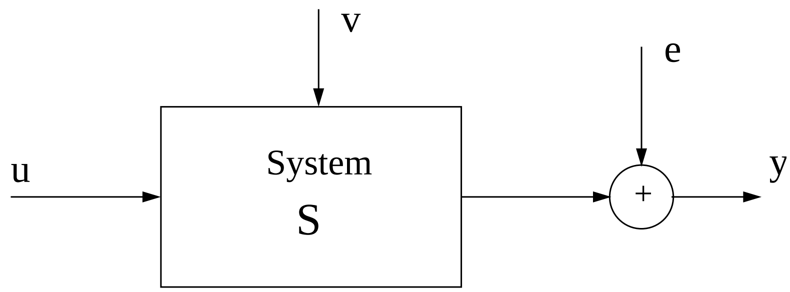
1 Introduction



Figur 1: System description.

- u is the input to the system (measurable).
- y is the output from the system (measurable).
- v is process noise (non measurable).
- e is measurement noise.

Problem: Given measurements $\{u(k),\, y(k)\}_{k=1}^{N}$ identify the system S.

Questions:

- Model structure?
- Choice of input?
- Identification method?
- How to deal with noise?

2 Stochastic variables - a short review

The following results are for reel and scalar variables. However, it is straightforward to generalize for complex and vector valued variables. Most of the results presented here are taken from the textbook *Discrete-time Stochastic Systems* by Torsten Söderström, *Springer*, 2002.

A *stochastic variable* y is characterized by its probability density function (pdf) $p_y(x)$. For example, a Gaussian distribution is described by

$$p_y(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-m)^2}{2\sigma^2}}, \quad \sigma > 0 \tag{2.1}$$

whereas a uniform distribution is given by

$$p_y(x) = \begin{cases} \frac{1}{b-a} & a < x \leq b \quad (b > a) \\ 0 & \text{otherwise} \end{cases} \tag{2.2}$$

Moments: To characterize a stochastic variable one often resorts to *moments*. The ith moment of a stochastic variable y is given by

$$E\, y^i = \int_{-\infty}^{\infty} x^i p_y(x)\, dx \tag{2.3}$$

where $E\{\cdot\}$ is the *expectation operator*. Notice the important fact that the expectation operator is a linear operator,*i.e.*, if $z = \sum_{i=1}^n a_i y_i$, where $\{y_i\}$ are random variables and a_i some constants, then it holds that

$$E\, z = \sum_{i=1}^{n} a_i E\, y_i \tag{2.4}$$

Two important and widely used concepts related to moments are the *mean value*

$$m_y \triangleq E\, y = \int_{-\infty}^{\infty} x p_y(x)\, dx \tag{2.5}$$

and the *variance*

$$\operatorname{var} y \triangleq E\,(y - m_y)^2 = E\, y^2 - m_y^2 \tag{2.6}$$

Correlation and dependence: Let ξ and η be two random variables with mean values m_ξ and m_η, respectively. The two variables are said to be *uncorrelated* if

$$E(\xi - m_\xi)(\eta - m_\eta) = 0 \tag{2.7}$$

The variables are said to be *independent* if

$$p_{\xi,\eta}(x, y) = p_\xi p_\eta \tag{2.8}$$

We have the following results:

- ξ, η independent $\Rightarrow$ ξ, η uncorrelated.
- ξ, η uncorrelated and Gaussian $\Rightarrow$ independent.

Notice, though, that for two general stochastic variables x and y

$$E\,xy \neq E\,xE\,y \tag{2.9}$$

However, if x and y are independent, then

$$E\,xy = E\,xE\,y \tag{2.10}$$

Gaussian distribution: A distribution of particular interest is the Gaussian pdf. A random variable y is said to be Gaussian distributed as

$$y \sim N(m, \sigma^2) \tag{2.11}$$

if its pdf is given by (2.1). Some properties are:

- If $y \sim N(m, \sigma^2)$ then $E\,y = m$ and var $y = \sigma^2$.
- If $y \sim N(m, \sigma^2)$ and $v = ay + b$, for some scalars a, b then $v \sim N(am + b, a^2\sigma^2)$.

The Gaussian distribution has several interesting features, including:

- Owing to the central limit theorem, the sum of many independent and equally distributed random variables can be well approximated by a Gaussian distribution. Hence, it seems feasible to model disturbances as Gaussian distributed random variables.
- Gaussian random variables have several attractive mathematical properties, see *e.g.* the results above.

3 Stochastic processes - some important concepts

A *stochastic process* $x(t)$ can for our purposes be considered as a sequence of random variables $\{x(t)\}_{t=0}^{k}$ (here we assume that $x(t)$ is scalar and real valued).

