Applied Analysis of Time-Series

A machine learning approach

G. Castellanos-Dominguez

Universidad Nacional de Colombia, *Signal Processing and Recognition Group*,
La Nubia, Manizales

Point estimate statistics

Given a set of measures $\{x_k : k \in K\}$ (termed the *sample*), *compute in an approximate way* (that is, **estimate**) its characteristic of randomness $\theta \in \mathbb{R}$:

$$\tilde{\theta} = g(\{x_k\} \subseteq X, \epsilon \in \mathbb{R}^+ | \theta)$$

where $\tilde{\theta}$ is the best guess, and ϵ is the criterion of proximity to the true value of θ. As a rule,

$$\epsilon^2 = \mathbb{E}\left\{\|(\tilde{\theta} - \theta)\|_2^2\right\}$$

then, $g(\cdot; \min\{\epsilon^2\}|\cdot)$ becomes the *Mean Square Error* (MSE) estimator.

Estimate Properties

Unbiasedness: An estimator is unbiased if the observed value is equal to the expected value:

$$\mathbb{E}\left\{\tilde{\theta}\right\} = \lim_{N\to\infty} \mathbb{E}\left\{g(X|\theta)\right\} = \theta$$

Consistency: Error between expected and obser- ved values reduces as the sample size increases

$$\lim_{N\to\infty} P\{|\tilde{\theta}\text{-}\theta|<\epsilon\}=1, \text{ then, } \lim_{N\to\infty} \mathbb{E}\left\{(\tilde{\theta}\text{-}\theta)^2\right\}$$

Point estimate: Method of moments

main assumptions
Let the distribution of X have unknown real-valued parameters $\{\theta_i\} \in \Theta$. Estimators are constructed by matching the sample moments with their corresponding distribution moments:

$$\int_{-\infty}^{\infty} x^n p_\xi(x)\, dx \approx \frac{1}{M} \sum_{m=1}^{M} x_m^n$$

$$\int_{-\infty}^{\infty} (x - m_{1x})^n p_\xi(x)\, dx \approx \frac{1}{M} \sum_{m=1}^{M} (x_m - m_{1x})^n$$

`Python notebook: 02 MomentsEst`

Approaches to parameter estimation

Generative models assume the distribution of entire available data have been generated with an objectively fixed parameter θ that is not a random variable.

Discriminative models assume that the distribution have been generated by random parameter θ, representing subjective uncertainty or subjective belief. That is, it fixes the data and instead provides possible values for θ.

Maximum Likelihood [*Generative model*]

The parameter estimate $\tilde{\theta}$ maximizes a functional with positive asymptote, $L_{\boldsymbol{x}}(\theta) \in \mathbb{R}^{+}$, for each value of the observed data $\boldsymbol{x}$ on a support subset Ξ, that is, the value that would have most likely produced the observed data:

$$L_{\boldsymbol{x}}(\theta) = p(\boldsymbol{x} \in \Xi; \theta \in \Theta), \text{ likelihood function}$$

Assuming that the loglikelihood $\ln L_{\boldsymbol{x}}(\boldsymbol{\theta})$ is differentiable on $\boldsymbol{\theta}$, $\tilde{\theta}_i$ is computed by solving:

$$\frac{\partial}{\partial \theta_i} \ln L_{\boldsymbol{x}}(\boldsymbol{\theta}) = 0, \quad i \in \{1, 2, \ldots, k\}$$

Assuming $\boldsymbol{x} = \{x_k : k \in K\}$ is a random sample from the fdp $p(\boldsymbol{x}; \theta)$, the loglikelihood is as follows:

$$L_{\boldsymbol{x}}(\boldsymbol{\theta}) = p(x_1; \theta) \ldots p(x_K; \theta),\ \boldsymbol{x} \in \Xi$$
$$\ln L_{\boldsymbol{x}}(\boldsymbol{\theta}) = \sum_{\forall k} p(x_k; \theta)$$

Let X have Gaussian fdp with unknown parameters $\boldsymbol{\theta} = \{m_{1x}, \sigma\}$ with the following loglikelihood:

$$L_{\boldsymbol{x}}(\boldsymbol{\theta}) = \frac{e^{(-(x_1 - m_{1x})/2\sigma^{2x})}}{\sqrt{2\pi}\sigma_x} \cdots \frac{e^{(-(x_K - m_{1x})/2\sigma^{2x})}}{\sqrt{2\pi}\sigma_x}$$
$$= \left(\frac{1}{\sqrt{2\pi}\sigma_x}\right)^{K/2} \exp\left(-\frac{1}{2\sigma_x^2} \sum_{\forall k} (x_k - m_{1x})^2\right)$$

$$\ln L_{\boldsymbol{x}}(\boldsymbol{\theta}) = -\frac{K}{2} \ln(2\pi) - \frac{K}{2} \ln(\sigma_x^2) - \frac{1}{2\sigma^2} \sum_{\forall k} (x_k - m_{1x})^2$$

