RR-73-8

A BAYESIAN APPROACH TO NONSTATIONARY PROCESSES

Robert L. Winkler

October 1973

A Bayesian Approach to Nonstationary Processes *

Robert L. Winkler **

I. Introduction

In the usual Bayesian approach to problems of statistical inference and decision concerning a parameter θ, Bayes' theorem can be expressed in the form

$$f(\theta|x) = f(\theta)\ f(x|\theta) / \int f(\theta)\ f(x|\theta)\ d\theta \quad ,$$

with the usual abuse of functional notation. That is, assuming that the prior information about θ can be expressed in the form of a prior distribution $f(\theta)$ and that new information x (sample information) concerning θ can be summarized by the likelihood $f(x|\theta)$, Bayes' theorem revises the prior distribution on the basis of the new information. The revised distribution $f(\theta|x)$ is called a posterior distribution. This provides a framework for inferences about θ, the uncertain quantity or parameter of interest, and for decisions which are related to θ. For detailed discussions of the Bayesian

---

*This paper will be published in the Proceedings of the Beaulieu Seminar on Decision Theory.

**Indiana University, U.S.A.; research scholar at the International Institute for Applied Systems Analysis.

approach to statistical inference and decision, see Raiffa and Schlaifer [8], DeGroot [4], and LaValle [6]. For a relatively non-technical discussion of the Bayesian approach see Winkler [13].

The above model for inference is a stationary model. That is, it assumes that θ takes on a single value, so that $f(\theta)$ and $f(\theta|x)$ represent uncertainty about what that value is. For instance, θ may represent the proportion of defective items produced by a certain manufacturing process, the proportion of consumers purchasing a given product, the mean daily sales at a given store, the rate at which cars arrive at a toll booth, the variance in the diameter of parts produced at a particular plant, and so on. In each case, θ is assumed to be fixed but unknown. In the Bayesian framework, information concerning θ is expressed in terms of a probability distribution. In general, the position taken in this paper is that of the subjective interpretation of probability (e.g. see de Finetti [3] and Savage [10]), so that $f(\theta)$ represents a quantification of the judgements of the statistician or of an expert consulted by the statistician (e.g. see Winkler [12] and Savage [11]). However, this assumption can be relaxed without loss of generality, since the mathematical results will not depend on the source of the prior distribution.

Many, if not most, real world data-generating processes are characterized by nonstationarity rather than stationarity.

For instance, the probability that an item produced by a manufacturing process is defective, and hence the proportion of defective items generated by the process, will generally vary over time, even over relatively short periods of time. This illustrates nonstationarity over time. The probability that a consumer will purchase a given product (hence the proportion of consumers purchasing the product) may vary over time, and in addition, the probability of purchasing the product at a given point in time may vary from consumer to consumer. This illustrates nonstationarity over time <u>and</u> nonstationarity at a given point in time. The other examples given in the preceding paragraph would obviously be more realistic if nonstationarity were assumed.

Despite the presence of nonstationarity in many real-world processes, formal Bayesian models involving nonstationarity have received little attention in the statistical literature. Exceptions are articles by Bather [1,2]. For example, in Bather [2], a nonstationary model is used in the study of control charts and the determination of optimal decision rules regarding the control charts. In Zellner [16], certain types of nonstationarity are considered within the framework of regression models. The objective of this paper is to present a brief preliminary report on an on-going research program, the aims of which are to develop formal models for handling nonstationarity within a Bayesian framework, to compare inferences from stationary and nonstationary models, and to investigate inferential and decision-theoretic applications involving nonstationarity.

II. The Development of Bayesian Models Incorporating Nonstationarity

If the process generating θ is nonstationary, then it is not particularly realistic to make inferences and decisions concerning θ as if θ only took on a single value. Instead, one should be concerned with a sequence $\theta_0, \theta_1, \theta_2, \ldots, \theta_i, \ldots$ of values of θ corresponding to different elements of the process, or members of the population of concern. If the concern is with a particular stochastic process over time, the subscripts represent different points in time; for instance, θ_i may be the value of θ during time period i (e.g. the probability that a given consumer will purchase a product during time period i). If the concern is with different elements at a particular point in time, the subscripts represent the elements; for instance, θ_i may be the value of θ for element i (e.g. the probability that consumer i will purchase a product at a particular point in time).

