May 22, 2009

Bayesian Changepoint Model With Gamma Prior

Tristan Fletcher

www.cs.ucl.ac.uk/staff/T.Fletcher/

1 Model Specification

Changepoint models assume one or more parameters defining the distribution of one or more observable variables are constant for a run of length r_t at time t but change when there is a changepoint and a new run starts, i.e. $r_t = 0$.

The aim in this example is to predict at time t the next observation x_{t+1} given the vector of all previous observations $\mathbf{x}_{1:t}$. In this example, we will assume that the observations are Gaussian distributed with mean μ and variance σ^2, i.e. $P(x_t) \sim N(\mu, \sigma^2)$ and that it is the variance σ^2 which changes at the changepoints.

We will also assume that the probability distribution of a run being of length r_t at time t given its length the previous time step will be defined as follows:

$$P(r_t|r_{t-1}) = \begin{cases} P_{cp} & \text{if } r_t = 0 \\ 1 - P_{cp} & \text{if } r_t = r_{t-1} + 1 \\ 0 & \text{otherwise} \end{cases} \tag{1.1}$$

where in general P_{cp} will be small so that $P(r_t = 0|r_{t-1})$, i.e. the probability of a change point occuring at time t, will be low.

2 Inference

If we assume that we can compute the the predictive distribution conditional on a given run length r_t, i.e. that we can calculate $P(x_{t+1}|r_t, \mathbf{x}_t^r)$, we can make a prediction about the next observation $P(x_{t+1}|\mathbf{x}_{1:t})$. We do this by summing over r_t the product of the predictive distribution and the posterior distribution of r_t given the observations so far observed:

$$P(x_{t+1}|\mathbf{x}_{1:t}) = \sum_{r_t} P(x_{t+1}|r_t, \mathbf{x}_t^r) P(r_t|\mathbf{x}_{1:t}) \tag{2.1}$$

where the posterior can be expressed as the joint probability distribution of the run length r_t and all the observations so far $\mathbf{x}_{1:t}$ divided by the probability of observing all these observations:

$$\begin{aligned} P(r_t|\mathbf{x}_{1:t}) &= \frac{P(r_t, \mathbf{x}_{1:t})}{P(\mathbf{x}_{1:t})} \\ &= \frac{P(r_t, \mathbf{x}_{1:t})}{\sum_{r_t} P(r_t, \mathbf{x}_{1:t})} \end{aligned} \tag{2.2}$$

This joint distribution can be expressed recursively:

$$
\begin{aligned}
P(r_t, \mathbf{x}_{1:t}) &= \sum_{r_{t-1}} P(r_t, r_{t-1}, \mathbf{x}_{1:t}) \\
&= \sum_{r_{t-1}} P(r_t, r_{t-1}, \mathbf{x}_{1:t-1}, x_t) \\
&= \sum_{r_{t-1}} P(r_t|x_t, r_{t-1}, \mathbf{x}_{1:t-1}) P(x_t|r_{t-1}, \mathbf{x}_{1:t-1}) P(r_{t-1}, \mathbf{x}_{1:t-1}) \\
P(x_t|r_{t-1}, \mathbf{x}_{1:t-1}) \equiv P(x_t|r_{t-1}, \mathbf{x}_{t-1}^{r_{t-1}}) \Rightarrow &\sum_{r_{t-1}} P(r_t|r_{t-1}) P(x_t|r_{t-1}, \mathbf{x}_{t-1}^{r_{t-1}}) P(r_{t-1}, \mathbf{x}_{1:t-1})
\end{aligned}
\tag{2.3}
$$

where $P(r_t|r_{t-1})$ is defined in (1.1), $P(r_{t-1}, \mathbf{x}_{1:t-1})$ is the result of (2.3) from the previous recursion and $P(x_t|r_{t-1}, \mathbf{x}_{t-1}^{r_{t-1}})$ is the predictive distribution conditional on a given run length, derived in the next section. $\mathbf{x}_{t-1}^{r_{t-1}}$ is the set of observations of x associated with the previous time step's run r_{t-1} and $\left|\mathbf{x}_{t-1}^{r_{t-1}}\right| = r_{t-1}$. In order to make the maths appear less cluttered, when r_{t-1} appears as a super-script, we will denote it by $r_{t'}$