ML Estimation of m_{ix}, σ_x^2

$$\frac{\partial}{\partial m_{1x}} \ln L_{\boldsymbol{x}}(m_{1x}, \sigma^2) = \frac{1}{\sigma^2} \sum_{\forall k} (x_i - m_{1x})$$

$$= \frac{1}{\sigma^2} \left(\sum_{\forall k} x_i - K m_{1x} \right)$$

$$\frac{\partial}{\partial \sigma^2} \ln L_{\boldsymbol{x}}(m_{1x}, \sigma^2) = -\frac{K}{2\sigma^2} + \frac{1}{2\sigma^4} \sum_{\forall k} (x_i - m_{1x})^2$$

equating to 0 the partial derivatives, we will have:

$$\theta_1 : \hat{m}_{1x} = \frac{1}{K} \sum_{\forall k \in K} x_k$$

$$\theta_2 : \hat{\sigma}_x^2 = \frac{1}{K} \sum_{\forall k \in K} (x_k - m_{1x})^2$$

MLE illustration

Example (Optimizing procedure)
Generate observed values $\boldsymbol{x} = \{x_k : k \in K\}$ of a random variable $\mathcal{N}_X(m_{1x}, \sigma^2)$.

Write the log-likelihood functions for the simulated random variables and verify that the simulated maximum likelihood estimates for $\hat{m}_{1x}$ and $\hat{\sigma}_x^2$ are reasonably close to the true parameters.

Produce side-by-side graphs of $\ln L_{\boldsymbol{x}}(m_{1x}, \sigma^2)$, indicating where the simulated maximum occurs in each graph.

`Python notebook: 03 MLExample`

Bayesian Estimation [*Conditional model*]

$P(\theta)$ - *prior*: height measured at room, $\boldsymbol{\theta} \in \mathbb{R}^{Nsample}$

$P(D)$ - *evidence*: records of height for the last N_y years

$P(D|\theta)$ - *likelihood*: The probability of data given θ

$P(\theta|d)$ - *posterior*: The probability of a parameter given the likelihood of some data, computed as follows:

$$P(\theta|D) = \frac{P(D|\theta)P(\theta)}{P(D)} = \frac{P(D|\theta)P(\theta)}{\sum_{\theta' \in \Theta} P(D|\theta')P(\theta')}$$

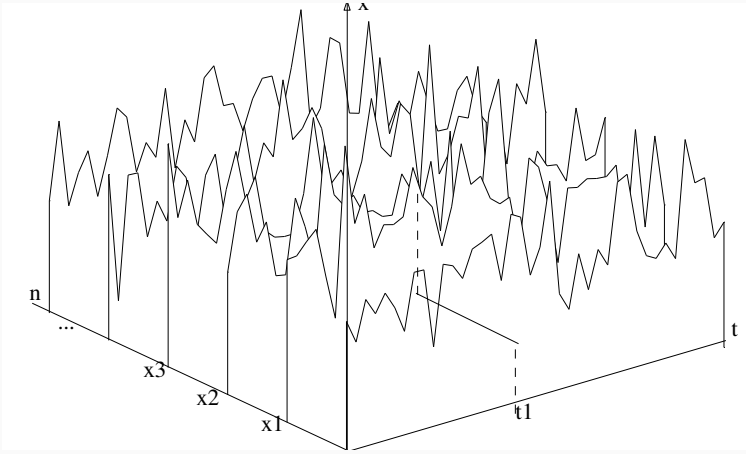
The evidence is replaced with an integral of the numerator because (1) $P(D)$ is extremely difficult to calculate, (2) $P(D)$ doesn't rely on θ, which is what we really care about, and (3) its usability as a normalizing factor can be substituted for the integral value, which ensures that the integral of the posterior distribution is 1.

Random function [process]

A variable $x(\cdot,\cdot,)$ spanned over a domain(s) like time, space, etc.

$x_i(t)$ - trajectory.
$\{x_i(t : \forall i \in N)\}$ - signal ensemble.

If the uncertainty model varies over time, $p(x,t) = \text{var}$, the random process is termed *nonstationary*.



