MATH4425 (T1A) – Tutorial 8

Kazovskaia Anastasiia

April, 11

Important information

- T1A: **Thursday 19:00 - 19:50** (Rm 1033, LSK Bldg)
- Office hours: **Wednesday 14:00 - 14:50** (Math support center, 3rd floor, Lift 3)
- Any questions to be addressed to **akazovskaia@connect.ust.hk**

1 Parameter Estimation, Diagnostic Checking and Model Selection. The method of moments

Let's assume $\dot{Z}_t := Z_t - \mu$ follows AR(p) process

$$\dot{Z}_t = \phi_1 \dot{Z}_{t-1} + \cdots + \phi_p \dot{Z}_{t-p} + a_t$$

We already know that

$$\rho_k = \phi_1 \rho_{k-1} + \cdots + \phi_p \rho_{k-p}$$

1.1 Definition

So, **Yule-Walker equations** follow immediately

$$\rho_1 = \phi_1 \rho_0 + \phi_2 \rho_{-1} + \cdots + \phi_p \rho_{-p} = \phi_1 + \phi_2 \rho_1 + \cdots + \phi_p \rho_p$$

Similarly,

$$\rho_2 = \phi_1 \rho_1 + \phi_2 + \cdots + \phi_p \rho_{p-1}$$

$$\rho_3 = \phi_1 \rho_2 + \phi_2 \rho_1 + \cdots + \phi_p \rho_{p-2}$$

$$\vdots$$

$$\rho_p = \phi_1 \rho_1 + \phi_2 \rho_2 + \cdots + \phi_p$$

Since ρ_k can be estimated as $\hat{\rho}_k$,

$$\hat{\phi} := \begin{pmatrix} \hat{\phi}_1 \\ \hat{\phi}_2 \\ \hat{\phi}_3 \\ \vdots \\ \hat{\phi}_p \end{pmatrix} = \begin{pmatrix} 1 & \hat{\rho}_1 & \hat{\rho}_2 & \cdots & \hat{\rho}_{p-2} & \hat{\rho}_{p-1} \\ \hat{\rho}_1 & 1 & \hat{\rho}_1 & \cdots & \hat{\rho}_{p-3} & \hat{\rho}_{p-2} \\ \hat{\rho}_2 & \hat{\rho}_1 & 1 & \cdots & \hat{\rho}_{p-4} & \hat{\rho}_{p-3} \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ \hat{\rho}_{p-1} & \hat{\rho}_{p-2} & \hat{\rho}_{p-3} & \cdots & \hat{\rho}_1 & 1 \end{pmatrix}^{-1} \begin{pmatrix} \hat{\rho}_1 \\ \hat{\rho}_2 \\ \hat{\rho}_3 \\ \vdots \\ \hat{\rho}_p \end{pmatrix}$$

$\hat{\phi}$ is a **Yule-Walker estimator**.

Moreover, using the observation

$$\operatorname{cov}(\dot Z_t, \dot Z_t) = \operatorname{cov}(\dot Z_t, \phi_1 \dot Z_{t-1} + \cdots + \phi_p \dot Z_{t-p} + a_t) =$$

$$\phi_1 \operatorname{cov}(\dot Z_t, \dot Z_{t-1}) + \cdots + \phi_p \operatorname{cov}(\dot Z_t, \dot Z_{t-p}) + \operatorname{cov}(\dot Z_t, a_t) \Leftrightarrow$$

$$\gamma_0 = \phi_1 \gamma_1 + \cdots + \phi_p \gamma_p + \sigma_a^2 \Leftrightarrow$$

$$\sigma_a^2 = \gamma_0 - \phi_1 \gamma_1 - \cdots - \phi_p \gamma_p$$

we can estimate σ_a^2 as

$$\hat\sigma_a^2 = \hat\gamma_0 (1 - \hat\phi_1 \hat\rho_1 - \cdots - \hat\phi_p \hat\rho_p)$$