2.1 Predictive Distribution

We will use τ to refer to the inverse of the observation variance, i.e. the precision. Furthermore, instead of using a point estimate for τ we will assume that it is distributed according to a Gamma distribution so that $\tau \sim Gam(\alpha, \beta)$. We can then express the predictive distribution $P(x_t|r_{t-1}, \mathbf{x}_{t-1}^{r_{t'}})$ as follows:

$$
\begin{aligned}
P(x_t|r_{t-1}, \mathbf{x}_{t-1}^{r_{t'}}) &= \int_0^\infty P(x_t|\tau^{-1}) P(\tau^{-1}|r_{t-1}, \mathbf{x}_{t-1}^{r_{t'}}) d\tau \\
&= \int_0^\infty P(x_t|\tau^{-1}) P(\tau^{-1}|\mathbf{x}_{t-1}^{r_{t'}}) d\tau \\
&= \int_0^\infty P(x_t|\tau^{-1}) \frac{P(\mathbf{x}_{t-1}^{r_{t'}}|\tau^{-1}) P(\tau^{-1})}{P(\mathbf{x}_{t-1}^{r_{t'}})} d\tau \\
&= \frac{1}{K} \int_0^\infty N(x_t|\mu, \tau^{-1}) N(\mathbf{x}_{t-1}^{r_{t'}}|\mu, \tau^{-1}) Gam(\tau|\alpha, \beta) d\tau \\
&= \frac{1}{K} \int_0^\infty N(\mathbf{x}_t^{r_{t'}}|\mu, \tau^{-1}) Gam(\tau|\alpha, \beta) d\tau \text{ where } \left|\mathbf{x}_t^{r_{t'}}\right| = r_{t-1} + 1 \\
&= \frac{1}{K} \int_0^\infty \left(\frac{\tau}{2\pi}\right)^{\frac{r_{t'}+1}{2}} \exp\left\{-\frac{\tau}{2} \sum_{i=0}^{r_{t'}} (x_{t-i} - \mu)^2\right\} \frac{\beta^\alpha \exp^{(-\beta\tau)} \tau^{\alpha-1}}{\Gamma(\alpha)} d\tau \\
&= \frac{1}{K} \frac{\beta^\alpha}{\Gamma(\alpha)} \left(\frac{1}{2\pi}\right)^{\frac{r_{t'}+1}{2}} \int_0^\infty \tau^{\alpha + \frac{r_{t'}}{2} - \frac{1}{2}} \exp\left\{-\tau \left[\sum_{i=0}^{r_{t'}} \frac{(x_{t-i} - \mu)^2}{2} + \beta\right]\right\} d\tau
\end{aligned}
$$

Substituting in $z = \tau\Delta$ where $\Delta = \sum_{i=0}^{r_{t'}} \frac{(x_{t-i} - \mu)^2}{2} + \beta$ so that $dz = d\tau\Delta$:

$$
\begin{aligned}
P(x_t|r_{t-1}, \mathbf{x}_{t-1}^{r_{t'}}) &= \frac{1}{K}\frac{\beta^\alpha}{\Gamma(\alpha)}\left(\frac{1}{2\pi}\right)^{\frac{r_{t'}+1}{2}} \Delta^{-\alpha-\frac{r_{t'}}{2}-\frac{1}{2}} \int_0^\infty z^{\alpha+\frac{r_{t'}}{2}-\frac{1}{2}} \exp(-z)dz \\
&= \frac{1}{K}\frac{\beta^\alpha}{\Gamma(\alpha)}\left(\frac{1}{2\pi}\right)^{\frac{r_{t'}+1}{2}} \Delta^{-\alpha-\frac{r_{t'}}{2}-\frac{1}{2}} \Gamma(\alpha + \frac{r_{t'}+1}{2}) \\
&= \frac{1}{K}\frac{\Gamma(\alpha + \frac{r_{t'}+1}{2})}{\Gamma(\alpha)}\beta^\alpha \left(\frac{1}{2\pi}\right)^{\frac{r_{t'}+1}{2}} \Delta^{-\alpha-\frac{r_{t'}}{2}-\frac{1}{2}}
\end{aligned}
$$

Substituting back in $\Delta = \sum_{i=0}^{r_{t'}} \frac{(x_{t-i} - \mu)^2}{2} + \beta$ and also $\alpha = \frac{\nu}{2}$ and $\beta = \frac{\nu}{2\lambda}$:

$$
\begin{aligned}
P(x_t|r_{t-1}, \mathbf{x}_{t-1}^{r_{t'}}) &= \frac{1}{K}\frac{\Gamma\left(\frac{\nu}{2} + \frac{r_{t'}+1}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)}\left(\frac{\nu}{2\lambda}\right)^{\frac{\nu}{2}}\left(\frac{1}{2\pi}\right)^{\frac{r_{t'}+1}{2}}\left(\frac{\nu}{2\lambda} + \sum_{i=0}^{r_{t'}}\frac{(x_{t-i}-\mu)^2}{2}\right)^{-\frac{\nu}{2}-\frac{r_{t'}+1}{2}} \\
&= \frac{1}{K}\frac{\Gamma\left(\frac{\nu}{2} + \frac{r_{t'}+1}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)}\left(\frac{\nu}{2\lambda}\right)^{\frac{\nu}{2}}\left(\frac{1}{2\pi}\right)^{\frac{r_{t'}+1}{2}}\left[\frac{\nu}{2\lambda}\left(1 + \frac{2\lambda}{\nu}\sum_{i=0}^{r_{t'}}\frac{(x_{t-i}-\mu)^2}{2}\right)\right]^{-\frac{(\nu+r_{t'}+1)}{2}} \\
&= \frac{1}{K}\frac{\Gamma\left(\frac{\nu}{2} + \frac{r_{t'}+1}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)}\left(\frac{1}{2\pi}\right)^{\frac{r_{t'}+1}{2}}\left(\frac{\nu}{2\lambda}\right)^{-\frac{r_{t'}+1}{2}}\left(1 + \frac{\lambda}{\nu}\sum_{i=0}^{r_{t'}}(x_{t-i}-\mu)^2\right)^{-\frac{(\nu+r_{t'}+1)}{2}} \\
&= \frac{1}{K}\frac{\Gamma\left(\frac{\nu}{2} + \frac{r_{t'}+1}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)}\left(\frac{\lambda}{\nu\pi}\right)^{\frac{r_{t'}+1}{2}}\left(1 + \frac{\lambda}{\nu}\sum_{i=0}^{r_{t'}}(x_{t-i}-\mu)^2\right)^{-\frac{(\nu+r_{t'}+1)}{2}}
\end{aligned}
\tag{2.4}
$$

The constant K is the Marginal Likelihood $P(\mathbf{x}_{t-1}^{r_{t'}})$ and is calculated in a very similar way:

$$
\begin{aligned}
P(\mathbf{x}_{t-1}^{r_{t'}}) &= \int_0^\infty P(\mathbf{x}_{t-1}^{r_{t'}}|\tau^{-1})P(\tau^{-1})d\tau \\
&= \int_0^\infty N(\mathbf{x}_{t-1}^{r_{t'}}|\mu, \tau^{-1})Gam(\tau|\alpha, \beta)d\tau \\
&= \int_0^\infty \left(\frac{\tau}{2\pi}\right)^{\frac{r_{t'}}{2}} \exp\left\{-\frac{\tau}{2}\sum_{i=1}^{r_{t'}}(x_{t-i}-\mu)^2\right\}\frac{\beta^\alpha \exp^{(-\beta\tau)}\tau^{\alpha-1}}{\Gamma(\alpha)}d\tau \\
&= \frac{\beta^\alpha}{\Gamma(\alpha)}\left(\frac{1}{2\pi}\right)^{\frac{r_{t'}}{2}}\int_0^\infty \tau^{\alpha+\frac{r_{t'}}{2}-1}\exp\left\{-\tau\left[\sum_{i=1}^{r_{t'}}\frac{(x_{t-i}-\mu)^2}{2} + \beta\right]\right\}d\tau
\end{aligned}
$$

Substituting in $z = \tau\Delta$ where $\Delta = \displaystyle\sum_{i=1}^{r_{t'}} \frac{(x_{t-i} - \mu)^2}{2} + \beta$ so that $dz = d\tau\Delta$:

$$\begin{aligned}
P(\mathbf{x}_{t-1}^{r_{t'}}) &= \frac{\beta^\alpha}{\Gamma(\alpha)} \left(\frac{1}{2\pi}\right)^{\frac{r_{t'}}{2}} \Delta^{-\alpha-\frac{r_{t'}}{2}} \int_0^\infty z^{\alpha+\frac{r_{t'}}{2}-1} \exp(-z) dz \\
&= \frac{\beta^\alpha}{\Gamma(\alpha)} \left(\frac{1}{2\pi}\right)^{\frac{r_{t'}}{2}} \Delta^{-\alpha-\frac{r_{t'}}{2}} \Gamma(\alpha + \frac{r_{t'}}{2}) \\
&= \frac{\Gamma(\alpha + \frac{r_{t'}}{2})}{\Gamma(\alpha)} \beta^\alpha \left(\frac{1}{2\pi}\right)^{\frac{r_{t'}}{2}} \Delta^{-\alpha-\frac{r_{t'}}{2}}
\end{aligned}$$

