Hidden Markov Models for Analysis of Defective Industrial Machine Parts

Pornpit Sirima and Premysl Pokorny

***Abstract*— Monthly counts of industrial machine part errors are modeled using two-state hidden Markov models (HMMs) in order to describe the effect of machine part error correction on the likelihood of the machine parts to be in a "defective" or "non-defective" state. A Bayesian framework is used for parameter estimation. The study finds that the machine part error correction does not improve the machine part status of individual part, but there is a very strong month-to-month dependence of machine part states. A comparison shows that the proposed HMM has a better performance than the traditional Poisson generalized estimating equations (GEE) that directly model the counts.**

***Index Terms*— Hidden Markov Models (HMMs), Machine parts errors, Defective and non-defective state, Bayesian framework**

I. Introduction

Industrial machine parts data have provided information useful for addressing questions regarding effectiveness of machine part error correction. Using machine part error records, it is possible to collect the number of machine part errors for each part over an entire time period. This data type has been useful in the study of machine part errors. A model for these count data can then be constructed and used to estimate the effect of the machine error correction.

In the study of industrial machine parts, it is reasonable to hypothesize an unobserved machine part state that governs individual errors with the normal error rate corresponding to a "non-defective" state and an excess errors corresponding to "defective" state. The probability of being in the "defective" or "non-defective" state for a particular part in a given month will differ depending on its past state in which that part was in and other possible covariates including specifically error correction. The terms "defective" and "non-defective" are used throughout this paper as labels for the two different states, but it is important to point out that the two states reflect periods of high and low machine errors which are the surrogates for the concepts of "defective" and "non-defective" respectively. As such there may be periods of frequent machine errors corresponding to a "defective" state being predicted by the model that in the reality of the machine part do not represent a "defective" period in the machine part and vice versa for low use and the "non-defective" state.

In this paper we propose the model for "defective" and "non-defective" unobserved machine states as a hidden Markov chain and since the observations are monthly counts of machine part errors, a two-state Poisson hidden Markov model (HMM) [1, 2] is used. The major research question is whether machine part error correction reduces the probability of subsequently entering the "defective" state as measured by the machine part errors.

We consider fitting the two-state hidden Markov model to the number of machine part errors per month. Estimation of parameters for the HMM can proceed within a frequentist or Bayesian framework. Within the maximum likelihood framework, the EM algorithm, also known as the Baum-Welch algorithm in the HMM literature, can be implemented by treating the hidden states as missing values, implementing the forward backward recursion in the E-step and finding the value of the parameters that maximize the likelihood in the M-step [3-4]. Within the Bayesian framework, the MCMC technique and Metropolis-Hastings algorithms can be used to sample from the posterior distribution of the parameters [5]. Reference [6] pointed out that MCMC methods for HMMs can also be improved by incorporating the forward-backward, likelihood and Viterbi recursive algorithms into the MCMC algorithm, improving convergence as well as computational efficiency. While these algorithms can be incorporated to improve the computational efficiency of the MCMC, it is important to note that the direct Gibbs sampling approach for the HMMs is computationally straightforward and intuitive. Moreover, direct Gibbs sampling can be implemented in the existing OpenBUGS software which, in this paper, is used for fitting the data within a Bayesian framework.

The methodology and application are given in section 2, including HMMs, Bayesian models, Gibbs sampling, accessing MCMC convergence, and an application. The results are presented in section 3. Section 4 and 5 give some discussion and conclusion, respectively.

II. Methodology and Application

A. Hidden Markov Models

Let $\mathbf{y} = (y_1, ..., y_T)^T$ be the vector of observed variables, indexed by time. HMMs [7-8] assume that the distribution of

Manuscript received December 22, 2013; revised January 29, 2014. This work was supported in part by the Department of Manufacturing Systems and Processes, the Faculty of Mechanical Engineering, Technical University of Liberec, Czech Republic.

Pornpit Sirima is with the Department of Manufacturing Systems and Processes, the Faculty of Mechanical Engineering, Technical University of Liberec, Czech Republic (e-mail: sirima_ie@hotmail.com).

Premysl Pokorny is with the Department of Manufacturing Systems and Processes, the Faculty of Mechanical Engineering, Technical University of Liberec, Czech Republic (corresponding author, phone:+420-48-535-3366; e-mail: premysl.pokorny@tul.cz).