The usual stationary model for inference and decision assumes that $\theta_i = \theta_j$ for all i and j, so that the common value can be treated as a single parameter. Another possibility is that θ_i and θ_j are related in a deterministic manner. But if the deterministic relationship between θ_i and θ_j is a one-to-one relationship for all i and j, then θ_i can be related to a single parameter θ for each i and the problem is once again reduced to one concerning a single parameter, although inference concerning the single parameter may be a more difficult problem than in the case in which $\theta_i = \theta_j$ for all i and j. In any event, the situation in which θ_i and θ_j are related in a deterministic manner is not considered in this paper.

Two types of nonstationary models will be considered here:

1. Models in which θ_i and θ_j are related in a stochastic manner, and
2. Models in which θ_i and θ_j are independent and identically distributed, conditional upon some "second-order" parameter(s).

The first type of model is likely to be applicable when non-stationarity over time is present, and the second type of model, while also applicable for nonstationarity over time, appears to be much more suitable than the first type of model handling nonstationarity at a given point in time.

If θ_i and θ_j are related in a stochastic manner, a formal treatment of the situation necessitates some assumptions about the stochastic relationship. In many cases, the specification of the stochastic relationship between successive values of θ is sufficient; if this relationship is stationary (this is a type of second-order stationarity), the stochastic relationship between θ_i and θ_{i+1} is the same as that between θ_j and θ_{j+1} for any i and j. If the relationship between θ_i and θ_{i+1} can be summarized by a parameter (or vector of parameters) ψ, then information concerning ψ is useful in making inferences concerning θ_i. For example, a production process may be nonstationary in the sense that the mean weight of the output of the process may change over time although the variance of weight is relatively constant. Moreover, shifts in the mean from one time period to the next may behave according to a random walk; that is, $\theta_{i+1} = \theta_i + \varepsilon_i$, where the ε_i are independent and identically distributed. If ε_i is normally distributed with mean m and variance v, then the parameters m and v summarize the stochastic relationship between θ_i and θ_{i+1}. In the notation introduced earlier in this paragraph, $\psi = (m,v)$. Bather [1] investigates this particular example with $m = 0$ and known $v > 0$.

If ψ is known, the Bayesian approach to the problem is straightforward. At the beginning of time period i, the information about θ_i can be expressed in the form of a probability distribution $f(\theta_i)$. During period i, θ_i is not observed, but some sample information x_i is observed, and x_i can be used to update the distribution of θ_i in the usual manner:

$$f(\theta_i|x_i) = f(\theta_i)\ f(x_i|\theta_i) / \int f(\theta_i)\ f(x_i|\theta_i)\ d\theta_i \quad .$$

This is identical to the procedure that is used in the stationary case. The revised distribution of θ_i can, in turn, be used to determine $f(\theta_{i+1})$, the distribution of θ_{i+1} at the beginning of time period i+1:

$$f(\theta_{i+1}) = \int f(\theta_{i+1}|\theta_i,\psi)\ f(\theta_i|x_i)\ d\theta_i \quad .$$

Thus, at the end of a period, it is necessary to take into account the sample information acquired during that period and the relationship between the value of θ during that period and the value of θ during the next period.

If ψ is not known, revisions involve θ_i and ψ:

$$f(\theta_i,\psi|x_i) = f(\theta_i,\psi)\ f(x_i|\theta_i,\psi) / \int\int f(\theta_i,\psi)\ f(x_i|\theta_i,\psi)\ d\theta_i\ d\psi$$

The resulting distribution of θ_{i+1} and ψ is

$$f(\theta_{i+1},\psi) = \int\int f(\theta_{i+1}|\theta_i,\psi)\ f(\theta_i,\psi|x_i)\ d\theta_i\ d\psi \quad .$$

This situation is conceptually not different from the case in which ψ is known, but it may be considerably more difficult to handle in practice because it involves a joint distribution of θ_i and ψ at each time period i.

