5 장　확률과정

확률과정(stochastic process)

확률법칙(probability laws)에 의해 생성되는 일련의 통계적인 현상

확률공간에서 정의되는 확률변수들의 모임 $Z_t, t \in T$

$T = (-\infty, \infty)$ 혹은 $T = [0, \infty)$: **연속형 확률과정**

$T = 0, \pm 1, \pm 2, \ldots$ 혹은 $T = 0, 1, 2, \ldots$: **이산형 확률과정**

시계열 : 확률과정 $Z_t(\omega), t = 1, 2, \ldots$의 관측값, $Z_t, t = 1, 2, \ldots$으로 표현

확률과정의 **실현값**(realization) 또는 **표본통로**(sample path)

5.1 정상확률과정

시계열 자료 $Z_1, Z_2, \ldots, Z_n$: 어느 특정한 확률과정의 실현된 값

정의 : 정상성(stationarity) :

시계열의 확률적인 성질들이 시간의 흐름에 따라 불변(time-invariant)

<그림 5.1>의 특징

① 뚜렷한 추세가 없다. 즉, 시계열의 평균이 시간 축에 평행하다.

② 시계열의 진폭(변동)이 시간의 흐름에 따라 일정하다

(1) 엄격한 의미의 정상성(strict stationarity)

임의의 자연수 $t_1, t_2, \ldots, t_n$과 k에 대해

$$f(Z_{t_1}, Z_{t_2}, \ldots, Z_{t_n}) = f(Z_{t_1+k}, Z_{t_2+k}, \ldots, Z_{t_n+k})$$

강한 의미의 정상성(strong stationarity)

$\mu_t = E(Z_t)$ **: 평균**

$\sigma_t^2 = Var(Z_t) = E[(Z_t - \mu_t)^2]$ **: 분산**

$\gamma_{s-t} = Cov(Z_t, Z_s) = E[(Z_t - \mu_t)(Z_s - \mu_s)]$ **: 자기공분산(autocovariance)**

(2) 약한 의미의 정상성(weak stationarity)

$E(Z_t) = \mu$, $Var(Z_t) = \sigma_z^2$: 상수로서 시간 t 에 관계없이 동일

$E[(Z_t - \mu)(Z_{t+k} - \mu)] = Cov(Z_t, Z_{t+k}) = \gamma_k$: 시차(lag) k 에만 의존하고 시점 t 와는 무관

2차 정상성(second-order stationarity) 또는 공분산정상성(covariance stationarity)

- 정규과정(Gaussian process)의 경우 정상성에 대한 두 정의가 동일

5.2 확률과정의 예

5.2.1 백색잡음과정(white noise process) : <그림 5.2>

서로 독립이고 동일한 분포를 따르는 i.i.d. 확률변수들의 계열로 구성된 확률과정

$Z_t = \epsilon_t$, $t = 1, 2, \ldots$, ϵ_t : 백색잡음과정 $WN(0, \sigma_\epsilon^2)$

$E(Z_t) = E(\epsilon_t) = 0$

$\gamma_0 = Var(Z_t) = Var(\epsilon_t) = E(\epsilon_t^2) = \sigma_\epsilon^2$, $\gamma_k = Cov(Z_t, Z_{t+k}) = Cov(\epsilon_t, \epsilon_{t+k}) = 0$.

시간 t 에 무관 => 정상확률과정

5.2.2 확률보행과정(random walk process)

$$Z_t = Z_{t-1} + \epsilon_t\,, \quad Z_0 = 0, \quad t = 1, 2, \ldots$$

: 절편(drift)이 없는 **확률보행과정** 또는 **임의보행과정** : <그림 5.3>

ϵ_t : 시점 t 에서 어떤 사람이 임의의 방향으로 움직이는 보폭

원점($Z_0 = 0$)에서 출발, $Z_1 = \epsilon_1$이고 $Z_t = \sum_{i=1}^{t} \epsilon_i$는 t 시간 후의 위치

$$E(Z_t) = 0\,,$$

$$Var(Z_t) = Var(\sum_{i=1}^{t} \epsilon_i) = t\sigma_\epsilon^2$$

$$\gamma_k = Cov(Z_t, Z_{t+k}) = E\left\{(\sum_{i=1}^{t} \epsilon_i) \cdot (\sum_{i=1}^{t+k} \epsilon_i)\right\} = E(\sum_{i=1}^{t} \epsilon_i^2) = t\sigma_\epsilon^2$$

