

Maximum entropy test for infinite order autoregressive models[†]

Sangyeol Lee¹ · Jiyeon Lee² · Jungsik Noh³

¹²³Department of Statistics, Seoul National University

Received 2 March 2013, revised 26 March 2013, accepted 1 April 2013

Abstract

In this paper, we consider the maximum entropy test in infinite order autoregressive models. Its asymptotic distribution is derived under the null hypothesis. A bootstrap version of the test is discussed and its performance is evaluated through Monte Carlo simulations.

Keywords: Goodness of fit test, infinite order autoregressive models, maximum entropy measure.

1. Introduction

The maximum entropy principle (Jaynes, 1963) is well known as a criterion for selecting a priori probabilities. Maximum entropy modeling has been successfully applied to diverse research fields. For a probability density function $f(x)$, Boltzmann-Shannon entropy is defined as

$$H(f) = - \int_{-\infty}^{\infty} f(x) \log(f(x)) dx. \quad (1.1)$$

Forte and Hughes (1988) proposed the function $-\sum p_i \log(p_i/(x_i - x_{i-1}))$ as a discrete analogue of (1.1), where $p_i = P[x_{i-1} < X \leq x_i] = \int_{x_{i-1}}^{x_i} f(x) dx$, $i = 1, \dots, n-1$ and $a = x_0 < \dots < x_n = b$.

Goodness of fit (gof) tests measure the degree of agreement between the distribution of an observed random sample and a theoretical statistical distribution. In time series analysis, the gof test problem has been a crucial issue for modeling time series. In particular, the normality test has attracted much attention from many researchers, since the normality of time series ensures several advantageous properties that non-normal time series do not possess. As relevant references, we refer to Lee and Wei (1999), Lee and Na (2001) and Lee

[†] This research was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education, Science and Technology (2011-0010936) grant number: 2011-0010936.

¹ Professor, Department of Statistics, Seoul National University, Seoul 151-747, Korea.

² Graduate student, Department of Statistics, Seoul National University, Seoul 151-747, Korea.

³ Corresponding author: Post doctor, Department of Statistics, Seoul National University, Seoul 151-747, Korea. E-mail: nohjssunny@gmail.com

et al. (2010). Recently, Lee *et al.* (2011) developed a maximum entropy test in iid settings and demonstrated its usefulness. Since the test outperforms several existing gof tests, in this study, we consider applying the maximum entropy test to infinite order autoregressive time series models. In Section 2, we introduce the test statistic and summarize its asymptotic distribution in iid settings. In Section 3, we apply the test to infinite order autoregressive models. In Section 4, we perform a simulation study. Particularly, a bootstrap method is employed.

2. Maximum entropy test

Let $Y_i, i = 1, \dots, n$, be a random sample from a distribution with unknown distribution function F and consider the following test of fit:

$$H_0 : F = F_0 \quad \text{vs.} \quad F \neq F_0. \quad (2.1)$$

Lee *et al.* (2011) considered the following generalization of Forte and Hughes (1988) entropy:

$$S^w(F) = - \sum_{i=1}^m w_i (F(s_i) - F(s_{i-1})) \log \left(\frac{F(s_i) - F(s_{i-1})}{s_i - s_{i-1}} \right), \quad (2.2)$$

where the w 's are appropriate weight functions with $0 \leq w_i \leq 1$ and $\sum_{i=1}^m w_i = 1$, m is the number of disjoint intervals for partitioning the data range, and $-\infty < a \leq s_1 \leq \dots \leq s_m \leq b < \infty$ are preassigned partition points. They proposed an entropy test and obtained the following result.

Theorem 2.1 Let Y_1, \dots, Y_n , be a random sample from a continuous distribution with cumulative distribution function F . Under H_0 given in (2.1), as $n \rightarrow \infty$, we have

$$\sqrt{n} \sup_{w \in W} |S^w(F_n)| \xrightarrow{d} \sup_{w \in W} \left| \sum_{i=1}^m w_i (\mathcal{B}(s_i) - \mathcal{B}(s_{i-1})) \right|, \quad (2.3)$$

where F_n is the sample distribution based on $U_i = F_0(Y_i)$, $\mathcal{B}(s)$ is a Brownian bridge on $[0,1]$, W denotes the space of bounded weights $w = (w_1, \dots, w_m)$ with $0 < w_i < 1$ and $\sum_{i=1}^m w_i = 1$, and $0 = s_0 \leq s_1 \leq \dots \leq s_m = 1$.

