On Sequentially Normalized Maximum Likelihood Models

Teemu Roos and Jorma Rissanen

Complex Systems Computation Group
Helsinki Institute for Information Technology HIIT
FINLAND

WITMSE-08, Tampere, Finland, August 18, 2008

HELSINKI
INSTITUTE FOR
INFORMATION
TECHNOLOGY

Universal Models

Given a sequence, $x^n = (x_1, \ldots, x_n)$, the best fitting model in a model class, $\mathcal{M}$, is the **maximum likelihood** model

$$\sup_{\theta \in \Theta} p(x^n \,;\, \theta) = p(x^n \,;\, \hat{\theta}(x^n)) \ .$$

Universal Models

Given a sequence, $x^n = (x_1, \ldots, x_n)$, the best fitting model in a model class, $\mathcal{M}$, is the **maximum likelihood** model

$$\sup_{\theta \in \Theta} p(x^n ; \theta) = p(x^n ; \hat{\theta}(x^n)) \ .$$

A **universal model** $q(\cdot)$ achieves almost as short a code-length as the ML model:

$$\lim_{n \to \infty} \frac{1}{n} \ln \frac{p(x^n ; \hat{\theta}(x^n))}{q(x^n)} = 0 \ ,$$

i.e., the log-likelihood ratio (**'regret'**) is allowed to grow sublinearly.

Universal Models

Given a sequence, $x^n = (x_1, \ldots, x_n)$, the best fitting model in a model class, $\mathcal{M}$, is the **maximum likelihood** model

$$\sup_{\theta \in \Theta} p(x^n \,;\, \theta) = p(x^n \,;\, \hat{\theta}(x^n)) \ .$$

A **universal model** $q(\cdot)$ achieves almost as short a code-length as the ML model:

$$\lim_{n \to \infty} \frac{1}{n} \ln \frac{p(x^n \,;\, \hat{\theta}(x^n))}{q(x^n)} = 0 \ ,$$

i.e., the log-likelihood ratio (**'regret'**) is allowed to grow sublinearly.

The minimax optimal (NML) model (Shtarkov, 1987):

$$p_{\mathrm{NML}}(x^n) = \frac{p(x^n \,;\, \hat{\theta}(x^n))}{C_n} \ , \quad C_n = \sum_{x^n \in \mathcal{X}^n} p(x^n \,;\, \hat{\theta}(x^n)) \ .$$

Approximations and Alternatives to NML

1. Approximations:
   - BIC: $\frac{k}{2} \ln n$

Approximations and Alternatives to NML

1. Approximations:
    - BIC: $\frac{k}{2} \ln n$
    - Fisher information: $\frac{k}{2} \ln \frac{n}{2\pi} + \ln \int_{\Theta} \sqrt{\det I(\theta)}\, d\theta$

Approximations and Alternatives to NML

1. Approximations:
   - BIC: $\frac{k}{2} \ln n$
   - Fisher information: $\frac{k}{2} \ln \frac{n}{2\pi} + \ln \int_{\Theta} \sqrt{\det I(\theta)}\, d\theta$
2. Monte Carlo methods

Approximations and Alternatives to NML

1. Approximations:
   - BIC: $\frac{k}{2} \ln n$
   - Fisher information: $\frac{k}{2} \ln \frac{n}{2\pi} + \ln \int_{\Theta} \sqrt{\det I(\theta)}\, d\theta$
2. Monte Carlo methods
3. Other forms of universal models:
   - two-part

Approximations and Alternatives to NML

1. Approximations:
   - BIC: $\dfrac{k}{2} \ln n$
   - Fisher information: $\dfrac{k}{2} \ln \dfrac{n}{2\pi} + \ln \displaystyle\int_{\Theta} \sqrt{\det I(\theta)}\, d\theta$
2. Monte Carlo methods
3. Other forms of universal models:
   - two-part
   - plug-in (predictive least squares (PLS), predictive MDL)