분산과 자기공분산이 시간의 함수 => **비정상(nonstationary) 확률과정**

절편이 있는 확률보행과정$(Z_0 = 0)$ **: <그림 5.4>**

$$Z_t = \delta + Z_{t-1} + \epsilon_t\,, \qquad t = 1, 2, \ldots$$

$$Z_1 = \delta + \epsilon_1, \cdots, \; Z_t = t\delta + \sum_{i=1}^{t} \epsilon_i$$

$E(Z_t) = t\delta$: 시간 t 의 함수 : 시간의 흐름에 따라 수준이 점차로 증가

$Var(Z_t) = t\sigma_\epsilon^2$: 시간 t 의 함수 : 시간의 흐름에 따라 분산이 점차로 증가

5.2.3 이동평균과정(moving average process)

$$Z_t = \mu + \epsilon_t - \theta\epsilon_{t-1}\,, \; t = 1, 2, \ldots$$

$$E(Z_t) = \mu$$

$$\gamma_k = Cov(Z_t, Z_{t+k}) = E[(\epsilon_t - \theta\epsilon_{t-1})(\epsilon_{t+k} - \theta\epsilon_{t+k-1})]$$

$$= \begin{cases} (1+\theta^2)\,\sigma_\epsilon^2, & k = 0 \\ -\theta\sigma_\epsilon^2, & k = 1 \\ \quad 0\,, & \text{다른경우} \end{cases}$$

: 평균과 분산 및 자기공분산은 시간 t 와는 무관, 시차 k 만의 함수

=> 정상확률과정, 1차-**이동평균과정**

5.2.4 선형과정(linear process)

$$Z_t = \mu + \epsilon_t + \psi_1 \epsilon_{t-1} + \psi_2 \epsilon_{t-2} + \cdots = \mu + \sum_{j=0}^{\infty} \psi_j \epsilon_{t-j}, \quad \psi_0 = 1$$

Z_t : 서로 독립인 확률변수들의 선형결합

$E(Z_t) = \mu$,

$$Var(Z_t) = \sigma_\epsilon^2 \sum_{j=1}^{\infty} \psi_j^2$$

$$\gamma_k = Cov(Z_t, Z_{t+k}) = E(\sum_{j=0}^{\infty} \sum_{i=0}^{\infty} \psi_i \psi_j \epsilon_{t-i} \epsilon_{t+k-j}) = \sigma_\epsilon^2 \sum_{i=0}^{\infty} \psi_i \psi_{i+k}$$

선형과정이 정상적이기 위한 조건(stationarity condition)

시점 t 와는 무관, 시차 k 만의 함수이기 위한 조건 : $\sum_{j=0}^{\infty} \psi_j^2 < \infty$

$$\because \ |\gamma_k| \le \sqrt{Var(Z_t) \cdot Var(Z_{t+k})} = \sigma^2 \sum_{j=0}^{\infty} \psi_j^2$$

: 무한이동평균과정(infinite moving average process)

Wold's decomposition(1938)

임의의 이산형 정상확률과정은 서로 상관이 없는 결정적인(deterministic)과정과 비결정적인 (indeterministic)과정의 합으로 나타낼 수 있으며 또한 모든 비결정적이고 약한 의미의 정상확률과정은 서로 상관이 없는 확률변수들의 선형결합(또는 선형과정)으로 나타낼 수 있다.

5.2.5 자기회귀과정(autoregressive process)

$$Z_t - \mu = \phi(Z_{t-1} - \mu) + \epsilon_t, \quad E(Z_t) = \mu < \infty$$

: 1차-자기회귀과정(first order autoregressive process)

$$Z_t = \delta + \phi Z_{t-1} + \epsilon_t \quad \text{또는} \quad \dot{Z}_t = \phi \dot{Z}_{t-1} + \epsilon_t, \quad \text{단, } \mu = \delta/(1-\phi) \text{ 이고 } \dot{Z}_t = Z_t - \mu$$

