Journal of the Korean Statistical Society (2000), 29: 1, pp 1-8

A Probabilistic Interpretation of the KL Spectrum

Seongbaek Yi[1] and ByoungSeon Choi[2]

ABSTRACT

A spectrum minimizing the frequency-domain Kullback-Leibler information number has been proposed and used to modify a spectrum estimate. Some numerical examples have illustrated the KL spectrum estimate is superior to the initial estimate, *i.e.*, the autocovariances obtained by the inverse Fourier transformation of the KL spectrum estimate are closer to the sample autocovariances of the given observations than those of the initial spectrum estimate. Also, it has been shown that a Gaussian autoregressive process associated with the KL spectrum is the closest in the time-domain Kullback-Leibler sense to a Gaussian white noise process subject to given autocovariance constraints. In this paper a corresponding conditional probability theorem is presented, which gives another rationale to the KL spectrum.

Keywords: KL spectrum; Conditional probability density; Autoregressive process; Local limit theorem

1. INTRODUCTION

Consider a second-order stationary time series $\{Y_t\}$. Its autocovariance function (ACVF) and spectrum are defined by

$$\begin{aligned} \sigma(j) &= Cov(Y_t, Y_{t+j}), \quad j = 0, 1, \ldots, \\ S(\lambda) &= \frac{1}{2\pi} \sum_{l=-\infty}^{\infty} \sigma(l) \exp(-i\lambda l), \quad -\pi \leq \lambda \leq \pi, \end{aligned}$$

respectively. It is known by Choi (1990) that the spectrum minimizing the frequency-domain KL information number

$$I(S; \tilde{S}) = \int_{-\pi}^{\pi} S(\lambda) \ln \frac{S(\lambda)}{\tilde{S}(\lambda)} d\lambda$$

[1] Division of Mathematical Sciences, Pukyong National University, Pusan, 608-737, Korea
[2] Department of Applied Statistics, Yonsei University, Seoul, 120-749, Korea

subject to the first $(p+1)$ autocovariance constraints

$$\sigma(0) = \alpha_0, \sigma(1) = \alpha_1, \ldots, \sigma(p) = \alpha_p$$

equals

$$S_{KL}(\lambda) = C\frac{S(\lambda)}{|\sum_{l=0}^{p} \phi_l \exp(-i\lambda l)|^2}, \quad \phi_0 = -1,$$

where the coefficients $\phi_1, \phi_2, \ldots, \phi_p$ and the positive constant C are the solutions of the simultaneous equations

$$\int_{-\pi}^{\pi} S_{KL}(\lambda) \exp(i\lambda j) d\lambda = \alpha_j, \quad j = 0, 1, \ldots, p.$$

It is called the KL spectrum. If $\{Y_t\}$ is an ARMA$(m-p, q)$ process $(m \geq p)$, then the KL spectrum of $\{Y_t\}$ subject to the first $(p+1)$ autocovariance constraints equals the spectrum of an ARMA(m, q) process. Thus, it is a generalization of a large family of spectrums including the maximum entropy spectrum and the ARMA spectrum.

The KL spectrum can be used for modifying spectrum estimates and corresponding time domain models to obtain revised estimates, which are more pertinent to the given observations than the initial estimates. Let $\{y_1, y_2, \cdots, y_N\}$ be an N-realization of the stationary process, and define the sample ACVF $\{\hat{\sigma}(j)\}$ as an estimate of the ACVF by

$$\hat{\sigma}(j) = \hat{\sigma}(-j) = \frac{1}{N}\sum_{t=1}^{N-j}(Y_t - \bar{Y})(Y_{t+j} - \bar{Y}), \quad j = 0, 1, \ldots,$$

where $\bar{Y} = \sum_{t=1}^{N} Y_t/N$. Also, let $\hat{S}_y(\lambda)$ be an estimate of the spectrum. If the estimates do not satisfy

$$\int_{-\pi}^{\pi} \hat{S}_y(\lambda) \exp(i\lambda j) d\lambda = \hat{\sigma}(j), \quad j = 0, 1, \ldots, p,$$

then we would rather use the KL spectrum estimate,

$$\hat{S}_{KL}(\lambda) = \hat{C}\frac{\hat{S}_y(\lambda)}{|\sum_{l=0}^{p} \hat{\phi}_l \exp(-i\lambda l)|^2}, \quad \hat{\phi}_0 = -1,$$

instead of the initial estimate $\hat{S}_y(\lambda)$. Here $\hat{\phi}_1, \hat{\phi}_2, \cdots, \hat{\phi}_p$ and $\hat{C}$ satisfy

$$\int_{-\pi}^{\pi} \hat{S}_{KL}(\lambda) \exp(i\lambda j) d\lambda = \hat{\sigma}(j), \quad j = 0, 1, \ldots, p.$$

