



BOOTSTRAPPING METHOD FOR CONSTRUCTING CONFIDENTIAL INTERVALS FOR ESTIMATED PARAMETERS OF THE HIDDEN MARKOV MODEL

Yosuke Inaba and Etsuo Miyaoka

Department of Information Sciences

Tokyo University of Science

1-3 Kagurazaka, Shinjuku-ku

Tokyo, Japan

e-mail: yosuke.inaba@ppdsnbl.co.jp

Abstract

The Hidden Markov Model (HMM) is a statistical model for analysing time series data where the observational data $X^{(T)} = (X_1, X_2, \dots, X_T)$ depends on unobservable states $C^{(T)} = (C_1, C_2, \dots, C_T)$. HMM is used in such fields as speech recognition and bioinformatics, and has recently gained popularity in life science as well. In HMM, we need to estimate the parameters, and constructing their confidential interval involves some difficulties. To solve this problem, we propose a bootstrapping method to construct confidential intervals and apply it to earthquake data. When applying this method, we need to choose an appropriate number of states, hence we calculated each AIC score, and then made comparisons.

Received: June 11, 2015; Accepted: July 15, 2015

2010 Mathematics Subject Classification: 62-XX.

Keywords and phrases: Hidden Markov Model, HMM, bootstrap, likelihood-function, Baum-Welch, forward-backward algorithm.

Communicated by K. K. Azad

1. Introduction

The Hidden Markov Model (HMM) is a statistical model for analysing time series data where the observational data $X^{(T)} = (X_1, X_2, \dots, X_T)$ depends on unobservable states $C^{(T)} = (C_1, C_2, \dots, C_T)$. HMM is used in such areas as speech recognition and bioinformatics, and has recently gained popularity in life science as well. In analysis of HMM, we need to estimate the parameters, and there exists a strong tool for estimating, that is, the Baum-Welch algorithm [3]. The Baum-Welch algorithm is an EM algorithm for HMM. Durbin et al. applied the Baum-Welch algorithm to biological sequence analysis [2]. Constructing confidential intervals involves some difficulties. In the maximum likelihood estimation, the estimated parameter usually has asymptotic normality, and as such, Konishi et al. described methods that use the Fisher information matrix [4]. However, in HMM, this method cannot be applied because HMM is not a regular model, and its estimated parameters do not have asymptotic normality [3]. To solve this problem, we proposed a bootstrapping method to construct confidential intervals and apply it to earthquake data. When applying HMM, we need a criterion to model comparison, and we use AIC [5] to compare model fitness for each number of states.

2. Definition

2.1. Markov chain

Let $\mathbf{C}^{(t)} = \{C_t \in \{1, \dots, m\} | t \in \mathbb{N}\}$ be a discrete sequence of random variables. $\mathbf{C}^{(t)}$ is called a *Markov chain* when it satisfies the following:

$$\Pr(C_{t+1} | \mathbf{C}^{(t)}) = \Pr(C_{t+1} | C_t) \text{ for any } t \text{ in } \mathbb{N}.$$

The parameter for the Markov chain $\gamma_{ij} := \Pr(C_{s+t} = j | C_s = i)$ is called the *transition probability*. A Markov chain is termed *homogenous* when γ_{ij} is independent of s . In matrix notation,

$$\mathbf{\Gamma} := (\gamma_{ij}) : \text{transition probability matrix.}$$

The first element of the sequence is special, because it does not have an adjacent element. As such, we have to define an initial distribution δ as $\delta := \{\Pr(C_1 = 1), \Pr(C_1 = 2), \dots, \Pr(C_1 = m)\}$, $\mathbf{1} := (1, \dots, 1)$, $\delta \mathbf{1}' = 1$. In particular, if $\delta \Gamma = \delta$, the Markov chain is deemed *stationary*.

2.2. Poisson distribution

A discrete random variable X is Poisson distribution for $\lambda > 0$ when its probability function is given by $p(X = k) = \frac{\lambda^k e^{-\lambda}}{k!}$. Note that the Poisson distribution has the reproductive property:

$$X \sim \text{Poisson}(\lambda), Y \sim \text{Poisson}(\mu) \Rightarrow X + Y \sim \text{Poisson}(\lambda + \mu).$$

2.3. HMM

Let $\mathbf{C}^{(t)}$ be a parameter process and $\mathbf{X}^{(t)}$ be the output. We call $\{\mathbf{C}^{(t)}, \mathbf{X}^{(t)}\}$ an *HMM* if it satisfies the following:

$$\Pr(C_t | \mathbf{C}^{(t-1)}) = \Pr(C_t | C_{t-1}),$$

$$\Pr(X_t | \mathbf{X}^{(t-1)}, \mathbf{C}^{(t)}) = \Pr(X_t | C_t).$$

The parameters for HMM are:

$\lambda := (\lambda_1, \dots, \lambda_m)$: state-dependent distribution parameter.

$\Gamma := (\gamma_{ij})$: transition probability matrix.

