Different approaches to statistical estimation of the response function

Vladimir Zaiats

Universitat de Vic, SPAIN

E-mail: vladimir.zaiats@uvic.es

My thanks are due to Prof. E. Masry (UCSD) for stimulating discussions

Talk at SAPS7, Le Mans, March 19, 2009

Outline of the talk

1 Linear systems with constant parameters

2 Basic dynamical characteristics

3 Estimation of the response function/transfer function

4 Estimation of the response function/transfer function in continuous time after discrete observations

1. Linear systems with constant parameters

Ideal system — a system having constant parameters and such that the two basic characteristics, the **input** and the **output** are related (black box)

A system has **constant parameters** if all basic properties of the system do not vary in time (time invariant)

A system is **linear** if its reaction $f(x)$ to input x is additive and homogeneous

Additive: $f(x_1 + x_2) = f(x_1) + f(x_2)$

Homogeneous: $f(c \cdot x) = c \cdot f(x)$, for any $c = \text{const}$

2. Basic dynamical characteristics

$h(s)$, $s \in \mathbb{R}$ — the **impulse response function** or the **transfer function in the time domain**

Meaning: reaction of the system to a unit impulse (a pulse of unit area and infinitesimal width) enjoyed by the system s units of time before the instant where the output is measured

In continuous time, this implies

$$y(t) = \int_{-\infty}^{\infty} h(s)x(t-s)ds = \int_{-\infty}^{\infty} x(s)h(t-s)ds, \qquad t \in \mathbb{R} \qquad (1)$$

Causal systems — those whose response at any instant does not depend on the future of the input, for any input

This means that (a necessary and sufficient condition for a system to be causal)

$$h(s) = 0, \quad s < 0, \tag{2}$$

and that the lower limit of integration in the first integral in (1) equals 0 instead of $-\infty$ for causal systems

Stable systems — those in which bounded inputs produce bounded outputs

If

$$\int_{-\infty}^{\infty} |h(s)| ds < \infty, \tag{3}$$

then

$$|y(t)| \leq \sup_{t \in \mathbb{R}} |x(t)| \cdot \int_{-\infty}^{\infty} |h(s)| ds < \infty.$$

It is well-known that condition (3) is necessary and sufficient for system (1) to be stable

3. Estimation of the response function/transfer function

A general linear model of a time-invariant **causal** system with additive stochastic disturbances can be written as follows:

$$y(t) = \sum_{k=1}^{\infty} h_k x(t-k) + \xi(t), \qquad t = 1, 2, \dots \tag{4}$$

in a discrete-time setting, and

$$y(t) = \int_0^{\infty} h(u) x(t-u) du + \xi(t), \qquad t \geq 0, \tag{5}$$

in a continuous-time setting. Here, x and y are the input and the output, respectively, at time t, and ξ is supposed to be a stationary zero-mean process. Models (4) and (5) are also called SISO, meaning single-input single-output.

In model (4), the complex-valued function

$$H(e^{i\omega}) = \sum_{k=1}^{\infty} h_k e^{-ik\omega}, \qquad -\pi \le \omega \le \pi, \tag{6}$$

is the so-called transfer function in the frequency domain. A continuous-time counterpart of (6) is

$$H(\lambda) = \int_0^{\infty} h(u) e^{-i\lambda u}, \qquad \lambda \in \mathbb{R}. \tag{7}$$

These transfer functions play a very important role for understanding the properties of systems (4) and (5)

In non-causal systems, the lower limits on the right-hand side of (6) and (7) should be changed to $-\infty$

The references dealing with estimation of the response/transfer function go back to the cornerstone statistical monographs:

Brillinger *Time Series: Data Analysis and Theory*

Jenkins & Watts *Spectral Analysis and Its Applications*

We will focus more on the results of J. & W. since they treated both discrete- and continuous-time models, while B. dealt with multi-dimensional discrete-time models

Jenkins & Watts (Chapter 10)

Consider both systems (4) and (5), always subtracting the mean from the input and the output

J. & W. claim that continuous-time model (5) is adequate for description of real-life systems if:

(i) x does not vary much,

(ii) there is no feedback in the system,

(iii) no external variables affect the output

In discrete-time model (4), the following useful considerations should be taken into account

Estimation of the h_k's from model (4) can be problematic due to two circumstances:

1. The number of weights h_k to be estimated can be rather large. An alternative is to search for a parametrization as follows:

$$y(t)-\alpha_1 y(t-1)-\ldots-\alpha_m y(t-m)=\beta_0 x(t)-\beta_1 x(t-1)-\ldots-\beta_l x(t-l)+\xi(t)$$

2. The least squares estimates of h_k obtained from (4) could be strongly correlated. This occurs similar to what happens to estimates of the auto- and cross-correlation functions.