Random Signal Ensemble

$$\text{Narrow-sense Stationarity: } p(\xi_i, t) = p(\xi_i, t + \Delta t),\ \forall \Delta t$$

$$\text{Wide-sense Stationarity: } m_{1x}(t) = m_{1x}, \quad \sigma_x(t) = \sigma_x, \forall t$$

Python notebook: TseriesGen

Wide-sense stationarity

$$\int_{-\infty}^{\infty} \xi^n p(\xi)\, d\xi \approx \lim_{T\to\infty} \frac{1}{T}\int_T \xi^n(t)\, dt \triangleq \overline{\xi^n(t)}, n \in \mathbb{N}$$

$$\int_{-\infty}^{\infty} (\xi - m_{1\xi})^n p(\xi)\, d\xi \approx \lim_{T\to\infty} \frac{1}{T}\int_T \left(\xi(t) - \overline{\xi(t)}\right)^n dt, n \geq 2,$$

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \xi\eta p(\xi,\eta)\, d\xi d\eta \approx \lim_{T\to\infty} \frac{1}{T}\int_T \xi(t)\eta^*(t+\tau)dt = R_{\xi\eta}(\tau)$$

$$R_\xi(\tau) = \lim_{T\to\infty} \frac{1}{T}\int_T \xi(t)\,\xi^*(t+\tau)\, dt$$

$$K_\xi(\tau) = \lim_{T\to\infty} \frac{1}{T}\int_T \left(\xi(t)\text{-}\overline{\xi(t)}\right)\left(\xi^*(t+\tau)\text{-}\overline{\xi(t)}\right) dt$$

$$= R_\xi(\tau) - m_{1\xi}^2$$

Correlation function: Properties

(a). *Parity.* $R_\xi(\tau) = R_\xi^*(-\tau)$, $K_\xi(\tau) = K_\xi(-\tau)$

(b). *Maximal value.* $|R_\xi(\tau)| \le R_\xi(0)$, $\quad |K_\xi(\tau)| \le K_\xi(0)$.

$$K_\xi(0) = \frac{1}{T}\int_0^T \left(\xi(t) - \overline{\xi(t)}\right)\left(\xi^*(t) - \overline{\xi(t)}\right) dt = \sigma_\xi^2,$$

$$R_\xi(0) = \frac{1}{T}\int_0^T \xi(t)\,\xi^*(t)\, dt = \overline{\xi^2(t)}$$

(c). *Periodicity.* If $\xi(t) = \xi(t-T)$, $\forall t \in T$, then $R_\xi(\tau) = R_\xi(\tau - T), \forall \tau \in T$.

(d). *Convergence.* If $\xi(t) \neq \xi(t-T)$, $\forall t \in T$ then

$$\lim_{|\tau|\longrightarrow\infty} R_\xi(\tau) = \overline{\xi^2(t)}, \qquad \lim_{|\tau|\to\infty} K_\xi(\tau) = 0,$$

Example (CF)

Find the CF of $\xi(t) = a\cos(\omega_c t + \phi)$, $a = \text{const}$, $\omega_c = \text{const}$, and $p(\phi) = 1/2\pi$ is the random phase.

$$R_\xi(\tau) = \lim_{T\to\infty} \frac{1}{T} \int_0^T a\cos(\omega t + \phi)\, a\cos(\omega t + \omega\tau + \phi)\, dt$$

$$= \lim_{T\to\infty} \frac{a^2}{T} \int_0^T \frac{1}{2}\left(\cos(\omega\tau + \phi) + \cos(2\omega t + \omega\tau + 2\phi)\right) dt$$

$$\int_0^T \cos\omega\tau dt = \cos\omega\tau$$

$$\int_0^T \cos(2\omega t + 2\phi + \omega\tau)\, dt = \cos(2\phi + \omega\tau) \int_0^T \cos 2\omega t dt$$

$$- \sin(2\phi + \omega\tau) \int_0^T \sin 2\omega dt$$

Since the following equality takes place:

$$\lim_{T\to\infty}\int_T \cos k\omega t dt = 0,\ \lim_{T\to\infty}\int_T \sin k\omega t dt = 0, \forall k \in \mathbb{Z}$$

then, the correlation function of $\xi(t)$ is computed as below:

$$R_\xi(\tau) = \frac{a^2}{2}\cos\omega\tau$$

Example (CF)
Compute CR of $\xi(t) = k_1\cos(\omega_c t + k_0\phi) + k_2\cos(t) + \eta(t)$

`Python notebook: 05 CorrFunction`

Correlation function: Basic Models

(a). $R_\xi(\tau) = N_0\delta(\tau)/2$. Then, $\xi(t)$ is a random process with values totally independent, that is, having the highest uncertainty.

(b). $\lim_{|\tau|\to\infty} R_\xi(\tau) = 0$. Then, $\xi(t)$ with fading dependency. The more distant the values, the stronger the independence between them.

(c). $R_\xi(\tau) = f(R_\xi(\tau - \tau_1), \ldots, R_\xi(\tau - \tau_m),)$, m-order Markovian process. $m = 1$ – plain Markovian process

(d). $\lim_{\tau_1\to\infty} R_\xi(\tau) = \text{const}$. Then, $\xi(t)$ is a random process with values entirely dependent; that is, there is no uncertainty at all.