The situation in which θ_i and θ_j are related in a stochastic manner can be investigated under various assumptions concerning the relationship. For instance, in Winkler and Barry [15], the situation in which shifts in the mean of a process behave according to a random walk is generalized to the multivariate case, where θ_i represents a vector of means, m represents a vector, and v represents a covariance matrix. Further generalizations might include the relaxation of the assumption that the variance of the process is stationary, so that $\theta_i = (\mu_i, \sigma_i)$ in the univariate case, or the relaxation of the assumption that shifts occur at regular intervals (e.g. the occurrence of shifts may behave like a Poisson process). Another option is to assume that θ_{i+1} depends on θ_{i-1} (other than simply through θ_i) as well as on θ_i.

The other type of nonstationary model to be discussed here requires the assumption that $\theta_0, \theta_1, \ldots, \theta_i, \ldots$ are independent and identically distributed, conditional upon some second-order parameters. Because of this assumption, the problem is reduced to one of making inferences about the distribution of θ_i, which might be called the "distribution of nonstationarity." For instance, if θ_i is the probability that consumer i will purchase a given product, the distribution of nonstationarity might represent the distribution of different values of θ across the population of consumers. If θ_i is the mean for day i of a stochastic process generating sales at a given store, the distribution of nonstationarity might represent the different values of θ over time (it might be assumed that the distribution of stationarity remains the same over time, thus avoiding the problem of second-order nonstationarity).

In many applications, it would be convenient to assume a particular model for the distribution of nonstationarity. If θ_i is the probability that consumer i will purchase a given product, then a convenient and reasonable model is the beta model, in which case the distribution of nonstationarity is a beta distribution with parameters α and β. (Note that if α and β are small enough, the distribution is U-shaped, which might seem reasonable in some cases). If θ_i is the mean for day i of the process generating sales at a given store, then a normal model might be applicable, in which case the distribution of nonstationarity is a normal distribution with parameters μ and σ^2. In general, the distribution of nonstationarity will have a parameter (or a vector of parameters) which will be denoted by ϕ, so that the distribution of nonstationarity can be written in the form $f(\theta_i|\phi)$ for all i.

The easiest situation to handle is that in which ϕ is known. Indeed, this may be compared to the usual Bayesian approach, in which the prior distribution is, for example, a beta distribution with fixed α and β. It should be emphasized, however, that the interpretation of the distribution is different and that since the process is nonstationary, the usual application of Bayes' theorem is meaningless unless the process is assumed to be stationary over short time periods. If we <u>know</u> that θ is nonstationary and we know the exact nature of the nonstationarity (i.e. since we know ϕ, we know the exact distribution of nonstationarity), then there is essentially no uncertainty involving the distribution of

θ_i. There is uncertainty about future sample outcomes, however, and this uncertainty at time i can be expressed in the form of a predictive distribution:

$$f(x) = \int f(\theta_i|\phi)\, f(x|\theta_i)\, d\theta_i \quad .$$

Predictive distributions such as these, although they are often ignored, are of great importance since important decisions may depend on a future sample outcome x rather than on a parameter θ_i (see Roberts [9]). For instance, decisions concerning a new product depend on x, the actual future purchase behavior, not on θ_i, the probability that consumer i purchases the product. The parameter θ_i is only relevant indirectly, in the sense that the predictive distribution f(x) depends on the distribution of θ_i, $f(\theta_i|\phi)$, and on $f(x|\theta_i)$.

A more difficult, but more realistic, situation occurs when ϕ is not known (i.e. the exact distribution of nonstationarity is not known). The recent work by Ferguson [5] concerning a Bayesian approach to nonparametric problems is in this spirit. Empirical Bayes methods (e.g. Maritz [7] assume that ϕ is not known but attempt to determine a point estimate of ϕ instead of a probability distribution for ϕ. In the Bayesian approach taken in this paper, uncertainty about ϕ can be formally expressed in terms of a probability distribution $f(\phi)$, which might be called the prior distribution of ϕ. New sample information x can now be used to revise the distribution of ϕ, yielding a posterior distribution $f(\phi|x)$, using Bayes' theorem:

$$f(\phi|x) = f(\phi)\ f(x|\phi) / \int f(\phi)\ f(x|\phi)\ d\phi \quad .$$

The "likelihood" in this application of Bayes' theorem is $f(x|\phi)$, which is the predictive distribution in the stationary Bayesian model (recall that in the stationary model, ϕ is known and the uncertainty concerns θ). This "likelihood" is related to the usual likelihood $f(x|\theta_i)$ and to the distribution of nonstationarity $f(\theta_i|\phi)$ as

$$f(x|\phi) = \int f(x|\theta_i)\ f(\theta_i|\phi)\ d\theta_i \quad .$$

The final distribution of interest in the nonstationary model is the new "predictive" distribution, which incorporates the uncertainty about ϕ (the predictive distribution in the stationary model assumes ϕ is known):

$$f(x) = \int f(x|\phi)\ f(\phi)\ d\phi \quad ,$$

or

$$f(x) = \int \int f(x|\theta_i)\ f(\theta_i|\phi)\ f(\phi)\ d\theta_i\ d\phi \quad .$$

The situation in which $\theta_0, \theta_1, \ldots, \theta_i, \ldots$ are independent and identically distributed can be investigated under various assumptions concerning the distribution of nonstationarity,

$f(\theta_i|\phi)$, and concerning $f(\phi)$. For instance, suppose that θ_i represents the mean of a normal data-generating process with known variance σ^2, that the distribution of nonstationarity is a normal distribution with unknown mean m and known variance v, and that the prior distribution of $\phi = m$ is a normal distribution. Then it can be shown that the posterior distribution of $\phi = m$ following samples from one or more of the "populations" (values of i) is also a normal distribution. In the same situation, if the distribution of nonstationarity is a normal distribution with known mean m and unknown variance v and if the prior distribution of $\phi = v$ is a translated-inverted-gamma distribution, then the posterior distribution of $\phi = v$ is also a translated-inverted-gamma distribution. The application of Bayes' theorem under various distributional assumptions such as these is now being studied.

The two types of models presented in this section are quite general and should be able, at least conceptually, to handle a great variety of situations for which nonstationarity is present. These nonstationary models are currently being developed in greater detail, and questions such as tractability and applicability are being investigated.

III. Work in Progress

The models discussed in this paper allow the Bayesian to formally introduce nonstationarity. Since stationarity assumptions are often quite unrealistic, the introduction of possible nonstationarity greatly increases the realism and the applicability of Bayesian procedures. The objective of the work in progress along these lines is to develop and investigate nonstationary Bayesian models, and this research involves several facets, some of which are briefly discussed in the following paragraphs.

A. Tractability

In the case where θ_i and θ_j are stochastically related, the choice of a model to represent the stochastic relationship has a direct bearing on how easy it is to make various inferential statements and decisions concerning the process of interest. Once certain models are assumed for $f(\theta_i)$ and $f(x_i|\theta_i)$, the determination of $f(\theta_i|x_i)$ proceeds along standard lines, but yet another distribution, $f(\theta_{i+1}|\theta_i,\psi)$, is needed to derive $f(\theta_{i+1})$. Corresponding to $f(\theta_i|x_i)$, it may be possible to find a family of distributions $f(\theta_{i+1}|\theta_i,\psi)$ that is tractable in the sense that $f(\theta_{i+1})$ is not difficult to determine analytically if $f(\theta_{i+1}|\theta_i,\psi)$ is a member of the given family. For instance, the example involving a random walk mentioned in Section 2 yields a tractable solution. Similarly, in the case where $\theta_0,\theta_1,\ldots,\theta_i,\ldots$ are independent and identically distributed, it may be possible to find families of distributions $f(\phi)$ that simplify the analysis somewhat if certain statistical models are assumed for $f(\theta_i|\phi)$ and $f(x|\theta_i)$, as in the standard Bayesian approach to stationary processes. This approach is analogous to the development of natural-conjugate families of distributions for θ in the stationary case (see Raiffa and Schlaifer [8]), but it may be more complex because interrelationships among several distributions are involved. In addition to the investigation of the possibility of tractable families, the use of numerical methods will also be considered. In situations where it is difficult or impossible to find tractable families or in situations where such families are not rich enough to provide realistic approximations, numerical methods should prove useful.