Particularly if the ARMA model is initially used to estimate a spectrum, then the KL spectrum estimate can be easily calculated as follows. Let $\hat{e}_1, \hat{e}_2, \ldots, \hat{e}_N$ be the residuals from the ARMA model associated with the initial spectrum estimate $\hat{S}_y(\lambda)$. If it is a good estimate, then the residuals should not be serially correlated. However, if the sample autocovariances of the residuals are not near 0 for some lags greater than 0, the spectrum estimate should be modified. Assume that p is the largest one among the lags whose autocovariances are significantly different from 0. Let

$$\frac{1}{N}\sum_{t=1}^{N}\hat{e}_t^2 = \alpha_0, \frac{1}{N}\sum_{t=1}^{N-1}\hat{e}_t\hat{e}_{t+1} = \alpha_1, \ldots, \frac{1}{N}\sum_{t=1}^{N-p}\hat{e}_t\hat{e}_{t+p} = \alpha_p.$$

Then, $\hat{\phi}_1, \hat{\phi}_2, \ldots, \hat{\phi}_p$ satisfy the Yule-Walker equations

$$\sum_{l=1}^{p}\hat{\phi}_l\alpha_{|j-l|} = \alpha_j, \quad j = 1, 2, \ldots, p.$$

By solving the equations using the Levinson-Durbin algorithm, we obtain the KL spectrum estimate

$$\hat{S}_{KL}(\lambda) = \hat{C}\frac{\hat{S}_y(\lambda)}{|\sum_{l=0}^{p}\hat{\phi}_l\exp(-i\lambda l)|^2}, \quad \hat{\phi}_0 = -1.$$

As an example, we generate 200 observations from the ARMA(2, 1) model

$$y_t = 1.3y_{t-1} - 0.6y_{t-2} + e_t - 0.8e_{t-1},$$

where $\{e_t\}$ is a Gaussian white-noise process with variance 1. When the observations are fitted to an ARMA(2, 1) model, the estimated model is

$$y_t = -0.0677 - 0.4979y_{t-1} + 0.4242y_{t-2} + \hat{e}_t + 0.9835\hat{e}_{t-1}.$$

The ACVF and the partial autocorrelation function (PACF) of the residuals due to the estimated model show that they are far from a white noise process. Moreover the modified portmanteau statistic is 55.9 with 9 degrees of freedom (DF). In order to modify the estimate using the KL spectrum theory, we should determine the AR order of the residuals $\{\hat{e}_t\}$. Applying the AIC, the BIC and the CAT criteria, we have concluded that the residuals be from an AR model of order 3, *i.e.*, AR(3) model. By the KL spectrum theory, we fit the observations $\{y_t\}$ to an ARMA(5, 1) model, and obtain the modified model

$$\begin{aligned} y_t &= -0.0319 + 1.0180y_{t-1} - 0.3869y_{t-2} - 0.1456y_{t-3} \\ &\quad - 0.0548y_{t-4} - 0.0086y_{t-5} + \hat{e}_t - 0.6083\hat{e}_{t-1}. \end{aligned}$$

The ACVF and the PACF of the estimated residuals show the randomness. Also, the modified portmanteau statistic is 7.6 with 6 DF. Thus, the KL model is more relevant to the observations than the initially estimated model. Since the t-ratio of the 5th AR coefficient is -0.07, we have tried to fit the observations to an ARMA$(4,1)$ model. But we have got a message from the program *SAS/ETS* that the model cannot be estimated with these data. We have also tried to fit the observations to an ARMA$(3,1)$ model and have realized that the residuals due to the estimated ARMA$(3,1)$ model are far from a white-noise process. The corresponding modified portmanteau statistic is 40.4 with 8 DF. More numerical examples in Choi (1990) have illustrated that the KL spectrum estimate $\hat{S}_{KL}(\lambda)$ is superior to the initial estimate $\hat{S}_y(\lambda)$, *i.e.*, the autocovariances obtained through the inverse Fourier transformation of the KL spectrum estimate are closer to the sample autocovariances of the given observations than those of the initial spectrum estimate.