(Co)variance Matrix

A square matrix that holds the first-order mixed moment between each pair of data elements, for which *variances* appear on the diagonal while *covariances* – on all other elements.

$$\mathrm{cov}_{x,y,z} = \begin{bmatrix} \sigma_x^2 & \sigma_{xy} & \sigma_{xz} \\ \sigma_{yx} & \sigma_y^2 & \sigma_{yz} \\ \sigma_{zx} & \sigma_{zy} & \sigma_z^2 \end{bmatrix}$$

where each semipositive-definite scalar value is estimated as:

$$\begin{aligned} \mathrm{cov}_{x,y} &= \mathbb{E}\{(X - \mathbb{E}\{X\})(Y - \mathbb{E}\{Y\})\} \\ &= \frac{\sum_{\forall x_i \in X, y_i \in Y} (x_i - \bar{x})(y_i - \bar{y})}{N - 1} \end{aligned}$$

`Python notebook: 0e CovarianceMatrix`

Power Spectral Density

For random signals, analysis is carried out within a long enough time segment, that is:

$$x_T(t) = \mathrm{rect}_T(t)\, x(t),$$

$$X_T(\omega) = \mathscr{F}\{x(t)\,\mathrm{rect}(t/T)\}$$

$$\frac{1}{2\pi}\int_{\mathbb{R}} X(\omega)X^*(\omega)d\omega = \int_{\mathbb{R}} x^2 dt, \quad \text{Parseval's Theorem}$$

$$\lim_{T\to\infty}\frac{1}{T}\mathbb{E}\left\{\int_T x^2 dt\right\} = \lim_{T\to\infty}\frac{1}{2\pi}\int_{\mathbb{R}}\frac{1}{T}\mathbb{E}\left\{X_T(\omega)X_T^*(\omega)\right\}d\omega$$

$$\text{assuming } m_{1x}=0, \quad m_{2x} = \frac{1}{2\pi}\int_{\mathbb{R}} S_x(\omega)d\omega,$$

$$S_x(\omega) = \lim_{T\to\infty}\frac{|X_T(\omega)X_T^*(-\omega)|}{T} \geq 0, \quad \text{PSD}$$

$$\text{Properties:} \quad S_x(\omega)\in\mathbb{R}^+, S_x(\omega) = S_x(-\omega)$$

Wiener-Jinchin Transform

$$
\begin{aligned}
S_x(\omega) &= \lim_{T\to\infty} \mathbb{E}\left\{X_T(\omega)\, X_T^*(-\omega)\right\}/T \\
&= \lim_{T\to\infty} \frac{1}{T}\mathbb{E}\left\{\int_T x_T(t_1)\, e^{(j\omega t_1)} dt_1 \int_T x_T(t_2)\, e^{(-j\omega t_2)} dt_2\right\} \\
&= \lim_{T\to\infty} \frac{1}{T}\mathbb{E}\left\{\int_T dt_2 \int_T e^{(-j\omega(t_2-t_1))} x_T(t_1)\, x_T(t_2)\, dt_1\right\} \\
&= \lim_{T\to\infty} \frac{1}{T}\int_T dt_2 \int_T e^{(-j\omega(t_2-t_1))}\mathbb{E}\left\{x_T(t_1)\, x_T(t_2)\right\} dt_1 \\
&= \lim_{T\to\infty} \frac{1}{T}\int_{-T-t_1}^{T-t_1} d\tau \int_T e^{-j\omega\tau} R_x(t_1, t_1+\tau) dt_1 \\
&= \int_{-\infty}^{\infty} \left(\lim_{T\to\infty} \frac{1}{T}\int_T R_x(t_1, t_1+\tau) dt_1\right) e^{-j\omega\tau} d\tau
\end{aligned}
$$

Wiener-Jinchin Transform

Therefore, we have:

$$S_x(\omega) = \mathscr{F}\left\{\mathbb{E}\left\{R_x(t, t+\tau)\right\}\right\}$$
$$\Rightarrow S_x(\omega) = \mathscr{F}\left\{R_x(\tau)\right\}$$

Likewise, it holds that:

$$\mathscr{F}^{-1}\left\{S_\xi(\omega)\right\} = \mathscr{F}^{-1}\left\{\mathscr{F}\left\{R_\xi(\tau)\right\}\right\}$$
$$= R_\xi(\tau)$$

Shape restriction. To be an implementable process, its Fourier Transform must fulfill the following condition:

$$\mathscr{F}\left\{R_\xi\left(\tau\right)\right\} \geq 0, \quad \forall\omega$$

White Gaussian Noise

An ergodic process holding all spectral components, each one with the same power in average, defined as:

$$S(\omega) = N_0/2, \quad \omega \in (-\infty, \infty).$$

Using the *Wiener-Jinchin Transform*, we obtain:

$$R(\tau) = \int_{-\infty}^{\infty} \frac{N_0}{2} e^{j2\pi ft} df = \frac{N_0}{2}\delta(\tau)$$