Approximations and Alternatives to NML

1. Approximations:
   - BIC: $\dfrac{k}{2} \ln n$
   - Fisher information: $\dfrac{k}{2} \ln \dfrac{n}{2\pi} + \ln \displaystyle\int_{\Theta} \sqrt{\det I(\theta)}\, d\theta$
2. Monte Carlo methods
3. Other forms of universal models:
   - two-part
   - plug-in (predictive least squares (PLS), predictive MDL)
   - mixtures (Bayes)

Approximations and Alternatives to NML

1. Approximations:
   - BIC: $\dfrac{k}{2} \ln n$
   - Fisher information: $\dfrac{k}{2} \ln \dfrac{n}{2\pi} + \ln \displaystyle\int_{\Theta} \sqrt{\det I(\theta)}\, d\theta$
2. Monte Carlo methods
3. Other forms of universal models:
   - two-part
   - plug-in (predictive least squares (PLS), predictive MDL)
   - mixtures (Bayes)
   - **sequential NML**

Sequential NML

Basic Idea

1. Maximize likelihood (like in NML).
2. Normalize over current observation, x_i.
3. Combine obtained conditionals.

Sequential NML

Basic Idea

1. Maximize likelihood (like in NML).
2. Normalize over current observation, x_i.
3. Combine obtained conditionals.

Always gives a stochastic process (unlike NML).

Sequential NML

Basic Idea

1. Maximize likelihood (like in NML).
2. Normalize over current observation, x_i.
3. Combine obtained conditionals.

Always gives a stochastic process (unlike NML).

Each conditional is “locally” minimax optimal.

Sequential NML

The sNML (variant 1) model is defined as

$$p_{\mathrm{sNML1}}(x^n) = \prod_{i=1}^{n} \frac{p(x_i \mid x^{i-1}\ ;\ \hat{\theta}(x^i))}{K_i(x^{i-1})}$$

Sequential NML

The sNML (variant 1) model is defined as

$$p_{\mathrm{sNML1}}(x^n) = \prod_{i=1}^{n} \frac{p(x_i \mid x^{i-1} \,;\, \hat{\theta}(x^i))}{K_i(x^{i-1})}$$

$$K_i(x^{i-1}) = \sum_{x_i} p(x_i \mid x^{i-1} \,;\, \hat{\theta}(x^i))$$

Sequential NML

The sNML (variant 1) model is defined as

$$p_{\mathrm{sNML1}}(x^n) = \prod_{i=1}^{n} \frac{p(x_i \mid x^{i-1} \; ; \; \hat{\theta}(x^i))}{K_i(x^{i-1})}$$

$$K_i(x^{i-1}) = \sum_{x_i} p(x_i \mid x^{i-1} \; ; \; \hat{\theta}(x^i))$$

Compare to the plug-in model:

$$p_{\mathrm{plug-in}}(x^n) = \prod_{i=1}^{n} p(x_i \mid x^{i-1} \; ; \; \hat{\theta}(x^{i-1}))$$

Sequential NML

The sNML (variant 1) model is defined as

$$p_{\mathrm{sNML1}}(x^n) = \prod_{i=1}^{n} \frac{p(x_i \mid x^{i-1}\,;\,\hat{\theta}(x^i))}{K_i(x^{i-1})}$$

$$K_i(x^{i-1}) = \sum_{x_i} p(x_i \mid x^{i-1}\,;\,\hat{\theta}(x^i))$$

Compare to the 'ordinary' NML model:

$$p_{\mathrm{NML}}(x^n) = \frac{p(x^n\,;\,\hat{\theta}(x^n))}{C_n}$$

$$C_n = \sum_{x^n \in \mathcal{X}^n} p(x^n\,;\,\hat{\theta}(x^n))$$

Sequential NML

The sNML (variant 1) model is defined as

$$p_{\mathrm{sNML1}}(x^n) = \prod_{i=1}^{n} \frac{p(x_i \mid x^{i-1} \,;\, \hat{\theta}(x^i))}{K_i(x^{i-1})}$$

$$K_i(x^{i-1}) = \sum_{x_i} p(x_i \mid x^{i-1} \,;\, \hat{\theta}(x^i))$$

The second variant of sNML is defined as

$$p_{\mathrm{sNML2}}(x^n) = \prod_{i=1}^{n} \frac{p(x^i \,;\, \hat{\theta}(x^i))}{K_i'(x^{i-1})}$$

$$K_i'(x^{i-1}) = \sum_{x_i} p(x^i \,;\, \hat{\theta}(x^i))$$

Computational Complexity

The only computational issue in applying NML/sNML in the discrete (multinomial) case is the normalization factor.