Substituting back in $\Delta = \displaystyle\sum_{i=1}^{r_{t'}} \frac{(x_{t-i} - \mu)^2}{2} + \beta$ and also $\alpha = \frac{\nu}{2}$ and $\beta = \frac{\nu}{2\lambda}$:

$$\begin{aligned}
P(\mathbf{x}_{t-1}^{r_{t'}}) &= \frac{\Gamma\left(\frac{\nu}{2} + \frac{r_{t'}}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)} \left(\frac{\nu}{2\lambda}\right)^{\frac{\nu}{2}} \left(\frac{1}{2\pi}\right)^{\frac{r_{t'}}{2}} \left(\frac{\nu}{2\lambda} + \sum_{i=1}^{r_{t'}} \frac{(x_{t-i} - \mu)^2}{2}\right)^{-\frac{\nu}{2}-\frac{r_{t'}}{2}} \\
&= \frac{\Gamma\left(\frac{\nu}{2} + \frac{r_{t'}}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)} \left(\frac{\nu}{2\lambda}\right)^{\frac{\nu}{2}} \left(\frac{1}{2\pi}\right)^{\frac{r_{t'}}{2}} \left[\frac{\nu}{2\lambda}\left(1 + \frac{2\lambda}{\nu} \sum_{i=1}^{r_{t'}} \frac{(x_{t-i} - \mu)^2}{2}\right)\right]^{-\frac{(\nu+r_{t'})}{2}} \\
&= \frac{\Gamma\left(\frac{\nu}{2} + \frac{r_{t'}}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)} \left(\frac{1}{2\pi}\right)^{\frac{r_{t'}}{2}} \left(\frac{\nu}{2\lambda}\right)^{-\frac{r_{t'}}{2}} \left(1 + \frac{\lambda}{\nu} \sum_{i=1}^{r_{t'}} (x_{t-i} - \mu)^2\right)^{-\frac{(\nu+r_{t'})}{2}} \\
&= \frac{\Gamma\left(\frac{\nu}{2} + \frac{r_{t'}}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)} \left(\frac{\lambda}{\nu\pi}\right)^{\frac{r_{t'}}{2}} \left(1 + \frac{\lambda}{\nu} \sum_{i=1}^{r_{t'}} (x_{t-i} - \mu)^2\right)^{-\frac{(\nu+r_{t'})}{2}} \qquad (2.5)
\end{aligned}$$

Substituting $K = P(\mathbf{x}_{t-1}^{r_{t'}})$ from (2.5) into (2.4) gives the final expression for $P(x_t | r_{t-1}, \mathbf{x}_{t-1}^{r_{t'}})$:

$$P(x_t | r_{t-1}, \mathbf{x}_{t-1}^{r_{t'}}) = \frac{\Gamma\left(\frac{\nu_n}{2} + \frac{1}{2}\right)}{\Gamma\left(\frac{\nu_n}{2}\right)} \frac{1}{\sqrt{\nu_n \pi \sigma_n^2}} \left[1 + \frac{1}{\nu_n}\left(\frac{x_t - \mu}{\sigma_n}\right)^2\right]^{\frac{\nu_n+1}{2}} \qquad (2.6)$$

where $\nu_n = \nu + r_{t'}$, $\beta_n = \beta + \frac{1}{2}\sum_{i=1}^{r_{t'}} (x_i - \mu)^2$ and $\sigma_n^2 = \frac{2\beta_n}{\nu_n}$. (2.6) is a Student's t-distribution where the paramters ν and σ are functions of the run length r_{t-1} i.e. $x_t \sim St(\nu^{r_{t'}}, \sigma^{r_{t'}})$.

3 Implementation

In order to carry out inference in the changepoint model, the following steps need to be carried out:

1. Assume that we are starting at time t with a zero run length, i.e. that $r_0 = 0$. Choose values for μ^0, ν^0 and λ^0.

2. Make a new observation x_t.

3. Using (2.4), evaluate Predictive Probabilities for all r_t where $0 < r_t < t$:
$P(x_t|r_{t-1}, \mathbf{x}_{t-1}^{r'}) = St(x_t|\nu^{r_{t'}}, \sigma^{r_{t'}})$

4. Using (2.3), calculate Growth Probabilities for all r_t where $0 < r_t < t$:
$P(r_t = r_t + 1, \mathbf{x}_{1:t}) = P(r_{t-1}, \mathbf{x}_{1:t-1})P(x_t|r_{t-1}, \mathbf{x}_{t-1}^{r'})(1 - P_{cp})$

5. Calculate Changepoint Probabilities for all r_t where $0 < r_t < t$:
$P(r_t = 0, \mathbf{x}_{1:t}) = \sum_{r_{t-1}} P(r_{t-1}, \mathbf{x}_{1:t-1})P(x_t|r_{t-1}, \mathbf{x}_{t-1}^{r'})P_{cp}$

6. Calculate Evidence:
$P(\mathbf{x}_{1:t}) = \sum_{r_t} P(r_t, \mathbf{x}_{1:t})$

7. Determine Run Length Distributions for all r_t where $0 < r_t < t$:
$P(r_t|\mathbf{x}_{1:t}) = \frac{P(r_t, \mathbf{x}_{1:t})}{P(\mathbf{x}_{1:t})}$

8. Repeat steps 2 to 7.