ISBN: 978-988-19253-3-6
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

each observed data point y_t depends on an unobserved (hidden) variable, denoted s, that takes on values from 1 to k. The hidden variable $\mathbf{s}=(s_1,\ldots,s_T)^T$ characterizes the "state" which the generating process is at any time t. HMMs further postulate a Markov Chain for the evolution of the unobserved state variable and, hence, the process for s_t is assumed to depend on the past realizations of $\mathbf{y}$ and $\mathbf{s}$ only through s_{t-1}:

$$p(s_t = j \mid s_{t-1} = i) = \lambda_{ij}\,, \tag{1}$$

where λ_{ij} is the generic element of the transition matrix $\mathbf{\Lambda}=(\lambda_{ij})$, with vector of stationary probability $\boldsymbol{\pi}$ satisfying $\boldsymbol{\pi}^T\mathbf{\Lambda}=\boldsymbol{\pi}^T$. Figure 1 illustrates the dependency structure in a HMM. Showing that each observation y_t is the conditionally independent of all other unobserved and observed data, given s_t.

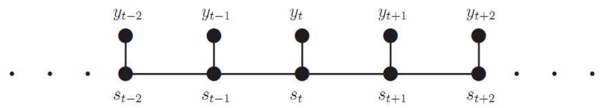

Fig. 1 Dependency structure in a HMM

B. Bayesian Models

Suppose $\mathbf{y}$ is a vector of observations, $\mathbf{y}=(y_1,\ldots,y_m)$, and $\boldsymbol{\theta}$ is a vector of parameters, $\boldsymbol{\theta}=(\theta_1,\ldots,\theta_k)$ that are not observable.

For Bayesian models [9], Let $\mathrm{f}(\mathbf{y}\mid\boldsymbol{\theta})$ represent the probability density function of $\mathbf{y}$ given $\boldsymbol{\theta}$, and $\pi(\boldsymbol{\theta})$ is a prior for $\boldsymbol{\theta}$. Then, the posterior probability density function of $\boldsymbol{\theta}$ is given by

$$\pi(\boldsymbol{\theta}\mid\mathbf{y})=\frac{\mathrm{f}(\mathbf{y}\mid\boldsymbol{\theta})\pi(\boldsymbol{\theta})}{\int \mathrm{f}(\mathbf{y}\mid\boldsymbol{\theta})\pi(\boldsymbol{\theta})\mathrm{d}\boldsymbol{\theta}}\,. \tag{2}$$

The goal of Bayesian inference is to get the posterior. In particular, some numerical summaries may be obtained from the posteriors. For example, to keep things simple, a Bayesian point estimator for a univariate θ is often obtained as the posterior mean:

$$\begin{aligned}\mathrm{E}(\theta\mid\mathbf{y}) &= \int \theta\pi(\theta\mid\mathbf{y})\mathrm{d}\theta \\ &= \frac{\int \theta \mathrm{f}(\mathbf{y}\mid\theta)\pi(\theta)\mathrm{d}\theta}{\int \mathrm{f}(\mathbf{y}\mid\theta)\pi(\theta)\mathrm{d}\theta}\,.\end{aligned} \tag{3}$$

The posterior variance, $\mathrm{var}(\theta\mid\mathbf{y})$, is often used as Bayesian measure of uncertainty. Markov Chain Monte Carlo (MCMC) methods are proposed to handle the computation.

C. Gibbs Sampling

The Gibbs sampling [10] decomposes the joint posterior distribution into full conditional distributions for each parameter in the model and then sample from them. The sampler can be efficient when the parameters are not highly dependent on each other and the full conditional distributions are easy to sample from. It does not require an instrumental proposal distribution as Metropolis methods do. However, while deriving the conditional distributions can be relatively easy, it is not always possible to find an efficient way to sample from these conditional distributions.

Suppose $\boldsymbol{\theta}=(\theta_1,\ldots,\theta_k)^T$ is the parameter vector, $\mathrm{p}(\mathbf{y}\mid\boldsymbol{\theta})$ is the likelihood, and $\pi(\boldsymbol{\theta})$ is the prior distribution. The full posterior conditional distribution of $\pi(\theta_i\mid\theta_j, i\neq j,\mathbf{y})$ is proportional to the joint posterior density; that is, $\pi(\theta_i\mid\theta_j, i\neq j,\mathbf{y})\propto \mathrm{p}(\mathbf{y}\mid\boldsymbol{\theta})\pi(\boldsymbol{\theta})$. For instance, the one-dimensional conditional distribution of θ_1 given $\theta_j=\theta_j^*$, $2\le j\le k$, is computed as

$$\begin{aligned}&\pi(\theta_1\mid\theta_j=\theta_j^*, 2\le j\le k,\mathbf{y}) \\ &\quad= \mathrm{p}(y\mid(\boldsymbol{\theta}=(\theta_1,\theta_2^*,\ldots,\theta_k^*)^T\pi(\boldsymbol{\theta}=(\theta_1,\theta_2^*,\ldots,\theta_k^*)^T)\,.\end{aligned} \tag{4}$$

