BAYESIAN PROCEDURE IN PHENOLOGICAL STUDY USING NON-INFORMATIVEPRIOR UNDER CONSTANT TIME SERIES MODEL

Vijay Kumar Pandey
Department of Statistics, University of Lucknow, Lucknow-226007 (INDIA)
* *vijay.pandey550@gmail.com*

ABSTRACT

The identification of changes in observational data relating to human induced climate change remains a topic of paramount importance. In particular, scientifically sound and rigorous methods for detecting changes are urgently needed. Here we will develop a Bayesian procedure in phenological study using non-informative prior under constant time series model.

Keywords: Bayes Analysis, phenology, constant time series model.

INTRODUCTION

The global average surface temperature has increased over the twentieth century by about 0.6 ± 0.2°C and is projected to continue to rise at a rapid rate. Many studies have been done of ecological impacts of this recent climate change. Penology is perhaps the simplest and most frequently used bio-indicator to track climate changes.

Bayesian statistical methods have been applied so far in climate change detection, analysis and attribution (e.g. Hobbs,1997; Hasselman, 1998; Leroy, 1998; Tol and De Vos, 1998; Barnett, 1999; Katz, 2002 and V. Dose and A.Menzel,2004).

In this paper we will focus to develop a probability model by using Bayesian concept in phenological study under time series model. Here we will use non-informative prior to develop the model. In the next section, we shortly introduced the Bayesian concepts.

Bayesian Procedure: Here we will introduce the Bayesian procedure and terminology which is necessary to develop the probability model for phonological studies. Bayesian probability theory is based on the application of two rules. The first is the conventional product rule for manipulating conditional probabilities . It allows a probability density function to be broken down depending on two (or more) variables $P\left(\vec{\theta}, \vec{d}/M, I\right)$ conditional on the model M that specifies the meaning of the parameter $\vec{\theta}$ and additional information I into simplified form as,

$$P\left(\vec{\theta}, \vec{d}/M, I\right) = P\left(\vec{\theta}/M, I\right) * P\left(\vec{d}/\vec{\theta}, M, I\right) \quad (1)$$

Where $P(\vec{\theta}/M, I)$ and $P(\vec{\theta}, \vec{d}/M, I)$ depends only on the single (vector)-variables $\vec{\theta}$ and $\vec{d}$ respectively.

Equation (1) may be expanded in an alternative way due to symmetry in the variable$\vec{\theta}$,$\vec{d}$ –

$$P\left(\vec{\theta},\vec{d}/M,I\right)=P\left(\vec{d}/,M,I\right)*P\left(\vec{\theta}/\vec{d},M,I\right) \quad (2)$$

Equating the right hand sides of equation (1) and (2), we find Bayes, theorem. that is,

$$P(\vec{\theta}/\vec{d},M) = \frac{P(\vec{\theta}/M,I)*P(\vec{d}/\vec{\theta},M,I)}{P(\vec{\theta}/M,I)} \quad (3)$$

The function on the left hand side is called the posterior density of the parameters $\vec{\theta}$ given data $\vec{d}$ and model M. It is equal to the prior density of the parameters$\vec{\theta}$, $P(\vec{\theta}/M,I)$ which encodes our information on $\vec{\theta}$, prior to considering the data $\vec{d}$ times the likelihood.

$P(\vec{d}/M,I)$is formly the normalisation for the posterior density,

$$P(\vec{d}/M,I) = \int d\vec{\theta}P(\vec{\theta}/M,I) *P(\vec{d}/\vec{\theta},M,I) \quad (4)$$

By inverse application of the product rule we arrive at the Bayesian marginalisation rule, which completes Bayes theory and has no counterpart in traditional statistics,

$$P(\vec{d}/M,I) = \int d\vec{\theta}P(\vec{\theta}/M,I)*P(\vec{d}/\vec{\theta},M,I) \quad (5)$$

Equation (5) allows for us an important interpretation. It is obviously the likelihood of the data $\vec{d}$ given the model M regardless of the numerical values of the parameter $\vec{\theta}$. Employing Bayes theorem to invert (5) we obtain-

$$P(\vec{d}/M,I) = \int d\vec{\theta}P(\vec{d},\vec{\theta}/M,I) \quad (6)$$

Equation (6) is then the probability of a model M out of a possible variety given the data $\vec{d}$.