It is common to test the composite null hypothesis that the unknown distribution belongs to a parametric family $\{F_\theta\}_{\theta \in \Theta}$, where Θ is an open subset in R^k . In this case, a consistent estimator $\hat{\theta}$ can be used to test $H_0 : F = F_\theta$ vs. $H_1 : \text{not } H_0$ by using $F_{\hat{\theta}}(Y_i) = \hat{U}_i$. As discussed in Durbin (1975), the limiting distribution is affected by the estimation of θ . The effect, though, may diminish when m is large and $\max_i (s_i - s_{i-1})$ is small.

3. Infinite order autoregressive models

In this section, we consider the maximum entropy test for the AR(∞) model satisfying the difference equation

$$X_t - \sum_{j=1}^{\infty} \beta_j X_{t-j} = \varepsilon_t, \quad (3.1)$$

where ε_t are iid r.v.'s with mean zero, unknown variance σ^2 and finite fourth moment, and the function $A(z) = 1 - \sum_{j=1}^{\infty} \beta_j z^j$ is analytic on an open neighborhood of the closed unit disk D in the complex plane and has no zeroes on D . This assumption implies that the coefficients β_j are geometrically bounded. The model (3.1) covers a broad class of stationary processes including causal and invertible ARMA(p, q) models.

Suppose that one wishes to test $H'_0 : \varepsilon_t \sim F_0(\cdot/\sigma)$, $\sigma > 0$ vs. $H'_1 : \text{not } H'_0$, where F_0 is strictly increasing and twice differentiable and satisfies $\int x dF_0(x) = 0$.

In order to perform a test, we fit an AR(q) model to the data, where $q = q_n$ is a sequence of positive integers that is no more than n and diverges to ∞ . Assume that X_1, \dots, X_n are observed. Solving the equation:

$$\frac{d}{d\beta_j} \sum_{t=q+1}^n (X_t - \sum_{j=1}^q \beta_j X_{t-j})^2 = 0, \quad j = 1, \dots, q,$$

we estimate $\beta_n = (\beta_1, \dots, \beta_q)'$ by $\hat{\beta}_n = (\hat{\beta}_1, \dots, \hat{\beta}_q)'$, where

$$\hat{\beta}_n = \left(\sum_{t=q+1}^n \mathbf{X}_{t-1} \mathbf{X}'_{t-1} \right)^{-1} \sum_{t=q+1}^n \mathbf{X}_{t-1} X_t.$$

We define the residuals $\hat{\varepsilon}_t = X_t - \hat{\beta}'_n \mathbf{X}_{t-1}$, where $\mathbf{X}_t = (X_t, \dots, X_{t-q+1})'$, and estimate σ^2 by $\hat{\sigma}_n^2 = \frac{1}{n-q} \sum_{t=q+1}^n \hat{\varepsilon}_t^2$. The following is very useful to measure the accuracy of the above LSE (cf. Lee and Wei (1999)).

Lemma 3.1 Suppose that q satisfy $n^{-1/2} q^2 \log n \rightarrow 0$ and $n^{7/4} q r^q \rightarrow 0$ for all $r \in (0, 1)$. Then, $\|\hat{\beta}_n - \beta_n\|^2 = O_P(q/n)$.

Remark 3.1 A typical q satisfying the assumptions in Lemma 3.1 is $c(\log n)^d$ with $c, d > 0$, or un^{-v} with $u > 0, 0 < v < 1/4$.