$\delta := (\delta_1, \dots, \delta_m)$: initial distribution.

2.4. Likelihood function for HMM

In the case of discrete observations, for $m = 1, 2, \dots, m$,

$$p_i(x) := \Pr(X_t = x | C_t = i).$$

We define p_i to be the probability *density* function of X_t if the Markov chain is in state i at time t .

In matrix notation

$$\mathbf{P}(x) := \begin{pmatrix} p_1(x) & & \\ & \ddots & \\ & & p_m(x) \end{pmatrix}.$$

The likelihood function L_T is then defined as follows:

$$L_T := \Pr(X^{(T)} = x^{(T)}) = \sum_{c_1, c_2, \dots, c_T=1}^m \Pr(X^{(T)} = x^{(T)} | C^{(T)} = c^{(T)}).$$

2.5. Forward and backward probabilities

We define vector \mathbf{a}_t , for $t = 1, 2, \dots, T$, as

$$\mathbf{a}_t := \delta \mathbf{P}(x_1) \mathbf{\Gamma P}(x_2) \mathbf{\Gamma P}(x_3) \cdots \mathbf{\Gamma P}(x_t) \quad t \in \{1, \dots, T\}.$$

The elements of \mathbf{a}_t are referred to as *forward probabilities*.

Then, the following holds:

$$\begin{aligned} \mathbf{a}_t &= \mathbf{a}_{t-1} \mathbf{\Gamma P}(x_t), \\ \alpha_t(j) &= \Pr(X^{(t)} = \mathbf{x}^{(t)}, C_t = j). \end{aligned}$$

We define the *backward probabilities* in the same way:

$$\mathbf{\beta}_t := \mathbf{\Gamma P}(x_{t+1}) \mathbf{\Gamma P}(x_{t+2}) \cdots \mathbf{\Gamma P}(x_T) \mathbf{1}' \quad t \in \{1, \dots, T-1\}.$$

Then, the following holds:

$$\beta_t(i) = \mathbf{P}(X_{t+1}^{(T)} = x_{t+1}^{(T)} | C_t = i).$$

In particular, for $t \in \{1, \dots, T\}$,

$$\alpha_t(i) \beta_t(i) = \mathbf{P}(X^{(T)} = x^{(T)}, C_t = i),$$

$$\mathbf{a}_t \mathbf{\beta}_t = \mathbf{P}(X^{(T)} = x^{(T)}) = L_T,$$

$$\mathbf{P}(C_t = j | X^{(T)} = x^{(T)}) = \alpha_t(j) \beta_t(j) / L_T,$$

$$\mathbf{P}(C_{t-1} = j, C_t = k | X^{(T)} = x^{(T)}) = \alpha_{t-1}(j) \gamma_{jk} p_k(x_t) \beta_t(k) / L_T.$$

3. Baum-Welch Algorithm

3.1. Parameter estimation

Baum-Welch algorithm is an EM algorithm for HMM that treats hidden states as missing data. To employ the EM algorithm, we define the following functions. Let c_1, c_2, \dots, c_T be the sequence of states. Then, we have

$$u_j(t) = \begin{cases} 1(c_t = j) \\ 0(c_t \neq j) \end{cases} \quad (t = 1, 2, \dots, T),$$

$$v_{jk}(t) = \begin{cases} 1(c_{t-1} = j \text{ and } c_t = k) \\ 0(c_{t-1} \neq j \text{ or } c_t \neq k) \end{cases} \quad (t = 2, 3, \dots, T).$$

We also define the complete data log likelihood (CDLL) as follows:

$$\begin{aligned} CDLL(\delta, \gamma, p) &:= \log(\Pr(\mathbf{X}^{(T)}, \mathbf{c}^{(T)} | \delta, \gamma, p)) \\ &= \log \left(\delta_{c_1} \prod_{t=2}^T \gamma_{c_{t-1}, c_t} \prod_{t=1}^T p_{c_t}(x_t) \right) \\ &= \log(\delta_{c_1}) + \sum_{t=2}^T \log(\gamma_{c_{t-1}, c_t}) + \sum_{t=2}^T \log(p_{c_t}(x_t)). \end{aligned}$$

The EM algorithm, in general, is an iterative method for performing maximum likelihood estimation. The algorithm for HMMs proceeds as follows:

- **E step.** Replace all quantities $v_{jk}(t)$ and $u_j(t)$ by their conditional expectations given observations $\mathbf{x}^{(T)}$ (and given current parameter estimates):

$$\begin{aligned} \hat{u}_j(t) &= \Pr(C_t = j | \mathbf{X}^{(T)} = \mathbf{x}^{(T)}) = \alpha_t(j)\beta_t(j)/L_T, \\ \hat{v}_{jk}(t) &= \Pr(C_{t-1} = j, C_t = k | \mathbf{X}^{(T)} = \mathbf{x}^{(T)}) \\ &= \alpha_{t-1}(j)\gamma_{jk}p_k(x_t)\beta_t(k)/L_T. \end{aligned}$$

- **M step.** Having replaced $v_{jk}(t)$ and $u_j(t)$ by $\hat{v}_{jk}(t)$ and $\hat{u}_j(t)$, maximize CDLL with respect to three sets of parameters: δ , γ , and p .