1.2 Drawbacks

The drawbacks of Yule-Walker estimation are

1) It is not robust if a_t are not normally distributed

2) It is not easy to find estimates for ARMA(p, q) model with $q \neq 0$

1.3 Example

Given the model $Z_t = \phi Z_{t-1} + a_t$ with $|\phi| < 1$, $a_t \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, \sigma_a^2)$, and $Z_1, Z_2, \ldots, Z_n$, the Yule-Walker estimators of ϕ and σ_a^2 can be found as follows:

$$\hat\phi = 1^{-1} \hat\rho_1 = \hat\rho_1$$

$$\hat\sigma_a^2 = \hat\gamma_0 (1 - \hat\phi \hat\rho_1) = \hat\gamma_0 (1 - \hat\rho_1^2)$$

2 Parameter Estimation, Diagnostic Checking and Model Selection. Maximum Likelihood (ML) Method

2.1 Conditional ML Estimation

Let's assume $\dot Z_t := Z_t - \mu$ follows ARMA(p, q) process

$$\dot Z_t = \phi_1 \dot Z_{t-1} + \cdots + \phi_p \dot Z_{t-p} + a_t - \theta_1 a_{t-1} - \cdots - \theta_q a_{t-q},$$

where $a_t \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, \sigma_a^2)$.

Given $Z_1, Z_2, \ldots, Z_n$, the joint density $f(Z_n, Z_{n-1}, \ldots, Z_1)$ is the **likelihood function** w.r.t. $\phi := (\phi_1, \ldots, \phi_p)^T$ and $\theta := (\theta_1, \ldots, \theta_q)^T$.

If the initial values $Z_* := (Z_{1-p}, Z_{2-p}, \ldots, Z_0)^T$ and $a_* := (a_{1-q}, a_{2-q}, \ldots, a_0)^T$ are used, then $f(Z_n, Z_{n-1}, \ldots, Z_1 \mid Z_*, a_*)$ is called the **conditional likelihood function**. It can be expressed as the product of conditional densities:

$$f(Z_n, Z_{n-1}, \ldots, Z_1 \mid Z_*, a_*) = \prod_{i=1}^{n} f(Z_i \mid Z_{i-1}, \ldots, Z_1, Z_*, a_*)$$

Therefore,

$$\dot Z_1 = \phi_1 \dot Z_0 + \phi_2 \dot Z_{-1} + \cdots + \phi_p \dot Z_{1-p} + a_1 - \theta_1 a_0 - \cdots - \theta_q a_{1-q} \Rightarrow$$

$$\dot Z_1 \mid (\dot Z_*, a_*) \sim \mathcal{N}(M_1, \sigma_a^2),$$

where

$$M_1 := \phi_1 \dot Z_0 + \phi_2 \dot Z_{-1} + \cdots + \phi_p \dot Z_{1-p} - \theta_1 a_0 - \cdots - \theta_q a_{1-q},$$

so,

$$f(\dot{Z}_1 \mid \dot{Z}_*, a_*) = \frac{1}{\sqrt{2\pi\sigma_a^2}} \exp\left(-\frac{(\dot{Z}_1 - M_1)^2}{2\sigma_a^2}\right)$$

Similarly,

$$\dot{Z}_2 = \phi_1 \dot{Z}_1 + \phi_2 \dot{Z}_0 + \cdots + \phi_p \dot{Z}_{2-p} + a_2 - \theta_1 a_1 - \cdots - \theta_q a_{2-q} \Rightarrow$$

$$\dot{Z}_2 \mid (\dot{Z}_1, \dot{Z}_*, a_*) \sim \mathcal{N}(M_2, \sigma_a^2),$$

where

$$M_2 := \phi_1 \dot{Z}_1 + \phi_2 \dot{Z}_0 + \cdots + \phi_p \dot{Z}_{2-p} - \theta_1 a_1 - \cdots - \theta_q a_{2-q},$$

so,

$$f(\dot{Z}_2 \mid \dot{Z}_1, \dot{Z}_*, a_*) = \frac{1}{\sqrt{2\pi\sigma_a^2}} \exp\left(-\frac{(\dot{Z}_2 - M_2)^2}{2\sigma_a^2}\right)$$