Computational Complexity

The only computational issue in applying NML/sNML in the discrete (multinomial) case is the normalization factor.

- In NML, we have a **sum of products**:

$$C_n = \sum_{x^n} p(x^n \,;\, \hat{\theta}(x^n)) = \sum_{x^n} \prod_{i=1}^{n} p(x_i \mid x^{i-1} \,;\, \hat{\theta}(x^n)).$$

Computational Complexity

The only computational issue in applying NML/sNML in the discrete (multinomial) case is the normalization factor.

- In NML, we have a **sum of products**:

$$C_n = \sum_{x^n} p(x^n \,;\, \hat{\theta}(x^n)) = \sum_{x^n} \prod_{i=1}^{n} p(x_i \mid x^{i-1} \,;\, \hat{\theta}(x^n)).$$

- In sNML, we have a **product of sums**:

$$Z_n(x^n) = \prod_{i=1}^{n} K_i(x^{i-1}) = \prod_{i=1}^{n} \sum_{x_i'} p(x_i' \mid x^{i-1} \,;\, \hat{\theta}(x^{i-1}, x_i')).$$

Computational Complexity

The only computational issue in applying NML/sNML in the discrete (multinomial) case is the normalization factor.

- In NML, we have a **sum of products**:

$$C_n = \sum_{x^n} p(x^n\,;\,\hat{\theta}(x^n)) = \sum_{x^n} \prod_{i=1}^{n} p(x_i \mid x^{i-1}\,;\,\hat{\theta}(x^n)).$$

- In sNML, we have a **product of sums**:

$$Z_n(x^n) = \prod_{i=1}^{n} K_i(x^{i-1}) = \prod_{i=1}^{n} \sum_{x_i'} p(x_i' \mid x^{i-1}\,;\,\hat{\theta}(x^{i-1}, x_i')).$$

Remarkably, we can evaluate both in $\mathcal{O}(n)$ time (Kontkanen & Myllymäki, 2007).

Computational Complexity

The only computational issue in applying NML/sNML in the discrete (multinomial) case is the normalization factor.

- In NML, we have a **sum of products**:

$$C_n = \sum_{x^n} p(x^n\,;\,\hat{\theta}(x^n)) = \sum_{x^n} \prod_{i=1}^{n} p(x_i \mid x^{i-1}\,;\,\hat{\theta}(x^n)).$$

- In sNML, we have a **product of sums**:

$$Z_n(x^n) = \prod_{i=1}^{n} K_i(x^{i-1}) = \prod_{i=1}^{n} \sum_{x_i'} p(x_i' \mid x^{i-1}\,;\,\hat{\theta}(x^{i-1}, x_i')).$$

Remarkably, we can evaluate both in $\mathcal{O}(n)$ time (Kontkanen & Myllymäki, 2007). In general, **NML is hard** but **sNML is easy**.

Properties of sNML: Bernoulli case

Both variants of sNML are universal:

Properties of sNML: Bernoulli case

Both variants of sNML are universal:

- sNML1 is identical to Laplace's "add one" rule:

$$P_{\mathrm{sNML1}}(1 \mid x^n) = P_{\mathrm{Lap}}(1 \mid x^n) = \frac{n_1 + 1}{n + 2}.$$

Properties of sNML: Bernoulli case

Both variants of sNML are universal:

- sNML1 is identical to Laplace's "add one" rule:

$$P_{\mathrm{sNML1}}(1 \mid x^n) = P_{\mathrm{Lap}}(1 \mid x^n) = \frac{n_1 + 1}{n + 2}.$$