3.2. Standard error and confidential interval

Although the point estimates are easy to compute, exact interval estimates are not available. It is known that, in the context of maximum likelihood estimation, when data volume is big enough, maximum likelihood estimators can be normal and efficient under some regulatory conditions [4]. However, this might not work in some situations, for example, when some parameters are on or near the boundary of their parameter process. We propose a bootstrapping method to solve this problem.

4. Bootstrapping Method

We apply the following bootstrapping method. Roughly speaking, the idea of a parametric bootstrap is to assess the properties of model Θ by using those of a model with parameters $\hat{\Theta}$. The following steps are performed to estimate the variance-covariance matrix of $\hat{\Theta}$.

1. Fit the model, that is, compute Θ .
2. Generate a sample $X^{(T)}$ using the parameters estimated above.
3. Fit the model again by the sample.
4. Repeat steps 2-3 2000 times.
5. Calculate the 5% and 95% percentile points for the parameters.

5. Application

In this section, we apply the above method to the series of annual count of major earthquakes (i.e., magnitude 7 or greater for years 1900-2006, which is downloaded from the National Earthquake Information Center (NEIC) site (<http://earthquake.usgs.gov/data>)). The data is shown in Figure 1.

First, we have to choose an appropriate number of states m , and we therefore compare AIC scores [3]:

$$\text{AIC} = -2\log L + 2p.$$

Result is shown in Table 1. $m = 3$ minimizes the score, and so we choose $m = 3$. Then, we estimate the HMM parameters using the Baum-Welch algorithm, and then estimate the confidence intervals. The result is shown in Figure 1.

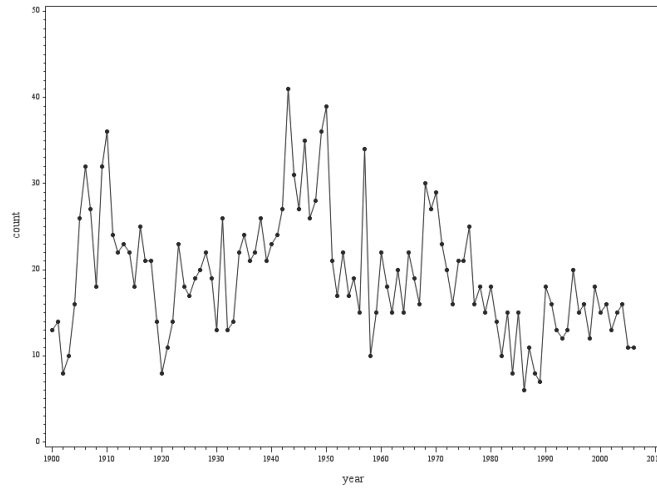


Figure 1. Major earthquakes (magnitude 7 or greater).

Table 1. AIC scores

m	$-\log L$	AIC
1	391.919	785.838
2	341.879	693.757
3	328.527	679.055
4	327.359	692.717
5	327.071	712.141
6	324.575	731.150

Table 2. Confidence intervals

parameter	Baum-Welch			
	est	90% conf.		S.E.
λ_1	13.133	11.807	14.584	0.995
λ_2	19.713	17.779	21.417	1.177
λ_3	29.709	25.079	33.651	2.86
γ_{11}	0.939	0.727	0.981	0.119
γ_{12}	0.032	0	0.177	0.092
γ_{13}	0.028	0	0.155	0.088
γ_{21}	0.04	0	0.167	0.073
γ_{22}	0.906	0.699	0.97	0.116
γ_{23}	0.053	0	0.197	0.089
γ_{31}	0	0	0.04	0.055
γ_{32}	0.19	0.055	0.736	0.211
γ_{33}	0.809	0.221	0.932	0.214

6. Conclusion

We construct confidence intervals for HMM parameters estimated using an EM algorithm. We compare them to existing parameters and confirm their applicability.

References

- [1] W. Zucchini and L. L. MacDonald, Hidden Markov Models for Time Series, CRC Press, 2009.

- [2] R. Durbin, S. Eddy, A. Korgh and G. Mitchison, Biological Sequence Analysis: Probabilistic Models of Proteins and Nucleic Acids, Cambridge University Press, 1998.
- [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, Ann. Math. Statist. 41 (1970), 164-171.
- [4] S. Konishi, Y. Ochi and H. Ohmori, Methods of Computational Statistics, Asakura Publishing Co. Ltd., 2008.
- [5] H. Akaike, Information theory and an extension of the maximum likelihood principle, Proceedings of the 2nd International Symposium on Information Theory, Akadimiai Kiado, Budapest, 1973, pp. 267-281.