Choi (1991) has shown that the Gaussian AR(p) process is the closest in the time-domain Kullback-Leibler sense to independently, identically and normally distributed random variables subject to the first $(p+1)$ autocovariance constraints. Thus, if the residuals found from parametric time series modeling or regression modeling are autocorrelated, then it would be better to regard them as from an AR model. It implies that the KL spectrum estimate is theoretically more reasonable than the initial spectrum estimate in the time-domain Kullback-Leibler sense as well as in the frequency-domain Kullback-Leibler sense. The purpose of this paper is to present a probabilistic assessment of this result.

2. THEOREM AND PROOF

Throughout this paper we assume that the stochastic proces $\{Y_t\}$ is stationary and regular with zero mean. Also, assume that for fixed N and p, the sequence is circular, *i.e.*, $Y_{N-p+1} = Y_1, Y_{N-p+2} = Y_2, \ldots, Y_N = Y_p$. Let

$$\tilde{\sigma}(j) = \tilde{\sigma}(-j) = \frac{1}{N-p} \sum_{t=p+1}^{N} Y_t Y_{t+j} \ , \ j = 0, 1, \ldots .$$

Also, let Σ_j be a $j \times j$ Toeplitz matrix with (r,s) element $\tilde{\sigma}(r-s)$. If $\tilde{\sigma}(0) = \alpha_0, \tilde{\sigma}(1) = \alpha_1, \ldots, \tilde{\sigma}(p) = \alpha_p$, where $\alpha_0, \alpha_1, \ldots, \alpha_p$ are given constants, and if A_{p+1} is the $(p+1) \times (p+1)$ symmetric Toeplitz matrix whose (i,j) element is $\alpha_{|i-j|}$, then the regularity of the process implies A_{p+1} is positive definite. Thus

the system of Yule-Walker quations,

$$\sum_{j=1}^{p} \phi_j \alpha_{|l-j|} = -\alpha_l, \quad l = 1, 2, \ldots, p,$$

has the unique solution $\{\phi_1, \phi_2, \ldots, \phi_p\}$, and then we can define

$$\alpha_l = \alpha_{-l} = -\sum_{j=1}^{p} \phi_j \alpha_{l-j}, \quad l = p+1, p+2, \ldots,$$

$$\sigma^2 = -\sum_{j=0}^{p} \phi_j \alpha_{-j}, \quad \phi_0 = -1.$$

For $j = 1, 2, \ldots$, let A_j be the $j \times j$ symmetric Toeplitz matrix whose (r, s) element is α_{r-s}. For any N-dimensional vector $\boldsymbol{z}_N = (z_1, z_2, \ldots, z_N)^t$ and any $1 \leq l < m \leq N$, let

$$\begin{aligned} \boldsymbol{z}_l &= (z_1, z_2, \ldots, z_l)^t, \\ \boldsymbol{z}_{l,m} &= (z_l, z_{l+1}, \ldots, z_m)^t, \\ \boldsymbol{z}_l^* &= (z_{N-l+1}, z_{N-l+2}, \ldots, z_N)^t. \end{aligned}$$

If $\boldsymbol{Z}_N$ has the normal distribution with mean $\boldsymbol{u}$ and covariance matrix Σ, then for $n \leq N$, the probability density function of $\boldsymbol{Z}_N$ and the conditional probability density function of $\boldsymbol{Z}_n$ given A are denoted by $\phi(\boldsymbol{z}_N; \boldsymbol{u}, \Sigma_N)$ and $\phi(\boldsymbol{z}_n | A; \boldsymbol{u}, \Sigma)$, respectively.

Lemma 1. Let $\{Y_1, Y_2, \cdots, Y_N\}$ be from a stationary Gaussian AR(k) model ($k \leq p$), *i.e.*,

$$Y_t = \beta_1 Y_{t-1} + \beta_2 Y_{t-2} + \cdots + \beta_k Y_{t-k} + E_t, \quad t = k+1, k+2, \ldots, N,$$

where $\{E_t\}$ is a sequence of independent and identically distributed normal random variables with means 0 and variances σ^2. Then, the conditional probability density funtion of $\boldsymbol{y}_N$ given $\boldsymbol{y}_p = \boldsymbol{y}_p^*$ is

$$\begin{aligned} & \phi(\boldsymbol{y}_N \mid \boldsymbol{y}_p = \boldsymbol{y}_p^*; \boldsymbol{0}, \Sigma_N) \\ = \; & \frac{1}{\{(2\pi)^N \det(\Sigma_N)\}^{1/2}} \exp\left\{ -\frac{1}{2} \boldsymbol{y}_p^t \Sigma_p^{-1} \boldsymbol{y}_p - \frac{N-p}{2\sigma^2} \sum_{r=0}^{p} \sum_{s=0}^{p} \phi_r \phi_s \tilde{\sigma}(r-s) \right\}. \end{aligned}$$

