Journal of Statistical Planning and Inference 143 (2013) 2128–2141

Contents lists available at ScienceDirect

Journal of Statistical Planning and Inference

journal homepage: www.elsevier.com/locate/jspi

Random-effect models with singular precision

Woojoo Lee [a], Youngjo Lee [b,*]

[a] Department of Statistics, Inha University, 235, Yonghyun-dong, Nam-gu, Incheon, 402-751, South Korea
[b] Department of Statistics, Seoul National University, 599 Gwanangno, Gwanak-gu, Seoul, 151-742, South Korea

ARTICLE INFO

Article history:
Received 16 June 2011
Received in revised form
22 April 2013
Accepted 1 July 2013
Available online 10 July 2013

Keywords:
Best linear unbiased predictor
Intrinsic autoregression
Random-effect models
Singular precision
Smoothing splines
State-space models

ABSTRACT

We show that smoothing spline, intrinsic autoregression (IAR) and state-space model can be formulated as partially specified random-effect model with singular precision (SP). Various fitting methods have been suggested for the aforementioned models and this paper investigates the relationships among them, once the models have been placed under a single framework. Some methods have been previously shown to give the best linear unbiased predictors (BLUPs) under some random-effect models and here we show that they are in fact uniformly BLUPs (UBLUPs) under a class of models that are generated by the SP of random effects. We offer some new interpretations of the UBLUPs under models of SP and define BLUE and BLUP in these partially specified models without having to specify the covariance. We also show how the full likelihood inferences for random-effect models can be made for these models, so that the maximum likelihood (ML) and restricted maximum likelihood (REML) estimators can be used for the smoothing parameters in splines, etc.

1. Introduction

Let A be a known $(n-k)\times n$ matrix and $\mathbf{y}$ an $n\times 1$ vector of the observed data. Suppose that we have a statistical model (or likelihood) available for $A\mathbf{y}$, but not for $\mathbf{y}$, yet the aim of statistical analysis is to make prediction $\hat{\mathbf{y}}$. Based only upon the model of the partial data $A\mathbf{y}$, can we make a good prediction, for example of the conditional mean of $\mathbf{y}$? If so, in what sense? In this paper we introduce random-effect models with singular precision (SP) to unify several approaches that have been developed to tackle this problem in various areas of statistics. They have given different remedies. With new random-effect models with SP, we study the relationships among them, by offering a common framework under which the previous works can be placed.

Random-effect models, smoothing splines (Green and Silverman, 1994), intrinsic autoregression (IAR) models (Besag and Kooperberg, 1995) and state-space models (Harvey, 1989; Durbin and Koopman, 2001), etc., have gained popularity over the years. Some connections among them have been addressed. For example, smoothing splines can be fitted by random-effect models (Wahba, 1990; Speed, 1991; Eilers and Marx, 1996; Verbyla et al., 1999; Lee et al., 2006) and by state-space models (Wecker and Ansley, 1983). Random-effect models, specified by the first-two moments of random effects, are parametric models that allow full likelihood inferences, while the rest of models, specified by the SP, are partially specified (semi-parametric) models, without having commonly agreed likelihood inferences yet. In fact, a random-effect model with SP is a class of models, composed of infinitely many parametric random-effect models. In this paper we study the relationships between these two types of models and show how to fit partially specified models by using fully parametric models such as

* Corresponding author. Tel.: +82 2 880 6568; fax: +82 2 883 6144.
E-mail address: youngjo@snu.ac.kr (Y. Lee).

0378-3758/$ - see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.jspi.2013.07.007

random-effect models. This means that the likelihood inferences can be made for these partially specified or semi-parametric models. We study the meaning of the resulting ML and REML estimators.

Consider the curve fitting problem

$$y_i = f(t_i) + e_i \quad i = 1, \ldots, n$$

where $a < t_1 < \cdots < t_n < b$ and $f(\cdot)$ is a smooth but unknown function to be estimated by minimizing

$$\sum_{i=1}^{n} (y_i - f(t_i))^2 + \xi \int_a^b (f''(u))^2 \, du. \tag{1}$$

One of the most popular methods to estimate $f(\cdot)$ involves the use of cubic smoothing splines (Green and Silverman, 1994). The solution to (1) is unique and can be obtained by the penalized least squares (PLS) of the regression model, as was done in Green and Silverman (1994)

$$\mathbf{y} = \mathbf{v} + \mathbf{e}, \tag{2}$$

where $\mathbf{v} = (v_1, \cdots, v_n)^T$ with the ith element $v_i = v(t_i)$ and $\mathbf{e} = (e_1, \ldots, e_n)^T \sim N(\mathbf{0}, \phi I_n)$ are white noise, $\mathbf{0} = (0, \ldots, 0)^T$, and I_n is the identity matrix of order n. In their approach $\hat{\mathbf{y}} = \hat{v}$ is estimating $\mathbf{f} = (f(t_1), \cdots, f(t_n))^T = \mathrm{E}(\mathbf{y}|\mathbf{v})$. They used the penalty $p(\mathbf{v}) = \mathbf{v}^T \Gamma \mathbf{v}$ for some singular matrix Γ (Green and Silverman, 1994, p. 18) where Γ is derived from the second term in (1). Besag and Kooperberg (1995) called Γ precision. Predictions can be made at different values of t_i.

