Special Issue of Engineering and Scientific International Journal ISSN 2394-7187(Online) ISSN 2394-7179 (Print)
*Proceedings of the UGC Sponsored - National Conference on "**Recent Statistical Computing Techniques and their Applications**"*
Department of Statistics and Department of Computer Science, Ramanujan College,University of Delhi, Delhi.
March 11-12, 2016. ***Published on January 2017***

A Comparative Study of Bayesian Procedure in Phenology using Non-Informative and Informative Prior under Constant Time Series Model

Vijay Kumar Pandey

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

Abstract— In this paper we will compare a Bayesian procedure in phenological study using non-informative and informative prior under constant time series model using concept of Bayes' factor.

Keywords— Bayesian Analysis, Phenology, Constant Time Series Model, Bayes' factor.

1. Introduction

The global average surface temperature has increased over the 20th century by about 0.6 ± 0.2℃ and is projected to continue to rise at a rapid rate. Many studies have been done in evidence of ecological impact of the recent climate change. Menzel And Estrella (2001), Sparks And Menzel (2002), Walther And Root (2003) have been reported shifts in plant and animal phenology for the boreal and temperature zones of the northern hemisphere. According to Schnelle (1995) And Schwartz (2003), phenology is a field of research which studies the annual rhythm of biological phenomena mainly related to climate. Centuries ago, especially for agricultural purpose, phenological knowledge improved the understanding of the variation in life cycle events.

Methodologically, trends in Time- Series are often analysed using simple linear regression where phonological data or temperatures are plotted against time for phenology. This method has been developed by Menzel And Fabian (1999), Walrher (2002), Parmesan And Yohe (2003) And Menzel (2006) for the study of phenological events. The slope of the regression equation indicates the average rate of change in phenology (days/year) or temperature (℃/year). This method can easily be applied to a large number of sites.

The main disadvantages of this least square method are their restriction to Time Series showing more or less linear performance, possibly poor extrapolation properties and sensitivity to outlier or extremes and boundary values. For inherently non-linear process it becomes difficult to find a linear model that fits the data as the range of the data increases. Therefore, we can say that while the least square method provides optimal estimates of unknown parameter, it is sensitive to the presence of outliers in the data used to fit a statistical model In consequence**,** Dose And Menzel (2004), introduced a Bayesian Approach of model comparison to evaluate the fit of a constant, a linear regression, and a change point model on Time Series. The Bayesian analysis has the great advantage of analysing varying changes, model probabilities and change point probabilities of Time Series. Menzel And Rutishauser (2008), Dose And Menzel (2006) has proposed the concept of correlation regression analysis for the study of environmental on plant phenology. In this chapter analogous to the procedure of Dose And Menzel (2004), we have used the Bayesian approach to describe long-term phenological and temperature Time Series with Constant Time Series Model. In Constant Time Series Model, there is no rate of change and represent just the mean value of the data.It has also been observed that the use of informative prior gives more consistent results than non-informative prior for the same time series dada in phenological study.

2. Bayesian Analysis in Time Series Modelling

The goals of Time Series Models include irregular series, forecasting series into the medium orlong-term future and casual modelling of variables moving in parallel through time. Dependency through time is the basis for extrapolation into the future, for example via autoregression of a metric variable y_t on previous values of the series y_{t-k} (k=1,2,.......) are based on known future values of predictor values x_t. Another goal of Time Series analysis is detecting changes in structure in the series possibly as a result of an 'intervention' such as economic policy, pollution incident or medical treatment. Gordan And Smith (1990), Wang, Zivot And Martin (2000) outlines Bayesian approaches to structural shifts in biochemical, interest rate and spending Time Series, respectively. Recently, much development, especially, perspective, has occurred in discrete Time Series Cargni Et Al. (1997), Czado And Muller (2004), Czado And Song (2001), State-Space Models Bass Et Al. (1997), Godsill Et Al. (2004), Multivariate Time Series Brandt And Freeman (2006), Waggoner And Zha (1999) And Model Selection Chen(1999); Koop And Potter (1999), Vermaak Et Al.(2004).

