INTERNATIONAL JOURNAL OF ADAPTIVE CONTROL AND SIGNAL PROCESSING
*Int. J. Adapt. Control Signal Process.* 2004; **18**:1–18

# When Has Estimation Reached A Steady State? The Bayesian Sequential Test

Miroslav Kárný, Jan Kracík, Ivan Nagy and Petr Nedoma

*Adaptive Systems Department*
*Institute of Information Theory and Automation*
*Academy of Sciences of the Czech Republic*
*P. O. Box 18, 182 08 Prague, Czech Republic*

SUMMARY

This paper is concerned with distributions of time series, which (i) are influenced by initial conditions (ii) are stimulated by an exogenous signal or (iii) are obtained by recursive estimation of underlying parameters and thus undergo a transient period.

In computer intensive applications, it is desirable to stop the processing when the transient period is *practically* over. This aspect is addressed here from a Bayesian perspective. Under an often met assumption that the model of a system's time series is recursively estimated anyway, the computational overhead of the constructed stopping rule is negligible. Algorithmic details are presented for important normal ARX models (auto-regression with exogenous variable) and models of discrete-valued, independent, identically distributed data. The latter case provides non-parametric Bayesian estimation of credibility interval with sequential stopping. 

KEY WORDS: Bayesian estimation, sequential stopping, ARX model, non-parametric estimation

## 1. INTRODUCTION

The analysis of a system's time series [1] is the corner stone of a broad range of dynamic data processing techniques used in a variety of applications, e.g. [2, 3, 4, 5, 6, 7]. Often, a large amount of data is available or can be cheaply generated in simulations widely used in numerical procedures [8, 9]. Then both the related analysis and numerical evaluations are computationally intensive and it is important to decide when the processing should be stopped.

In this paper, the above query is answered for tasks in which the stopping makes sense when the inspected series overcomes the transient period. Typically, this happens when the transient period is caused by (i) non-trivial initial conditions, (ii) stimulation by an exogenous signal that itself becomes stationary or (iii) unfinished recursive parameter estimation.

The proposed solution uses Bayesian decision-making [10, 11] and elementary ideas of sequential stopping dating back to [12]. It extends and makes more systematic the solution given in [13]. It is restricted to cases in which the time series model is recursively estimated. The computational overhead related to the evaluation of the stopping rule is negligible. The class of such problems is narrower than that involved with caring about stationary behavior. It covers, however, practically significant examples such as mixture estimation on large data sets [14] or prior tuning of adaptive controllers [9, 15].

The paper is organized as follows. Basic facts on Bayesian estimation within an exponential family (EF) are recalled in Section 2. The stopping problem is formulated and solved in Section 3. The solution is specialized to normal ARX models in Section 4 and to models describing discrete-valued, mutually independent, identically distributed data in Section 5. ARX models are widely used for

*Correspondence to: Adaptive Systems Department
Institute of Information Theory and Automation
Academy of Sciences of the Czech Republic
P. O. Box 18, 182 08 Prague, Czech Republic

adaptive predictors and controllers. The discrete-data case is especially suitable for Bayesian non-parametric estimation [16]. The particular application to non-parametric estimation of credibility intervals with stopping is presented in Section 6. Then possible applications are demonstrated on simulation examples, Section 7, followed by concluding remarks in Section 8.

## 2. PRELIMINARIES

We use the following common notations. The symbol $f$ is reserved for a probability density function (pdf) distinguished by identifiers in its arguments. Integrals $\int$ are definite, are generally *multiple*, and are evaluated over the domain of the integrand. The symbol $d(t)$ denotes the sequence of data records $d_1, \ldots, d_t$. The non-numerical left superscript $^a$ of an object $B$ is also used. The expectation $\mathcal{E}[\bullet|*]$ is taken over all uncertain quantities in $\bullet$ except for those fixed by the condition $*$.

The collection of pdfs $f(d_t|d(t-1),\Theta)$, parameterized by an unknown finite-dimensional parameter $\Theta$ serves for a tailored data description. The data records $d_t \equiv [y_t, u_t]$ consist of the measured system output $y_t$ and the known system input $u_t$. However, the system input $u_t$ need not be present.

The parameter $\Theta$ is supposed to be unknown to the input generator (controller, human being), i.e., the generator meets natural conditions of control (NCC) [17]. The formal expression of this assumption reads

$$f(u_t|d(t-1),\Theta) = f(u_t|d(t-1)) \text{ or equivalently } f(\Theta|u_t, d(t-1)) = f(\Theta|d(t-1)). \tag{1}$$

Under NCC, the joint pdf of closed-loop data, up to the time horizon $T$, and of the unknown $\Theta$ then becomes

$$f(d(T),\Theta) = \underbrace{\prod_{t=1}^{T} \underbrace{f(y_t|u_t, d(t-1),\Theta)}_{\text{parameterized model}}}_{\mathcal{L}(d(T),\Theta)\equiv\text{likelihood function}} \times \underbrace{\prod_{t=1}^{T} \underbrace{f(u_t|d(t-1))}_{\text{input generator}}}_{\mathcal{R}(d(T))\equiv\text{control-strategy realization}} \times \underbrace{f(\Theta)}_{\text{prior pdf}}. \tag{2}$$

NCC and Bayes rule [17] imply that the posterior pdf has the form

$$f(\Theta|d(T)) = \frac{\mathcal{L}(d(T),\Theta) f(\Theta)}{\int \mathcal{L}(d(T),\Theta) f(\Theta)\, d\Theta} \equiv \frac{\mathcal{L}(d(T),\Theta) f(\Theta)}{\underbrace{\mathcal{I}(d(T))}_{\text{normalizing integral}}}. \tag{3}$$

Thus, the posterior pdf depends on the input generator through the measured data $d(T)$ only.

Parameterized models belonging to the exponential family (EF) [18] are considered. Their use is (almost) inevitable in the considered data intensive applications as their likelihood functions depend on finite-dimensional sufficient statistics. Models in EF have the form

$$f(y_t|u_t, d(t-1),\Theta) = A(\Theta) \exp\left\{\text{tr}\left[B'(\Psi_t)C(\Theta)\right]\right\}, \tag{4}$$

where $A(\Theta)$ is a non-negative function $\Theta$, and the symbol $'$ denotes transposition. $B(\Psi_t)$ is a matrix-valued function of the data vector

$$\Psi_t \equiv [y_t', \psi_t']' \equiv [y_t', u_t', d_{t-1}', \ldots, d_{t-\partial}', 1]' \equiv [d_t', \ldots, d_{t-\partial}', 1]'$$

of a finite order $\partial \geq 0$. $C(\Theta)$ is a vector or matrix-valued function of $\Theta$ with dimensions compatible with $B(\Psi)$. Recall, that tr denotes the trace of a matrix.

Note that in the static case obtained for $\partial = 0$, EF covers a range of distributions such as Poisson, multinomial, exponential, Wishart etc. In the dynamic case, the family reduces to a multivariate normal distribution with linearly entering regression coefficients and Markov chains of finite order. Luckily enough, general dynamic models can be "universally" [19] approximated by finite probabilistic mixtures [20] of pdfs (components) formed by normal and Markov factors [21]. For them, the derived stopping rule is applied component wise.

For a member of EF, the likelihood function has the form

$$\mathcal{L}(d(T),\Theta) \equiv \mathcal{L}(V_T, \nu_T, \Theta) \equiv A^{\nu_T}(\Theta) \exp\left\{\text{tr}\left[V_T' C(\Theta)\right]\right\}.$$

The sufficient statistic, describing the likelihood function without information loss, is formed by the

evaluated recursively, using $V_t = V_{t-1} + B(\Psi_t),\ \nu_t = \nu_{t-1} + 1, t = 1, \ldots, T$, starting with $V_0 = 0$ and $\nu_0 = 0$.