Colored noise assumes $\Delta\omega < \infty$:

$$\begin{cases} S(\omega) &= N_0/2, (-\Delta\omega < \omega < \Delta\omega), \\ R(\tau) &= N_0\Delta\omega \operatorname{sinc}(2\Delta\omega\tau)/2, \end{cases}$$

`Python notebook: 06 ExNoiseColored`

Examples of $S(\omega)$ and $R(\tau)$

Daily temperatures, `Python: 07 ExTemperatures`

MEG recordings, `Python: 07a ExEGG`

Stochastic Modeling

A real valued (one-dimensional domain) stochastic process is a family of random variables $\{X_t : t \in I\}$ defined on a probabilistic space, $X_t : \text{Observation set} \to \mathbb{R}, t \in I \subseteq \mathbb{R}^+$

$\{X_t\}$ is a discrete-state process if its values are countable. Otherwise, it is a continuous-state process. State space – the set $S \subseteq \mathbb{R}$ whose elements are the values of the process.

The model is ruled by Stochastic Differential Equations that can be viewed as stochastic process from two points of view:

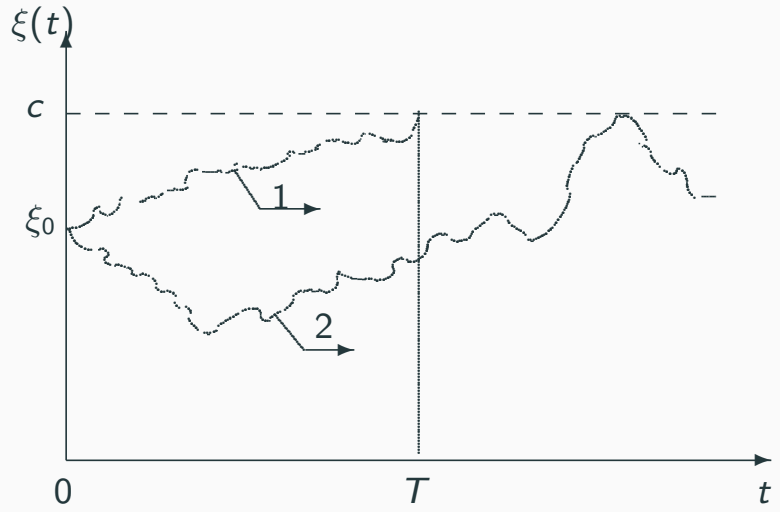
The changes (evolution) of randomness between neighboring states becomes the stochastic process (**trial-based analysis**)

An stochastic process with a probability distribution across-time (**ensemble-based analysis**)

State evolution over time: Random Walk

A path holds a succession of random steps on a state space. The particle, if it is in state i, can in a single transition either stay in i or move to one of the neighboring states $i-1$, $i+1$:

$$Pr\{\xi_{n+1} = i+1 \mid \xi_n = i\} = p_i$$
$$Pr\{\xi_{n+1} = i-1 \mid \xi_n = i\} = q_i$$
$$Pr\{\xi_{n+1} = i \mid \xi_n = i\} = r_i$$



stochastic paths

`Python: 08 RandomWalk`

Changes over time (continuous model)

Brownian motion: A differential equation perturbed by noise:

$$d\xi(t) = \mu\xi(t)dt + \sigma\xi(t)d\eta(t)$$
$$\xi(T) - \xi(0) = \mu\int_0^T \xi(t)dt + \sigma\int_0^T \xi(t)d\eta(t)$$

where $\xi(t)$ is the meaningful variable at time t, $\mu > 0$ (drift) and $\sigma > 0$ (volatility) are the drift and diffusion parameters, $\eta(t)$ is the Gaussian white noise term considered the derivative of Brownian motion.

`Python notebook: 08a Brownian-Motion`

Changes modeled by linear responses

Let x and y be the input and output os a linear system, then:

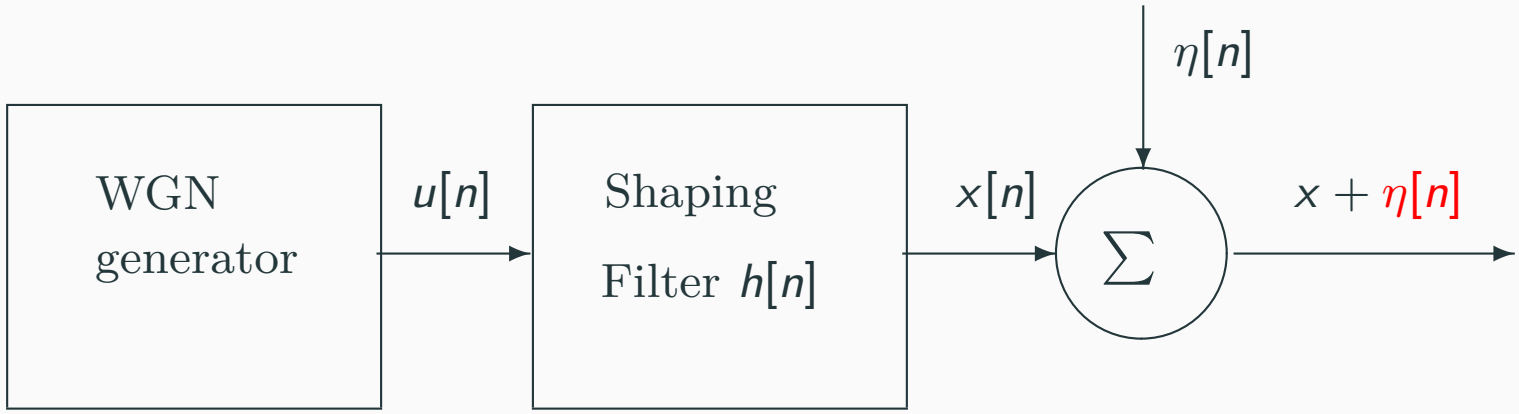
$$S_y(z) = S_x(z)H(z)H^*(1/z^*)$$
$$S_x(z) = \mathscr{Z}\{r_x[m]\},\ S_y(z) = \mathscr{Z}\{r_y[m]\}, H(z) = \mathscr{Z}\{h[m]\}$$

Let $x = \eta$ be WGN with zero-mean and power σ_η^2, so that:

$$S_x(z) = \mathscr{Z}\{\sigma_\eta^2\delta[m]\} = \sigma_\eta^2$$
$$\text{Then, } S_y(z) = \sigma_\eta^2 H(z)H^*(1/z^*)$$
$$\sim H(z)H^*(1/z^*) = \frac{B(z)B^*(1/z^*)}{A(z)A^*(1/z^*)}$$

$$A(z) = 1 + \sum_{k=1}^{p} a_k z^{-k}, \quad B(z) = 1 + \sum_{k=1}^{q} b_k z^{-k},\ z^{-k} \text{ backshift}$$

Linear Regressive Models



Parametric modeling generation of a discrete stochastic process.

At *state* $x[n]$, an *autoregressive model* encodes linear relationships between p, q delayed samples, as follows:

$$x[n] = \sum_{k=1}^{p} a_k x[n-k] + \sum_{k=0}^{q} b_k u[n-k] + \eta[n], \quad a_k \in \mathbb{R},\ b_k \in \mathbb{R}$$

where the input $\boldsymbol{u}$ is an unobservable excitation of WGN with unity intensity.

Differential equation modeling

Model	*Iterative equation*
AR	$x[n] = -\sum_{k=1}^{p} x[n-k] + u[n]$
MA	$x[n] = u[n] + \sum_{k=1}^{q} b_k u[n-k]$
ARMA	$x[n] = \sum_{k=1}^{p} a_k x[n-k] + u[n] + \sum_{k=1}^{q} b_k u[n-k],$

Autoregressive models provide a parsimonious description of a stochastic process through two polynomials: the input one for the autoregression (AR) and the output for the moving average (MA).

Implementation of Autoregressive Models

Assumption: the underlying data comes from a second-order stationary process. Otherwise, data transforming is used.

Order Selection: Information criteria measuring the goodness of fit with the estimated parameters:

$$\text{Akaike Information Criterion AIC} = -2\log(L) + 2k,$$
$$\text{Bayesian Information Criterion BIC} = k\log(n) - 2\log(L)$$

L is the likelihood estimate for the data, k - the number of estimated parameters, n - sample size.

Probability evolution: Markov processes

Evolution of time-point probabilities, $P(\xi(t) \leq \xi \mid \xi(t_0) = \xi_0)$, through a differential operator $\mathscr{K}$:

$$\frac{d}{dt} P = \mathscr{K} \{P\}$$

A process $\xi(t)$ is *Markovian*, if for a fixed state $\xi(u)$, other discrete values $\xi(t)$, $t > u$, do not depend on $\xi(s), s < u$.

That is, the conditional PDF of the last value $\xi(t_n)$:

$$\begin{aligned} P\left(\xi(t_n) \leq \xi_n | \xi(t_1) = \xi_1, \ldots, \xi(t_{n-1}) = \xi_{n-1}\right) \\ = P\left(\xi(t_n) \leq \xi_n | \xi(t_{n-1}) = \xi_{n-1}\right) \end{aligned}$$

A Markov process is a random walk with a selected probability for making a move that is independent of the previous history of the system. The combination of displacements from neighboring time positions is a sequence of steps thet forms a chain.

Examples

`09c ArmaGen 09ca regModels`

`09d MarkovPtransition`

Predecir el precio del peso con respecto al dólar mediante modelos AR, MA y ARMA. Inspección visual de la función de autocorrelación y fase sobre dependencias lineales entre muestra actual y pasadas. `09e modelosregresivos`

Ajustar una serie de tiempo generada por un proceso autoregresivo utilizando procesos de Markov donde los estado para deducir la matriz de transición y las probabilidades iniciales se calculen utilizando diferentes estrategias parar estimación de estados: `09f MarkovChains`

Time series analysis

Extraction of meaningful summary and statistical information from data points arranged in chronological sequence.