Stochastic dependence in consecutive observations themselves is widely observed COX et al. (1996), and observation-driven models are the most commonly used for long-term forecasting. For example, Helfenstein (1991) cites time dependencies in environmental medicine, while Time Series of economic indicators such as prices and output levels also usually show autocorrelation over time. Another sort of dependency takes the form of regular seasonal or cyclic fluctuations, as in many climatic or biomedical series.

A major class of models for stationary Time Series data is the autoregressive integrated moving average models of Box And Jenkins (1970), where stationary is based on removing trends, cyclical or seasonal regularities. Discrete data Time Series Models may also try to replicate features of Box-Jenkins metric data models, as in integer-valued autoregressive models (Mc Cab And Martin (2005).

However, many observed series exhibit clear upward or downward trends, and require transformation or differencing to achieve stationarity, so that the stationarity and non- stationarity are assessed as alternative models for the data series according to Berger And Yang (1994), Naylor And Marriott(1996). The alternative structural model approach focuses on the observed components of series, such as trends, seasonal cycles or changing impacts of predictors. Thus, a typical time series may consist of up to four components:

y_t = *Trend* + *Seasonal effects* + *Regression Term* + *Irregular Effects*

One option for modelling these effects is by a set of fixed constants, e.g a polynomial in time to describe the trend in the level of the series, and seasonal dummies to represent seasonal factors. Such a model places equal weight on all observations when predicting the future. A more flexible approach is provided by structural time series models that allow time varying coefficients such that forecasts place more weights on recent observations Harvey (1989). The closely related Bayesian methodology for state-space time series modelling has been denoted dynamic linear modelling West And Harrison (1997), through such approaches readily extend to nonlinear and non-Gaussian data Carlin Et.Al. (1992), Tanizaki And Mariano (1998).

The univariate methods may be extended to modelling multivariate dependence through time. For example, autoregressive observational dependence would mean each series depending both on its own past and on the past values of one or more of the other series Sims (1980), Sims And Zha (1998), and often extends to panel data. One advantage of simultaneously modelling several series is the possibility of pooling information to improve precision and out-of-sample forecasts.The Bayesian Analysis concept in Time Series Models has been studied for climate change prediction in phenology**.** Dose And Menzel (2004) proposed a Bayesian method to investigate phenological changes, and these have been helpful in detecting different parts of a single time series that show different patterns. Climate change detection employing non-parametric Bayesian function estimation is especially useful for studies of climate change impacts in natural systems where conditions are prescribed to change.

3. Informative Prior in Bayesian Methods

An informative prior expresses specific, definite information about a variable. An example of a prior distribution is the temperature at noon tomorrow. A reasonable approach is to make the prior a normal temperature with expected value equal to today's noontime temperature with variance equal to the day to day variance of atmospheric temperature or a distribution of the temperature for that of the year. According to Bayes' theorem, we have $\pi\ (\theta/x)\ \ \alpha f\ (x/\theta)\pi\ (\theta)$

Here $\pi(\theta/x)$is called the posterior distribution of θ, which is used for doing inference, $f(x/\theta)$ is called the likelihood function and $\pi(\theta)$ is called the prior distribution ofθ. If $\pi(\theta)$ that is the prior distribution of θ consists the complete and valid information about the parameter θ then it is called the informative prior distribution of θ**.** Sometimes it may not be possible to obtain $\pi(\theta/x)$in a closed form. To avoid problem of integration statisticians use conjugate prior distributions. Let X has the likelihood function$f(x/\theta)$ and $\pi(\theta)$ be the prior distribution of θ, then $\pi(\theta)$ is said to be a conjugate prior family if the corresponding posterior distribution $\pi(\theta/x)$ belongs to same family as $\pi(\theta)$.