A model in EF possesses the conjugated prior pdf $f(\Theta) \equiv f(\Theta|d(0))$ whose form coincides with that of the likelihood function. Its considered choice makes the functional descriptions of the prior and posterior pdfs (related by (3)) identical, that is

$$\begin{aligned} f(\Theta|d(t)) &\equiv f(\Theta|V_t, \nu_t) \equiv \frac{A^{\nu_t}(\Theta) \exp\left\{\mathrm{tr}\left[V_t' C(\Theta)\right]\right\}}{\mathcal{I}(V_t, \nu_t)},\ t = 0, 1, \ldots, T, \\ \mathcal{I}(V_t, \nu_t) &= \int A^{\nu_t}(\Theta) \exp\left\{\mathrm{tr}\left[V_t' C(\Theta)\right]\right\}\, d\Theta, \\ V_t &= V_{t-1} + B(\Psi_t),\ \nu_t = \nu_{t-1} + 1, \quad V_0,\ \nu_0 \text{ a priori chosen.} \end{aligned} \tag{5}$$

The corresponding predictive pdf reads, cf. (2),

$$f(d(t)) = f(d(t)|V_0, \nu_0) = \frac{\mathcal{I}(V_t, \nu_t)}{\mathcal{I}(V_0, \nu_0)} \mathcal{R}(d(t)). \tag{6}$$

*The pdfs (5) and (6) describe completely the Bayesian estimation and prediction with (i) parameterized model in EF (ii) conjugate prior pdf and (iii) control strategy meeting NCC (1).*

The final preparatory step concerns the factorized version of Bayesian estimation. This version simplifies treatment of multi-output systems in EF. It also makes modelling more flexible and suitable for a joint description of outputs with discrete and continuous entries. It rests on the following parameterization implied by the chain rule

$$\begin{aligned} f(y_t|u_t, d(t-1), \Theta) &= \prod_{i=1}^{m} \underbrace{f(y_{i;t}|y_{i+1;t}, \ldots, y_{m;t}, u_t, d(t-1), \Theta)}_{i-th\ factor} \equiv \\ &\equiv \prod_{i=1}^{m} A(\Theta_i) \exp\left\{\mathrm{tr}\left[B'(\Psi_{i;t}) C(\Theta_i)\right]\right\}. \end{aligned} \tag{7}$$

The $i$-th data vector $\Psi_{i;t}$ is defined recursively

$$\begin{aligned} \Psi_{i;t} &= [y_{i;t}, \Psi_{i+1;t}']',\ i = 1, \ldots, m \equiv \text{the number of } y \text{ entries and} \\ \Psi_{m+1;t}' &\equiv \psi_t \equiv [u_t', d_{t-1}', \ldots, d_{t-\delta}', 1]'. \end{aligned}$$

The parameter $\Theta_i$ is a reduction of $\Theta$ to those entries that influence the $i$-th *factor* defined as the pdf predicting the $i$-th entry $y_{i;t}$ of the output $y_t$.

The product form (7) allows us to select a conjugate prior for each factor and update their posterior pdfs in parallel. The conjugate posterior pdf of the $i$-th factor is characterized by its individual sufficient statistics $V_{i;t}$, $\nu_{i;t}$.

## 3. SEQUENTIAL STOPPING

The design of a strategy that decides sequentially whether to stop the estimation or not is addressed here. The design is driven by the wish to obtain a computationally simple and still reasonably justified strategy.

First we show that the posterior pdfs on unknown parameters converge almost surely. Then, we demonstrate how this convergence influences the *Kullback-Leibler (KL)* divergence [22] of a pair of successive posterior pdfs. Recall that the KL divergence $\mathcal{D}(f_1||f_2)$ of a pair of pdfs $f_i(x)$ on a common domain $\{x\}$ is defined

$$\mathcal{D}(f_1||f_2) = \int f_1(x) \ln\left(\frac{f_1(x)}{f_2(x)}\right) dx. \tag{8}$$

It then holds that

$$\mathcal{D}(f_1||f_2) \geq 0\ \&\ \mathcal{D}(f_1||f_2) = 0 \Leftrightarrow f_1(x) = f_2(x) \text{ for almost all } x. \tag{9}$$

To avoid technicalities, we assume that only a finite number of $\Theta$ values is a priori probable.

**Proposition 1 (Asymptotic behavior of posterior pdfs)** *Let NCC (1) hold and the prior pdf is non-zero on a finite number of values of unknown parameters $\Theta_n$, $n = 1, \ldots, N < \infty$. Then, for any fixed $\Theta_n$, the sequence $\{f(\Theta_n|d(t))\}_{t\geq 1}$ converges almost surely and the "conditional" KL divergence*

$$Q(d(t)) \equiv \mathcal{D}(f(\Theta|d(t))||f(\Theta|d(t-1))|d(t)) \equiv \int f(\Theta|d(t)) \ln\left(\frac{f(\Theta|d(t))}{f(\Theta|d(t-1))}\right) d\Theta$$

*Proof:* First we show that $f(\Theta_n|d(t))$ is non-negative bounded martingale with respect to $d(t)$ for any fixed $\Theta_n$ with $f(\Theta_n) > 0$. Then, martingale theory [23] implies that almost surely $f(\Theta_n|d(t)) \rightarrow_{t\rightarrow\infty} f(\Theta_n|d(\infty)) \in [0,1]$.

Indeed, $f(\Theta_n|d(t)) \in [0,1]$ due to finiteness of $N$. Thus, it remains to show the basic martingale property, i.e. the equality $\mathcal{E}[f(\Theta_n|d(t))|d(t-1)] = f(\Theta_n|d(t-1))$. Its validity can be seen as follows

$$\begin{aligned}
&\mathcal{E}[f(\Theta_n|d(t))|d(t-1)] \equiv \int f(\Theta_n|d(t))f(y_t|u_t,d(t-1))f(u_t|d(t-1))\,dy_t du_t = \\
= &\int \frac{f(y_t|u_t,d(t-1),\Theta_n)f(\Theta_n|d(t-1))}{f(y_t|u_t,d(t-1))} f(y_t|u_t,d(t-1))f(u_t|d(t-1))\,dy_t du_t = \\
= &\int f(y_t|u_t,d(t-1),\Theta_n)f(\Theta_n|d(t-1))f(u_t|d(t-1))\,dy_t du_t = \\
= &\; f(\Theta_n|d(t-1)) \int f(u_t|d(t-1))\,du_t = f(\Theta_n|d(t-1)).
\end{aligned}$$

The Bayes rule implies that $f(\Theta_n|d(t-1)) = 0 \Rightarrow f(\Theta_n|d(t)) = 0$. Thus, the conditional KL divergence has non-zero finite contributions only from those $\Theta_n$ for which $f(\Theta_n|d(t-1)) > 0$. Due to this and the convergence of $f(\Theta_n|d(t-1))$, the relevant ratios $\frac{f(\Theta|d(t))}{f(\Theta|d(t-1))}$ forming the argument of logarithms converge to unity and thus the whole KL divergence converges to zero. □

Non-negativity and convergence to zero of the conditional KL divergence imply that it makes sense to stop estimation at time moment $t$, determined by data $d(t)$, when

$$Q(d(t)) \equiv \int f(\Theta|d(t)) \ln\left(\frac{f(\Theta|d(t))}{f(\Theta|d(t-1))}\right) d\Theta \leq \varepsilon, \tag{10}$$

where $\varepsilon > 0$ is a stopping threshold close to zero.