Having identified the appropriate model to explain the data we are left with the determination of the parameters, which specify the model. The full information on the parameters, is of course contained in the posterior distribution (3).It is sufficiently simple meaning that $P(\theta/D,I)$ resembles a Gaussian function, then it may be summarized in terms of mean and variance,

$$\langle\theta\rangle = \int \theta P(\theta/D,I)d\theta)$$

$$\langle\Delta\theta^2\rangle = \int(\theta-\langle\theta\rangle)2P(\theta/D,I)d\theta \quad (7)$$

This completes a Bayesian analysis if the problem is model selection and best estimate of the parameters specify the model.

Model: The likelihood function for this model must incorporate the data$\vec{d}$, the years $\vec{x}$, the scatter of the data will be characterized by a variable σ and the constant f that we choose to define the 'no trend' on the data.

The model becomes,

$d_i - f_i = \in_i \quad \forall i$ (8)

Where $\in_i$ is i.i.d. and follow normal distribution with mean zero and variance σ^2 . Hence

$$P(\vec{d}/\vec{x},\sigma,I) = \left(\frac{1}{\sigma\sqrt{2\pi}}\right) N \exp\left[-\frac{1}{2\sigma^2}\sum_{i=1}^{N}(d_i - f)^2\right] \quad (9)$$

From the equation (9), we must now calculate the evidence P $(\vec{d}/\vec{x},c,I)$, where C denotes that the constant model. From the marginalization theorem,(4):

$$P(\vec{d}/\vec{x},c,I) = \int P(\vec{d},f,\sigma/\vec{x},c,I)\, df\, d\sigma \quad (10)$$

$$= \int P(f,\sigma/\vec{x},c,I)\; P(\vec{d}/\vec{x},f,\sigma,c,I)\, df\, d\sigma \quad (11)$$

The first distribution under the integral in(11) is logically independent of $\vec{x}$and c simplifies to,

$$P(f,\sigma/I) = P(\sigma/I) \quad (12)$$

The prior distribution P(f/I) on f is chosen(weekly informative) to be constant over the range 2γ,

$$P(f/I) = \frac{1}{2\gamma} (13)$$

The range γcan be estimated from the variance of the data.

Similar choice is made for P(σ/I). We choose a normalised form of Jeffrey's prior:

$$P(\sigma/\beta, I) = \frac{1}{2\ln\beta}\frac{1}{\sigma} \quad ; \frac{1}{\beta} < \sigma < \beta (14)$$

This leads to the marginal likelihood from equation (11),

$$P(\vec{d}/\vec{x},c,I) = \left(\frac{1}{2\pi}\right) N/2\, \frac{1}{2\gamma}\frac{1}{2\ln\beta}\int \frac{d\sigma}{\sigma}\frac{1}{\sigma^N}\exp\left[-\frac{1}{2\sigma^2}\sum_{i=1}^{N}(d_i - f)^2\right] df$$

Now, $\sum_{i=1}^{N}(d_i - f)^2 = \sum_{i=1}^{N}(d_i - \bar{d} + \bar{d} - f)^2$

$$= N\,(f - \bar{d})2 + N\overline{\Delta d}2$$

Where, $\bar{d} = \frac{1}{N}\sum_{i=1}^{N} d\mathrm{i}$ and $\overline{\Delta d}2 = \frac{1}{N}\sum_{i=1}^{n}(d\mathrm{i} - \bar{d})2$.