For example let X~b (n, θ), 0 <θ< 1 and $\pi(\theta)$ be a beta distribution with parameters (α, β), then,

$$g\ (\theta/x) = \frac{\theta^{x+\alpha-1}(1-\theta)^{\beta+n+x-1}}{\int_0^1 \theta^{x+\alpha-1}(1-\theta)^{\beta+n+x-1}} d\theta$$

$$= \frac{\theta^{\alpha+\alpha-1}(1-\theta)^{\beta+n+x-1}}{\boldsymbol{B}(x+\alpha,\beta+n+x)} \quad ; \quad 0<\theta<1$$

which is also a beta distribution.

Thus the family of beta distribution is a conjugate family of priors for θ**.** Raiffa And Schlaifer (1961) noted the following conditions for choosing a class G of priors.

- Posterior distribution should be easy to determine.
- It should be possible to find expectation of some 'utility functions' with respect to members in G.
- G should be analytically tractable in three respects:
  - G should be closed in the sense that if a prior belongs to G, the posterior should also belong to G.
  - G should be rich, i.e., there should be exist a member of G capable of expressing decision maker's belief and prior information.
  - G should be parameterisable in a manner which can be readily interpreted, so that it will be easy to verify that the chosen members of the decision maker's prior judgements about Θ.

Suppose that the likelihood function f (x/θ) can be written as, f (x/θ) = K(x/θ) ρ(x) where, $\rho(x)$ is independent of θ and $x = x_1, x_2, x_3, \ldots \ldots \ldots \ldots \ldots, x_n$ be a random sample.
The function K(x/θ) is called the Kernal of the likelihood and $\rho(x)$ is the residual of the likelihood. If Y = $\emptyset(x)$ is a sufficient statistics for θ then, f (x/θ) = K(y/θ).ρ(x). Instead of considering the kernel K (y / θ)as a function of y with parameter θ on the sample space of Y, we can consider it as a function K (θ/y) of θ with parameter y.
The density function:
g (θ/y) = K(θ/y) N(y), θ∈ Θ, Where N(y) is a normalising constant and is given by the expretion,
N (y) =$[\int K(\theta/y) d\theta]^{-1}$
Here the limit of integration is taken on the basis of given limitθ and y is treated as a parameter is called the 'Natural conjugate of likelihood' f (y/θ) of y. For example let $X_1, X_2, X_3, \ldots \ldots \ldots \ldots \ldots, X_n$ are i.i.d. N (μ, σ) variables. Then,

$$f(x/\mu,\sigma) = \left[\sqrt{\frac{n}{2\pi\sigma}} e^{-\frac{n}{2}(\bar{x}-\mu)^2}\right] \rho(x)$$

$$f(x/\mu,\sigma) = \rho(n,\bar{x})\, \rho(x)$$

So that the statistics (n,$\bar{x}$) is sufficient for μ.
Considering K (n,$\bar{x}/\mu,\sigma$) as a function of μ, we let density function of μ over (-∞,∞) for given y = (n,$\bar{x}$) looked upon as, g (μ/y) α K(y/μ,σ^2) i.e.

g (μ/y) = $\sqrt{\frac{n}{2\pi\sigma}} \exp[-n(\mu-\bar{x})^2]$ is a density function of μ.

This density function is a natural conjugate of likelihood f (x / μ). The families of natural conjugate priors with parameter y (value of the sufficient statistics) in the range of Y have the properties of tractability and closedness. Richness of families of natural conjugate densities can be greatly enhanced by increasing the domain of the parameter to include a value for which

- K(y/θ) is non-negative.
- $\int_\Theta K(\theta/y)$ converges

4. Non-Informative Prior in Bayesian Methods

Thedistribution of the parameter θ is called 'Prior Distribution of θ. If this distribution consists complete and valid information about the parameter θ then it is called the informative prior of θ. A non- informative prior is one in which little new explanatory Power about the unknown parameter is provided by intention. It is not easy to identify non- informative (or objective) prior distribution which may represent prior ignorance or vague prior knowledge and provides solely data dependent conclusions.

In fact, every prior specification has some information or predictive implications and therefore, 'vague' is not a very useful term to represent lack of knowledge. There is no non- informative prior that may represent total ignorance. From practical point of view, the notation of vague prior may be considered as a prior which has minimal effect relative to the data on the final decision.