Such a stopping implies that $\ln\left(\frac{f(\Theta|d(t))}{f(\Theta|d(t-1))}\right) \approx \frac{f(\Theta|d(t))}{f(\Theta|d(t-1))} - 1$ is expected to be in the interval $[-\varepsilon,\varepsilon]$ for $\varepsilon \approx 0$. This interprets $\varepsilon$ as a "relative error" caused by substituting $f(\Theta|d(t-1))$ instead of $f(\Theta|d(t))$ and helps us to select specific values of $\varepsilon$. The typical choice $\varepsilon = 0.01$ corresponds to the expected relative error 1% .

Obviously, the inequality (10) is only the *necessary* condition for a proper stopping. In order to guarantee bounded expected relative errors for further moments, we should require

$$\begin{aligned}
&\sum_{\kappa=1}^{k} \mathcal{E}\left[\ln\left(\frac{f(\Theta|d(t+\kappa))}{f(\Theta|d(t+\kappa-1))}\right)|d(t)\right] \equiv \\
\equiv &\int f(\Theta, d_{t+1},\ldots,d_{t+k}|d(t)) \ln\left(\frac{f(\Theta|d(t+k))}{f(\Theta|d(t))}\right) d(d_{t+1},\ldots,d_{t+k})\,d\Theta \leq \varepsilon
\end{aligned}$$

for all $k \geq 0$. It needs, however, computationally intensive multi-step predictions, which pay off only when the acquiring of new data items is rather expensive. Here, we assume that computational cost is the main reason for stopping and thus we can stop estimation when the simplest necessary condition (10) is met.

## 4. APPLICATION TO ARX MODELS

Application of the stopping rule (10) to ARX models requires evaluation of the test statistic $Q(d(t))$ in (10). For it, basic information on the estimation of ARX models has to be recalled. Estimation of the factorized model (7) implies that it is sufficient to deal with a single factor. In the multi-output case, the posterior pdf $f(\Theta|d(t)) = \prod_{i=1}^{m} f(\Theta_i|d(t))$ and the overall test statistic is simply the sum of the test statistics $Q_i(d(t))$ connected with individual factors. This is obvious from the definition of $Q(d(t))$ as the conditional KL divergence of posterior pdfs in the product form.

The factor of the normal ARX model predicting a scalar entry $y_t$ has the form

$$\begin{aligned}
&f(y_t|\text{tail entries of current } y, u_t, d(t-1), \Theta) = \mathcal{N}_{y_t}(\theta'\psi_t, r) \equiv \\
\equiv &\; (2\pi r)^{-0.5} \exp\left[-0.5 r^{-1}(y_t - \theta'\psi_t)^2\right] = \underbrace{(2\pi r)^{-1/2}}\exp\left\{\mathrm{tr}\left[\underbrace{\Psi_t\Psi_t'}\left(-\frac{[-1\ \theta']'[-1\ \theta']}{2r}\right)\right]\right\}.
\end{aligned}$$

It is parameterized by $\Theta$ composed of regression coefficients $\theta$ and noise variance $r$. The indicated correspondence to EF (4) determines directly the form of the conjugate prior pdf

$$f(\Theta|V_0,\nu_0) = \frac{(2\pi r)^{-0.5(\nu_0+\gamma+2)} \exp\left\{-\mathrm{tr}\left(V_0 \frac{[-1\ \theta']'[-1\ \theta']}{2r}\right)\right\}}{\mathcal{I}(V_0,\nu_0)}, \tag{11}$$

where $\gamma$ denotes the number of regression coefficients. It is known as Gauss-inverse-Wishart ($GiW$) pdf. The extended information matrix $V_0$ employed to determine it has to be positive definite. Bayesian estimation in EF implies the updating formula $V_t = V_{t-1} + \Psi_t \Psi_t'$. Numerically, it is reasonable to update factors of its decomposition $V = L'DL$ [24]. In it, $L$ is a lower triangular matrix with unit diagonal and $D$ is a diagonal matrix with positive diagonal entries. Among others, this decomposition allows us to perform efficient evaluation of the normalization integral. By splitting

$$\begin{aligned}
L &\equiv \begin{bmatrix} 1 & 0 \\ {}^{y\psi}L & {}^{\psi}L \end{bmatrix},\ D \equiv \mathrm{diag}\left[{}^{y}D,\ {}^{\psi}D\right],\ {}^{y}D = \text{scalar, we get} \\
\mathcal{I}(V,\nu) &\equiv \mathcal{I}(L,D,\nu) = \Gamma(0.5\nu)\,{}^{y}D^{-0.5\nu}\left|{}^{\psi}D\right|^{-0.5}\pi^{-0.5\nu}(2\pi)^{0.5\gamma} \text{ with} \\
\Gamma(z) &\equiv \int_0^\infty x^{z-1}\exp(-x)\,dx,\ z>0.
\end{aligned}$$

Updating of the factorized extended information matrix is equivalent to recursive least squares [17] and it holds that

$$\begin{aligned}
\hat{\theta} &\equiv {}^{\psi}L^{-1}\,{}^{y\psi}L = \mathcal{E}[\theta|L,D,\nu] = \text{ least-squares estimate of } \theta \\
\mathrm{cov}(\theta|r) &= r\,{}^{\psi}L^{-1}\,{}^{\psi}D^{-1}\left({}^{\psi}L'\right)^{-1} \\
\hat{r} &\equiv \frac{{}^{y}D}{\nu-2} = \mathcal{E}[r|L,D,\nu] = \text{ point estimate of } r \\
\mathcal{E}[r^{-1}|L,D,\nu] &= \frac{\nu}{{}^{y}D} \\
\mathcal{E}[\ln(r)|L,D,\nu] &= \ln({}^{y}D) - \ln(2) - \frac{\partial \ln(\Gamma(0.5\nu))}{\partial(0.5\nu)}.
\end{aligned} \tag{12}$$

Use of $L'DL$ decomposition gives the $GiW$ pdf (11) the form

$$f(\Theta|L,D,\nu) \equiv GiW_{\theta,r}(L,D,\nu) \equiv \frac{(2\pi r)^{-0.5(\nu+\gamma+2)} \exp\left\{-\left(\frac{[-1\ \theta']L'D([-1\ \theta']L')'}{2r}\right)\right\}}{\mathcal{I}(L,D,\nu)}. \tag{13}$$

The predictive pdf has the general form (6). It defines the one-step-ahead output predictor as the Student pdf [17]

$$\begin{aligned}
& f(y_t|u_t,d(t-1)) \equiv \\
\equiv\ & S_{y_t}\left(\hat{y}_t,\ {}^{y}D_{t-1}(1+\zeta_t),\nu_t\right) \equiv \frac{\Gamma(0.5\nu_t)\left[{}^{y}D_{t-1}(1+\zeta_t)\right]^{-0.5}}{\sqrt{\pi}\Gamma(0.5(\nu_t-1))\left(1+\frac{\hat{e}_t^2}{{}^{y}D_{t-1}(1+\zeta_t)}\right)^{0.5\nu_t}} \\
\hat{e}_t &\equiv y_t - \hat{y}_t \equiv y_t - \hat{\theta}_{t-1}'\psi_t \equiv \text{ prediction error} \\
\zeta_t &\equiv \psi_t'\,{}^{\psi}L_{t-1}^{-1}\,{}^{\psi}D_{t-1}^{-1}\left({}^{\psi}L_{t-1}'\right)^{-1}\psi_t.
\end{aligned} \tag{14}$$