: **마코프(Markov)과정**

$$\dot{Z}_t = \phi(\phi \dot{Z}_{t-2} + \epsilon_{t-1}) + \epsilon_t = \phi^2(\phi \dot{Z}_{t-3} + \epsilon_{t-2}) + \epsilon_t + \phi\epsilon_{t-1}$$

$$\vdots$$

$$= \epsilon_t + \phi\epsilon_{t-1} + \phi^2\epsilon_{t-2} + \cdots = \sum_{j=0}^{\infty} \phi^j \epsilon_{t-j}.$$

정상성의 조건

$$\gamma_0 = Var(Z_t) = E(Z_t - \mu)^2 = \sum_{j=0}^{\infty} \phi^{2j} E(\epsilon_{t-j}^2) = \sigma_\epsilon^2 \sum_{j=0}^{\infty} \phi^{2j}$$

$$\gamma_k = Cov(Z_t, Z_{t+k}) = E(\mathrm{Z_t} - \mu)(\mathrm{Z_{t+k}} - \mu) = \mathrm{E}(\dot{\mathrm{Z}}_{\mathrm{t}} \dot{\mathrm{Z}}_{\mathrm{t+k}})$$

$$= E[(\sum_{j=0}^{\infty} \phi^j \epsilon_{t-j})(\sum_{i=0}^{\infty} \phi^i \epsilon_{t+k-i})]$$

$$= \sigma_\epsilon^2 \sum_{j=0}^{\infty} \phi^j \phi^{j+k} = \phi^k \sigma_\epsilon^2 \sum_{j=0}^{\infty} \phi^{2j} (= \phi^k \gamma_0)$$

$$|\phi| < 1 \ \ ==> \ \sum_{j=0}^{\infty} \phi^{2j} < \infty$$

$\gamma_0 = \sigma_\epsilon^2 (1 - \phi^2)$, $\gamma_k = \phi^k \gamma_0$, $k = 1, 2, ...$: 시차 k 만의 함수

참고 : $Z_t = \mu + \sum_{j=0}^{\infty} \phi^j \epsilon_{t-j}$: $\psi_j = \phi^j$ 인 선형과정

$|\phi| < 1$ => 선형과정의 정상성 조건 $\sum_{j=0}^{\infty} \psi_j^2 = \sum_{j=0}^{\infty} \phi^{2j} < \infty$ 이 만족

5.3 자기상관함수(autocovariance function : ACF)

자기공분산함수 : 시간에 따른 상관정도의 척도

$$\gamma_k = Cov(Z_t, Z_{t+k}) = E[(Z_t - \mu)(Z_{t+k} - \mu)]$$

자기상관함수(autocorrelation function)

$$\rho_k = Corr(Z_t, Z_{t+k}) = \frac{Cov(Z_t, Z_{t+k})}{\sqrt{Var(Z_t) \cdot Var(Z_{t+k})}}$$

$$\gamma_0 = Var(Z_t) = E[(Z_t - \mu)^2] \quad ==> \quad \rho_k = \frac{\gamma_k}{\gamma_0}$$

자기공분산함수와 자기상관함수의 성질

$(a)\ \gamma_0 = Var(Z_t)\ ;\ \rho_0 = 1$

$(b)\ |\gamma_k| \le \gamma_0\ ;\ |\rho_k| \le 1,\ k = 1, 2, \ldots$

$(c)\ \gamma_k = \gamma_{-k}\ ;\ \rho_k = \rho_{-k},\ k = 1, 2, \ldots$

표본자기공분산함수(sample autocovariance function)

$$\hat{\gamma}_k = \frac{\sum_{t=1}^{n-k}(Z_t - \overline{Z})(Z_{t+k} - \overline{Z})}{n}, \quad k = 0, 1, 2, \ldots \quad \text{단, } \overline{Z} = \sum_{t=1}^{n} Z_t / n$$

표본자기상관함수(sample autocorrela -tion function: $SACF$)

$$\hat{\rho}_k = \frac{\hat{\gamma}_k}{\hat{\gamma}_0}, \ k = 1, 2, \ldots \ , \quad E(\hat{\rho}_k) = \rho_k$$

Bartlett(1946) : n 이 크면 $\hat{\rho}_k$: 점근적으로 정규분포

$\rho_k = 0 (k > q)$인 경우 : $Var(\hat{\rho}_k) \approx \frac{1}{n}(1 + 2\sum_{i=1}^{q} \rho_i^2)$