A variety of criteria are suggested for comparing methods of producing a non- informative prior. The traditional approach to construct non- informative priors used by Laplace (1805), Jeffreys (1983), Lindley (1956), and others, is that the method would be modified or adjusted in order to obtain solutions to a problem for which the method fails. Historically, Bayes (1763) And Laplace (1805) suggested the uniform prior distribution which was later found to depend upon the choice of the parameter. Jeffreys (1983) approached the problem by suggesting rules to construct invariant priors.

Difficulties for multi-parameter problems were later resolved by Jeffreys (1983) through ad-hoc modifications to the prior. A prior probability distribution that represents perfect ignorance or indifference would produce a posterior probability distribution that represents what one should need about the parameter θ on the basis of the evidence (data) X alone. Such a prior is called ''neutral'' or non- informative prior: Royall (1997). According to Jeffreys (1983), non- informative priors provide a formal way of ignorance of the value of the parameter over the permitted range.

There is neither one accepted definition of the word non-informative prior nor one accepted criterion for choosing a non- informative prior. Laplace's principle of insufficient reason to select a Uniform prior, g(θ)=1 for all $\theta \in \Theta$, has been criticized since the times of Boole and Venn due to lack of transformation invariance property.
Lindley (1956), Zellner (1977), Bernardo (1979) And Akaike (1983) have used information theoretic considerations to construct non- informative priors. Zellner's prior is often considered impractical in the sense that it is difficult to interpret priors in relation to the posterior distribution.

Another commonly used procedure developed by Novic (1969), depends on limiting forms for the conjugate prior distributions. In this procedure, the hyperparameter of the conjugate prior is made to approach some limiting value.
For example, if $X_1, X_2, \ldots \ldots \ldots, X_n$ is a random sample from N(θ, r),precision r known, and the conjugate prior for θ is N(μ, τ) then the posterior distribution of θ is N$\left(\frac{\tau\mu + nr\bar{x}}{\tau + nr}, \tau + nr\right.$ which tends to NX, nr as the prior precision $\tau \to 0$.It should be noted that this limiting posterior distribution cannot be derived from any proper prior distribution. In case, both θ and r are unknown, the normal-Gamma prior with hyper-parameters, μ, τ, α *and* β , we have the conditional posterior distribution of θ ,given r, as N$\left(\frac{\tau\mu + n\bar{x}}{\tau + n}, (\tau + n)r\right)$ which tends to N$(\bar{X}, nr)$ as $\tau \to 0$, and as $\alpha \to$-1/2, $\beta \to$0 the marginal posterior distribution of r tends to Gamma $\left[\frac{n-1}{2}, \frac{1}{2}\left(\sum_{i=1}^{n}(X_i - \bar{X})^2\right)\right]$, provided n≥ 2.

This limiting posterior distribution results from the improper prior g $(\theta,r) \propto \frac{1}{r}$. It should be noted that the condition $\alpha > 0$ for the hyperparameter of the Gamma prior must be violated. Such posterior distributions are sometimes called 'nil' posterior distributions.

In fact noninformative priors are not unique. Sometimes, improper priors may lead to behaved posteriors and paradoxes like marginalisation paradox of David, Stone And Zidek (1973). Thomas Bayes (1763) And Laplace (1774) expressed complete ignorance by assigning uniform prior probability distribution for the unknown parameter (s) of the model. Laplace said ''When the probability of simple event is unknown, we may suppose all values between 0 and 1 equally likely.''Box And Tiao (1973) And Bernado (1979) have argued that a non-informative prior should be regarded as a reference prior, i.e., a prior which is convenient to use a standard in analysing statistical data.