**Proposition 2 (Stopping for a normal ARX factor)** *Let us perform a single step of recursive least squares, i.e. updating the statistics $L_{t-1}, D_{t-1}, \nu_{t-1}$ determining the conjugate pdf $GiW_{\theta,r}(L_{t-1},D_{t-1},\nu_{t-1})$ by the data vector $\Psi_t$. Let us store ${}^{y}D_{t-1}$ and quantities $\zeta_t, \hat{e}_t$ obtained as a by-product. Let us evaluate the following*

$$\begin{aligned}
F(\nu_t) &\equiv 2\ln(\Gamma(0.5(\nu_t-1))) - 2\ln(\Gamma(0.5\nu_t)) + \frac{\partial \ln(\Gamma(0.5\nu_t))}{\partial(0.5\nu_t)} \\
G(\zeta_t) &\equiv \ln(1+\zeta_t) - \frac{\zeta_t}{1+\zeta_t} \\
\rho_t &\equiv \frac{\hat{e}_t^2}{{}^{y}D_{t-1}(1+\zeta_t)} \\
H(\nu_t,\rho_t,\zeta_t) &\equiv (\nu_t-1)\ln(1+\rho_t) - \nu_t\frac{\rho_t}{(1+\rho_t)(1+\zeta_t)}.
\end{aligned} \tag{15}$$

*Then, the estimation meets the stopping criterion (10) if*

*Proof:* It holds that $Q(d(t)) \equiv$

$$
\begin{aligned}
&\equiv \int f(\Theta|d(t)) \ln\left(\frac{f(\Theta|d(t))}{f(\Theta|d(t-1))}\right) d\Theta = \\
&= -\ln(f(y_t|u_t, d(t-1)) + \int f(\Theta|d(t)) \ln(f(y_t|u_t, d(t-1), \Theta))\, d\Theta \equiv \\
&\equiv -\ln\left[S_{y_t}\left(\hat{\theta}_{t-1}\psi_t,\ {}^{y}D_{t-1}(1+\zeta_t), \nu_t\right)\right] + \\
&+ \int GiW_{\theta,r}(V_t, \nu_t)\left\{-0.5\ln(2\pi r) - 0.5\frac{(y_t - \theta'\psi_t)^2}{r}\right\} dr\, d\theta = \\
&= 0.5\ln(\pi) + \ln(\Gamma(0.5(\nu_t - 1))) - \ln(\Gamma(0.5\nu_t)) + 0.5\ln\left[{}^{y}D_{t-1}(1+\zeta_t)\right] + \\
&+ 0.5\nu_t \ln\left(1 + \frac{\hat{e}_t^2}{{}^{y}D_{t-1}(1+\zeta_t)}\right) - 0.5\ln(2\pi) - 0.5\ln\left({}^{y}D_t\right) + 0.5\ln(2) + \\
&+ 0.5\frac{\partial \ln\left(\Gamma(0.5\nu_t)\right)}{\partial(0.5\nu_t)} - 0.5\nu_t \frac{\left(y_t - \hat{\theta}_t'\psi_t\right)^2}{{}^{y}D_t} - 0.5\psi_t\, {}^{\psi}L_t^{-1}\, {}^{\psi}D_t^{-1}\left({}^{\psi}L_t'\right)^{-1}\psi_t = \\
&= 0.5\Bigg\{\underbrace{2\ln(\Gamma(0.5(\nu_t - 1))) - 2\ln(\Gamma(0.5\nu_t)) + \frac{\partial \ln\left(\Gamma(0.5\nu)\right)}{\partial(0.5\nu)}}_{F(\nu_t)} + \underbrace{\ln(1+\zeta_t) - \frac{\zeta_t}{1+\zeta_t}}_{G(\zeta_t)} + \\
&+ \underbrace{(\nu_t - 1)\ln\left(1 + \frac{\hat{e}_t^2}{{}^{y}D_{t-1}(1+\zeta_t)}\right) - \nu_t \frac{\hat{e}_t^2}{{}^{y}D_{t-1}(1+\zeta_t)\left(1 + \frac{\hat{e}_t^2}{{}^{y}D_{t-1}(1+\zeta_t)}\right)(1+\zeta_t)}}_{H\left(\nu_t, \frac{\hat{e}_t^2}{{}^{y}D_{t-1}(1+\zeta_t)}, \zeta_t\right)}\Bigg\},
\end{aligned}
$$

where we use identities known in connection with recursive least squares

$$
\begin{aligned}
y_t - \hat{\theta}_t'\psi_t &= \frac{\hat{e}_t}{1+\zeta_t} \\
\psi_t'\, {}^{\psi}L_t^{-1}\, {}^{\psi}D_t^{-1}\left({}^{\psi}L_t'\right)^{-1}\psi_t &= \zeta_t/(1+\zeta_t), \quad {}^{y}D_t = {}^{y}D_{t-1} + \frac{\hat{e}_t^2}{1+\zeta_t}.
\end{aligned}
$$

□

The above algorithm uses quantities evaluated during estimation anyway. The function $F(\nu_t)$ can be cheaply numerically evaluated using standard approximations of $\ln(\Gamma(\nu))$ as well as its derivative [25]. It is also possible to construct its direct numerical approximation without approximating internal functions. Functions $G$ and $H$ are also simple and if need be they can be approximated by simpler functions, too. The fact that $\zeta_t$ and $\frac{\hat{e}_t^2}{1+\zeta_t}$ are obtained as by-product of the updating of the $L'DL$ decomposition of the matrix $V$ by the data vector $\Psi$ can be seen as follows

$$
1 + \zeta_t = \frac{\left|{}^{\psi}L_t'\, {}^{\psi}D_t\, {}^{\psi}L_t\right|}{\left|{}^{\psi}L_{t-1}'\, {}^{\psi}D_{t-1}\, {}^{\psi}L_{t-1}\right|} = \prod_{i=2}^{\gamma+1}\frac{D_{i;t}}{D_{i;t-1}}, \quad \frac{\hat{e}_t^2}{1+\zeta_t} = {}^{y}D_t - {}^{y}D_{t-1} = D_{1;t} - D_{1;t-1}. \tag{17}
$$

The discussed stopping rule is directly applicable whenever an ARX model is estimated. Extensions to a generalized regression model predicting a known non-linear transformation of the output [17] or to ARX models working on a filtered regression vector [26] make its potential use quite wide.

Sometimes, the time series is expected to converge and its ARX or other model is not estimated. Then, it is useful to estimate very simple version of ARX model just for stopping purposes. Whenever the Proposition 1 is applicable, the simple model converges almost surely and the stopping rule can be applied to it. For instance, by applying the stopping rule to the simplified models

$$
\mathcal{N}_{y_t}(\theta_1 y_{t-1} + \theta_2 u_t + \theta_3,\ {}^{s}r) \text{ or } \mathcal{N}_{y_t}(\theta_1 y_{t-1},\ {}^{s}r) \text{ etc.},
$$

## 5. APPLICATION TO DISCRETE-VALUED DATA

The parameterized model

$$\begin{aligned} f(y_t|u_t, d(t-1), \Theta) &\equiv f(y_t|\Theta) \equiv \Theta_{y_t} = \exp\left[\sum_{y=1}^{M} B_y(y_t)\ln(\Theta_y)\right] \\ y_t &= 1, \ldots, M \equiv \text{the number of } y \text{ values} \\ B_y(y_t) &= \left\{ \begin{array}{ll} 1 & \text{if } y = y_t \\ 0 & \text{otherwise} \end{array} \right. , \ \Theta = [\Theta_1, \ldots, \Theta_M] \in \left\{ \Theta_y \geq 0, \ \sum_{y=1}^{M} \Theta_y = 1 \right\} \end{aligned} \tag{18}$$

is assumed here. It is parameterized by unknown time-invariant probabilities $\Theta_{y_t}$ assigned to possible outputs $y_t$. This model has a widespread use in Bayesian non-parametric estimation [16]. In Section 6, its engineering version is applied to the Bayesian estimation of a credibility interval.

