}}\right)^{-\tilde{\alpha}-1}$$

with

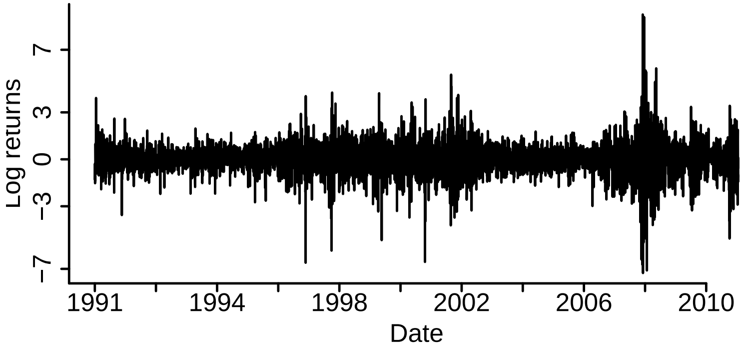
$$\tilde{\alpha} = \alpha + 10 - j + 1, \quad \tilde{\beta} = \beta + \sum_{i=j}^{10} Y_i$$

Multiple Change Points

So far, we have considered only the case where the data contains at most a single change point.

This is the simplest version of the change point problem. However in practice the series may contain multiple change points
Remember the Dow Jones data:

Dow Jones



Multiple Change Points

Suppose there are k change points dividing the series into $k+1$ segments. Let θ_j denote the (unknown) value of θ in the j^{th} segment. Se the change point model is:

$$Y_i = \begin{cases} p(Y_i|\theta_1) & \text{if } i \le \tau_1 \\ p(Y_i|\theta_2) & \text{if } \tau_1 < i \le \tau_2 \\ p(Y_i|\theta_3) & \text{if } \tau_2 < i \le \tau_3 \\ \dots & \\ p(Y_i|\theta_{k+1}) & \text{if } \tau_k < i \le n \end{cases}$$

We need to estimate the change point locations $\tau_1, \dots, \tau_k$.

Conceptually, everything we here in this case is **exactly the same** as in the single change point case

Multiple Change Points

We estimate the change point locations using Bayes Theorem as always:

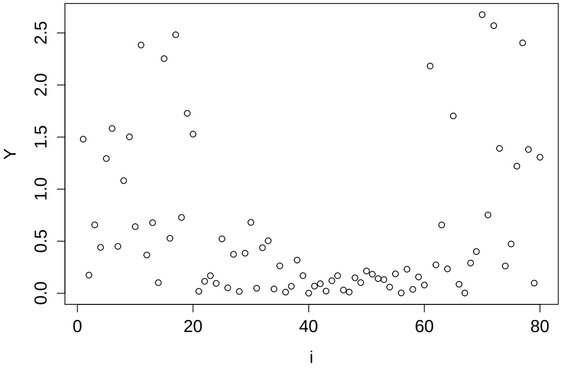
$$p(\tau_1, \ldots, \tau_k | Y_1, \ldots, Y_n) = \frac{p(Y_1, \ldots, Y_n | \tau_1, \ldots, \tau_k) p(\tau_1, \ldots, \tau_k)}{p(Y_1, \ldots, Y_n)}$$

Multiple Change Points - Exponential Distribution

So for example in the case where the observations have an Exponential distribution:

$$Y_i = \begin{cases} Exponential(\lambda_1) & \text{if } i \leq \tau_1 \\ Exponential(\lambda_2) & \text{if } \tau_1 < i \leq \tau_2 \\ Exponential(\lambda_3) & \text{if } \tau_2 < i \leq \tau_3 \\ \ldots & \\ Exponential(\lambda_{k+1}) & \text{if } \tau_k < i \leq n \end{cases}$$

Multiple Change Points - Exponential Distribtuion



Multiple Change Points - Prior

As in the single change point case, we usually assume the change points are uniformly distributed over $1, 2, \ldots, n$.