Therefore,

$$P(\vec{d}/\vec{x},c,I) = \left(\frac{1}{2\pi}\right) N/2\, \frac{1}{2\gamma}\frac{1}{2\ln\beta}\int \frac{1}{\sigma}\frac{1}{\sigma^N}\exp\left(\frac{-N\overline{\Delta d}^2}{2\sigma^2}\right) d\sigma \int_{-\infty}^{\infty} exp\left\{\frac{N}{2\sigma^2}(f - \bar{d})^2\right\} df \quad (15)$$

Now $\int_{-\infty}^{\infty} exp\left\{-\frac{N}{2\sigma^2}(f - \bar{d})^2\right\} df = \sigma\sqrt{\frac{2\pi}{N}}$

Hence, equation (15) can be written as,

$$P(\vec{d}/\vec{x},c,I) = \sqrt{\frac{2\pi}{N}}\left(\frac{1}{2\pi}\right)^{N/2}\frac{1}{2\gamma}\frac{1}{2ln\beta}\int_0^\infty \frac{1}{\sigma}\frac{1}{\sigma^{N-1}}\exp\left(\frac{-N\overline{\Delta d}^2}{2\sigma^2}\right)d\sigma.$$

Now by taking substitution $x = \frac{1}{\sigma^2}$,we find

$$\int_0^\infty \frac{1}{\sigma}\frac{1}{\sigma^{N-1}}\exp\left(\frac{-N\overline{\Delta d}^2}{2\sigma^2}\right)d\sigma = \frac{1}{2}\frac{\Gamma\frac{N-1}{2}}{(N\overline{\Delta d}^2)^{\frac{N-1}{2}}} \qquad (16)$$

Collecting terms the evidence for the constant model becomes,

$$P(\vec{d}/\vec{x},c,I) = \frac{1}{2}\left(\frac{1}{\pi}\right)^{\frac{N-1}{2}}\frac{1}{2\gamma}\frac{1}{2\,ln\beta}\frac{\Gamma\frac{N-1}{2}}{(N\overline{\Delta d}^2)^{\frac{N-1}{2}}}\frac{1}{\sqrt{N}} \qquad (17)$$

The residual sum of square of the model (17) is given by the expression,

$R = N\overline{\Delta d}^2$

Where $\overline{\Delta d}^2 = \frac{1}{N}\sum_{i=1}^{N}(d_i - \bar{d})^2$

$= \frac{1}{N}\sum_{i=1}^{N} {d_i}^2 - \bar{d}^2$

For an illustration, we have the mean temperature data of Faizabad district in ℃ from the year 1990-1991 to 1999-2000 in which,

$\sum d = 254.55$

$\bar{d} = 25.45$

$\bar{d}^2 = 647.98$

$\sum {d_i}^2 = 6484.96$

$\overline{\Delta d}^2 = \frac{1}{N}\sum_{i=1}^{N} {d_i}^2 - \bar{d}^2$

$= 648.96 - 647.98$

$= 0.5\ 16$

Therefore,

Residual sum of the model (17) is

$R = N\overline{\Delta d}^2$

$= 10 * 0.516$

$= 5.16$

ISSN 2229-5518

ACKNOWLEDGEMENT

We thank to Professor Rohit Pandey of Narendra Dev Agriculture University, Kumarganj, Faizabad, Uttar Pradesh, India to provide such type of data.

CONCLUSION

From the above discussion we can say that, by using the model (17) we can describe the Time Series Data and assessment of their functional behaviour. We can measure the rates of change with uncertainty margins as well as evolution of independent treatment of Time Series triggering parameters and affected systems.

REFFERENCES

[1] Barnett, V.(1982): Comparative Statistical Inference
[2] Berger, J. O.(1985): Statistical Decision Theory And Bayesian Analysis.
[3] Box, G.E.P. and Tiao, G.C. (1973): Bayesian Inference in Statistical Analysis.
[4] Broemeling,L.D.(1985): Bayesian Analysis of Linear Models
[5] Lindley,D.V.(1965): Introduction to Probability And Statistics From a Bayesian Point of View