Generally,

$$M_n := \phi_1 \dot{Z}_{n-1} + \phi_2 \dot{Z}_{n-2} + \cdots + \phi_p \dot{Z}_{n-p} - \theta_1 a_{n-1} - \cdots - \theta_q a_{n-q},$$

$$f(\dot{Z}_n \mid \dot{Z}_{n-1}, \ldots, \dot{Z}_1, \dot{Z}_*, a_*) = \frac{1}{\sqrt{2\pi\sigma_a^2}} \exp\left(-\frac{(\dot{Z}_n - M_n)^2}{2\sigma_a^2}\right)$$

2.1.1 Definition

Let

$$L_*(\phi, \mu, \theta, \sigma_a^2) = f(Z_n, Z_{n-1}, \ldots, Z_1 \mid Z_*, a_*) =$$

$$\prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma_a^2}} \exp\left(-\frac{(\dot{Z}_i - M_i)^2}{2\sigma_a^2}\right) \Leftrightarrow$$

$$\ln L_*(\phi, \mu, \theta, \sigma_a^2) = -\frac{n}{2} \ln(2\pi\sigma_a^2) - \frac{S_*(\phi, \mu, \theta)}{2\sigma_a^2},$$

where $S_*(\phi, \mu, \theta) := \sum_{t=1}^{n} a_t^2(\phi, \mu, \theta)$ and $a_t(\phi, \mu, \theta) := \dot{Z}_t - M_t$.

Note: Do not confuse $a_t(\phi, \mu, \theta)$ with a_t, which is actually $a_t(\phi_0, \mu_0, \theta_0)$, where ϕ_0, μ_0, θ_0 are **true parameters**.

$\ln L_*(\phi, \mu, \theta, \sigma_a^2)$ is called the **conditional log-likelihood function**. Its *maximizer* $(\hat{\phi}, \hat{\mu}, \hat{\theta}, \hat{\sigma}_a^2)$ is called the **conditional maximum likelihood estimator (CMLE)** of $(\phi, \mu, \theta, \sigma_a^2)$.

Actually, $(\hat{\phi}, \hat{\mu}, \hat{\theta})$ is also the *minimizer* of $S_*(\phi, \mu, \theta)$. So, $(\hat{\phi}, \hat{\mu}, \hat{\theta})$ is called **conditional least squares estimator (CLSE)** of (ϕ, μ, θ).

2.1.2 Calculation

In practice, we first find $(\hat{\phi}, \hat{\mu}, \hat{\theta})$ by minimizing $S_*(\phi, \mu, \theta)$. Then calculate $\hat{\sigma}_a^2$ by

$$\hat{\sigma}_a^2 = \frac{S_*(\hat{\phi}, \hat{\mu}, \hat{\theta})}{n - p - q - 1}$$

Note: Generally, it might be very hard to find a closed form solution of this optimization problem.

2.1.3 Remarks

1) For AR(p) model, CMLE $\hat{\phi}$ is equivalent to Yule-Walker $\hat{\phi}$ and is called **ordinary least squares (OLS) estimator**. Basically, it means that you may treat it as a regression with the noise term distributed as $\mathcal{N}(0, \sigma_a^2)$

2) Usually, we take initial values $Z_* = a_* = \mathbf{0}$ or $Z_* = \boldsymbol{\mu}, a_* = \mathbf{0}$. It can be proved that these initial values do not affect the *quality* of the estimators.

   For example, in the MA(1) model:

$$a_{100}(\theta) = Z_{100} + \theta a_{99}(\theta) =$$

$$Z_{100} + \theta Z_{99} + \theta^2 Z_{98} + \cdots + \theta^{100} Z_0 + \ldots$$

3) When $a_t \overset{\text{i.i.d.}}{\sim} (0, \sigma_a^2) \neq \mathcal{N}(0, \sigma_a^2)$, the method is called **quasi-MLE (QMLE)**. The estimators remain consistent and asymptotically normal, however the estimator might not be the best