$Z_1, Z_2, \ldots, Z_n$: 서로 독립이고 동일한 분포를 따를 경우

$\hat{\rho}_k$: $\approx$ $N(0, \frac{1}{n})$

==> 시계열자료의 독립(백색잡음)여부의 검정 $H_0 : \rho_k = 0$에 이용

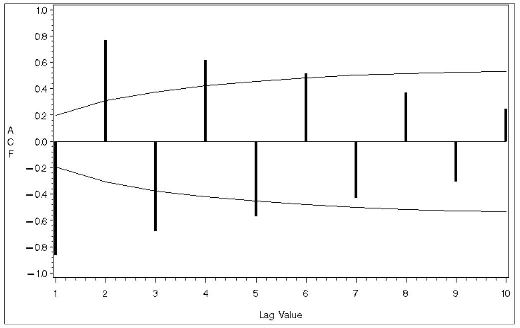
95 % 근사신뢰구간 : $(\pm 2/\sqrt{n})$

표본상관도표표(sample correlogram)

$SACF\ \hat{\rho}_k$, $k=0,1,2,\ldots$ 와 표본부분자기상관계수($SPACF$)의 그림

: 주어진 시계열자료가 어느 확률과정의 모형으로부터 생성된 것인지를 판단하는데 이용되며 $ARMA(p,q)$ 모형의 적합시 모형의 차수 p 와 q 의 값을 결정하는 데 주로 이용됨

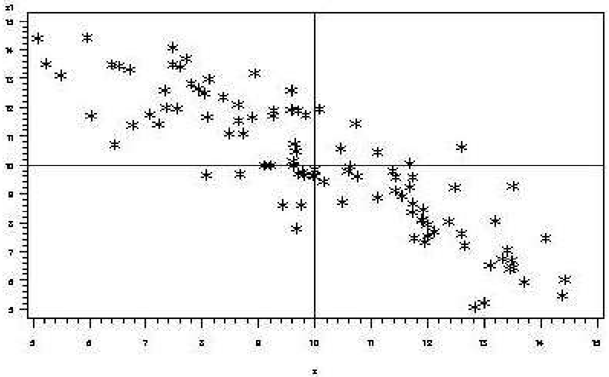
예제 5.1 **표본자기상관함수($SACF$) :** $Z_t = 19 - 0.9Z_{t-1} + \epsilon_t$, ϵ_t : 백색잡음과정 $N(0,1)$

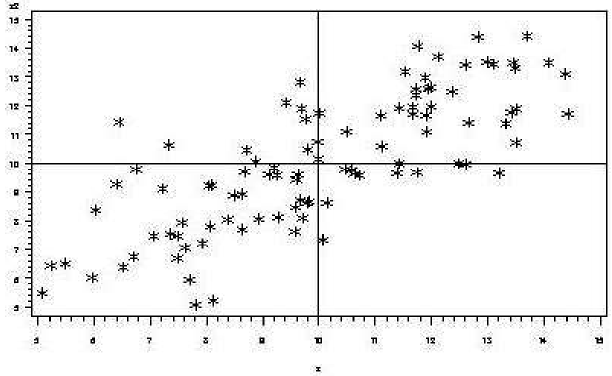


<그림 5.5> $Z_t = 19 - 0.9Z_{t-1} + \epsilon_t$,

$\hat{\rho}_1$ 이 -1에 가까운 경우 : Z_t의 시계열그림은 평균 $E(Z_t)$를 중심으로 위아래로 움직이는 모양

$\hat{\rho}_1$이 1에 가까운 경우 : Z_t의 시계열그림은 평균 $E(Z_t)$ 의 한쪽 방향에 한동안 머무르는 모양





<그림 5.6> (Z_t, Z_{t-1})**의 산점도** **<그림 5.7>** (Z_t, Z_{t-2})**의 산점도**

시차 1인 경우의 $SACF$: $\hat{\rho_1} = \widehat{Corr}(Z_t, Z_{t-1}) = -0.86723$

: (Z_t, Z_{t-1})의 산점도인 <그림 5.6>이 음의 기울기를 가지고 있음을 의미

시차 2에서의 $SACF$: $\hat{\rho_2} = \widehat{Corr}(Z_t, Z_{t-2}) = 0.76753$

: (Z_t, Z_{t-2}) 의 산점도 <그림 5.7>이 양의 기울기를 가지고 있으며 또한 <그림 5.6>보다 더 직선에서 이탈하고 있음을 보여줌

