時系列解析入門 - 共分散関数

1. 時系列の分布と定常性

時系列解析においては、時系列の変動の特徴を捉えるためには y_n の分布だけではなく y_n と y_{n-k} の同時分布(Joint Distribution)を調べる必要がある。
時系列を $\{y_1,\ldots,y_N\}$ とするとき、 y_n の期待値 $\mu_n = \mathrm{E}(y_n)$ を時系列 y_n の平均値関数(mean value function)と呼ぶ。
また、時系列 y_n と y_{n-1} との共分散、

$$\mathrm{Cov}(y_n, y_{n-k}) = \mathrm{E}\{(y_n - \mu_n)(y_{n-k} - \mu_{n-k})\}$$

を自己共分散(autocovariance)と呼ぶ。とくに $k = 0$ とおくと時系列の分散関数 $Var(y_n)$ が得られる。

ここでは、平均、分散、共分散が時間をシフトしても変化しない場合を考える。l を時間のシフト量を示す任意の整数とするとき、

$$\mathrm{E}(y_n) = \mathrm{E}(y_{n-l})$$
$$\mathrm{var}(y_n) = \mathrm{var}(y_{n-l})$$
$$\mathrm{Cov}(y_n, y_m) = \mathrm{Cov}(y_{n-l}, y_{m-l})$$

が成り立つ時系列は弱定常（二次定常、weakly stationary、covariance stationary）といわれる。

一般に、平均・分散・共分散だけで時系列の特徴を捉えられないので、 $y_1,\ldots,y_N$ の同時確率密度関数(joint probability density function) $f(y_1,\ldots,y_N)$ を調べなければならない。そのためには、任意の整数 k および任意の時刻 $i_1 < i_2 < \ldots < i_k$ について $y_{i_1},\ldots,y_{i_k}$ の同時確率分布 $f(y_{i_1},\ldots,y_{i_k})$ を定めればよい。とくにこの分布が k 変量の正規分布で表される場合には、ガウス型時系列と呼ばれ、平均ベクトルと分散共分散行列で特徴を捉えることができる。

ある時系列の変化が時間の変化によって不変で、その確率分布を時間軸方向に移動しても変化しない場合、その時系列は強定常(strongly stationary)であるという。すなわち時系列が強定常であるとは、任意の時間シフト量 l および任意の $i_1,.i_2,\ldots,i_k$ について、

$$f(y_{i_1},\ldots,y_{i_k}) = f(y_{i_1-l},\ldots,y_{i_k-l})$$

が成り立つことである。ガウス型の時系列において弱定常性は強定常性と同じである。

2.　定常時系列の自己共分散関数

定常時系列では平均値関数μ_nは時刻nに依存しないので、$\mu = \mathrm{E}(y_n)$と表し、時系列y_nの平均と呼ぶ。またy_nとy_{n-k}の共分散$\mathrm{Cov}(y_n, y_{n-k})$は時間差$k$だけに依存する量となるので、

$$C_k = \mathrm{Cov}(y_n, y_{n-k}) = \mathrm{E}((y_n - \mu)(y_{n-k} - \mu))$$

と表し、定常時系列の自己共分散関数(autocovariance function)という。kはラグ(lag)とも呼ばれる。とくに$k=0$のときには自己共分散関数はy_nの分散に等しい。自己共分散関数は偶関数で$C_l = C_{-l}$で$|C_k| \le C_0$が成り立つ。

また、y_nとy_{n-k}の相関係数、

$$R_k = \frac{\mathrm{Cov}(y_n, y_{n-k})}{\sqrt{\mathrm{var}(y_n)\mathrm{var}(y_{n-k})}}$$

をラグkの関数とみなしたものを自己相関関数(autocorrelation function)と呼ぶ。
定常系列の場合は、

$$\mathrm{var}(y_n) = \mathrm{var}(y_{n-k}) = C_0$$

なので、

$$R_k = \frac{C_k}{C_0}$$

と表せる。

時系列y_nが相関のない確率変数の実現値で、その自己共分散関数が

$$C_k = \begin{cases} \sigma^2, & k = 0\text{のとき} \\ 0, & k \ne 0\text{のとき} \end{cases}$$