Harold Jeffreys suggested a thumb rule specifying non-informative prior for a scalar parameter θ as follows:

Rule: If $\theta \in$[a,b], where a and b are finite numbers or $\theta \in(-\infty,\infty)$, then

g(θ)= constant

***Rule2*:** If $\theta \in(0,\infty)$, assume log θ to be uniformly distributed over the whole real line. This implies, using transformation of variables,

$$g(\theta) \propto \frac{1}{\theta}$$

The improper prior g(θ) $\propto \frac{1}{\theta}$, $\theta \in(0,\infty)$, has the following properties:

(i) $\int_0^\infty \frac{d\theta}{\theta} = \infty$, which indicates that ∞ represents certainty as in the earlier case

(ii) $\int_0^a \frac{d\theta}{\theta} = \infty$, and

(iii) $\int_a^\infty \frac{d\theta}{\theta} = \infty$.

Jeffreys (1946) proposed a formal rule for obtaining a non-informative prior as:

Case (1) - If θ is a K-vector valued parameter, g(θ) $\propto \sqrt{|detI(\theta)|}$,

Where I(θ) is a KxK Fisher's (information) matrix whose (i,j)th element is E$\left[\frac{\partial^2 logl(\theta/x)}{\partial\theta_i\partial\theta_j}\right]$

Case (2) - In particular, if θ is a scalar parameter, JEFFREYS' non-informative prior for θ is

g $(\theta) \propto \sqrt{I(\theta)}$.

Berger (1980) And Robert (1994) gave the concept of non-informative prior. Using this concept by the following examples we observe:For the first case let X be distributed as f(x-θ) where θ is a location parameter. A location invariant density is invariant to linear transformations. This means that Y=X+a is distributed as f(y-$\emptyset$) with $\emptyset = \theta + a$ i.e., X and y have same distribution, but with different location parameters. Since the model is location invariant, the prior distribution should be location invariant. Therefore;

$\pi(\theta) = \pi(\theta$ - a) $\forall$ a; where a is a constant.

$\Rightarrow$ $\pi(\theta)$ = 1.

Thus an invariant non-informative prior for a location parameter is the uniform distribution. Another argument leading to the same result, is that since θ and $\emptyset$ are location parameters in the same model, they should have the same prior.

For the second case let X be distributed as $\frac{1}{\sigma}f\left(\frac{x}{\sigma}\right)$ which is a scale invariant density with scale parameter σ. That the distribution is scale invariant, means that Y= CX has the same distribution as X, but with a different scale parameter. Since the density is scale variant, the prior distribution should be scale variant.

π(A) = π(A/C) $\forall$ A $\in$ (0, +∞) and C >0

This leads to $\pi(\sigma) = \sigma^{-1} = \frac{1}{\sigma}$ which is an improper distribution. Thus, we see that in both cases, the invariant non-informative prior is improper. A difficulty with this method is that all problems do not have an invariant structure. Kass And Wasserman (1996) stated two different interpretations of non-informative priors:

- Non-informative priors are formal representations of ignorance
- There is no objective, instead non-informative priors are chosen by public agreement much like units of length and weight. In the second interpretation priors are the ''default'' to use where there is insufficient information to otherwise define the prior. Today, no one use the first interpretation to claim that one particular prior is truly non-informative.Pericchi And Walley (1991**)** have a quite different view. They said that no single probability distribution can model ignorance satisfactory. Therefore large classes of distribution are needed. They used the interpretation of Kass and Wasserman (1996), but they realise that a single distribution is not enough.

5. Constant Time Series Model using Informative Prior under Phenology

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.

Then the model becomes,

$$d_i - f = \epsilon_i \quad \forall i \qquad (1)$$

Where ϵ_i is i.i.d. and follow normal distribution with mean zero and variance σ^2. Hence using least square theory we get

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

The same expression follows if we use the conept of maximum entropy theory which was given by Jaynes (1957) and Kapur And Kesavan (1992). From the equation (3.4.2), we must now calculate the evidence. P ($\vec{d}/\vec{x}$, c,I). An additional c has been split off from the general conditional background I to make explicit that we treat the constant model here. From the marginalization theorem in equation (1.7.4) of chapter 1, we get,

$$P(\vec{d}/\vec{x}, c, I) = \int P(\vec{d}, f, \sigma/\vec{x}, c, I)\, df\, d\sigma \quad (3)$$

