Maximum Likelihood Estimation.

- Suppose that $\{x_t\}$ is a Gaussian time series with mean 0 and autocovariance function $\kappa(i,j) = E[x_i x_j]$.
- Let $X_T = (x_1, \cdots, x_T)^T$ and $\hat{X}_T = (\hat{x}_1, \cdots, \hat{x}_T)^T$, where as before, $\hat{x}_1 = 0$ and $\hat{x}_n = E[x_n \mid x_1, \cdots, x_{n-1}] = P_{n-1} x_n$, for $n \geq 2$.
- Let Γ_T denote the covariance matrix, $\Gamma_T = E[X_T X_T^T]$, and assume that Γ_T is nonsingular.
- The likelihood of X_T is given by:

$$L(\Gamma_T) = (2\pi)^{-T/2} \, |\Gamma_T|^{-1/2} \exp\left(-\tfrac{1}{2} X_T^T \Gamma_T^{-1} X_T\right).$$

- We now show that, the direct calculation of $|\Gamma_T|$ and Γ_T^{-1} can be avoided by expressing this in terms of the innovation $x_n - \hat{x}_n$.

Innovations.

- Recall that $X_T = C_T(X_T - \hat{X}_T)$.
- We also note that, components of $X_T - \hat{X}_T$ uncorrelated with a diagonal covariance matrix

$$\mathcal{D}_T = \operatorname{diag}(v_0, \cdots, v_{T-1}).$$

- Clearly,

$$\Gamma_T = \operatorname{Cov}(X_T) = C_T \operatorname{Cov}(X_T - \hat{X}_T) C_T^T = C_T \mathcal{D}_T C_T^T.$$

- So,

$$X_T^T \Gamma_T^{-1} X_T = (X_T - \hat{X}_T)^T C_T^T (C_T^T)^{-1} \mathcal{D}_T^{-1} C_T^{-1} C_T (X_T - \hat{X}_T)$$
$$= (X_T - \hat{X}_T)^T \mathcal{D}_T^{-1} (X_T - \hat{X}_T).$$

Also

$$|\Gamma_T| = |C_T|^2 |\mathcal{D}_T| = \prod_{i=0}^{T-1} v_i.$$

(Note that C_T is a lower triangular matrix with diagonal entries equal to 1.)

Likelihood.

- The likelihood of X_T is given by:

$$L(\Gamma_T) = \frac{1}{\sqrt{(2\pi)^T v_0 \cdots v_{T-1}}} \exp\left\{-\frac{1}{2}\sum_{n=1}^{T} (x_n - \hat{x}_n)^2 / v_{n-1}\right\}.$$

- If Γ_T is expressible in terms of finite number of unknown parameters, the maximum likelihood estimators of the parameters are those values that maximize L for the given data set.
- When $\{X_t\}$ is not Gaussian, it often makes sense to use the Gaussian likelihood even through it is wrong.
- The maximum Gaussian likelihood estimators of ARMA coefficients has the same large-sample distribution for $\{Z_t\} \sim WN(0, \sigma^2)$, regardless of whether or not $\{Z_t\}$ is Gaussian.

One step predictors for ARMA(p,q) processes.

- Let $\{X_t\}$ be a ARMA(p,q) process, $m = \max(p,q)$.

- We know that

$$\hat{X}_{n+1} = \begin{cases} \sum_{j=1}^{n} \theta_{nj} (X_{n+1-j} - \hat{X}_{n+1-j}) & 1 \leq n \leq m. \\ \phi_1 X_n + \cdots + \phi_p X_{n+1-p} + \sum_{j=1}^{q} \theta_{nj} (X_{n+1-j} - \hat{X}_{n+1-j}), & n \geq m \end{cases}$$

and

$$E\left[(X_{n+1} - \hat{X}_{n+1})^2\right] = \sigma^2 v_n,$$

Where θ_{nj} and v_n are determined from the Innovations algorithm.

The mle for ARMA(p,q).

- The maximum likelihood estimators are given by:

$$\hat{\sigma}^2 = \frac{1}{n} S(\hat{\phi}, \hat{\theta}), \text{ where } S(\hat{\phi}, \hat{\theta}) = \sum_{j=1}^{T} (x_j - \hat{x}_j)^2 / v_{j-1}$$

and $\hat{\phi}, \hat{\theta}$ are the values ~~which~~ that maximise,

$$l(\phi, \theta) = \log\left(\frac{1}{n} S(\hat{\phi}, \theta)\right) + 1/n \sum_{j=1}^{n} \log v_{j-1}.$$

- The minimisation must be done numerically.