The formulas (18) indicate that the model belongs to EF (4) with $A(\Theta) = 1$ and $C(\Theta) = [\ln(\Theta_1), \ldots, \ln(\Theta_M)]'$. The conjugate prior pdf can then be determined in the self-reproducing Dirichlet form

$$f(\Theta|V) \equiv \frac{\prod_{y=1}^{M} \Theta_y^{V_y - 1}}{\mathcal{B}(V)}, \ \mathcal{B}(V) \equiv \frac{\prod_{y=1}^{M} \Gamma(V_y)}{\Gamma\left(\sum_{y=1}^{M} V_y\right)}, \ V_y > 0$$

and the estimation reduces simply to the counting formula $V_{y;t} = V_{y;t-1} + B_y(y_t)$. The identities $\Gamma(x+1) = x\Gamma(x)$, $f(y_t|u_t, d(t-1)) = \mathcal{B}(V_t)/\mathcal{B}(V_{t-1})$ imply the following form of the predictive probability

$$f(y_t|u_t, d(t-1)) \equiv f(y_t|y(t-1)) = \frac{V_{y_t;t-1}}{\nu_{t-1}} \equiv \hat{\Theta}_{y_t;t-1}, \ \nu_{t-1} \equiv \sum_{y=1}^{M} V_{y;t-1}.$$

Note that the recursion for the statistic $V_t$ can be converted into a recursion for the predictor

$$\begin{aligned} \hat{\Theta}_{y;t} &= \hat{\Theta}_{y;t-1} + \nu_t^{-1}\left(B_y(y_t) - \hat{\Theta}_{y;t-1}\right), \ \hat{\Theta}_{y;0} = \frac{V_{y;0}}{\nu_0}, \ y = 1, \ldots, M, \\ \nu_t &= \nu_{t-1} + 1, \ \nu_0 = \sum_{y=1}^{M} V_{y;0}. \end{aligned}$$

The final relationship needed for specializing the stopping test (10) reads

$$\mathcal{E}[\ln(\Theta_y)|V] = \frac{\partial}{\partial V_y}\ln(\Gamma(V_y)) - \frac{\partial}{\partial \nu}\ln(\Gamma(\nu)).$$

Validity of this formula can be simply seen by taking derivatives of the integral defining $\mathcal{B}(V)$ with respect to its "parameter" $V_y$.

Thus, the necessary condition for stopping at time $t$ reads

$$Q(V_t) \equiv -\ln\left(\frac{V_{y_t;t-1}}{\nu_{t-1}}\right) + \frac{\partial}{\partial V_{y_t;t}}\ln(\Gamma(V_{y_t;t})) - \frac{\partial}{\partial \nu_t}\ln(\Gamma(\nu_t)) < \varepsilon$$

with $\varepsilon > 0$ being the chosen stopping threshold. This gives an overall estimation algorithm combined with stopping.

**Algorithm 1 (Estimation with stopping for the Dirichlet model)**

Initial phase

- *Select prior statistics* $V_0 \equiv [V_{1;0}, \ldots, V_{M;0}]$, $V_{y;0} > 0$.
- *Evaluate* $\nu_0 = \sum_{y=1}^{M} V_{y;0}$.
- *Select a stopping threshold* $\varepsilon > 0$ *and set* $t = 0$.
- *Specify the largest number* $T$ *of measurements to be done.*

Sequential phase

1. *Set* $t = t + 1$ *and measure* $y_t$.
2. *Set* $Q = -\ln\left(\frac{V_{y_t;t-1}}{\nu_{t-1}}\right)$.
3. *Perform the estimation step* $V_{y_t;t} = V_{y_t;t-1} + 1$, $\nu_t = \nu_{t-1} + 1$.
4. *Complete evaluation of the test statistic* $Q = Q + \frac{\partial}{\partial V_{y_t;t}}\ln(\Gamma(V_{y_t;t})) - \frac{\partial}{\partial \nu_t}\ln(\Gamma(\nu_t))$.

## 6. ESTIMATION OF CREDIBILITY INTERVALS

We demonstrate the usefulness of Algorithm 1 for stopping in the non-parametric estimation of credibility intervals $[\underline{y}, \overline{y}]$ assigned to an unknown uni-modal pdf $f(y)$. The credibility bounds $\underline{y}, \overline{y}$ searched for fulfil

$$\int_{\underline{y}}^{\overline{y}} f(y)\, dy = \beta \in (0.5, 1) \;\; \text{given} \;, \;\; \overline{y} - \underline{y} \to \min .$$

The estimates are based on $t$ mutually independent, real-valued, samples $y_t \sim f(y_t)$.

A pair of bounds $-\infty < \underline{y} < \overline{y} < \infty$ splits the real line containing support of $f(y)$ into three intervals $I_1 \equiv [-\infty, \underline{y}]$, $I_2 \equiv (\underline{y}, \overline{y})$ and $I_3 \equiv [\overline{y}, \infty]$. The relevant, finitely parameterized, model is then

$$f(y_t \in I_n | \underline{y}, \overline{y}, y(t-1), \Theta) = \prod_{i=1}^{3} \Theta_i^{\delta_{in}}, \tag{19}$$

where $\delta_{in} = 1$ if $n = i$ and zero otherwise. The probabilities $\Theta_i \geq 0$, $\sum_{i=1}^{3} \Theta_i = 1$ are unknown (as the pdf $f(y)$ is unknown) and are constant for a fixed bounds $\underline{y}, \overline{y}$. Construction of the credibility interval coincides with the selection of these bounds so that the estimate of $\Theta_2$ is close to the given credibility level $\beta \in (0.5, 1)$ while the length of the interval $I_2$ is as small as possible. In other words, we require $\mathcal{E}[\Theta_2 | \underline{y}, \overline{y}, y(t)] \approx \beta$ with the smallest $\overline{y} - \underline{y}$. At the same time, we decide whether to continue collecting new data by applying the stopping rule (10) for discrete observations. It defines indirectly the precision with which the probability of the credibility interval can be tuned. For simplicity, we take the uniform prior pdf on $\Theta$, irrespective of the chosen $\underline{y}, \overline{y}$. Thus,

$$\mathcal{E}[\Theta_i | \underline{y}, \overline{y}, y(t)] \;\; = \;\; \frac{\text{the number of } y\text{s within } I_i \; +1}{t+3}. \tag{20}$$

For fixed observations $y(t)$, the credibility bounds are obtained by inspecting observations $y(t)$ ordered in an ascending manner. Loosely speaking, the candidates for credibility bounds are then the nearest observed values among which the portion of the observed data is the closest one to the pre-specified credibility level $\beta \in (0.5, 1)$.

With this choice, we have all the ingredients needed for applying the stopping algorithm (1). The overall algorithm of sequential non-parametric estimation of credibility interval is as follows.

**Algorithm 2 (Sequential estimation of the credibility interval)**
Initial phase

- *Select credibility level* $\beta \in (0.5, 1)$.
- *Specify the largest number* $T$ *of measurements to be taken.*
- *Specify the stopping threshold* $\varepsilon > 0$.
- *Define the smallest number of observations* $\underline{t}$ *to be processed without stopping. Formula (20) implies that if all* $\underline{t}$ *measurements fall into* $I_2$ *then* $\mathcal{E}[\Theta_2 | y(\underline{t})] \geq \beta$ *is equivalent to* $\underline{t} \geq (3\beta - 1)/(1 - \beta)$. *Note that the right hand side of this inequality is at least 1 for a meaningful* $\beta > 0.5$.
- *Measure data* $y(\underline{t})$ *and order them into* $\tilde{y}(\underline{t})$ *in an ascending way.*

Sequential processing *for* $t = \underline{t} + 1, \underline{t} + 2, \ldots, T$.