- (Takimoto and Warmuth, 2000): The worst-case regret of sNML2 is bounded by

$$\sup_{x^n} \ln \frac{p(x^n \,;\, \hat{\theta}(x^n))}{p_{\mathrm{sNML2}}(x^n)} \leq \frac{1}{2} \ln(n+1) + \frac{1}{2}.$$

Properties of sNML: Bernoulli case

Both variants of sNML are universal:

- sNML1 is identical to Laplace's "add one" rule:

$$P_{\mathrm{sNML1}}(1 \mid x^n) = P_{\mathrm{Lap}}(1 \mid x^n) = \frac{n_1 + 1}{n + 2}.$$

- (Takimoto and Warmuth, 2000): The worst-case regret of sNML2 is bounded by

$$\sup_{x^n} \ln \frac{p(x^n\,;\,\hat{\theta}(x^n))}{p_{\mathrm{sNML2}}(x^n)} \leq \frac{1}{2}\ln(n+1) + \frac{1}{2}.$$

Is the sun going to rise? $x^n = 111\ldots1$.

$$\left(P_{\mathrm{Lap}}(1 \mid x^n)\right)_{n=0}^{\infty} = \left(\frac{1}{2}, \frac{2}{3}, \frac{3}{4}, \frac{4}{5}, \ldots\right).$$

Both variants of sNML are universal:

- sNML1 is identical to Laplace's "add one" rule:

$$P_{\mathrm{sNML1}}(1 \mid x^n) = P_{\mathrm{Lap}}(1 \mid x^n) = \frac{n_1 + 1}{n + 2}.$$

- (Takimoto and Warmuth, 2000): The worst-case regret of sNML2 is bounded by

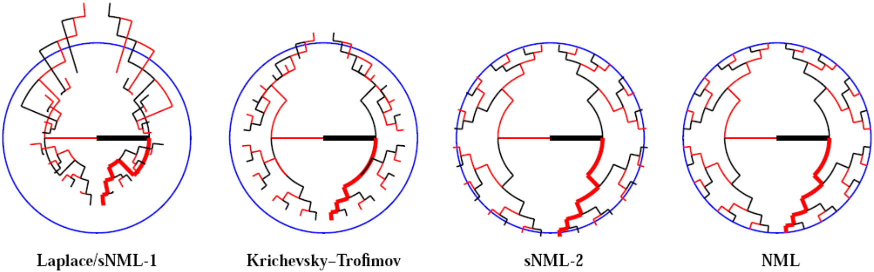
$$\sup_{x^n} \ln \frac{p(x^n\,;\,\hat{\theta}(x^n))}{p_{\mathrm{sNML2}}(x^n)} \leq \frac{1}{2} \ln(n+1) + \frac{1}{2}.$$

Is the sun going to rise? $x^n = 111\ldots1$.

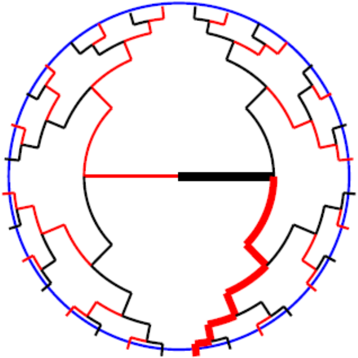
$$(P_{\mathrm{Lap}}(1 \mid x^n))_{n=0}^{\infty} = \left(\frac{1}{2}, \frac{2}{3}, \frac{3}{4}, \frac{4}{5}, \ldots\right).$$

$$(P_{\mathrm{sNML2}}(1 \mid x^n))_{n=0}^{\infty} = \left(\frac{1}{2}, \frac{4}{5}, \frac{27}{31}, \frac{256}{283}, \ldots\right).$$

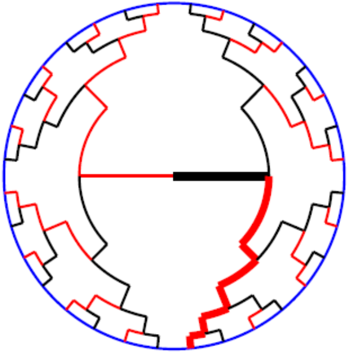
Regrets Visualized



Regrets Visualized



sNML-2



NML

Linear-Quadratic Models

Linear model $y_t = \beta' \bar{x}_t + \epsilon_t$ with Gaussian errors $\epsilon_t \sim \mathcal{N}(0, \sigma^2)$.