5.4 부분자기상관함수(partial autocorrelation function : $PACF$)

부분상관계수(partial correlation) :

$$\rho_{XY,Z} = \frac{E[X - E(X|Z) \cdot Y - E(Y|Z)]}{\sqrt{EX - E(X|Z)^2 \cdot EY - E(Y|Z)^2}}$$

$E(X|Z)$: X를 Z에 회귀시킨 최적선형(best linear) 예측값
$E(Y|Z)$: Y를 Z에 회귀시킨 최적선형 예측값

$\rho_{X^*Y^*}$: 변수 Z에 관하여 수정(adjust)한 후의 X 와 Y의 부분상관계수 $\rho_{XY,Z}$

$X^* = X - E(X|Z)$는 X를 Z에 회귀시킨 후의 잔차
$Y^* = Y - E(Y|Z)$ 는 Y를 Z에 회귀시킨 후의 잔차

부분자기상관함수 $PACF$ ϕ_{kk}

Z_t 와 Z_{t+k} 에서 $Z_{t+1}, Z_{t+2}, \ldots, Z_{t+k-1}$ 의 효과를 제거한 후의 상관계수

정상확률과정 Z_t , $E(Z_t)=0$

Z_t 를 $Z_{t+1}, Z_{t+2}, \ldots, Z_{t+k-1}$ 에 회귀시킨 최적선형예측

$$E(Z_t \mid Z_{t+1}, Z_{t+2}, \ldots\ Z_{t+k-1}) = \alpha_1 Z_{t+1} + \alpha_2 Z_{t+2} + \cdots + \alpha_{k-1} Z_{t+k-1}$$

$Z_{t+k-1}, Z_{t+k-2}, \ldots, Z_{t+1}$ 에 기초한 Z_{t+k} 의 최적선형예측

$$E(Z_{t+k} \mid Z_{t+k-1}, \ldots, Z_{t+1}) = \beta_1 Z_{t+k-1} + \beta_2 Z_{t+k-2} + \cdots + \beta_{k-1} Z_{t+1}$$

$PACF \quad \phi_{kk} = Corr\left\{ Z_t^* , Z_{t+k}^* \right\}$

$$Z_t^* = Z_t - (\alpha_1 Z_{t+1} + \alpha_2 Z_{t+2} + \cdots + \alpha_{k-1} Z_{t+k-1})$$

$$Z_{t+k}^* = Z_{t+k} - (\beta_1 Z_{t+k-1} + \beta_2 Z_{t+k-2} + \cdots + \beta_{k-1} Z_{t+1})$$

단, α_i와 β_i $(1 \leqq i \leqq k-1)$는

$$E(Z_t - \alpha_1 Z_{t+1} - \cdots - \alpha_{k-1} Z_{t+k-1})^2 \text{와 } E(Z_{t+k} - \beta_1 Z_{t+k-1} - \cdots - \beta_{k-1} Z_{t+1})^2$$