2.1.4 Example

Given the model $Z_t = \phi_1 Z_{t-1} + \cdots + \phi_p Z_{t-p} + a_t$ with the condition that all roots of $\phi(z) = 0$ lie outside the unit circle, $a_t \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, \sigma_a^2)$, and $Z_1, Z_2, \ldots, Z_n$, the minimizer of $S_*(\phi)$ can be found as follows:

$$a_t(\phi) = Z_t - \phi_1 Z_{t-1} - \cdots - \phi_p Z_{t-p} =: Z_t - \tilde{Z}_{t-1}^T \phi$$

$$S_*(\phi) = \sum_{t=1}^{n} a_t^2(\phi)$$

$$\frac{\partial S_*}{\partial \phi} = \sum_{t=1}^{n} \left(\frac{\partial a_t}{\partial \phi} a_t + a_t \frac{\partial a_t}{\partial \phi} \right) = 2 \sum_{t=1}^{n} \frac{\partial a_t}{\partial \phi} a_t = 2 \sum_{t=1}^{n} (-\tilde{Z}_{t-1})(Z_t - \tilde{Z}_{t-1}^T \phi) =$$

$$-2 \sum_{t=1}^{n} \tilde{Z}_{t-1} Z_t + 2 \sum_{t=1}^{n} \tilde{Z}_{t-1} \tilde{Z}_{t-1}^T \phi$$

$$\frac{\partial S_*}{\partial \phi}(\hat{\phi}) = 0 \Leftrightarrow$$

$$\hat{\phi} = \left(\sum_{t=1}^{n} \tilde{Z}_{t-1} \tilde{Z}_{t-1}^T \right)^{-1} \left(\sum_{t=1}^{n} \tilde{Z}_{t-1} Z_t \right)$$

To make sure that found $\hat{\phi}$ is the *minimizer*, we need to calculate the Hessian:

$$\frac{\partial^2 S_*}{\partial \phi \partial \phi^T} = 2 \sum_{t=1}^{n} \tilde{Z}_{t-1} \tilde{Z}_{t-1}^T$$

Let's check if it is positive definite:

$$x \frac{\partial^2 S_*}{\partial \phi \partial \phi^T} x^T = 2 \sum_{t=1}^{n} x \tilde{Z}_{t-1} \tilde{Z}_{t-1}^T x^T = 2 \sum_{t=1}^{n} (x \tilde{Z}_{t-1})(x \tilde{Z}_{t-1})^T > 0$$

Finally,

$$\hat{\sigma}_a^2 = \frac{1}{n-p} \sum_{t=1}^{n} (Z_t - \tilde{Z}_{t-t}^T \phi)^2$$

2.2 Unconditional ML Estimation

Let's assume $\dot{Z}_t := Z_t - \mu$ follows ARMA(p, q) process

$$\dot{Z}_t = \phi_1 \dot{Z}_{t-1} + \cdots + \phi_p \dot{Z}_{t-p} + a_t - \theta_1 a_{t-1} - \cdots - \theta_q a_{t-q},$$

where $a_t \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, \sigma_a^2)$.

2.2.1 Definition

Unconditional log-likelihood function is

$$\ln L(\phi, \mu, \theta, \sigma_a^2) = -\frac{n}{2} \ln(2\pi\sigma_a^2) - \frac{S(\phi, \mu, \theta)}{2\sigma_a^2},$$

where $S(\phi, \mu, \theta) = \sum_{t=-\infty}^{n} \left[\mathbb{E}(a_t \mid \phi, \mu, \theta, Z)\right]^2$, $Z = (Z_1, Z_2, \ldots, Z_n)$.

Its *maximizer* $(\hat{\phi}, \hat{\mu}, \hat{\theta}, \hat{\sigma}_a^2)$ is called the **unconditional maximum likelihood estimator (UMLE)** of $(\phi, \mu, \theta, \sigma_a^2)$.

Actually, $(\hat{\phi}, \hat{\mu}, \hat{\theta})$ is also the *minimizer* of $S(\phi, \mu, \theta)$. So, $(\hat{\phi}, \hat{\mu}, \hat{\theta})$ is called **unconditional least squares estimator (ULSE)** of (ϕ, μ, θ).