Based on the result of Lemma 3.1, Lee and Wei (1999) showed that under some regularity conditions, the residual empirical process

$$\hat{\mathcal{V}}_n(s) = \sqrt{n}(\hat{F}_n(s) - s), \quad 0 \leq s \leq 1$$

with

$$\hat{F}_n(s) = \frac{1}{n} \sum_{t=1}^n I(F_0(\hat{\varepsilon}_t/\hat{\sigma}_n) \leq s)$$

can be expressed as

$$\hat{\mathcal{V}}_n(s) = \mathcal{V}_n(s) + \frac{\sqrt{n}(\hat{\sigma}_n^2 - \sigma_0^2)}{2\sigma_0^2} f_0(F_0^{-1}(s)) F_0^{-1}(s) + \Delta_n(s), \tag{3.2}$$

where σ_0 denotes the true value of σ under H'_0 , $\mathcal{V}_n(s) = \sqrt{n}(F_n(s) - s)$, $f_0 = F'_0$, $F_n(s) = \frac{1}{n} \sum_{t=1}^n I(F_0(\varepsilon_t/\sigma_0) \leq s)$, and $\sup_s |\Delta_n(s)| = o_P(1)$. Furthermore, we assume that

$$\lim_{|x| \rightarrow \infty} |x f_0(x)| = 0 \text{ and } \sup_x |f'_0(x)| < \infty. \tag{3.3}$$

Then, we have the following result.

Theorem 3.1 Under H'_0 and (3.3), if $\max_{1 \leq i \leq m} |s_i - s_{i-1}| \rightarrow 0$ as $m \rightarrow \infty$, we have that for large m , as $n \rightarrow \infty$,

$$\hat{T}_n := \sqrt{n} \sup_{w \in W} |S_{\max}^w(\hat{F}_n)| \stackrel{d}{\approx} \sup_{w \in W} \left| \sum_{i=1}^m w_i (\mathcal{B}(s_i) - \mathcal{B}(s_{i-1})) \right|,$$

where the symbol $A_n \stackrel{d}{\approx} A$ as $n \rightarrow \infty$ indicates that the limiting distribution of A_n is approximately the same as the distribution of A as n tends to ∞ .

Proof: Express

$$S_{\max}^w(\hat{F}_n) = - \sum_{i=1}^m w_i \left(\hat{F}_n(s_i) - \hat{F}_n(s_{i-1}) \right) \cdot \log \left(\frac{\hat{F}_n(s_i) - \hat{F}_n(s_{i-1})}{s_i - s_{i-1}} - 1 + 1 \right).$$

Recall that $\hat{F}_n(s) \rightarrow s$ in probability under the null. Then, by using the fact that $|\log(1+x) - x| \leq x^2$ for $|x| < 1/2$, due to (3.2) (see the arguments in and below (8) of Lee *et al.* (2011)), we have

$$\begin{aligned} S_{\max}^w(\hat{F}_n) &= - \sum_{i=1}^m w_i \left(\frac{\hat{F}_n(s_i) - \hat{F}_n(s_{i-1})}{s_i - s_{i-1}} \right) \\ &\quad \cdot \left[\left(\hat{F}_n(s_i) - s_i \right) - \left(\hat{F}_n(s_{i-1}) - s_{i-1} \right) \right] + o_P(1/\sqrt{n}) \\ &= \frac{-1}{\sqrt{n}} \sum_{i=1}^m w_i \left(\frac{\hat{F}_n(s_i) - \hat{F}_n(s_{i-1})}{s_i - s_{i-1}} \right) \left(\hat{\mathcal{V}}_n(s_i) - \hat{\mathcal{V}}_n(s_{i-1}) \right) + o_P(1/\sqrt{n}), \end{aligned}$$

so that since $\mathcal{V}_n \xrightarrow{w} \mathcal{B}$, for large m , as $n \rightarrow \infty$,

$$\begin{aligned} &\sup_{w \in W} \left| \sqrt{n} S_{\max}^w(\hat{F}_n) + \sum_{i=1}^m w_i \frac{\sqrt{n}(\hat{\sigma}_n^2 - \sigma_0^2)}{2\sigma_0^2} \left\{ f_0(F_0^{-1}(s_i))F_0^{-1}(s_i) - f_0(F_0^{-1}(s_{i-1}))F_0^{-1}(s_{i-1}) \right\} \right| \\ &\stackrel{d}{\approx} \sup_{w \in W} \left| \sum_{i=1}^m w_i (\mathcal{V}_n(s_i) - \mathcal{V}_n(s_{i-1})) \right| \stackrel{d}{\approx} \sup_{w \in W} \left| \sum_{i=1}^m w_i (\mathcal{B}(s_i) - \mathcal{B}(s_{i-1})) \right|. \end{aligned}$$