Alternatively, Lee and Nelder (2006) showed that the solution to (1) can also be found by fitting the random-effect model

$$\mathbf{y} = X_0 \beta_0 + \mathbf{v} + \mathbf{e}, \tag{3}$$

where $X_0 = (\mathbf{1}, \mathbf{t}), \mathbf{1} = (1, \ldots, 1)^T$, $\mathbf{t} = (t_1, \ldots, t_n)^T$, and $\mathrm{E}(\mathbf{v}) = \mathbf{0}$, and $\mathrm{Cov}(\mathbf{v}) = \Gamma^+$, where throughout this paper + denotes the Moore–Penrose inverse. In their approach $\hat{\mathbf{y}} = X_0 \hat{\beta}_0 + \hat{\mathbf{v}}$ is estimating $\mathbf{f} = (f(t_1), \ldots, f(t_n))^T$. In the PLS approach the so-called smoothing parameter $\xi = \phi/\lambda$ is not a model parameter, so that it is often estimated by a numerical method such as cross validation, while in random-effect models it is the ratio of variance components, to be estimated by likelihood methods such as ML or REML. Speed (1991) pointed out that the generalized maximum likelihood estimator proposed in Wahba (1985) is identical to REML estimator. Incidentally, Verbyla et al.'s (1999) model can be shown to be identical to Lee and Nelder's (2006) random-effect model. Thus, it is of interest to investigate the relationships among estimators from these two apparently different models, namely the PLS of (2) and the random-effect models of Lee and Nelder (2006).

In spatial statistics, the following random-effect model of the general form is often considered:

$$\mathbf{y} = X\beta + \mathbf{v} + \mathbf{e} \tag{4}$$

where β is a $p \times 1$ vector of fixed unknown parameters, X is an $n \times p$ model matrix and an $n \times 1$ vector of random effects $\mathbf{v}$ is multivariate normal with $\mathrm{E}(\mathbf{v}) = \mathbf{0}$ and nonsingular covariance $\Sigma = \mathrm{Cov}(\mathbf{v}) = \Gamma^{-1}$. We have the density function

$$\{\det(\Gamma)\}^{1/2} \exp(-\mathbf{v}^T \Gamma \mathbf{v}/2), \tag{5}$$

which leads to the conditional distribution

$$v_i | v_{-i} \sim N\left(\sum_j \gamma_{ij} v_j, \kappa_i\right) \tag{6}$$

where $v_{-i} = \{v_j : j \neq i\}$, $\gamma_{ii} = 0$, $\gamma_{ij} = -\Gamma_{ij}/\Gamma_{ii} (i \neq j)$, and $\kappa_i = 1/\Gamma_{ii}$. In spatial statistics, this model is referred to as the conditional autoregressive model (Besag, 1974). Besag and Kooperberg (1995) proposed IAR models with the constraint $\Gamma \mathbf{1} = \mathbf{0}$, leading to a SP Γ of $\mathbf{v}$. For the IAR models, Besag and Kooperberg (1995) noted that conditional distribution (6) remains valid and proposed using the posterior mean under the improper prior proportional to

$$\exp(-\mathbf{v}^T \Gamma \mathbf{v}/2) = \exp\left(-\sum_{i<j} \Gamma_{ij} (v_i - v_j)^2/2\right),$$

This is equivalent to using the penalty $p(\mathbf{v}) = \mathbf{v}^T \Gamma \mathbf{v}$. However, $\exp(-\mathbf{v}^T \Gamma \mathbf{v}/2)$ cannot be normalized to be a density function of $\mathbf{v}$.

Let $\Gamma = A^T A/\lambda$ for some $(n-k) \times n$ matrix A with rank $n-k$. Here, k is the rank deficiency of Γ. Then, random effects $\mathbf{v}$ can be characterized by the state-transition equation

$$A\mathbf{v} = \mathbf{u} \sim N(\mathbf{0}, \lambda I_{n-k}). \tag{7}$$

Because A does not have a full rank, $\mathbf{v}$ cannot be recovered from $\mathbf{u}$ and thus the model for $(\mathbf{y}, \mathbf{v})$ is not completely specified, which gives rise to model with the SP Γ of $\mathbf{v}$. For inferences about β, Besag and Kempton (1986) proposed the use of the random-effect model for the reduced data (RD):

$$A\mathbf{y} = AX\beta + \mathbf{u} + A\mathbf{e}. \tag{8}$$

Based on a completely specified model of $A\mathbf{y}$, the likelihood inferences can be made for fixed unknown parameters. A question is whether using $A\mathbf{y}$ only can give a fully efficient analysis of parameters as obtainable under using the whole data $\mathbf{y}$. When it comes to prediction, of interest is $\hat{\mathbf{y}}$ rather than $A\hat{\mathbf{y}}$ and we need methods not based on RD, such as Kalman