Pattern decomposition splits a time series into several components, each representing the following underlying pattern components: Base Level */+ Trend */+ Seasonality */+ Noise (+ – *add decomposition*,* – *mult decomposition*)

Time series modeling/filtering: Transformation, extraction, or removal of stochastic pattern components.
Smoothing/detrending/deseasonalizing/stationarization

Machine Learning for extracting patterns from time-series:
Unsupervised learning infers input patterns (Clustering).
Supervised learning maps inputs to outputs based on example input-output pairs (Classification – categorical output, Regression/Prediction – continuous valued output).

Pattern decomposition

Patterns are temporal structures.

Stationary condition: the statistics (e.g. mean and variance) of the underlying signals remain constant over time.

Seasonality relates to pseudo-periodic structures or short-term cycles of time series.

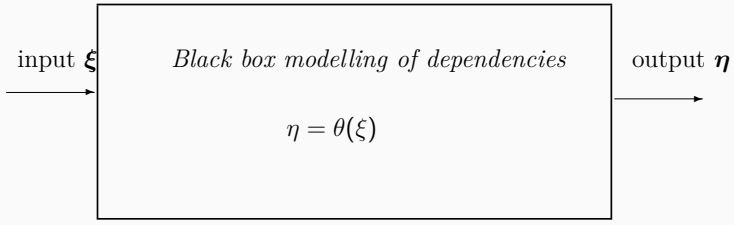
Trend concerns whether the time series has decreasing, constant, or increasing behavior over time.

Noise is the remaining variability in the data barely explained by the model.

`1a TseriesGen` `1b Preprocessing`

Building dependency from experimental data

Assessing of relationship between sets of experimental values assumed to be interacting among them. Though the precise interaction mechanism is unknown, the dependency association $\theta(\cdot)$ is supposed to be ruled by uncertainty principles, represented by a black-box model, having at the input the representative observations (*independent variable*) $\boldsymbol{\xi}$ together with the response outcomes $\boldsymbol{\eta}$ (*dependent variable*) at the output.



dependency between experimental data

The experiment provides measurements with independent values $\boldsymbol{X} \in \mathbb{R}^{s \times M}$, and dependent values $\boldsymbol{Y} \in \mathbb{R}^{r \times M}$:

$$\boldsymbol{x} = \begin{bmatrix} x_{11} & \cdots & x_{1M} \\ \cdots & \cdots & \cdots \\ x_{sM} & \cdots & x_{sM} \end{bmatrix}, \boldsymbol{Y} = \begin{bmatrix} y_{11} & \cdots & y_{1M} \\ \cdots & \cdots & \cdots \\ y_{rM} & \cdots & y_{rM} \end{bmatrix}$$

Issues to be solved for building dependencies:

- What class of uncertainty measures are to be used for evaluating the dependency (if any)?
- Is there a relationship associating the whole set of involved variables, or just part of them is contributing? How to assess the contribution of variables to the model?
- How to evaluate the effectiveness of dependency relation model? Does the build model supply an explainable association in terms of determining the state of an object under consideration?

Problem statement of experimental dependence

Let $\boldsymbol{x} \in \mathbb{R}^p$ and $\boldsymbol{y} \in \mathbb{R}^q$ be a couple of random vectors with the corresponding observation sets $(x_1, x_2, \ldots, x_n) \subset X$ and $(y_1, y_2, \ldots y_n) \subset Y$, for which the mutual dependence is assessed $\tilde{\boldsymbol{y}} = \theta(\boldsymbol{x} = X)$. Let $d(\tilde{\boldsymbol{y}}, \boldsymbol{y})$ be a metric in $\mathbb{R}^q$, so that there is a function $\theta(\cdot) : \mathbb{R}^p \mapsto \mathbb{R}^q$ for which the conditional mean reaches its minimum value in terms of the Euclidean norm ℓ_2-norm, resulting in the minimizing solution, termed *regression* of y on x, defined as follows:

$$\tilde{\boldsymbol{y}} = \mathbb{E}\{\boldsymbol{y} = Y \mid \boldsymbol{x} = X\} \tag{1}$$

Regression model

The assumed relationship results in the regression model with following expression:

$$\boldsymbol{y} = \theta(\boldsymbol{x} = X) + \boldsymbol{\epsilon}(\boldsymbol{x} = X)$$

where it holds that $\mathbb{E}\{\epsilon | \boldsymbol{x} = X\} = 0$.

The function θ is hardly known, mostly because of the complexity of characterizing the real-world systems.