In order to fully characterize a stochastic process its distribution function

$$P(x(t_1) < x_1,\, x(t_2) < x_2,\, \ldots, x(t_k) < x_k) \tag{3.1}$$

is needed for arbitrary k, x_i and t_i $(i = 1, \ldots, k)$. Often this is too cumbersome and one resorts to using moments (first and second order moments).
Mean:

$$m_x(t) \triangleq E\, x(t) \tag{3.2}$$

Covariance function:

$$r_x(t,s) \triangleq E\,\big(x(t) - m_x(t)\big)\big(x(s) - m_x(s)\big) \tag{3.3}$$

Stationarity: A stochastic process $x(t)$ is *strictly stationary* if its distribution is time-invariant. It is *weakly stationary* if the mean and the covariance functions are time-invariant. Let $x(t)$ be a (weakly) stationary stochastic process. Then, apparently the mean is constant and the covariance function depends only on the difference between its time arguments. Hereafter $x(t)$ is a (weakly) stationary stochastic process, and we have
Mean:

$$m_x = E\, x(t) \tag{3.4}$$

Covariance function:

$$r_x(\tau) = E\,\big(x(t+\tau) - m_x\big)\big(x(t) - m_x\big) \tag{3.5}$$

Results:

- $r_x(0) = E\, x^2(t)$ the variance
- $r_x(\tau) \leq r_x(0)$
- $r_x(-\tau) = r_x(\tau)$ (real-valued processes)

In practice one often use the following estimates

$$\hat{m}_x = \frac{1}{N}\sum_{k=1}^{N} x(k) \tag{3.6}$$

$$\hat{r}_x(\tau) = \frac{1}{N}\sum_{k=1}^{N} \big(x(k+\tau) - \hat{m}_x\big)\big(x(k) - \hat{m}_x\big) \tag{3.7}$$

Ergodicity: $\hat{m}_x \to m_x$ and $\hat{r}_x(\tau) \to r_x(\tau)$, as $N \to \infty$ (number of data goes to infinity).

Instead of characterizing a signal using its mean value and covariance function (time description), we can choose to describe the signal in the frequency domain using the spectral density.

Spectral density:

$$\phi_x(\omega) = \frac{1}{2\pi}\sum_{n=-\infty}^{\infty} r_x(n)e^{-in\omega} \qquad \left(r_x(k) = \int_{-\pi}^{\pi} \phi_x(\omega)e^{ik\omega}\, d\omega \right) \tag{3.8}$$

Interpretation: The covariance function describes the coupling between $x(t)$ and $x(t+\tau)$. If $\tau = 0$ we get the variance. The spectral density describes how the energy in the signal is distributed over frequency (on the average).

White noise: A sequence of independent (and identically) distributed random variables is called *white noise.*

Results: If $e(t)$ is white noise with zero mean and variance λ^2. Then

- $r_e(\tau) = E\, e(t+\tau)e(t) = \lambda^2 \delta_{\tau,0}$, where the Kronecker delta $\delta_{t,s} = \begin{cases} 1 & t = s \\ 0 & \text{otherwise} \end{cases}$.
- $\phi_e(\omega) = \frac{\lambda^2}{2\pi}$, it is constant!

Often, the white noise $e(t)$ is assumed to be Gaussian distributed, *i.e.*,

$$e(t) \sim N(0, \lambda^2) \tag{3.9}$$

3.1 Filtering white noise

Let $e(t)$ be white noise with zero mean and variance λ^2. Then $y(t)$, defined as

$$y(t) + a_1 y(t-1) + \cdots + a_n y(t-n) = e(t) + c_1 e(t-1) + \cdots + c_m e(t-m) \tag{3.10}$$

is referred to as an *ARMA-process.* Often (3.10) is written abbreviated as

$$A(q)y(t) = C(q)e(t) \tag{3.11}$$

where

$$A(q) = z^n + a_1 z^{n-1} + \ldots + a_n \tag{3.12}$$
$$C(q) = z^m + c_1 z^{m-1} + \ldots + c_m \tag{3.13}$$

AR-process: A special case of (3.10), $A(q)y(t) = e(t)$ $(c_i = 0)$.
MA-process: A special case of (3.10), $y(t) = C(q)e(t)$ $(a_i = 0)$.

Properties:

- Mean: $E\, y = H(1) E\, e = 0$.
- Spectrum : $\phi_y(z) = H(z)H(z^{-1})\frac{\lambda^2}{2\pi}$. Note that if $z = e^{i\omega}$, the spectral density is obtained.
- Recall that the spectrum $\phi_y(z)$ can approximate any continuous spectrum arbitrarily close. Hence, the use of filtered white noise is a good way to describe various signals.