을 최소로하는 최소제곱추정량으로 $\alpha_i = \beta_i$

예 : $\phi_{11}(=\alpha_1=\beta_1)=\rho_1$

$k=2$ 인 경우

$$\phi_{22}=Corr\,(Z_t-\alpha_1 Z_{t+1}\,,Z_{t+2}-\beta_1 Z_{t+1})$$

$$=Corr\,(Z_t-\rho_1 Z_{t+1}\,,Z_{t+2}-\rho_1 Z_{t+1})$$

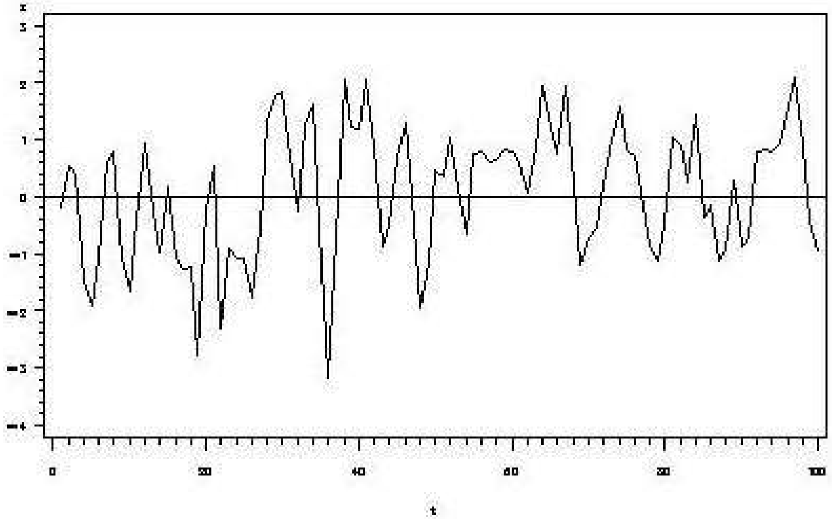
$$=\frac{Cov\,(Z_t-\rho_1 Z_{t+1}\,,Z_{t+2}-\rho_1 Z_{t+1})}{\sqrt{Var\,(Z_t-\rho_1 Z_{t+1})\cdot Var\,(Z_{t+2}-\rho_1 Z_{t+1})}}$$

$$=\frac{\gamma_2-2\rho_1\gamma_1+\rho_1^2\gamma_0}{\gamma_0-2\rho_1\gamma_1+\rho_1^2\gamma_0}$$

$$=\frac{\rho_2-2\rho_1^2+\rho_1^2}{1-2\rho_1^2+\rho_1^2}$$

$$=\frac{\rho_2-\rho_1^2}{1-\rho_1^2}$$

예제 5.2 표본부분자기상관계수($SPACF$)

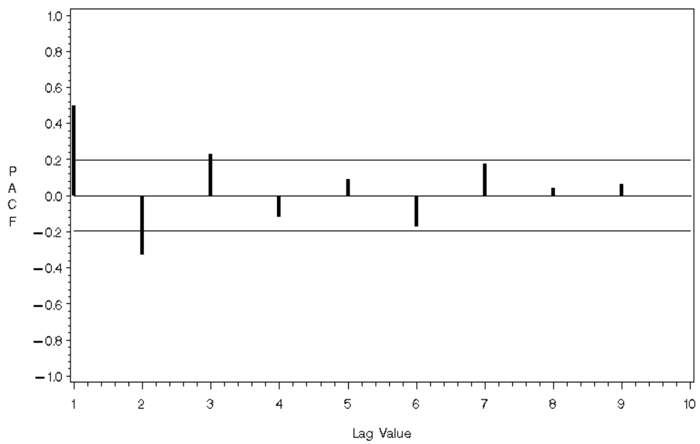


<그림 5.8> $Z_t = \epsilon_t + 0.8\epsilon_{t-1}$ 에서 생성된 모의실험자료의 시계열그림

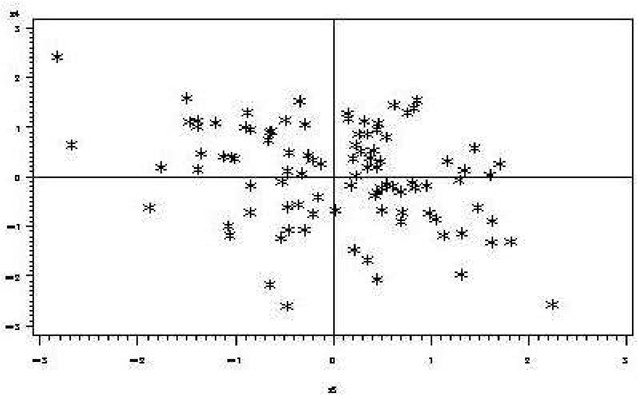
$$\widehat{\phi_{11}} = \widehat{\rho_1} = 0.4991$$

$$\widehat{\phi_{22}} = \widehat{Corr}(Z_t - \hat{Z}_t,\ Z_{t+2} - \hat{Z}_{t+2})$$

$$= \widehat{Corr}(Z_t - 0.4991 Z_{t+1}\ ,\ Z_{t+2} - 0.4991 Z_{t+1}) = -0.3308$$