1. *Measure* $y_t$ *and order* $\tilde{y}(t-1)$ *with* $y_t$ *into ascending* $\tilde{y}(t)$.
2. *Set the number of data that will not belong into the credibility interval* ${}^{o}j = floor[t(1-\beta) + 1 - 3\beta]$, *where* $floor$ *gives the nearest integer smaller than its argument. This choice of* ${}^{o}j$ *guarantees (see below) that* $\hat{\Theta}_{2;t} \geq \beta$. *It is meaningful for* $t \geq \underline{t}$.
3. *Initialize the indices of data omitted left* ${}^{l}j = 1$ *and right* ${}^{r}j = t$ *from the constructed credibility interval.*
4. *Search for credibility bounds for*
   $j = 1, \ldots, {}^{o}j - 1$
   *if* $\tilde{y}\left({}^{l}j + 1\right) - \tilde{y}\left({}^{l}j\right) > \tilde{y}\left({}^{r}j\right) - \tilde{y}\left({}^{r}j - 1\right)$ *set* ${}^{l}j = {}^{l}j + 1$ *else set* ${}^{r}j = {}^{r}j - 1$
   *end* % *of the* $j$ *cycle*
   $\underline{y} = \tilde{y}\left({}^{l}j\right)$, $\overline{y} = \tilde{y}\left({}^{r}j\right)$ % *lower and upper credibility bounds*
5. *Specify estimation and test statistic*
   $Q(d(t)) = \ln(t+2)$ % *test statistic*
   $\nu_t = t + 3$ % *degrees of freedom*

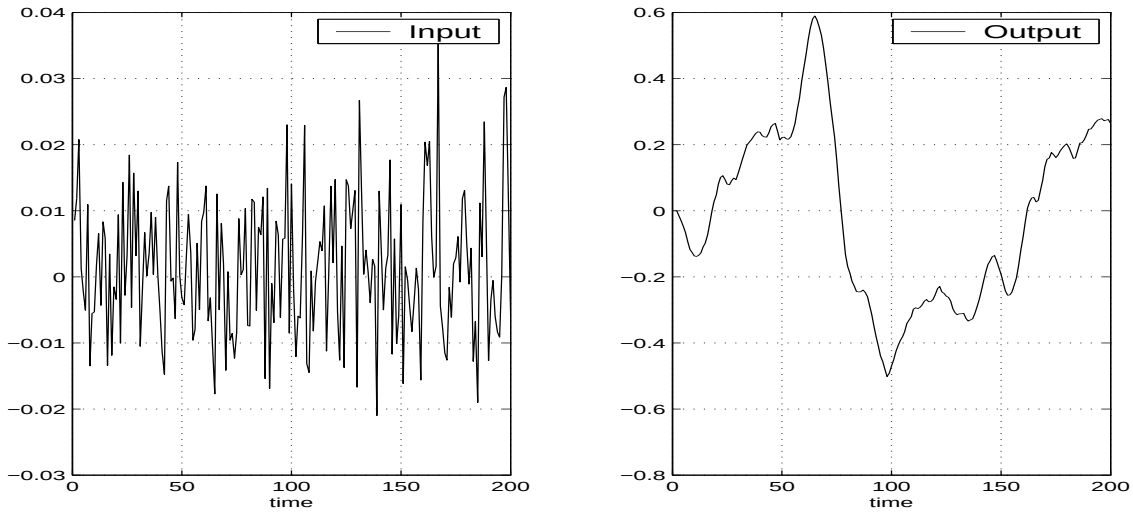



Figure 1. Simulated inputs and outputs

$V_{1;t} = {}^{l}j + 1$, $V_{3;t} = t - {}^{r}j + 2$, $V_{2;t} = \nu_t - V_{1;t} - V_{3;t}$
*if* $y_t < \underline{y}$
$\quad Q(d(t)) = Q(d(t)) - \ln(V_{1;t} - 1) + \frac{\partial}{\partial V_{1;t}} \ln(\Gamma(V_{1;t}))$
*else*
$\quad$ *if* $y_t \leq \overline{y}$
$\quad\quad Q(d(t)) = Q(d(t)) - \ln(V_{2;t} - 1) + \frac{\partial}{\partial V_{2;t}} \ln(\Gamma(V_{2;t}))$
$\quad$ *else*
$\quad\quad Q(d(t)) = Q(d(t)) - \ln(V_{3;t} - 1) + \frac{\partial}{\partial V_{3;t}} \ln(\Gamma(V_{3;t}))$
*end* *% end of the test where $y_t$ is*
$\hat{\Theta}_{3;t} = \frac{V_{3;t}}{\nu_t}$, $\hat{\Theta}_{1;t} = \frac{V_{1;t}}{\nu_t}$, $\hat{\Theta}_{2;t} = 1 - \hat{\Theta}_{3;t} - \hat{\Theta}_{1;t}$ *% estimates, unnecessary for stopping*
*if* $\varepsilon > Q(d(t))$
$\quad$ *stop and take the found credibility bounds as the final ones*
*else*
$\quad$ *continue in* Sequential processing

## 7. ILLUSTRATIVE EXAMPLES

The simulated examples presented here should help to give an insight into the properties of the proposed stopping rule.

### *7.1. Recursive estimation of the dynamic ARX factor*

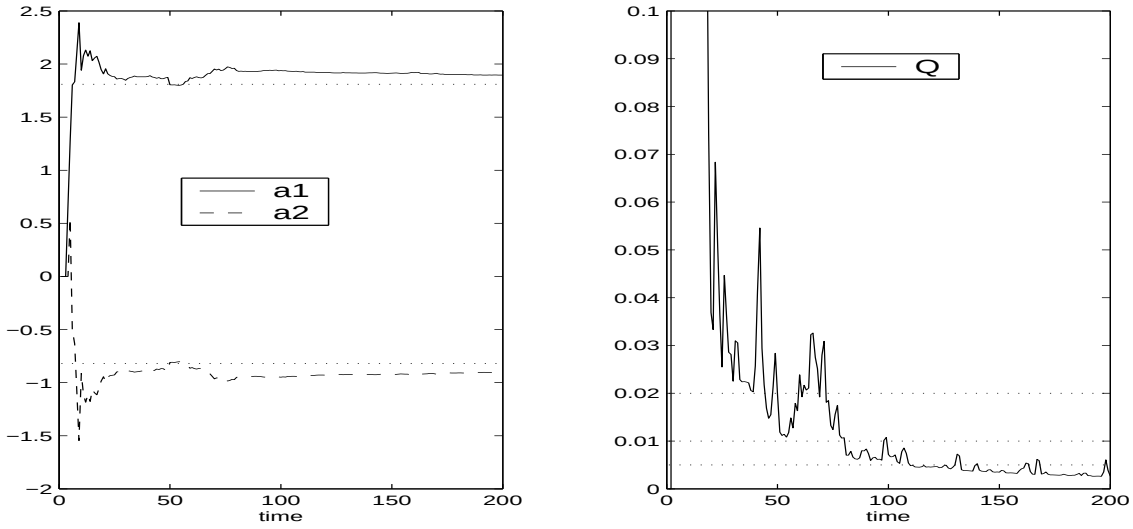
Recursive estimation stopping of a single-input single-output ARX model is presented in this subsection.

The simulated system is the discrete transformation of the continuous system with transfer function $F(s) = 1/(1+s)^2$. Its discrete version, obtained with a sampling period of 0.1 seconds, gives

$$y_t = 1.81y_{t-1} - 0.8187y_{t-2} + 0.00468u_t + 0.00438u_{t-1} + e_t,$$

in which $e_t$ is white zero-mean normal noise with variance $r = 0.0001$. Meanwhile the input $u_t$ is an independent white zero-mean normal noise with the same variance 0.0001.