2.2.2 Calculation

$S(\phi, \mu, \theta)$ is approximated by

$$S(\phi, \mu, \theta) = \sum_{t=-M}^{n} \left[\mathbb{E}(a_t \mid \phi, \mu, \theta, Z)\right]^2,$$

where $M = M(\epsilon)$ is an integer large enough so that

$$|\mathbb{E}(a_t \mid \phi, \mu, \theta, Z) - \mathbb{E}(a_{t-1} \mid \phi, \mu, \theta, Z)| \leq \epsilon \quad \forall t \leq -M - 1$$

$\mathbb{E}(a_t \mid \phi, \mu, \theta, Z)$ can be calculated using **Backcasting method**. For this, note that ARMA(p, q) model can be represented in both forward and backward forms using the *same* coefficients:

$$(1 - \phi_1 B - \phi_2 B^2 - \cdots - \phi_p B^p) Z_t = (1 - \theta_1 B - \theta_2 B^2 - \cdots - \theta_p B^q) a_t \quad \text{(forward form)}$$
$$(1 - \phi_1 F - \phi_2 F^2 - \cdots - \phi_p F^p) Z_t = (1 - \theta_1 F - \theta_2 F^2 - \cdots - \theta_p F^q) e_t, \quad \text{(backward form)}$$

where $BZ_t := Z_{t-1}$ and $FZ_t := Z_{t+1}$.

Finally, σ_a^2 can be estimated by

$$\hat{\sigma}_a^2 = \frac{S(\hat{\phi}, \hat{\mu}, \hat{\theta})}{n}$$

Note: Generally, it might be very hard to find a closed form solution of this optimization problem.

2.2.3 Example

Given the model $Z_t = \phi Z_{t-1} + a_t$ with $|\phi| < 1$, $a_t \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, \sigma_a^2)$, and $Z_1, Z_2, \ldots, Z_n$, the $S(\phi)$ can be found as follows:

$$\mathbb{E}(a_n \mid \phi, Z) = \mathbb{E}(Z_n - \phi Z_{n-1} \mid \phi, Z) = \mathbb{E}(Z_n \mid \phi, Z) - \phi \mathbb{E}(Z_{n-1} \mid \phi, Z) = Z_n - \phi Z_{n-1}$$

$$\vdots$$

$$\mathbb{E}(a_2 \mid \phi, Z) = \mathbb{E}(Z_2 - \phi Z_{n-1} \mid \phi, Z) = \mathbb{E}(Z_2 \mid \phi, Z) - \phi\mathbb{E}(Z_1 \mid \phi, Z) = Z_2 - \phi Z_1$$
$$\mathbb{E}(a_1 \mid \phi, Z) = \mathbb{E}(Z_1 - \phi Z_0 \mid \phi, Z) = \mathbb{E}(Z_1 \mid \phi, Z) - \phi\mathbb{E}(Z_0 \mid \phi, Z) = Z_1 - \phi\mathbb{E}(Z_0 \mid \phi, Z)$$
$$\mathbb{E}(a_0 \mid \phi, Z) = \mathbb{E}(Z_0 - \phi Z_{-1} \mid \phi, Z) = \mathbb{E}(Z_0 \mid \phi, Z) - \phi\mathbb{E}(Z_{-1} \mid \phi, Z)$$
$$\mathbb{E}(a_{-1} \mid \phi, Z) = \mathbb{E}(Z_{-1} - \phi Z_{-2} \mid \phi, Z) = \mathbb{E}(Z_{-1} \mid \phi, Z) - \phi\mathbb{E}(Z_{-2} \mid \phi, Z)$$
$$\vdots$$
$$\mathbb{E}(a_{-M} \mid \phi, Z) = \mathbb{E}(Z_{-M} - \phi Z_{-M-1} \mid \phi, Z) = \mathbb{E}(Z_{-M} \mid \phi, Z) - \phi\mathbb{E}(Z_{-M-1} \mid \phi, Z)$$