The second difficulty can be overcome by switching to the frequency domain

Transfer function (7) could be estimated by means of the ratio

$$\widehat{H}(\lambda) = \frac{\widehat{f}_{xy}(\lambda)}{\widehat{f}_{xx}(\lambda)}, \qquad \lambda \in \mathbb{R} \tag{8}$$

Here, $\widehat{f}_{xy}$ is a smoothed estimator of the input-output cross-spectrum, $\widehat{f}_{xx}$ is a smoothed estimate of the input spectrum

The values of estimate (8) on neighbour frequencies are uncorrelated. Therefore the estimate of the impulse function h obtained by taking the inverse Fourier transform of (8) will be more smooth than that obtained by a direct estimation of h

Since window shrinking in spectral analysis self-adapts to local properties of the frequency characteristics, one may expect that estimation in the frequency domain will require less parameters than that in the time domain

However, spectral analysis still deals with estimation of more parameters than that in an appropriately fitted parametric model. Therefore parametric estimation seems to be a final objective in this type of estimation

The main role of spectral methods in this type of analysis of systems is that it is **useful for guessing possible models**

Anyway, spectral analysis has some advantages as compared to the parametric approach:

1. It can easily be extended to multi-dimensional systems

2. In practical applications, people are often interested in describing the underlying functions within a quite limited frequency bandwidth, while parametric models enable describing it in a much wider bandwidth

3. Since spectral analysis is flexible, time series could be filtered out into components that would correspond to different frequency bandwidths. Further analysis is carried out separately in each of these bands. This is useful in some applications when there is no reason to assume that the same parametric model is valid for all frequency bandwidths

Further progress

Further efforts in system modeling for the above described problems were concentrated on obtaining models involving a finite number of unknown parameters, that would provide a “reasonable” approximation to the observed data, rather than to estimate parameters of the “true” system

Ljung & Yuan (1985) and Ljung & Wahlberg (1992) studied asymptotic properties of the least square estimator for finite impulse response approximation. They showed that the dimension d of the model should increase with the number n of observations in order to ensure convergence

In the closely related problem of autoregressive approximation, similar results were obtained using various order selection procedures, Shibata (1980), An *et al.* (1982), Bhansali (1986), Hannan & Kavalieris (1986), Gerencsér (1992)

These methods select, among a collection of parametric models, the model minimizing a certain information criterion. A good survey of model selection methods can be found in

Hannan & Deistler *The Statistical Theory of Linear Systems*

In the context of nonparametric regression estimation, the problem of model selection we considered, for example, by Efroimovich & Pinsker (1984), Polyak & Tsybakov (1990) for estimators based on orthogonal series, and in Barron *et al* (1995) for minimum contrast estimators on sieves

Goldenshluger (1998) focused on a nonparametric minimax approach to estimation of the transfer function in discrete time. His assumption is that the transfer function we would like to estimate, belongs to a certain family, and the quality of an estimator is measured by the worst case error over the family.

With this approach, estimation accuracy and the order of the optimal model are basically determined by the rate at which the "true" impulse response sequence tends to zero.

Juditsky *et al* (1995) and Sjöberg *et al* (1995) proposed a unified approach to nonlinear black-box modeling in system identification

They pointed out that nonlinear structures could be seen as a concatenation of a mapping from observed data to a regression vector and a nonlinear mapping from the regressor space to the output space. The latter mapping was usually formed as a basis function expansion

A particular problem was to deal with a large number of potentially necessary parameters. This problem was handled by means of making the number of 'used' parameters considerable smaller than that of 'offered' parameters, by regularization, shrinking, pruning, or regressor selection

4. Estimation of the response function/transfer function in continuous time after discrete observations

Chapter 10.3 in Jenkins & Watts deals with application of spectral methods for the estimation of frequency characteristics in continuous-time model (5)

First, the system is **no longer** assumed to be causal. They apply the least squares method, both in the time domain and in the frequency domain, in order to obtain estimate (8)