Two hundred of input-output data pairs employed for estimation are displayed in Fig. 1 (note differences in scales). Bayesian recursive estimation of the ARX model of correct structure is illustrated by trajectories of the point estimates of the auto-regression coefficients a1=1.81, a2=-0.8187 in the left subplot of Fig. 2. True values of the auto-regression coefficients are marked by a dotted line. The right subplot of Fig. 2 shows the trajectory of the test statistic $Q(d(t))$ (10). Possible thresholds 0.5%, 1% and 2 % are marked by dotted lines. The test statistic $Q(d(t))$ reaches the threshold 2% after




run example view code

Figure 2. Left: trajectories of point estimates of auto-regression coefficients. Right: the trajectory of the test statistic $Q(d(t))$.

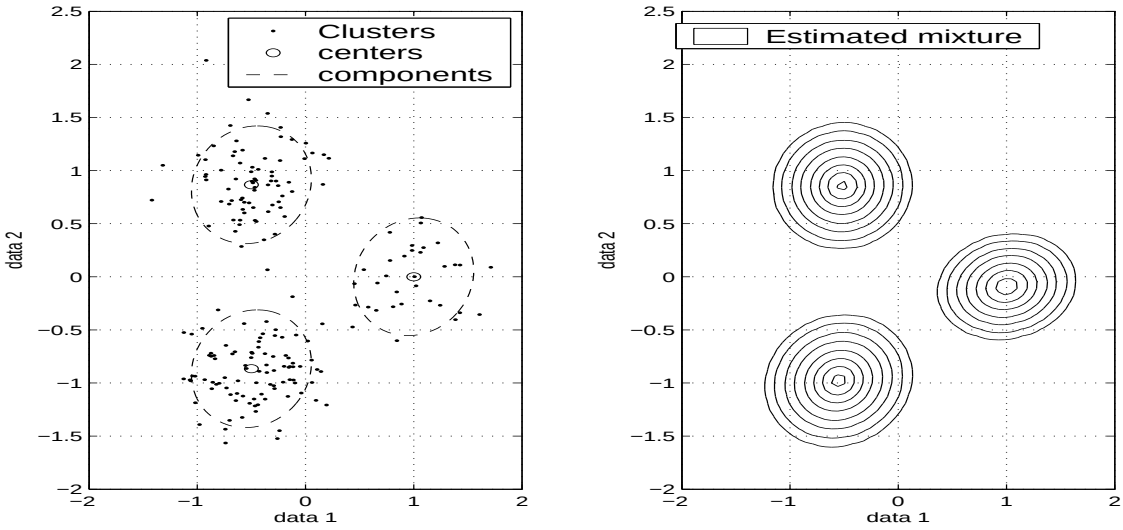



Figure 3. Data clusters and estimated mixture

data samples. The increase of $Q(d(t))$ above thresholds, observed without stopping, reflects both the suboptimal and stochastic nature of the stopping rule. Practically, the threshold values 0.05%, 1% are plausible as confirmed by a range of other experiments.

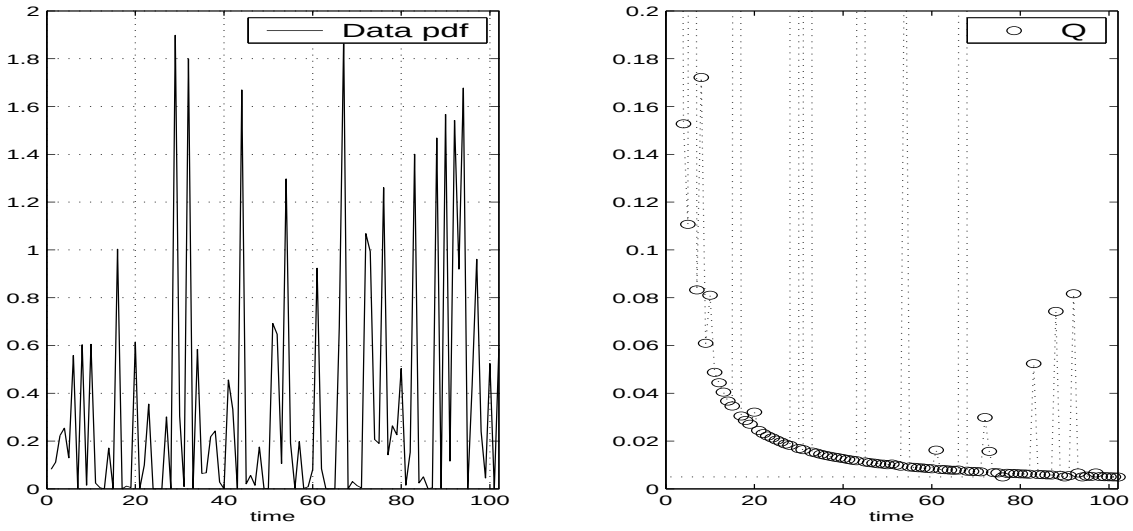
### *7.2. Estimation of a probabilistic mixture*

The example presented here illustrates how the stopping rule applied to a simplified model may be employed in complex estimation tasks when a large amount of data for processing is available.

Mutually independent two-dimensional data items were generated by a probabilistic mixture with the normal components (pdfs) $\mathcal{N}_{d_t}([1,0],R)$, $\mathcal{N}_{d_t}([-0.5,0.866],R)$, $\mathcal{N}_{d_t}([-0.5,-0.866],R)$, $R = \begin{bmatrix} 0.1 & 0.01 \\ 0.01 & 0.1 \end{bmatrix}$, mixed with probabilities $1/6, 1/3, 1/2$.

The data clusters and contours at simulated 95% equiprobability levels are displayed in the left subplot of Fig. 3.

The mixture is estimated using a novel projection algorithm [27, 28]. The resulting contours of the




run example view code

Figure 4. Trajectories of posterior data likelihood and test statistic

Recursive projection estimation [27] is the basic building block in complete mixture estimation that includes initialization of the mixture estimation [29] and estimation of the mixture structure. The repetitive estimation is then computationally intensive especially in higher-dimensional cases. Then, the use of sequential stopping is vital.

It can be shown [29] that the pdf of processed data $f(d(t)|v)$ is an adequate measure of the inspected model variant $v$. Thus, it makes sense to stop when its values reach steady state. This makes us to estimate together with mixture a static ARX log-normal model $\mathcal{N}_{\ln(f(d(t)|v))}(\theta, r)$ and to stop when this auxiliary simple estimation stabilizes. A trajectory of $f(d_t|d(t-1), v)$ is plotted in the left part of Fig. 4 and the corresponding test statistic $Q(d(t))$ is displayed in the right part of this figure. In this case, a sample size slightly above 100 is sufficient to reach stationarity. It corresponds to a threshold value of 0.5%. The sample size to reach stationarity is surprisingly low, which will bring substantial computational speed-up of the overall mixture estimation and extend applicability to a significantly larger set of practical problems.

### *7.3. Estimation of credibility intervals*

The result of the previous paragraph indicates the possibility of substantial mixture estimation speeding up. To be sure that this promising result is not random, we evaluate here a credibility interval for the moments of stopping. For it, the estimation previously done is repeated on independent realizations. With each run, the time needed to reach a stationary state is recorded and Algorithm 2 is used for checking the need for a further realization. The repetition was stopped when the test statistics reached the threshold level 0.5%.