This equation is an identity. The integral in (3.4.3) will now be expanded using the product rule:

$$P(\vec{d}/\vec{x}, c, I) = \int P(f, \sigma/\vec{x}, c, I) \cdot P(\vec{d}/\vec{x}, f, \sigma, c, I)\, df\, d\sigma \quad (4)$$

The first distribution under the integral in (3.4.4) is logically independent of $\vec{x}$ and c simplifies to,

$$P(f, \sigma/I) = P(\sigma/I) \quad (5)$$

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

$$P(f/I) = \frac{1}{2\gamma} \quad (6)$$

The range γ can be estimated from the variance of the data. Also, $\gamma < d_{max} - d_{min}$.

A possible K dimensional generalization of (6) would be $(\frac{1}{2\gamma})^K$.

A more possible choice of the prior volume is however, hypersphere

V_s (k, γ) ;

$$V_s(K,\gamma) = \frac{\gamma^K (\sqrt{\pi})^K}{\Gamma\left(\frac{K+2}{2}\right)} = P(\vec{f}|\gamma, K, I) \quad (7)$$

In general hupercube is converted into hypersphere if there are more than three dimensions.

Let the Conjugate Prior Density for σ^2 is the inverted – Gamma (α, λ) having the p.d.f.

$$g(\sigma^2) = \frac{\lambda^\alpha}{\Gamma\alpha}\left(\frac{1}{\sigma^2}\right)^{\alpha+1} \exp\left(\frac{-\lambda}{\sigma^2}\right) \quad ; \sigma^2>0 \quad (8)$$

Now,

$$\sum_{i=1}^N (d_i - f)^2 = \sum_{i=1}^N \left(d_i - \bar{d} + \bar{d} - f\right)^2$$
$$= \sum_{i=1}^N \left(f - \bar{d}\right)^2 + \sum_{i=1}^N \left(d_i - \bar{d}\right)^2$$
$$+2 \sum_{i=1}^N \left(f - \bar{d}\right)\left(d_i - \bar{d}\right)$$
$$= N\left(f - \bar{d}\right)^2 + N\overline{\Delta d^2}$$

Where $\bar{d} = \frac{1}{N}\sum_{i=1}^N d_i$,

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

The product term vanishes since algebric sum of deviations from mean is zero.

Therefore,

$$P(\vec{d}/\vec{x}, c, I) = \left(\frac{1}{2\pi}\right)^{N/2} \frac{1}{2\gamma} \frac{\lambda^\alpha}{\Gamma\alpha} \cdot$$
$$\int_0^\infty \left(\frac{1}{\sigma^2}\right)^{\alpha+1} \frac{1}{\sigma^N} \exp\left\{\frac{-1}{\sigma^2}\left(\frac{N\overline{\Delta d}^2}{2} + \lambda\right)\right\} d\sigma \cdot$$
$$\int_{-\infty}^\infty \exp\left\{-\frac{N}{2\sigma^2}(f - \bar{d})^2\right\} df \quad (9)$$

Now

$$\int_{-\infty}^\infty \exp\left\{-\frac{N}{2\sigma^2}(f - \bar{d})^2\right\} df = \sigma\sqrt{\frac{2\pi}{N}} \quad (10)$$

Now using equation (9) and 10) we can write

$$P(\vec{d}/\vec{x}, c, I) = \left(\frac{1}{2\pi}\right)^{\frac{N-1}{2}} \frac{1}{2\gamma} \cdot \frac{\lambda^\alpha}{\Gamma\alpha} \cdot$$
$$\int_0^\infty \left(\frac{1}{\sigma^2}\right)^{\alpha+1} \frac{1}{\sigma^{N-1}} \cdot \exp\left\{\frac{-1}{\sigma^2}\left(\frac{N\overline{\Delta d}^2}{2} + \lambda\right)\right\} d\sigma$$