So, we need to find $\mathbb{E}(Z_{-t} \mid \phi, Z)$:

$$Z_t = \sum_{i=0}^{\infty} \phi^i F^i e_t = e_t + \phi e_{t+1} + \phi^2 e_{t+2} + \cdots \Rightarrow$$

$$e_{-t} \text{ is independent of } Z \quad \forall t \geq 0$$

$$\mathbb{E}(Z_0 \mid \phi, Z) = \mathbb{E}(e_0 + \phi Z_1 \mid \phi, Z) = \mathbb{E}(e_0 \mid \phi, Z) + \phi\mathbb{E}(Z_1 \mid \phi, Z) = \phi Z_1$$
$$\mathbb{E}(Z_{-1} \mid \phi, Z) = \mathbb{E}(e_{-1} + \phi Z_0 \mid \phi, Z) = \mathbb{E}(e_{-1} \mid \phi, Z) + \phi\mathbb{E}(Z_0 \mid \phi, Z) = \phi^2 Z_1$$
$$\vdots$$
$$\mathbb{E}(Z_{-M} \mid \phi, Z) = \mathbb{E}(e_{-M} + \phi Z_{-M+1} \mid \phi, Z) = \mathbb{E}(e_{-M} \mid \phi, Z) + \phi\mathbb{E}(Z_{-M+1} \mid \phi, Z) = \phi^{M+1} Z_1$$

Therefore,

$$\mathbb{E}(a_1 \mid \phi, Z) = Z_1 - \phi\mathbb{E}(Z_0 \mid \phi, Z) = Z_1 - \phi^2 Z_1 = (1-\phi^2) Z_1$$
$$\mathbb{E}(a_0 \mid \phi, Z) = \mathbb{E}(Z_0 \mid \phi, Z) - \phi\mathbb{E}(Z_{-1} \mid \phi, Z) = \phi Z_1 - \phi^3 Z_1 = (1-\phi^2)\phi Z_1$$
$$\mathbb{E}(a_{-1} \mid \phi, Z) = \mathbb{E}(Z_{-1} \mid \phi, Z) - \phi\mathbb{E}(Z_{-2} \mid \phi, Z) = \phi^2 Z_1 - \phi^4 Z_1 = (1-\phi^2)\phi^2 Z_1$$
$$\vdots$$
$$\mathbb{E}(a_{-M} \mid \phi, Z) = \mathbb{E}(Z_{-M} \mid \phi, Z) - \phi\mathbb{E}(Z_{-M-1} \mid \phi, Z) = \phi^{M+1} Z_1 - \phi^{M+3} Z_1 = (1-\phi^2)\phi^{M+1} Z_1$$

and, finally,

$$S(\phi) = \sum_{t=-M}^{n} \left[\mathbb{E}(a_t \mid \phi, Z)\right]^2 = \sum_{t=2}^{n} \left[Z_t - \phi Z_{t-1}\right]^2 + \sum_{t=-M}^{1} \left[(1-\phi^2)\phi^{1-t} Z_1\right]^2$$

2.3 Exact ML Estimation

Let's assume $\dot{Z}_t := Z_t - \mu$ follows ARMA(p, q) process

$$\dot{Z}_t = \phi_1 \dot{Z}_{t-1} + \cdots + \phi_p \dot{Z}_{t-p} + a_t - \theta_1 a_{t-1} - \cdots - \theta_q a_{t-q},$$

where $a_t \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, \sigma_a^2)$.

2.3.1 Definition

Given $Z_1, Z_2, \ldots, Z_n$, the joint density $f(Z_n, Z_{n-1}, \ldots, Z_1)$ is the **(exact) likelihood function**. Let

$$L_(\phi, \mu, \theta, \sigma_a^2) = f(Z_n, Z_{n-1}, \ldots, Z_1)$$

Its *maximizer* $(\hat{\phi}, \hat{\mu}, \hat{\theta}, \hat{\sigma}_a^2)$ is called the **(exact) maximum likelihood estimator (MLE)** of $(\phi, \mu, \theta, \sigma_a^2)$.

2.3.2 Calculation

Note: Even to find the exact likelihood function, i.e. the function to optimize, is very hard.