Linear-Quadratic Models

Linear model $y_t = \beta' \bar{x}_t + \epsilon_t$ with Gaussian errors $\epsilon_t \sim \mathcal{N}(0, \sigma^2)$.

Least squares parameters $b_t' = \arg\min_\beta \|\beta' X_t - y^t\|^2$.

Linear-Quadratic Models

Linear model $y_t = \beta' \bar{x}_t + \epsilon_t$ with Gaussian errors $\epsilon_t \sim \mathcal{N}(0, \sigma^2)$.

Least squares parameters $b'_t = \arg\min_\beta \|\beta' X_t - y^t\|^2$.

Consider the following three representations:

$$
\begin{aligned}
y_t &= b'_{t-1} \bar{x}_t + e_t & (1) \quad & \text{“plug-in”} \\
y_t &= b'_n \bar{x}_t + \hat{\epsilon}_t(n) & (2) \quad & \text{“least-squares”} \\
y_t &= b'_t \bar{x}_t + \hat{e}_t & (3) \quad & \text{“sNML”}
\end{aligned}
$$

Linear-Quadratic Models

Linear model $y_t = \beta' \bar{x}_t + \epsilon_t$ with Gaussian errors $\epsilon_t \sim \mathcal{N}(0, \sigma^2)$.

Least squares parameters $b'_t = \arg\min_\beta \|\beta' X_t - y^t\|^2$.

Consider the following three representations:

$$
\begin{aligned}
y_t &= b'_{t-1}\bar{x}_t + e_t && (1) \quad \text{“plug-in”} \\
y_t &= b'_n\bar{x}_t + \hat{\epsilon}_t(n) && (2) \quad \text{“least-squares”} \\
y_t &= b'_t\bar{x}_t + \hat{e}_t && (3) \quad \text{“sNML”}
\end{aligned}
$$

Representation (1) corresponds to the **predictive least squares** (PLS) model selection criterion: $\sum_{i=m+1}^{n} e_t^2$.

Linear-Quadratic Models

Linear model $y_t = \beta' \bar{x}_t + \epsilon_t$ with Gaussian errors $\epsilon_t \sim \mathcal{N}(0, \sigma^2)$.

Least squares parameters $b'_t = \arg\min_\beta \|\beta' X_t - y^t\|^2$.

Consider the following three representations:

$$
\begin{aligned}
y_t &= b'_{t-1}\bar{x}_t + e_t && (1) \quad \text{"plug-in"} \\
y_t &= b'_n \bar{x}_t + \hat{\epsilon}_t(n) && (2) \quad \text{"least-squares"} \\
y_t &= b'_t \bar{x}_t + \hat{e}_t && (3) \quad \text{"sNML"}
\end{aligned}
$$

Representation (1) corresponds to the **predictive least squares** (PLS) model selection criterion: $\sum_{i=m+1}^{n} e_t^2$.

Representation (2) leads to the **AIC, BIC,** and **NML** criteria.

Linear-Quadratic Models

Linear model $y_t = \beta' \bar{x}_t + \epsilon_t$ with Gaussian errors $\epsilon_t \sim \mathcal{N}(0, \sigma^2)$.

Least squares parameters $b'_t = \arg\min_\beta \|\beta' X_t - y^t\|^2$.

Consider the following three representations:

$$
\begin{aligned}
y_t &= b'_{t-1}\bar{x}_t + e_t && (1) \quad \text{“plug-in”} \\
y_t &= b'_n \bar{x}_t + \hat{\epsilon}_t(n) && (2) \quad \text{“least-squares”} \\
y_t &= b'_t \bar{x}_t + \hat{e}_t && (3) \quad \text{“sNML”}
\end{aligned}
$$

Representation (1) corresponds to the **predictive least squares** (PLS) model selection criterion: $\sum_{i=m+1}^{n} e_t^2$.