Since the prior distribution is uniform, this means that it does not depend on the $\tau_1, \ldots, \tau_k$ values (as before

As such, it can be ignored when it comes to computing the posterior since it will be rolled into the normalising constant when we normalise the posterior

Multiple Change Points - Likelihood

Recall that in the single change point case if the λ parameters were known, the likelihood was:

$$p(Y|\tau) = \prod_{i=1}^{\tau} p(Y_i|\theta_1) \prod_{i=\tau+1}^{n} p(Y_i|\theta_2)$$

In the multiple change point case this becomes:

$$p(Y|\tau_1, \ldots, \tau_k) = \prod_{i=1}^{\tau_1} p(Y_i|\theta_1) \prod_{i=\tau_1+1}^{\tau_2} p(Y_i|\theta_2) \prod_{i=\tau_2+1}^{\tau_3} p(Y_i|\theta_3) \ldots \prod_{i=\tau_k+1}^{n} p(Y_i|\theta_{k+1})$$

i.e. we are still breaking the likelihood up into segments as before, we just have $k+1$ segments now rather than two

Multiple Change Points - Likelihood

Similarly in the single change point case when the parameters θ were unknown we have to integrate over them:

$$p(Y|\tau) = \int \prod_{i=1}^{\tau} p(Y_i|\theta_1)p(\theta_1)d\theta_1 \int \prod_{i=\tau+1}^{n} p(Y_i|\theta_2)p(\theta_2)d\theta_2$$

In the k change point case this becomes:

$$p(Y|\tau_1,\ldots,\tau_k) = \left(\int \prod_{i=1}^{\tau_1} p(Y_i|\theta_1)p(\theta_1)d\theta_1\right) \times \left(\int \prod_{i=\tau_1+1}^{\tau_2} p(Y_i|\theta_2)p(\theta_2)\right) \times$$

$$\times \left(\int \prod_{i=\tau_2+1}^{\tau_3} p(Y_i|\theta_3)p(\theta_3)d\theta_3\right) \times \ldots \times \left(\int \prod_{i=\tau_k+1}^{n} p(Y_i|\theta_{k+1})p(\theta_{k+1})d\theta_{k+1}\right)$$

Multiple Change Points - Example

Recall that with the Exponential sequence in the single change point case when we did these integrals, the likelihood of the observations to the left of the change point was

$$p(Y_1, \ldots, Y_\tau | \tau) = \frac{\beta^\alpha}{\Gamma(\alpha)} \frac{\Gamma(\alpha + \tau)}{(\beta + S_1)^{\alpha + \tau}}, \quad S_1 = \sum_{i=1}^{\tau} Y_i$$

In the multiple change point case, an identical argument can be used to do the integrals above. If we focus on (e.g.) the second segment (between τ_1 and τ_2) we get:

$$p(Y_{\tau_1+1}, \ldots, Y_{\tau_2} | \tau) = \frac{\beta^\alpha}{\Gamma(\alpha)} \frac{\Gamma(\alpha + \tau_2 - \tau_1)}{(\beta + S_2)^{\alpha + \tau_2 - \tau_1}}, \quad S_2 = \sum_{i=\tau_1+1}^{\tau_2} Y_i$$

Multiple Change Points - Example

So we have:

$$p(Y_1, \ldots, Y_n | \tau_1, \ldots, \tau_k) = \left[\frac{\beta^\alpha}{\Gamma(\alpha)}\right]^{k+1} \times \frac{\Gamma(\alpha + \tau_1)}{(\beta + S_1)^{\alpha+\tau_1}} \times \frac{\Gamma(\alpha + \tau_2 - \tau_1)}{(\beta + S_2)^{\alpha+\tau_2-\tau_1}}, \times$$

$$\times \frac{\Gamma(\alpha + \tau_3 - \tau_2)}{(\beta + S_3)^{\alpha+\tau_3-\tau_2}} \times \ldots \times \frac{\Gamma(\alpha + n - \tau_k)}{(\beta + S_{k+1})^{\alpha+n-\tau_k}},$$

Multiple Change Points - Example

Remember the posterior is:

$$p(\tau_1, \ldots, \tau_k | Y_1, \ldots, Y_n) = \frac{p(Y_1, \ldots, Y_n | \tau_1, \ldots, \tau_k) p(\tau_1, \ldots, \tau_k)}{p(Y_1, \ldots, Y_n)}$$

We can ignore the $p(Y)$ and $p(\tau_1, \ldots, \tau_k)$ terms since they don't depend on the $\tau_i' s$ (recall the latter is constant since the prior was uniform)

So, the unnormalised posterior is equal to the likelihood on the previous slide

Multiple Change Points

As before, we normalise the posterior by evaluating it at each possibly combination of the change points, and dividing through by the sum