Let $x = \frac{1}{\sigma^2}$

Then $x = \sigma^{-2}$

$$\Rightarrow \frac{dx}{d\sigma} = -\frac{1}{2x^{3/2}}$$

Now,

$$\frac{d\sigma}{\sigma} \frac{1}{\sigma^{N-1}} = -\frac{1}{2x^{3/2}} dx . \frac{1}{x^{-1/2} x^{-(N-1)/2}} = -\frac{dx}{2x^{3/2} x^{-N/2}} = -\frac{dx}{2x^{-(N-3)/2}}$$

Therefore,

$$\int_0^\infty \left(\frac{1}{\sigma^2}\right)^{\alpha+1} \frac{1}{\sigma^{N-1}} \cdot \exp\left\{\frac{-1}{\sigma^2}\left(\frac{N\overline{\Delta d}^2}{2} + \lambda\right)\right\} d\sigma$$
$$= \frac{1}{2}\Gamma\frac{2\alpha+N}{2} \frac{1}{\left(\frac{N\overline{\Delta d}^2}{2} + \lambda\right)^{\frac{2\alpha+N}{2}}} \quad (11)$$

Collecting terms the evidence for the constant model becomes,

$$P(\vec{d}/\vec{x}, c, I) = \frac{1}{2}\left(\frac{1}{\pi}\right)^{\frac{N-1}{2}} \frac{1}{2\gamma} \frac{\lambda^\alpha}{\Gamma\alpha} \cdot \frac{\Gamma\frac{2\alpha+N}{2}}{\left(\frac{N\overline{\Delta d}^2}{2} + \lambda\right)^{\frac{2\alpha+N}{2}}} \frac{1}{\sqrt{N}} \quad (12)$$

Now if we use non-informative prior i.e.

$$P(\sigma/\beta, I) = \frac{1}{2ln\beta} \cdot \frac{1}{\sigma}, \qquad \frac{1}{\beta} < \sigma < \beta \quad (13)$$

And in this case, $P_1\left(\vec{d}/\vec{x}, C, I\right) = \frac{1}{2}\left(\frac{1}{\pi}\right)^{\frac{N-1}{2}} \cdot \frac{1}{2\gamma} \frac{1}{2ln\beta} \cdot$

$$\frac{\Gamma\left(\frac{N-1}{2}\right)}{\left(N\overline{\Delta d^2}\right)^{\frac{N-1}{2}}} \cdot \frac{1}{\sqrt{N}} \quad (14)$$

This completes the analysis of constant time series model using informative prior in phenology. Now we compare the suitability of the derived constant time series model (12) by calculating the Bayes' factor in favour of constant time series model using informative prior against the constant time series model using non- informative prior which is given by

B=

$$\frac{\lambda^\alpha\ 2\ ln\beta}{\Gamma\alpha} \cdot \frac{\Gamma\frac{2\alpha+N}{2}\left(N\overline{\Delta d^2}\right)^{\frac{N-1}{2}}}{\left(\frac{N\overline{\Delta d}^2}{2} + \lambda\right)^{\frac{2\alpha+N}{2}} \left(N\overline{\Delta d^2}\right)^{\frac{N-1}{2}}} \quad (15)$$

For known value of α, β and λ we observe the Bayes' factor in favour of the model (12) numerically. Next, we calculate the mean and variance of σ^2as,

$$E(\sigma^2) = \int_0^\infty \sigma^2 \frac{\lambda^\alpha}{\Gamma\alpha}\left(\frac{1}{\sigma^2}\right)^{\alpha+1} \exp\left(\frac{-\lambda}{\sigma^2}\right) d\sigma^2 = \frac{\lambda}{\alpha-1}$$

provided $\alpha > 1$ (16)

and $E(\sigma^2)^2$ =

$$\int_0^\infty \sigma^4 \frac{\lambda^\alpha}{\Gamma\alpha}\left(\frac{1}{\sigma^2}\right)^{\alpha+1}\exp\left(\frac{-\lambda}{\sigma^2}\right) d\sigma^2 = \frac{\lambda^2}{(\alpha-1)^2(\alpha-2)} + \left(\frac{\lambda}{\alpha-1}\right)^2$$

Hence Variance of σ^2 is given by the formula

$$\text{Var}(\sigma^2) = \text{E}(\sigma^4) - (\text{E}(\sigma^2))^2 = \frac{\lambda^2}{(\alpha-1)^2(\alpha-2)} \text{ provided } \alpha > 2 \qquad (17)$$

Solving the eqations (16) and (17), we get the value of $\alpha$and λ.

