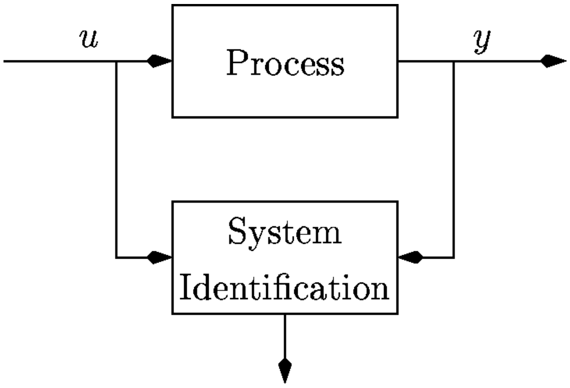
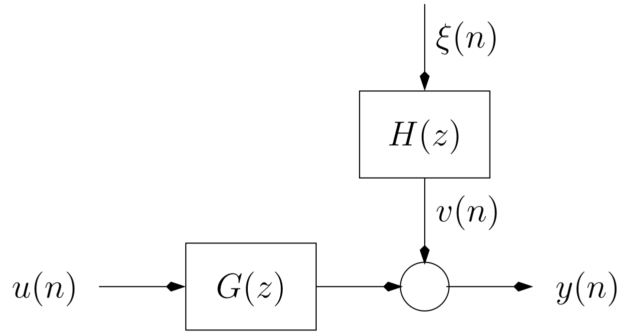
1. Identification

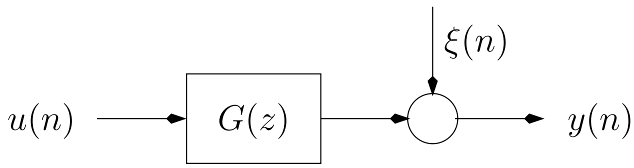
System Identification Problem:



Want to find $G(z)$ and $H(z)$:



First find G only, assuming white noise:



2. Stochastic Processes and Time Series

- A stochastic process is a statistical phenomenon that evolves in time according to probabilistic laws.
- A realization is a sample of the many possibilities that a process can take (population).
- A time-series is a set of values sampled from the process sequentially. A time-series is a particular realization of the process.
  - A discrete time-series is a set of observations made at discrete-times.
  - A deterministic time-series is one whose future values can be generated by a known (mathematical) function
  - A Stochastic time-series is one whose future values can be described only by some probabilistic distribution.

3. Stochastic Processes and Time Series

Consider a discrete time-series $\{u(t_1), u(t_2), \cdots, u(t_N)\}$ comprising N observations of a stochastic process.

- The joint probability distribution function (pdf) of $\{u(t_k)\}_{k=1}^{N}$ describes the probability that the random variable U takes on the values $U = u(t_k)$ jointly in a sequence of samples.
- A stochastic process is said to be strictly stationary if the joint pdf associated with the N observations taken at $t_1, t_2, \cdots, t_N$ is identical to the joint pdf associated with another set of N observations taken at times $t_1 + k, t_2 + k, \cdots, t_N + k$.
- In other words, a discrete-time is strictly stationary if the joint pdf of a set of observations remains unaffected by shifting all the times of observation forward or backward by an integer amount k.
- As a result, stationarity implies that properties of the pdf such as mean, variance, and higher-order moments do not change with time.

4. Properties of Stationary Time Series

The mean of a stationary stochastic process is defined as

$$\mu_u = \mathscr{E}(u_t) = \int_{-\infty}^{\infty} up(u)\,du$$

Often, one does not know the pdf $p(u)$. Estimate of mean:

$$m_u = \lim_{N\to\infty} \frac{1}{2N+1} \sum_{n=-N}^{N} u(n)$$

The next property of interest is variance. For stationary process, variance:

$$\sigma_u^2 = \mathscr{E}((u_t - \mu_u)^2) = \int_{-\infty}^{\infty} (u_t - \mu_u)^2 p(u)\,du$$

When pdf is not available, go for estimate of variance:

$$\hat{\sigma}_u^2 = \lim_{N\to\infty} \frac{1}{2N+1} \sum_{k=-N}^{N} (u(t_k) - m_u)^2$$

5. Auto Covariance Function (ACF)

ACF helps understand the inter-dependence of samples of time-series.

- The ACF for a general stochastic time series is defined as

$$\gamma(t, s) = \mathscr{E}((u_t - \mu_t)(u_s - \mu_s))$$

- For stationary time-series, the mean is constant and the dependence is only a function of the lag $l = t - s$.
- The ACF of a stationary process and its estimate are given by

$$\gamma_{uu}(t, s) = \gamma_{uu}(l) = E((u_t - \mu_u)(u_{t+l} - \mu_u))$$
$$r_{uu}(l) = \lim_{N\to\infty} \frac{1}{2N+1} \sum_{k=-N}^{N} (u(k) - m_u)(u(k-l) - m_u)$$

- ACF of a time-series is symmetric about the lag l.
- It takes the largest value at lag $l = 0$.
- In practice, a normalized function known as the auto-correlation function is used

$$\rho_{uu}(l) = \frac{\gamma_{uu}(l)}{\gamma_{uu}(0)}$$

6. Cross Covariance Function (CCF)

- The Cross-Covariance function (CCF) is a measure of dependence between samples of a time-series $\{u(k)\}$ and samples of another time-series $\{y(s)\}$.
- For stationary time-series, this is only a function of the distance between samples, *i.e.*, the lag l.
- The CCF of a stationary process is given by

$$\gamma_{uy}(t, s) = \gamma_{uy}(l) = \mathscr{E}((u_t - \mu_u)(y_{t+l} - \mu_y))$$

while, its estimate is given by

$$r_{uy}(l) = \lim_{N\to\infty} \frac{1}{2N+1} \sum_{k=-N}^{N} (u(k) - m_u)(y(k-l) - m_y)$$

CCF of two time-series satisfy $\gamma_{uy}(l) = \gamma_{yu}(-l)$. The largest value occurs at lag where the dependency is strongest. In practice, we use a normalized function known as the cross-correlation function.

$$\rho_{uy}(l) = \frac{\gamma_{uy}(l)}{\sqrt{\gamma_{uu}(0)}\sqrt{\gamma_{yy}(0)}},$$

7. Uses of ACF and CCF

- The ACF is used in detecting the underlying process, *i.e.*, whether it is a periodic, integrating, moving average, auto-regressive, independent, etc.
- ACF is also used in the estimation of periodicity of an oscillatory process and order of certain stochastic models.
- The CCF assumes the largest value when two time-series have strongest correlation. This has wide application in determining the time-delay of a system (important step in identification)
- The cross-correlation function is the output of an LTI system whose input sequence is the auto-correlation function.

$$r_{yu}(k) = g(k) * r_{uu}(k)$$

  This relationship is used in the estimation of impulse response coefficients.

8. Ergodicity

- To determine statistical properties, need many realizations
- But in reality, we usually have only one realization of the random signal.
- Moreover, we may not have information on PDF.
- A random signal $\tilde{u}(n)$ is said to be ergodic if all statistical averages can be determined from a single realization.
- For ergodic processes, time averages obtained from a single realization are equal to the statistical averages.
- For ergodic processes, the estimates approach the actual statistical properties when sufficiently large number of samples are taken while evaluating the summation.
- In view of this, we assume our processes to be ergodic and we will calculate only time averages.
- We will assume that we have sufficient number of samples and hence will take the time averages to be good approximations of statistical averages.