The Gibbs sampler works as follows:

1. Set $t=0$, and choose an arbitrary initial value of $\boldsymbol{\theta}^0=(\theta^0{}_1,\ldots,\theta_k{}^0)$.
2. Generate each component of $\boldsymbol{\theta}$ as follows:

   draw $\theta_1^{(t+1)}$ from $\pi(\theta_1\mid\theta_2^{(t)},\ldots,\theta_k^{(t)},\mathbf{y})$

   draw $\theta_2^{(t+1)}$ from $\pi(\theta_2\mid\theta_1^{(t+1)},\theta_3^{(t)}\ldots,\theta_k^{(t)},\mathbf{y})$

   …

   draw $\theta_k^{(t+1)}$ from $\pi(\theta_k\mid\theta_1^{(t+1)},\theta_3^{(t+1)}\ldots,\theta_k^{(t+1)},\mathbf{y})$.
3. Set $t=t+1$. If $t<T$, the number of desired samples, return to step 2. Otherwise, stop.

In the MCMC, there are other related processes, called convergence, which are described in the following topics.

D. Assessing MCMC convergence

Simulation-based Bayesian inference requires using simulated draws to summarize the posterior distribution or calculate any relevant quantities of interest. We have to decide whether the Markov chain has reached its stationary, or the desired posterior distribution and to determine the number of iterations to keep after the Markov chain has reached stationarity. Convergence diagnostics help to resolve these issues. Reference [11] discuss about convergence diagnostics. The common ones are visual analysis via trace plots and kernel density plots.

E. An Application

The data were collected from a thermo plastic injection molding machine in a car bumper auto parts manufacturer in Liberec city, Czech Republic. Altogether 27 machine parts were randomly chosen for this study, during the time period from January 2012 to November 2012. The number of machine parts errors were recorded.

A Hidden Markov model for the number of machine part errors Z_{it} is

$$\begin{aligned} & Z_{it} \mid \theta_{it} \sim Pois(\theta_{it}) \\ & \log(\theta_{it}) = \lambda_0 + \lambda_1 C_{it} \\ & C_{it} \mid C_{i(t-1)} \sim \text{Bin}(\text{logit}^{-1}(\beta_0 + \beta_1 C_{i(t-1)} + \beta_2 X_{it}), 1) \\ & C_{it_0} \sim \text{Bin}(1, \pi_1) \end{aligned} \quad (5)$$

where θ_{it}, which can be viewed as the mean of the Poisson, is determined by the unobserved machine part state C_{it}. This unobserved machine parts state follows a Markov chain, with transition probability modeled by a logistic regression with the previous health state $C_{i(t-1)}$. The parameter π_1 represents the initial probability of being in the "defective" machine state at the first month t_{i1}, i.e. $\Pr(C_{it_{i1}} = 1)$. The dummy variables $X_{it} = 1$ indicates the status of month t for part i as being after error correction (with before correction as the reference group, $X_{it} = 0$). Thus, the estimate for the coefficient of X_{it} is of primary interest to see if the probability of being in the "defective" state has significantly decreased after correction.

The hidden Markov model (5) is illustrated in Figure 1. The total number of machine parts error Z_{it} in a particular month t, is governed by the two state latent variable C_{it}. More specifically, Z_{it} comes from a two state Poisson distribution where the two different means of the Poisson distribution correspond to the two different values of the latent variable C_{it} which in turn depends on the previous state $C_{i(t-1)}$. To make the unobserved states identifiable, we assume that the lower mean corresponds to $C_{it} = 0$ and the higher mean corresponds to $C_{it} = 1$, which is operationalized by constraining λ_1 to be larger than zero. Thus $C_{it} = 0$ corresponds to the "non-defective" state and $C_{it} = 1$ corresponds to the "defective" state.

F. Parameter Estimation

The MCMC Gibbs sampling for parameter estimation was done in a Bayesian framework using MCMC techniques via OpenBUGS software. The joint posterior is broken into the full conditional posterior distribution with respect to each parameter and the Gibbs sampler [18] is used. Once the chain converges, the empirical joint posterior distribution for all the parameters can be used to obtain the posterior mean and the 2.5% and 97.5% quantiles can be used as the credible interval for all the parameters. The priors were chosen to be as noninformative as possible. In the total visits model, $N(0,10^5)$ priors were used for λ_0, β_0, β_1, β_2, and $N(0,10^5)$ with positive value restriction was used for λ_1.