B. Comparisons of Inferences from Stationary and Nonstationary Models

Such comparisons might indicate situations in which the nature of the nonstationarity is such that its formal inclusion in the model has little effect on the ultimate inferences which are drawn. In some cases it might be especially valuable to introduce nonstationarity formally, whereas in other situations it may contribute very little. By considering various situations, it may be possible to draw some conclusions regarding the conditions under which nonstationary models are particularly valuable.

C. Applicability

To investigate the applicability of nonstationary models, various specific applications will be studied. These might be drawn from areas such as finance (it might be assumed that the mean daily change in the price of a security is nonstationary (see Winkler [14] and Winkler and Barry [15]), marketing (simple Bernoulli and Markov models might be adapted to the nonstationarity case to study purchase behavior), and production (manufacturing processes might be considered to be nonstationary with respect to some parameters of interest--see Bather [1]). Some work concerning nonstationary Bayesian models for forecasting future security prices is currently being conducted.

D. Implications for Decision Theory

Often decisions must be made in the face of nonstationarity, and it should be useful to investigate the effect of the formal representation of nonstationarity on the resulting decisions. For instance, with regard to the nonstationary model for forecasting future security prices, implications for the selection of an optimal portfolio of securities are of interest. The study of such implications involves dynamic programming with revision of probability distributions under a nonstationary model. With regard to nonstationary models of purchase behavior, implications for marketing decisions are of interest. In a more general framework, the effect of nonstationarity on the solutions to certain "standard" decision making problems (e.g. finite-action problems with linear payoff functions) should be of interest.

In summary, a formal Bayesian approach to nonstationarity may have important implications for statistical inference and decision. The ongoing research program described in this paper is intended to study Bayesian models for nonstationary processes and to investigate some inferential and decision-theoretic implications of these models.

References

[1] Bather, J.A. "Control Charts and the Minimization of Costs," Journal of the Royal Statistical Society B, 25 (1963), 49-70.

[2] Bather, J.A. "Invariant Conditional Distributions," Annals of Mathematical Statistics, 36 (1965), 829-846.

[3] de Finetti, B. "La Prévision: Ses Lois Logiques, Ses Sources Subjectives," Annales de l'Institut Henri Poincaré, 7 (1937), 1-68.

[4] De Groot, M.H. Optimal Statistical Decisions. New York; McGraw-Hill, 1970.

[5] Ferguson, T.S. "A Bayesian Analysis of Some Nonparametric Problems," Annals of Statistics, 1 (1973), 209-230.

[6] LaValle, I.H. An Introduction to Probability, Decision, and Inference. New York; Holt, Rinehart and Winston, 1970.

[7] Maritz, J.S. Empirical Bayes Methods. London; Methuen, 1970.

[8] Raiffa, H., and Schlaifer, R. Applied Statistical Decision Theory. Boston; Division of Research, Harvard Business School, 1961.

[9] Roberts, H.V. "Probabilistic Prediction," Journal of the American Statistical Association, 60 (1965), 50-62.

[10] Savage, L.J. The Foundations of Statistics. New York; Wiley, 1954.

[11] Savage, L.J. "Elicitation of Personal Probabilities and Expectations," Journal of the American Statistical Association, 66 (1971), 783-801.

[12] Winkler, R.L. "The Assessment of Prior Distributions in Bayesian Analysis," Journal of the American Statistical Association, 62 (1967), 776-800.

[13] Winkler, R.L. An Introduction to Bayesian Inference and Decision. New York; Holt, Rinehart and Winston, 1972.

[14] Winkler, R.L. "Bayesian Models for Forecasting Future Security Prices," Journal of Financial and Quantitative Analysis, 8 (1973), in press.

[15] Winkler, R.L., and Barry, C.B. "Nonstationary Means in a Normal Process," in preparation (to be presented in session on Decision Theory at ASA meetings in New York, December 1973).

[16] Zellner, A. An Introduction to Bayesian Inference in Econometrics. New York, Wiley, 1971.