を満たすとき、分散σ^2の白色雑音（ホワイトノイズ、white noise）と呼ばれる。

定常系列$\{y_1, \ldots, y_N\}$が与えられたとき、平均μ、自己共分散関数C_k、自己相関関数R_kの推定値は以下の式により求められる。

$$\hat{\mu} = \frac{1}{N}\sum_{n=1}^{N} y_n$$

$$\hat{C}_k = \frac{1}{N}\sum_{n=k+1}^{N}(y_n - \hat{\mu})(y_{n-k} - \hat{\mu})$$

$$\hat{R}_k = \frac{\hat{C}_k}{\hat{C}_0}$$

$\hat{\mu}$ を標本平均(sample mean)、$\hat{C}_k$ を標本自己共分散関数(sample autocovariance function)、$\hat{R}_k$ を標本自己相関関数(sample autocorrelation function)と呼ぶ。

3.　相互共分散関数および相互相関関数

多変量時系列 $y_n = \left(y_n(1),\ldots,y_n(l)\right)^T$ の場合、平均ベクトル、相互共分散関数および相互相関関数が時系列の特徴を記述する基本的な統計量になる。
i 番目の時系列 $y_n(i)$ の平均は、

$$\mu(i) = \mathrm{E}\left(y_n(i)\right)$$

によって定義される。このとき $\mu = \left(\mu(1),\ldots,\mu(l)\right)^T$ を多変量時系列 y_n の平均ベクトル(mean vector)と呼ぶ。また $y_n(i)$ と時系列 $y_{n-k}(j)$ の共分散は、

$$C_k(i,j) = \mathrm{Cov}\left(y_n(i), y_{n-k}(j)\right) = \mathrm{E}\left\{\left(y_n(i)-\mu(i)\right)\left(y_{n-k}(j)-\mu(j)\right)^T\right\}$$

この時、$l \times l$ 行列

$$C_k = \begin{bmatrix} C_k(1,1) & .. & C_k(1,l) \\ \ldots & .. & .. \\ C_k(l,1) & .. & C_k(l,l) \end{bmatrix}$$

はラグ k の相互共分散と呼ばれる。また $C_k\left(k = 0,1,2,\ldots\right)$ をラグ k の関数と見るとき、相互共分散関数(cross-covariance)と呼ばれる。相互共分散関数の対角成分 $C_k(i,i)$ は i 番目の時系列 $y_n(i)$ の自己共分散関数となっている。
また、時系列 $y_n(i)$ と $y_{n-k}(j)$ の相関係数を、

$$R_k(i,j) = \frac{\mathrm{Cov}\left(y_n(i), y_{n-k}(j)\right)}{\sqrt{C_0(i,i)C_0(j,j)}} = \frac{C_k(i,j)}{\sqrt{C_0(i,i)C_0(j,j)}}$$

で表すとき、

$$R_k = \begin{bmatrix} R_k(1,1) & .. & R_k(1,l) \\ \ldots & .. & .. \\ R_k(l,1) & .. & R_k(l,l) \end{bmatrix}$$

は相互相関関数(cross-correlation function)と呼ばれる。
自己共分散関数および自己相関関数は偶関数で $C_{-k} = C_k$ 、$R_{-k} = R_k$ が成り立つが、相互共分散関数および相互相関関数については以下がなりたつ。

$$C_{-k} = C_k^{\ T}, R_{-k} = R_k^{\ T}$$

そのため、多変量時系列の場合にも C_k および R_k は $k \geq 0$ について求めればよい。
長さ N の多変量時系列 $\{y_1(j),\ldots,y_n(j)\}(j = 1,\ldots,l)$ が観測されたとき、平均 $\mu(i)$、相互共分散関数 $C_k(i,j)$ および相互相関関数 $R_k(i,j)$ の推定値は次の式により求められる。

$$\hat{\mu}(i) = \frac{1}{N}\sum_{n=1}^{N} y_n(i)$$

$$\hat{C}_k(i,j) = \frac{1}{N}\sum_{n=k+1}^{N}(y_n(i) - \hat{\mu}(i))(y_{n-k}(j) - \hat{\mu}(j))$$

$$\hat{R}_k(i,j) = \frac{\hat{C}_k(i,j)}{\sqrt{\hat{C}_0(i,i)\hat{C}_0(j,j)}}$$

l 次元ベクトル $\hat{\mu}$ は標本平均ベクトル(sample mean vector)、$l \times l$ 行列 $\hat{C}$ は標本相互共分散関数(sample cross-covariance function)、 $\hat{R}$ は標本相互相関関数(sample cross-correlation function)と呼ばれる。

誤植等はご指摘ください。
（この項目・了）