Representation (2) leads to the **AIC, BIC,** and **NML** criteria.

Representation (3) is new. $\Rightarrow$ **sequentially normalized least squares** (SNLS)

Sequentially Normalized Least Squares

Fixed variance $\hat{\sigma}_t^2 = \sigma^2$ case:

Non-normalized conditional:

$$f(y_t \mid y^{t-1}, X_t; \sigma^2, b_t) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(y_t - \hat{y}_t)^2}{2\sigma^2}\right),$$

where $\hat{y}_t = b_t' \bar{x}_t$.

Fixed variance $\hat{\sigma}_t^2 = \sigma^2$ case:

Non-normalized conditional:

$$f(y_t \mid y^{t-1}, X_t; \sigma^2, b_t) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(y_t - \hat{y}_t)^2}{2\sigma^2}\right),$$

where $\hat{y}_t = b_t' \bar{x}_t$.

Normalized conditional:

$$f_{\mathrm{SNLS}}(y_t \mid y^{t-1}, X_t; \sigma^2) = \frac{1}{\sqrt{2\pi\tau}} \exp\left(-\frac{(y_t - b_{t-1}' \bar{x}_t)^2}{2\tau}\right),$$

where $\tau = (1 + c_t)^2 \sigma^2$, $c_t = \bar{x}_t' (X_t X_t')^{-1} \bar{x}_t = \mathcal{O}(1/t)$.

Fixed variance $\hat{\sigma}_t^2 = \sigma^2$ case:

Non-normalized conditional:

$$f(y_t \mid y^{t-1}, X_t; \sigma^2, b_t) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(y_t - \hat{y}_t)^2}{2\sigma^2}\right),$$

where $\hat{y}_t = b_t' \bar{x}_t$.

Normalized conditional:

$$f_{\mathrm{SNLS}}(y_t \mid y^{t-1}, X_t; \sigma^2) = \frac{1}{\sqrt{2\pi\tau}} \exp\left(-\frac{(y_t - b_{t-1}' \bar{x}_t)^2}{2\tau}\right),$$

where $\tau = (1 + c_t)^2 \sigma^2$, $c_t = \bar{x}_t' (X_t X_t')^{-1} \bar{x}_t = \mathcal{O}(1/t)$.

Fixed variance $\hat{\sigma}_t^2 = \sigma^2$ case:

Non-normalized conditional:

$$f(y_t \mid y^{t-1}, X_t; \sigma^2, b_t) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(y_t - \hat{y}_t)^2}{2\sigma^2}\right),$$

where $\hat{y}_t = b_t' \bar{x}_t$.

Normalized conditional:

$$f_{\mathrm{SNLS}}(y_t \mid y^{t-1}, X_t; \sigma^2) = \frac{1}{\sqrt{2\pi\tau}} \exp\left(-\frac{(y_t - b_{t-1}' \bar{x}_t)^2}{2\tau}\right),$$

where $\tau = (1 + c_t)^2 \sigma^2$, $c_t = \bar{x}_t' (X_t X_t')^{-1} \bar{x}_t = \mathcal{O}(1/t)$.

Free variance case:

Consider the maximization problem

$$\sup_{\sigma^2} \prod_{t=m+1}^{n} f(y_t \mid y^{t-1}, X_t; \sigma^2, b_t).$$

Sequentially Normalized Least Squares

Free variance case:

Consider the maximization problem

$$\sup_{\sigma^2} \prod_{t=m+1}^{n} f(y_t \mid y^{t-1}, X_t; \sigma^2, b_t).$$

The maximizing variance is given by $\hat{\tau}_n = \frac{1}{n-m} \sum_{t=m+1}^{n} (y_t - \hat{y}_t)^2$, and the resulting non-normalized joint density is

$$(2\pi e \hat{\tau}_n)^{-(n-m)/2}.$$

Sequentially Normalized Least Squares

The SNLS criterion is given by

$$\begin{aligned}
&\mathrm{SNLS}(n, k) \\
&= \frac{n-m}{2} \ln \hat{\tau}_n - \frac{1}{2} \ln \hat{e}_{m+1} - \ln \frac{\Gamma\left(\frac{n-m}{2}\right)}{\Gamma(1/2)} + \ln \prod_{t=m+2}^{n} \frac{\sqrt{\pi}}{1-d_t} \\
&= \frac{n-m}{2} \ln(2\pi e \hat{\tau}_n) + \sum_{t=m+1}^{n} \ln(1+c_t) + R_n,
\end{aligned}$$

where the remainder term R_n is insignificant.