Their results are sometimes a bit empiric, but give a good insight into the problem

Our idea is to consider a mixed framework: the input and the output are assumed to be continuous-time stochastic processes, but both of them are observed at regularly spaced discrete locations stepped by ς. This idea was applied by Buldygin, Utzet and Z. (2004) for cross-correlogram estimators of the response function in Gaussian models

One of particular results in this paper was that ς should go to zero not slower than a certain rate in order to make the estimates consistent

On the other hand, the sampling scheme **need not be regular**. Moreover, irregular sampling has been shown to be useful in estimation of the correlation function and the spectral density of a stationary stochastic process

The idea behind this approach is that, for example, the correlation function of a process observed at regularly spaced discrete locations stepped by ς can consistently be estimated only at lags that are multiples of ς

See papers by Masry and co-authors (1978–)

Therefore irregular sampling seems for us to be an irreplaceable tool in functional estimation when we would like to estimate a function over the **whole domain of definition** of this function

On the other hand, irregular sampling appears to be crucial in applications. E.g., in seismic data analysis where the data may be incomplete due to the lack of observations or because of a more economic data acquisition, see Herrmann & Hennenfert (2008)

This kind of application leads us in a natural way to the extension of models (4) and (5) to spatial settings

Results

Let x be a real, stationary, measurable, 4th-order process with mean zero, continuous correlation function C and fourth order cumulant function $Q(u_1, u_2, u_3)$

The sampling instants form a stationary, orderly, 4th-order point process $\{\tau_n\}$ on $\mathbb{R}$ independent of x

Denoting by $N(\cdot)$ the counting process of $\{\tau_n\}$, we assume that:

(i) the joint distribution of $\{N(B_1 + h), \ldots, N(B_n + h)\}$ is independent of $h \in \mathbb{R}$ for any collection $\{B_1, \ldots, B_n\}$ of Borel sets in $\mathbb{R}$

(ii) $\mathsf{P}\{N((0,h]) \geq 2\} = o(h),\ h \downarrow 0$

(iii) $\mathsf{E}[N^4(B)] < \infty$ for any bounded Borel set B

The first result is that the mean of an appropriately chosen estimate (8) does nod depend on the sampling rate β

Second, if the input processes satisfies

$$\int_{-\infty}^{+\infty} (1+u)|C(u)|du < \infty,$$

the input spectral density f_{xx} exists having a bounded second derivative, if

$$\int_{-\infty}^{+\infty} |Q(u_1, u_2, u_3)|du_1 du_2 du_3 < \infty,$$

then, under some extra technical conditions, the estimates of the transfer function h obtained by the inverse Fourier transform of (8) are asymptotically normal with known mean and variance

Final remarks

1. We would like to make a look on whether or not the assumptions providing asymptotic normality are the best

In BUZ (2004), we have dealt with Gaussian inputs in a model similar to (5) but having no noise. (Therefore the outputs have also been Gaussian.) In that setting, the basic assumption on the unknown transfer function h is as follows:

$$\int_{-\infty}^{+\infty} h^2(\lambda) d\lambda < \infty \tag{9}$$

This is a minimum a priori assumption on h, otherwise there will be no convergence of finite-dimensional distributions of our estimate at all!

On the other hand, (9) admits non-stable systems, recall (3)

The estimate in BUZ (2004) has been constructed as a sample input-output cross-correlogram. Maybe the possibility of obtaining asymptotic normality under very mild assumptions (these assumptions seem to have a definite form) is due to the nature of this estimate?

2. The minimax approach should be tested in our framework

Goldenshluger's result (1998) for model (4) uses quite restrictive, as compared to assumptions (3) and (9), classes of functions:

$$\mathcal{G}_q(l, L) = \left\{ h : \sum_{k=1}^{\infty} |h_k|^q k^{lq} \leq L^q \right\}, \quad q \in [1, \infty]$$

$$\mathcal{H}(\rho, L) = \left\{ h : |h_k| \leq L\rho^{-k}, \ \rho > 1 \right\}$$

The estimate is based on a regularized input-output cross-correlogram, i.e. it is not a ratio of spectral estimates

3. Applications to real data — or at least simulations — should be carried out

4. It is interesting to have a look at extensions of the model to spatial data, e.g. in seismology, in order to understand the importance of irregular sampling in this setting

5. A guess on how the nonlinear case should be handled would also be desirable