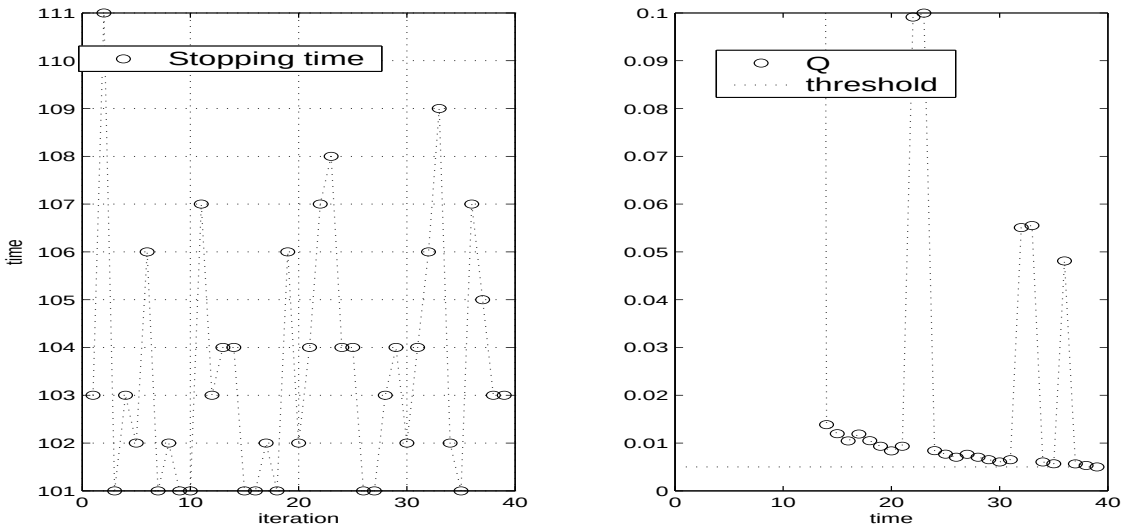
The stopping moments of individual runs are shown in Fig. 5, the left subplot, the value of the test statistics is in the right subplot.

For the chosen credibility level of $\beta = 0.7$, the credibility interval found is $< 101, 105 >$ steps.

In order to check this result further, we made 100 repetitions of the above experiment giving 100 independent realizations of the number of iterations recommended by the analyzed stopping rule. All runs stopped after 39 or 40 realizations. Both these results indicate the reliability of the sequential estimates obtained.

## 8. CONCLUDING REMARK

Sequential testing is an old but underestimated direction that can substantially shift the boundary of practically solvable estimation and simulation problems. The test elaborated in this paper serves as a practically useful confirmation of this claim. A wide range of applications are especially seen in




run example view code

Figure 5. Independent realizations of stopping moments and the trajectory of the test statistic

offered here is expected to increase their efficiency substantially.

## REFERENCES


1. T.W. Anderson, *An Introduction to Multivariate Statistical Analysis*, John Wiley, 1958.
2. L. Ljung, *System Identification: Theory for the User*, Prentice-Hall, London, 1987.
3. T. Bohlin, *Interactive System Identification: Prospects and Pitfalls*, Springer-Verlag, Berlin, Heidelberg, New York, 1991.
4. M. Basseville and I.V. Nikiforov, *Detection of abrupt changes: theory & applications*, Prentice Hall, Englewood Cliffs, New Jersey, 1993, ISBN 0 13 126780 9.
5. E. L. Sutanto, "Use of cluster analysis for the study of machine processes", in *Colloquium on Intelligent Manufacturing Systems*, London, December 1995, number 95/238, pp. 9/1–9/5.
6. R. Murray-Smith and T.A. Johansen, *Multiple Model Approaches to Modelling and Control*, Taylor & Francis, London, 1997.
7. M. Kano, S. Hasebe, I Hashimoto, and H. Ohno, "A new multivariate process monitoring method using principal component analysis", *Computers and Chemical Engineering*, vol. 25, no. 7-8, pp. 1103–1113, August 2001.
8. W.R. Gilks, S. Richardson, and D.J. Spiegelhalter, *Markov Chain Monte Carlo in practice*, Chapman & Hall, London, 1997, ISBN 0 412 05551 1.
9. J. Bůcha, M. Kárný, P. Nedoma, J. Böhm, and J. Rojíček, "Designer 2000 project", in *International Conference on Control '98*, London, September 1998, pp. 1450–1455, IEE.
10. M.H. DeGroot, *Optimal Statistical Decisions*, McGraw-Hill Company, New York, 1970.
11. J.M. Bernardo and A.F.M. Smith, *Bayesian theory*, John Wiley & Sons, Chichester, New York, Brisbane, Toronto, Singapore, 1997, 2nd edition.
12. A. Wald, *Statistical Decision Functions*, John Wiley & Sons, New York, London, 1950.
13. J. Rojíček and M. Kárný, "A sequential stopping rule for extensive simulations", in *Preprints of the 3rd European IEEE Workshop on Computer-Intensive Methods in Control and Data Processing*, J. Rojíček, M. Valečková, M. Kárný, and K. Warwick, Eds., Praha, September 1998, pp. 145–150, ÚTIA AV ČR.
14. A. Quinn, P. Ettler, L. Jirsa, I. Nagy, and P. Nedoma, "Probabilistic advisory systems for data-intensive applications", *International Journal of Adaptive Control and Signal Processing*, vol. 17, no. 2, pp. 133–148, 2003.
15. M. Kárný and A. Halousková, "Pre-tuning of self-tuners", in *Advances in Model-Based Predictive Control*, D. Clarke, Ed., pp. 333–343. Oxford University Press, Oxford, 1994.
16. T.S. Fergusson, "A Bayesian analysis of some nonparametric problems", *The Annals of Statistics*, vol. 1, pp. 209–230, 1973.
17. V. Peterka, "Bayesian system identification", in *Trends and Progress in System Identification*, P. Eykhoff, Ed., pp. 239–304. Pergamon Press, Oxford, 1981.
18. O. Barndorff-Nielsen, *Information and exponential families in statistical theory*, Wiley, New York, 1978.
19. S. Haykin, *Neural networks: A comprehensive foundation*, Macmillan College Publishing Company, New York, 1994.
20. D.M. Titterington, A.F.M. Smith, and U.E. Makov, *Statistical Analysis of Finite Mixtures*, John Wiley & Sons, Chichester, New York, Brisbane, Toronto, Singapore, 1985, ISBN 0 471 90763 4.
21. M. Kárný, I. Nagy, and J. Novovičová, "Mixed-data multi-modelling for fault detection and isolation", *Adaptive control and signal processing*, , no. 1, pp. 61–83, 2002.
22. S. Kullback and R. Leibler, "On information and sufficiency", *Annals of Mathematical Statistics*, vol. 22, pp. 79–87, 1951.

24. G.J. Bierman, *Factorization Methods for Discrete Sequential Estimation*, Academic Press, New York, 1977.
25. M. Abramowitz and I.A. Stegun, *Handbook of mathematical functions*, Dover Publications, Inc., New York, 1972.
26. V. Peterka, "Real-time parameter estimation and output prediction for ARMA-type system models", *Kybernetika*, vol. 17, pp. 526–533, 1981.
27. J. Andrýsek, "Initiation Options in Estimation of Dynamic Mixtures", Tech. Rep. 2030, ÚTIA AV ČR, Praha, 2001.
28. J. Andrýsek, "Approximate recursive Bayesian estimation of dynamic probabilistic mixtures", in *Multiple Participant Decision Making*, J. Andrýsek, M. Kárný, and J. Kracík, Eds., pp. 39–54. Advanced Knowledge International, Magill, Adelaide, 2004.
29. M. Kárný, P. Nedoma, I. Nagy, and M. Valečková, "Initial description of multi-modal dynamic models", in *Artificial Neural Nets and Genetic Algorithms. Proceedings*, V. Kůrková, R. Neruda, M. Kárný, and N. C. Steele, Eds., Wien, April 2001, pp. 398–401, Springer.
30. P. Nedoma, J. Böhm, T. V. Guy, L. Jirsa, M. Kárný, I. Nagy, L. Tesař, and J. Andrýsek, "Mixtools: User's Guide", Tech. Rep. 2060, ÚTIA AV ČR, Praha, 2002.