<그림 5.9> $Z_t = \epsilon_t + 0.8\epsilon_{t-1}$에서 생성된 모의실험자료의 SPACF



<그림 5.10> $(Z_t - 0.4991Z_{t+1}\,,\ Z_{t+2} - 0.4991Z_{t+1})$의 산점도

회귀모형을 이용하여 $PACF$를 구하는 방법

종속변수 Z_{t+k} 를 $Z_{t+k-1}, Z_{t+k-2}, \ldots, Z_t$ 에 회귀시킨 회귀식

$$Z_{t+k} = \phi_{k1} Z_{t+k-1} + \phi_{k2} Z_{t+k-2} + \cdots + \phi_{kk} Z_t + \epsilon_{t+k}$$

양변에 $Z_{t+k-j}\,(j \geq 1)$ 를 곱한 후 기대값을 구하여 보면

$$\gamma_j = \phi_{k1}\gamma_{j-1} + \phi_{k2}\gamma_{j-2} + \cdots + \phi_{kk}\gamma_{j-k}$$

양변을 γ_0 로 나누면

$$\rho_j = \phi_{k1}\rho_{j-1} + \phi_{k2}\rho_{j-2} + \cdots + \phi_{kk}\rho_{j-k}$$

$j = 1, 2, \ldots, k$에 대하여

$$\begin{cases} \rho_1 = \phi_{k1}\rho_0 + \phi_{k2}\rho_1 + \cdots + \phi_{kk}\rho_{k-1} \\ \rho_2 = \phi_{k1}\rho_1 + \phi_{k2}\rho_0 + \cdots + \phi_{kk}\rho_{k-2} \\ \quad\vdots \\ \rho_k = \phi_{k1}\rho_{k-1} + \phi_{k2}\rho_{k-2} + \cdots + \phi_{kk}\rho_0 \end{cases}$$

연립방정식을 Cramer 공식을 이용

$$\phi_{11} = \rho_1, \quad \phi_{22} = \frac{\begin{vmatrix} 1 & \rho_1 \\ \rho_1 & \rho_2 \end{vmatrix}}{\begin{vmatrix} 1 & \rho_1 \\ \rho_1 & 1 \end{vmatrix}} = \frac{\rho_2 - \rho_1^2}{1 - \rho_1^2}, \quad \phi_{33} = \frac{\begin{vmatrix} 1 & \rho_1 & \rho_1 \\ \rho_1 & 1 & \rho_2 \\ \rho_2 & \rho_1 & \rho_3 \end{vmatrix}}{\begin{vmatrix} 1 & \rho_1 & \rho_2 \\ \rho_1 & 1 & \rho_1 \\ \rho_2 & \rho_1 & 1 \end{vmatrix}}$$

$$\phi_{kk} = \frac{\begin{vmatrix} 1 & \rho_1 & \rho_2 & \cdots & \rho_{k-2} & \rho_1 \\ \rho_1 & 1 & \rho_1 & \cdots & \rho_{k-3} & \rho_2 \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ \rho_{k-1} & \rho_{k-2} & \rho_{k-3} & \cdots & \rho_1 & \rho_k \end{vmatrix}}{\begin{vmatrix} 1 & \rho_1 & \rho_2 & \cdots & \rho_{k-2} & \rho_{k-1} \\ \rho_1 & 1 & \rho_1 & \cdots & \rho_{k-3} & \rho_{k-2} \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ \rho_{k-1} & \rho_{k-2} & \rho_{k-3} & \cdots & \rho_1 & 1 \end{vmatrix}}$$

표본부분자기상관함수(sample partial autocorrelation function: $SPACF$) $\hat{\phi}_{kk}$

위의 ϕ_{kk} 식에서 ρ_j 대신에 $SACF$ $\hat{\rho}_j$ 를 대입하여 얻음

k가 커질 경우 행렬식의 계산을 하는 대신

Durbin-Levinson 알고리즘, Durbin(1960), 이용

$$\hat{\phi}_{11} = \hat{\rho}_1$$

$$\hat{\phi}_{k+1,k+1} = \frac{\hat{\rho}_{k+1} - \sum_{j=1}^{k} \hat{\phi}_{kj}\,\hat{\rho}_{k+1-j}}{1 - \sum_{j=1}^{k} \hat{\phi}_{kj}\,\hat{\rho}_j}$$

$$\hat{\phi}_{k+1,j} = \hat{\phi}_{kj} - \hat{\phi}_{k+1,k+1}\hat{\phi}_{k,k+1-j}, \qquad j = 1,2,\ldots,k$$