Hence, $\hat{T}_n \stackrel{d}{\approx} \sup_{w \in W} \left| \sum_{i=1}^m w_i (\mathcal{B}(s_i) - \mathcal{B}(s_{i-1})) \right|$. This completes the proof. □

Remark 3.2 In actual implementation, we introduce iid $w_i^{(l)}, l = 1, \dots, L$ from $U[0, 1]$, where L is a fixed positive integer. Then, by putting $w_{li} = \frac{w_i^{(l)}}{w_1^{(l)} + \dots + w_m^{(l)}}$ and by taking $s_i = i/m, i = 1, \dots, m$, we finally employ the test:

$$\hat{T}_n = \sqrt{n} \max_{1 \leq l \leq L} \left| \sum_{i=1}^m w_{li} (\hat{F}_n(i/m) - \hat{F}_n((i-1)/m)) \right|. \tag{3.4}$$

For more details, see (9)-(10) and Remark 2 of Lee *et al.* (2011).

Remark 3.3 Theorem 3.1 only provides an asymptotic result and cannot be directly applied to small samples. Therefore, in this study, we recommend to use a bootstrap method as in Stute *et al.* (1993):

- (1) Based on the data X_1, \dots, X_n , obtain the LSE $\hat{\beta}_n$ and $\hat{\sigma}_n^2$.
- (2) Generate $\epsilon_1^*, \dots, \epsilon_n^*$ that follow $F_0(\cdot/\hat{\sigma}_n)$ and construct X_1^*, \dots, X_n^* by using the equation: $X_t^* = \hat{\beta}_n' \mathbf{X}_{t-1}^* + \epsilon_t^*$ with β replaced by its LSE with $q = [\log n]^2$ by letting $X_i^* = 0$ for all $i \leq 0$. Then, calculate \hat{T}_n^* in (3.4) with a preassigned m based on these r.v.s.
- (3) Repeat the above procedure B times and calculate the $100(1 - \alpha)\%$ percentile of the obtained B number of \hat{T}_n^* values.
- (4) Reject H_0 if the \hat{T}_n value based on the original observations is larger than the obtained $100(1 - \alpha)\%$ percentile in (3).

The above bootstrap method is easy to implement and gives satisfactory results as seen in the next section.

4. Simulation results

In this section, we evaluate the performance of the maximum entropy test in (3.4) through a simulation study comparing with the Kolmogorov-Smirnov (KS) test. In this simulation study, we consider the ARMA(1,1) model where ϵ_t are assumed to be iid $N(0, 1)$ for the null hypothesis H_0 . For alternative, we consider that ϵ_t are assumed to follow a normal mixture distribution and the Student's t -distribution:

$$\begin{aligned} H_1^{(1)} : \epsilon_t &\sim p N(0, 1) + (1 - p) N(0, \sigma_0^2), & p = 0.9, \sigma_0^2 = 10, \\ H_1^{(2)} : \epsilon_t &\sim p N(0, 1) + (1 - p) N(0, \sigma_0^2), & p = 0.9, \sigma_0^2 = 25, \\ H_1^{(3)} : \epsilon_t &\sim t(4). \end{aligned}$$

We examine the empirical sizes and powers of the test with sample sizes n and m as $(n, m) = (100, 5), (300, 5),$ and $(500, 5)$ at the nominal levels 0.01, 0.05, and 0.1. The empirical sizes and powers are calculated as the number of rejections of the null hypothesis H_0 out of 1000 repetitions. In performing a test, we consider the bootstrap test discussed in Remark 3.3 with $B=500$ and explore the sizes and powers for the AR coefficient and MA coefficient $(\phi, \theta) = (0.5, 0.3)$. Table 4.1 shows that the empirical sizes are close to the nominal levels and Table 4.2 shows that the bootstrap test produces reasonably good powers. The figures in the parentheses are for the KS test. Although \hat{T}_n does not produce good powers for sample sizes $n=100$, the powers increase remarkably either as n increases or as σ_0^2 increases. We do not report details, we also evaluated \hat{T}_n for various AR coefficients and MA coefficients. The results appear to be similar to the ones reported here. Overall, our result strongly suggests that the maximum entropy test performs appropriately.