Lemma 1 can be easily proved using the joint probability density of the Gaussian AR(p) process. Applying the factorization theorem to Lemma 1, we know

that $\boldsymbol{Y}_p$, $\tilde{\sigma}(0), \tilde{\sigma}(1), \ldots, \tilde{\sigma}(p)$ are jointly sufficient as long as the process is circular. Thus, under the circularity assumption, the conditional probability density function of $\boldsymbol{Y}_N$ given $\boldsymbol{Y}_p, \tilde{\sigma}(0), \tilde{\sigma}(1), \ldots, \tilde{\sigma}(p)$ does not depend on the parameters $\beta_1, \beta_2, \ldots, \beta_k, \sigma^2$. Thus, for any $\mathbf{x}_p$,

$$\begin{aligned} & \phi(\boldsymbol{y}_N \mid \boldsymbol{Y}_p = \boldsymbol{Y}_p^* = \boldsymbol{x}_p, \tilde{\sigma}(0) = \alpha_0, \tilde{\sigma}(1) = \alpha_1, \ldots, \tilde{\sigma}(p) = \alpha_p; \mathbf{0}, \Sigma_N) \\ = \; & \phi(\boldsymbol{y}_N \mid \boldsymbol{Y}_p = \boldsymbol{Y}_p^* = \boldsymbol{x}_p, \tilde{\sigma}(0) = \alpha_0, \tilde{\sigma}(1) = \alpha_1, \ldots, \tilde{\sigma}(p) = \alpha_p; \mathbf{0}, A_N). \end{aligned}$$

Theorem 1. Let $\{Y_t \mid t = \cdots, -1, 0, 1, \cdots\}$ be from a stationary Gaussian AR(k) model ($k \leq p$). For positive integers N_1 and N_2, let

$$\tilde{\sigma}_b(j) = \tilde{\sigma}_b(-j) = \frac{1}{N_1 + N_2 - p} \sum_{t=-N_1+p+1}^{N_2} Y_t Y_{t+j}, \quad j = 0, 1, \ldots.$$

Then, the conditional probability density function of $\boldsymbol{Y}_M$ ($M > p$) given

$$\tilde{\sigma}_b(0) = \alpha_0, \ldots, \tilde{\sigma}_b(p) = \alpha_p, Y_{-N_1+1} = Y_{N_2-p+1}, \ldots, Y_{-N_1+p} = Y_{N_2}$$

tends to $\phi(\boldsymbol{y}_M; \mathbf{0}, A_M)$ as $N_1 \to \infty$ and $N_2 \to \infty$.

Proof. For $\boldsymbol{x}_p = (x_1, x_2, \ldots, x_p)^t$, let

$$\begin{aligned} T_b &= \{\tilde{\sigma}_b(j) = \alpha_j, \quad j = 0, 1, \ldots, p\}, \\ F_b &= \{Y_{-N_1+j} = x_j, \quad j = 1, 2, \cdots, p\}, \\ E_b &= \{Y_{N_2-p+j} = x_j, \quad j = 1, 2, \ldots, p\}, \\ G_b &= T_b \bigcap F_b \bigcap E_b, \\ N &= \min(N_1, N_2). \end{aligned}$$

Lemma 1 implies

$$\phi(\boldsymbol{y}_M \mid G_b; \mathbf{0}, \Sigma_N) = \phi(\boldsymbol{y}_M \mid G_b; \mathbf{0}, A_N),$$

which equals

$$\begin{aligned} & \frac{\phi(\boldsymbol{y}_M, T_b, F_b, E_b; \mathbf{0}, A_N)}{\phi(T_b, F_b, E_b; \mathbf{0}, A_N)} \\ = \; & \left\{ \frac{\phi(T_b \mid \boldsymbol{y}_M, F_b, E_b; \mathbf{0}, A_N)}{\phi(T_b \mid F_b, E_b; \mathbf{0}, A_N)} \right\} \left\{ \frac{\phi(\boldsymbol{y}_M, F_b, E_b; \mathbf{0}, A_N)}{\phi(F_b, E_b; \mathbf{0}, A_N)} \right\}. \end{aligned}$$