2.3.3 Example

Given the model $Z_t = \phi Z_{t-1} + a_t$ with $|\phi| < 1$, $a_t \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, \sigma_a^2)$, and $Z_1, Z_2, \ldots, Z_n$, the exact likelihood function can be found as follows:

$$s_1 := Z_1 = \sum_{j=0}^{\infty} \phi^j a_{1-j} \sim \mathcal{N}\left(0, \frac{\sigma_a^2}{1-\phi^2}\right)$$

Consider independent (but not identically distributed) $s_1, a_2, \ldots, a_n$. Their joint density is

$$f(s_1, a_2, \ldots, a_n) = f(s_1) \times \prod_{i=2}^{n} f(a_i) = \frac{\sqrt{1-\phi^2}}{\sqrt{2\pi\sigma_a^2}} \exp\left(-\frac{s_1^2(1-\phi^2)}{2\sigma_a^2}\right) \times \left(\frac{1}{\sqrt{2\pi\sigma_a^2}}\right)^{n-1} \exp\left(-\frac{\sum_{i=2}^{n} a_i^2}{2\sigma_a^2}\right)$$

Using the transformation

$$\begin{aligned} Z_1 &= s_1 \\ Z_2 &= \phi Z_1 + a_2 \\ &\vdots \\ Z_n &= \phi Z_{n-1} + a_n, \end{aligned}$$

we can directly obtain the joint density $f(Z_1, Z_2, \ldots, Z_n)$. The Jacobian of the transformation is

$$J = \begin{vmatrix} 1 & 0 & . & . & . & 0 & 0 & 0 \\ -\phi & 1 & 0 & . & . & 0 & 0 & 0 \\ 0 & -\phi & 1 & 0 & . & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & . & . & . & . & 0 & -\phi & 1 \end{vmatrix} = 1$$

Thus,

$$L(\phi, \sigma_a^2) = f(Z_1, Z_2, \ldots, Z_n) = J f(s_1, a_2, \ldots, a_n) =$$

$$\frac{\sqrt{1-\phi^2}}{\sqrt{2\pi\sigma_a^2}} \exp\left(-\frac{Z_1^2(1-\phi^2)}{2\sigma_a^2}\right) \times \left(\frac{1}{\sqrt{2\pi\sigma_a^2}}\right)^{n-1} \exp\left(-\frac{\sum_{i=2}^{n}(Z_i - \phi Z_{i-1})^2}{2\sigma_a^2}\right)$$

$$\ln L(\phi, \sigma_a^2) = -\frac{n}{2}\ln(2\pi\sigma_a^2) + \frac{1}{2}\ln(1-\phi^2) - \frac{S(\phi)}{2\sigma_a^2},$$

where $S(\phi) := Z_1^2(1-\phi^2) + \sum_{i=2}^{n}(Z_i - \phi Z_{i-1})^2$.

Note: You can use $f(Z_1, Z_2, \ldots, Z_n) = f(Z_n \mid Z_{n-1}) \times \cdots \times f(Z_2 \mid Z_1) \times f(Z_1)$ expansion and $Z_i = \phi Z_{i-1} + a_i \sim \mathcal{N}(\phi Z_{n-1}, \sigma_a^2)$ observation to imply the same result.

2.4 How To Find The Estimate?

1) In CLS we are looking for $(\hat{\phi}, \hat{\mu}, \hat{\theta})$ minimizing $S_*(\phi, \mu, \theta) = \sum_{t=1}^{n} a_t^2(\phi, \mu, \theta)$

2) In ULS, for example, for AR(1) model we are looking for $\hat{\phi}$ minimizing $S(\phi) = \sum_{t=-M}^{n} \left[\mathbb{E}(a_t \mid \phi, Z)\right]^2 = \sum_{t=2}^{n} \left[Z_t - \phi Z_{t-1}\right]^2 + \sum_{t=-M}^{1} \left[(1-\phi^2)\phi^{1-t} Z_1\right]^2$

3) In ML, for example, for AR(1) model we are looking for $\hat{\phi}$ minimizing $S(\phi) := Z_1^2(1-\phi^2) + \sum_{i=2}^{n}(Z_i - \phi Z_{i-1})^2$

All these tasks eventually lead us to the Mathematical optimization. Some iterative methods such as Gradient descent and Newton's method can be applied to find the estimate $(\hat{\phi}, \hat{\mu}, \hat{\theta})$.