Table 4.1 Empirical sizes based on the bootstrap method with $B = 500$

(ϕ, θ)	$m = 5, n = 100$			$m = 5, n = 300$			$m = 5, n = 500$		
	0.01	0.05	0.10	0.01	0.05	0.10	0.01	0.05	0.10
(0.5, 0.3)	0.014 (0.002)	0.052 (0.023)	0.105 (0.063)	0.012 (0.003)	0.056 (0.027)	0.107 (0.067)	0.011 (0.003)	0.055 (0.031)	0.102 (0.081)
(-0.5, 0.3)	0.011 (0.003)	0.055 (0.022)	0.115 (0.051)	0.013 (0.002)	0.047 (0.025)	0.091 (0.057)	0.013 (0.004)	0.045 (0.034)	0.090 (0.076)
(0.5, -0.3)	0.010 (0.003)	0.057 (0.024)	0.106 (0.050)	0.010 (0.002)	0.052 (0.027)	0.093 (0.060)	0.014 (0.005)	0.042 (0.030)	0.091 (0.076)
(-0.5, -0.3)	0.013 (0.002)	0.048 (0.020)	0.100 (0.061)	0.011 (0.002)	0.051 (0.019)	0.092 (0.064)	0.009 (0.004)	0.043 (0.033)	0.093 (0.078)

Table 4.2 Empirical powers based on bootstrap method with $B = 500$

(ϕ, θ)	$m = 5, n = 100$			$m = 5, n = 300$			$m = 5, n = 500$		
	0.01	0.05	0.10	0.01	0.05	0.10	0.01	0.05	0.10
	$H_1^{(1)}$								
(0.5, 0.3)	0.064 (0.007)	0.167 (0.065)	0.260 (0.134)	0.336 (0.079)	0.594 (0.316)	0.726 (0.507)	0.689 (0.328)	0.877 (0.670)	0.927 (0.824)
	$H_1^{(2)}$								
(0.5, 0.3)	0.295 (0.078)	0.499 (0.241)	0.597 (0.380)	0.955 (0.757)	0.987 (0.946)	0.993 (0.978)	1.000 (0.986)	1.000 (1.000)	1.000 (1.000)
	$H_1^{(3)}$								
(0.5, 0.3)	0.051 (0.014)	0.142 (0.053)	0.232 (0.112)	0.296 (0.098)	0.518 (0.280)	0.642 (0.438)	0.593 (0.277)	0.805 (0.574)	0.886 (0.721)

References

- Durbin, J. (1975). Weak convergence of the sample distribution function when parameters are estimated. *The Annals of Statistics*, **1**, 279–290.
- Forte, B. and Hughes, W. (1988). The maximum entropy principle: A tool to define new entropies. *Reports on Mathematical Physics*, **26**, 227–235.
- Jaynes, E. T. (1963). Information theory and statistical mechanics. In *Statistical Physics*, edited by K. Ford, Benjamin, New York, 181–218.
- Lee, S. and Na, S. (2001). On the Bickel-Rosenblatt test for first-order autoregressive models. *Statistics and Probability Letters*, **56**, 23–35.
- Lee, S., Na, O. and Na, S. (2003). On the cusum of squares test for variance change in nonstationary and nonparametric time series models. *Annals of the Institute of Statistical Mathematics*, **55**, 467–485.
- Lee, S., Park, S. and Lee, T. (2010). A note on the Jarque-Bera normality test for GARCH innovations. *Journal of the Korean Statistical Society*, **39**, 93–102.
- Lee, S., Vonta, I. and Karagrigoriou, A. (2011). A maximum entropy type test of fit. *Computational Statistics and Data Analysis*, **55**, 2635–2643.
- Lee, S. and Wei, C. Z. (1999). On residual empirical processes of stochastic regression models with applications to time series. *The Annals of Statistics*, **27**, 237–261.
- Stute, W., Manteiga, W. G. and Quindimil, M. P. (1993). Bootstrap based goodness-of-fit tests. *Metrika*, **40**, 243–256.