A local limit theorem of the sample autocovariances (see, *e.g.*, Theorem 2 of Taniguchi (1986)) implies that the first fraction of the RHS tends to 1 as $N \to \infty$. Also,

$$\begin{aligned} & \lim_{N\to\infty} \frac{\phi(\boldsymbol{y}_M, F_b, E_b; \mathbf{0}, A_N)}{\phi(F_b, E_b; \mathbf{0}, A_N)} \\ = & \lim_{N\to\infty} \phi(\boldsymbol{y}_M \mid F_b, E_b; \mathbf{0}, A_N) \\ = & \phi(\boldsymbol{y}_M; \mathbf{0}, A_M), \end{aligned}$$

where the last equality holds by the well-known properties about conditional and marginal distributions of multivariate normal random vectors (see, *e.g.*, Morrison (1990)) and $\lim_{l\to\infty} \alpha_l = 0$.

3. CONCLUDING REMARKS

From an information theoretic point of view, the limiting probability density function $\phi(\boldsymbol{y}_N; \mathbf{0}, A_N)$ in Theorem 1 is asymptotically the closest in the Kullback-Leibler sense to the joint probability density function of a Gaussian AR(k) process ($k \leq p$) among all the probability density functions satisfying the empirical constraints $\tilde{\sigma}_b(j) = \alpha_j,\ j = 0, 1, \ldots, p$. Of course, a white noise process can be regarded as an AR(0) process, and then Theorem 1 can be applied to it. It can be shown using a similar method to Csiszár, Cover and Choi's (1987) that the transition probability function $\phi(\boldsymbol{y}_{p+1,M} \mid \boldsymbol{y}_p; \mathbf{0}, \Sigma_N)$ is the closest in the Kullback-Leibler sense to the transition probability function $\phi(\boldsymbol{y}_{p+1,M} \mid \boldsymbol{y}_p; \mathbf{0}, A_N)$ subject to the autocovariance constraints

$$\sigma(0) = \alpha_0, \sigma(1) = \alpha_1, \ldots, \sigma(p) = \alpha_p.$$

Thus, Theorem 1 is a generalization of Van Campenhout and Cover's (1981) conditional limit theorem about Sanov's large deviation problem and gives another rationale to the KL spectrum. If we modify initial estimates of the spectrum and the corresponding ARMA model based on the KL spectrum theory, then the AR order always increases. Thus, it is possible that the modified spectrum estimate may have too many peaks. However, we can reduce both the AR and the MA orders by the cancellation law, *i.e.*, deleting common factors of the AR and the MA characteristic functions. For this purpose, we try to fit the observations to ARMA($p-1, q-1$) model if the KL spectrum is obtained by an ARMA(p, q) model. In case of the example in Section 1, we fit the observations

to an ARMA$(4,0)$ model and obtain the following

$$y_t = -0.0597 + 0.4570y_{t-1} - 0.0918y_{t-2} - 0.1873y_{t-3} - 0.2269y_{t-4} + \hat{e}_t.$$

The corresponding modified portmanteau statistic is 9.6 with 8 DF. Due to the principle of parsimony we would rather use the ARMA$(4,0)$ model than the ARMA$(5,1)$ model, the AR and the MA characteristic functions of which have $(1 - 0.61z)$ as an approximate common factor.

REFERENCES

Choi, B. S. (1990). "An information theoretic spectral density," *IEEE Trans. Acoustics, Speech and Signal Processing*, ASSP-38, 717-721.

Choi, B. S. (1991). "A time-domain interpretation of the KL spectrum," *IEEE Trans. Signal Processing*, SP-39, 721-723.

Csiszár, I., Cover, T. M. and Choi, B. S. (1987). "Conditional limit theorems under Markov conditioning," *IEEE Trans. Information Theory*, IT-33, 788-801.

Morrison, D. F. (1990). *Multivariate Statistical Methods*, 3rd ed., New York: McGraw-Hill.

Taniguchi, M. (1986). "Berry-Esseen theorems for quadratic forms of Gaussian stationary processes," *Prob. Th. Rel. Fields*, 72, 185-194.

Van Campenhout, J. M. and Cover, T. M. (1981). "Maximum entropy and conditional probability," *IEEE Trans. Information Theory*, IT-27, 483-489.