6. Numerical Illustration

To demonstrate how the bayes' factor of the model (3.4.12) can be calculatedthe following 10 years average temperature data has been taken from the Acharya Nrendra Deva Aggricultre and Technology, University of Faizabad (India) as,

Table 1: Average temperature

Years	Average temperature of the years	Production of rice in tones
2000	26.26	21.13
2001	26.32	20.576
2002	25.47	20.022
2003	26.1	19.468
2004	24.16	18.914
2005	24.5	18.360
2006	25.112	18.935
2007	25.87	19.510
2008	25.85	20.085
2009	24.9	20.660

Here N = number of years of the data =10

$\sum d_i$ = Sum of the average temperature data of the years = 254.555. Now, $\bar{d} = \frac{245.555}{10} = 25.455$,
$\bar{d}^2 = (25.455)^2 = 647.98$
Therefore, $\Delta\bar{d}^2 = \frac{1}{N}\sum_{i=1}^{N} {d_i}^2 - \bar{d}^2 = \frac{1}{10}(6484.96) - 647.98 = 0.516$. To find the value of α and λ numerically we have the following variance table as,

Table 1: Variance of verage temperature

Years	Variance of the Average temperature of the years
2000	1.66
2001	0.45
2002	1.65
2003	0.32
2004	1.66
2005	0.31
2006	1.64
2007	0.24
2008	1.78
2009	0.35

From the table (2) we have $\text{E}(\sigma^2) = 1.006 = \frac{\lambda}{\alpha-1}$

and $\text{Var}(\sigma^2) = 0.5058 = \frac{\lambda^2}{(\alpha-1)^2(\alpha-2)}$

This imlies that $\alpha = 4$ and $\lambda = 3.018$

Also, we have

$$\text{Var (d)} = 0.516 = \sigma^2$$
$$\text{S.D. (d)} = 0.7183 = \sigma$$

Since, $\frac{1}{\beta} < \sigma < \beta \Rightarrow \beta = 1.5$

Hence putting above values in equation (12) and (14) we get,

$$\text{P}(\vec{d}/\vec{x}, c, \text{I}) = 0.0000256 \text{ and } P_1(\vec{d}/\vec{x}, c, \text{I}) = 0.0000012$$

Now, the Bayes' factor in favour of the constant time series model using informative prior against the constant time series model using non- informative prior is given by

$$\text{B} = \frac{0.0000256}{0.0000012} = 21.333$$

7. Conclusion

Since the Bayes' factor in favour of the constant time series model using informative prior against the constant time series model using non- informative prior is greater than unity.

References

[1] Berliner, L.M, Bayesian Climate Change Assesment, Jour. Climate, 13, 3805-3820, 2000.
[2] Berger, J. O. Statistical Decision Theory And Bayesian Analysis, 1985.
[3] Box, G.E.P. and Tiao, G.C. Bayesian Inference in Statistical Analysis, 1973.
[4] Dose, V. and Menzel, ABayesian Analysis of Climate Change Impacts in Phenology, Global Change Biol., 10, 259-272,2004.
[5] Ferguson, T.SA Bayesian Analysis of Some nonparametric, Problems, The Annals of Statistics 1, 209-230, 1973.
[6] Kapoor, J.N. and Kesvan, H.K. Entrophy Optimization Principles with applications, Academic Press, Boston. 1992.
[7] Kass, R.E. and Raftery, A.EBayes Factor and Model Uncertainity,Jour.Amr. Statist. Assoc. 90, 773-795, 1995.
[8] Lindley,D.V,Introduction to Probability And Statistics From a Bayesian Point of View,1965.