Sequentially Normalized Least Squares

Theorem: If the data is generated by a k-parameter linear-quadratic model (either non-random X_n, or AR model), then we have

$$\mathrm{SNLS}(n,k) = \frac{n-m}{2}\ln(2\pi e\hat{\tau}_n) + \frac{2k+1}{2}\ln n + o(\ln n),$$

Sequentially Normalized Least Squares

Theorem: If the data is generated by a k-parameter linear-quadratic model (either non-random X_n, or AR model), then we have

$$\mathrm{SNLS}(n,k) = \frac{n-m}{2}\ln(2\pi e\hat{\tau}_n) + \frac{2k+1}{2}\ln n + o(\ln n),$$

and

$$\mathrm{SNLS}(n,k) = \frac{n-m}{2}\ln(2\pi e\hat{\sigma}_n^2) + \frac{k+1}{2}\ln n + o(\ln n)$$

almost surely for almost all β and σ^2.

Sequentially Normalized Least Squares

Theorem: If the data is generated by a k-parameter linear-quadratic model (either non-random X_n, or AR model), then we have

$$\mathrm{SNLS}(n,k) = \frac{n-m}{2}\ln(2\pi e\hat{\tau}_n) + \frac{2k+1}{2}\ln n + o(\ln n),$$

and

$$\mathrm{SNLS}(n,k) = \frac{n-m}{2}\ln(2\pi e\hat{\sigma}_n^2) + \frac{k+1}{2}\ln n + o(\ln n)$$

almost surely for almost all β and σ^2.

Note that the effective *number of parameters is doubled.*

Experiment: AR Model Order Estimation

		sample size, n						
		50	100	200	400	800	1600	3200
$k = 1$	AIC	70.5	71.3	72.0	70.0	71.4	70.8	70.9
	BIC	**93.5**	**96.9**	**97.9**	**98.0**	**99.4**	**99.5**	**99.4**
	PLS	75.8	86.3	91.1	93.5	96.7	97.8	98.1
	NML	82.5	88.3	89.7	91.5	94.3	95.9	96.6
	SNLS	78.5	87.5	92.2	93.9	97.0	98.1	98.3
$k = 4$	AIC	42.8	52.5	60.1	63.3	65.4	66.5	67.5
	BIC	**45.7**	59.6	67.8	76.5	**82.6**	88.3	91.4
	PLS	42.1	58.3	68.5	**77.0**	82.5	88.3	91.9
	NML	45.0	**60.2**	68.0	76.7	82.5	88.0	91.6
	SNLS	42.4	59.2	**69.4**	**77.0**	82.4	**88.5**	**92.0**
$k = 7$	AIC	**33.7**	45.4	55.3	59.6	63.6	65.7	67.3
	BIC	29.2	43.4	59.1	69.5	77.9	82.8	88.6
	PLS	30.0	44.7	60.5	70.0	78.5	82.9	88.6
	NML	28.8	44.2	59.8	69.8	78.3	83.0	88.4
	SNLS	30.1	**46.5**	**61.2**	**70.6**	**79.4**	**83.2**	**88.9**
$k = 10$	AIC	**28.5**	**43.9**	51.5	59.3	64.2	67.1	67.7
	BIC	20.6	35.7	51.0	66.1	74.4	81.4	85.5
	PLS	20.1	35.7	50.7	65.0	73.4	80.8	84.8
	NML	20.2	37.1	51.9	**66.8**	74.6	81.4	**85.8**
	SNLS	21.4	37.9	**52.3**	66.5	**74.8**	**81.8**	85.6