The visual analysis, history plots and kernel density plots are used for the MCMC convergence diagnostics. We performed 25,000 MCMC iterations with 5,000 burn-in iterations.

To evaluate the model performance, the proposed model is compared with the traditional Poisson generalized estimating equations (GEE) that directly model the counts, using mean square errors (MSE).

The GEE is expressed as:

$$\begin{aligned} & Z_{it} \mid \theta_{it} \sim Pois(\theta_{it}) \\ & \log(\theta_{it}) = \beta_0 + \beta_1 X_{it} \ , \end{aligned} \quad (5)$$

where β_0 and β_1 are regression coefficients, the dummy variables $X_{it} = 1$ indicates the status of month t for part i as being after error correction. We use SPSS software for the GEE parameter estimation.

III. Results

The mean number of machine part error was 2.35 per one part per one month. As the observed data represent the count of machine parts errors, direct modeling of the data via GEE is considered. The visual analysis is used for MCMC convergence diagnostics. The trace plots are shown in Fig. 2-6 and the kernel density plots are shown in Fig. 7-11. The chains moving around the parameter spaces and the kernel densities looking like their distributions indicate that each parameter is converged to a stationary density.

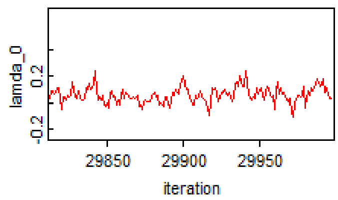

Fig. 2 Trace of λ_0

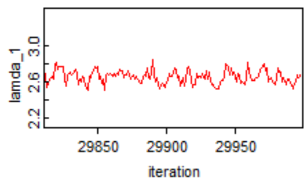

Fig. 3 Trace of λ_1

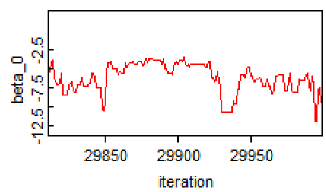

Fig. 4 Trace of β_0

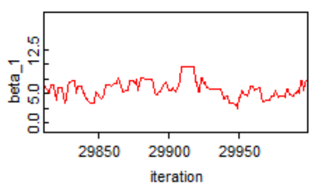

Fig. 5 Trace of β_1

ISBN: 978-988-19253-3-6
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

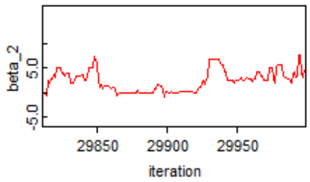

Fig. 6 Trace of β_2

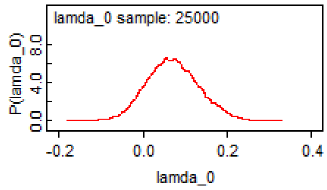

Fig. 7 Kernel density of λ_0

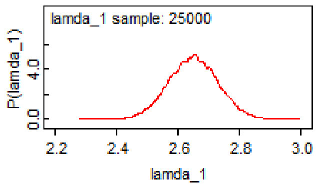

Fig. 8 Kernel density of λ_1

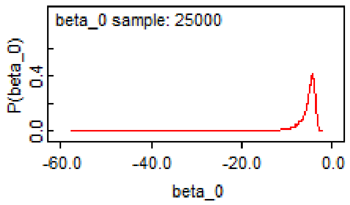

Fig. 9 Kernel density of β_0

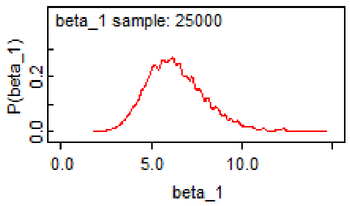

Fig 10. Kernel density of β_1

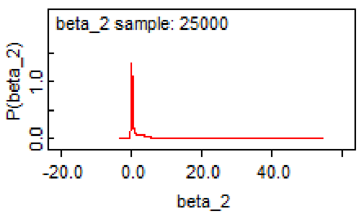

Fig 11. Kernel density of β_2

The posterior summary of the estimated parameters is shown in Table 1.

TABLE I
PARAMETER ESTIMATES FROM THE HMM

Parameter	Mean	SD	95% Credible Interval	
λ_0	0.071	0.063	-0.045	0.198
λ_1	2.649	0.080	2.492	2.808
β_0	-5.469	2.491	-12.08	-3.336
β_1	6.393	1.66	3.655	10.14
β_2	1.284	2.593	-0.3057	7.868

Table 1 shows the results of the HMM fit to the machine part error counts per month. The estimate of β_2 (1.284) implies that the odds of transitioning to or remaining in the "defective" state in any given month after correction is exp(1.284) = 3.611 of what it was before correction. This provides evidence in favor of the machine part error correction not improving the machine part status of individual part.

In addition to this main finding, the results from the model also characterize a very strong month-to-month dependence of machine part states ($\beta_1 =$ 6.393) where the odds of remaining in the "defective" state in the current month if an machine part was in the "defective" state the last month is estimated to be exp(6.393) = 597.647 times the odds of newly transitioning to the "defective" state if an individual part was "non-defective" in the previous month.

For the model comparison, the mean square errors (MSE) of the proposed HMM (3.193) is smaller than the GEE (4.948), indicating that it has a better performance.

IV. DISCUSSION

We propose a HMM for machine part errors. The model assumes there are unobserved machine part states that govern the machinery care utilization of a particular machine part, and the machine part state is governed only by the frequency of errors.

The main goal of the HMM is to model changing machine part states over time not necessarily modeling the changing number of errors. In the HMM, the observed machine errors are really only a surrogate for "machine parts status". Measurement error is allowed between the observed machine errors and the underlying machine state. For the GEE, the goal is to model the changing numbers of machine part themselves. As being seen from the model comparison using mean square errors (MSE), the GEE does not fit as well. Reference [12] give a review and description of several different Markov and latent (hidden) Markov models. The proposed HMMs can be applied to other similar problems and can be extended to multivariate Poisson data.

ISBN: 978-988-19253-3-6
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

V. CONCLUSION

The objective of this study is to propose a two-state hidden Markov model in order to describe the effect of machine part error correction on the likelihood of the machine parts to be in a “defective” or “non-defective” state. A Bayesian framework is used for parameter estimation. The study finds that the machine part error correction does not improve the machine part status of individual part, and there is a very strong month-to-month dependence of machine part states. Using root mean square errors, the proposed HMM is compared to the GEE that directly model the counts. The results show that the proposed model has a better performance.

REFERENCES

[1] W. M. Melanie, and L. Ran, “Multiple indicator hidden Markov model with an application to medical utilization data,” *Stat Med*, vol. 28 no.2, pp. 293–310, 2009.

[2] I. L. MacDonald, and W. Zucchini, *Hidden Markov and other Models for Discrete-valued Time Series*. New York: Chapman and Hall, 1997.

[3] L. E. Baum, T. Petrie, G. Soules, and N. Weiss, “A maximization technique occurring in the statistical analysis of probabilistic functions of Markov chains,” *The Annuals of Mathematical Statistics*, vol. 41, pp. 164–171, 1970.

[4] J. A. Bilmes, “A Gentle Tutorial of the EM Algorithm and its Application to Parameter Estimation for Gaussian Mixture and Hidden Markov Models,” *Technical Report, University of Berkeley, ICSITR*, pp. 97-121, 1998.

[5] J. H. Albert, S. Chib, “Bayes Inference via Gibbs Sampling of Autoregressive Time Series Subject to Markov Mean and Variance Shifts,” *Journal of Business and Economic Statistics*, vol. 11, pp. 1–15, 1993.

[6] S. L. Scott, “Bayesian Methods for Hidden Markov Models: Recursive Computing in the 21st Century,” *Journal of the American Statistical Association*, vol. 97, pp. 337–351, 2002.

[7] A. M. Poritz, “Hidden Markov models: a guided tour,” *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing* , vol. 1, pp. 7-13, 1988.

[8] L. R. Rabiner, “A tutorial on hidden Markov models and selected applications in speech recognition,” *Proceedings of the IEEE*, vol. 77, pp. 257-285, 1989.

[9] P. Congdon, *Bayesian Statistical Modelling*. 2nd ed. New York: John Wiley and Sons, 2006.

[10] S. Geman, D. Geman, “Stochastic Relaxation Gibbs Distributions, and the Bayesian Restoration of Images,” *IEEE Transactions of Pattern Analysis and Machine Intelligence*, vol. 6, pp. 721-74, 1984.

[11] S. P. Brooks, and G. O. Roberts, “Assessing Convergence of Markov Chain Monte Carlo Algorithms,” *Statistics and Computing*, vol. 8, pp. 319-335, 1998.

[12] R. Langeheine, F. van de Pol, *Latent Markov Chains*, *Applied Latent Class Analysis*, Cambridge University Press; pp. 304-341, 2002.