Two approaches to include the pairwise relationship in the model **??**:

> An approximating function is assumed, relying on certain empirical evidence,
> $\theta(\cdot)$ is learned from an observation set using data-driven approaches.

Approximating Regression using Gaussian pdfs

A probability density function $\mathbb{E}\{\boldsymbol{y}=Y|\boldsymbol{x}=X\}\approx p(\boldsymbol{x}=X;\boldsymbol{\pi})$ is assumed to approximate the regression model in **??** by ruling the parameter set $\tilde{\boldsymbol{\pi}}$:

$$\boldsymbol{y} = p(X;\tilde{\boldsymbol{\pi}}) + \boldsymbol{\epsilon}(X), \quad \text{s.t.: } \tilde{\boldsymbol{\pi}} = \min_{\boldsymbol{\pi}}\{\boldsymbol{\epsilon}(X)\} \tag{2}$$

Without losing generality, the pdf of variables X and Y are Gaussian, that is: $p(Y)=\mathcal{N}(m_{1y},\sigma_y)$, $p(X)=\mathcal{N}(m_{1x},\sigma_x)$, so that their joint is also Gaussian, as follows:

$$p(Y,X) = \frac{1}{\sqrt{(2\pi)^2(\sigma_y^2\sigma_x^2-\sigma_{yx}^2)}}$$
$$\exp\{\frac{1}{2(\sigma_y^2\sigma_x^2-\sigma_{yx}^2)}(\sigma_x^2\bar{y}^2 - 2\sigma_{yx}^2\bar{y}\bar{x} + \sigma_y^2\bar{x}^2)\}$$

where $\bar{z} = z - m_{1z}$ is the centralized mean value.

The conditional pdf needed in Eq. (1) is computed as $p(Y \mid X) = p(Y, X)/p(X)$, yielding the Gaussian expression:

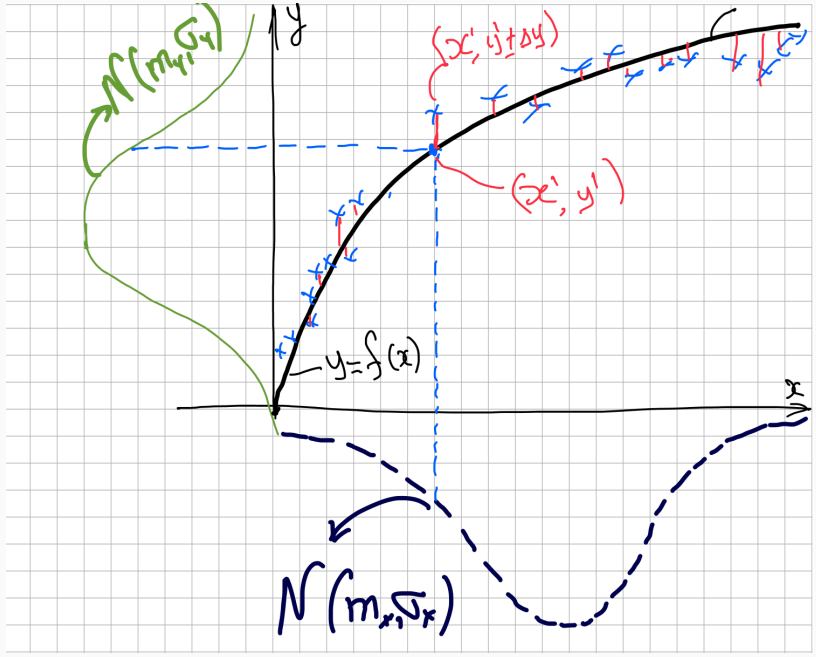
$$\frac{\sigma_x^2}{2(\sigma_y^2\sigma_x^2 - \sigma_{yx}^2)}(y - m_{1y} - \sigma_{yx}^2\sigma_x^{-2}(x - m_{1x}))^2$$

The term $m_{1y} + \sigma_{yx}^2\sigma_x^{-2}(x - m_{1x})$ is the Gaussian conditional mean value implementing the regression in Eq. (1):

$$\begin{aligned}\mathbb{E}\{\boldsymbol{y} = Y \mid \boldsymbol{x} = X\} &= m_{1y} + \sigma_{yx}^2\sigma_x^{-2}(x - m_{1x}) \\ &= m_{1y} + r_{yx}\sqrt{\frac{\sigma_x^2}{\sigma_y^2}}(x - m_{1x})\end{aligned}$$

where $r_{yx} = \sigma_{yx}^2/\sqrt{\sigma_y^2\sigma_x^2}$ is the correlation index between random variables X and Y.

The correlation index ranges within $-1 \leq r_{yx} \leq 1$, meaning that whenever $r_{yx} = 0$, the conditional mean does not depend on the input variable X. In other words, a couple of uncorrelated Gaussian random variables implies their independence.



x.

Furthermore, the regression Y on X turns to be linearly dependent on the value X. Therefore, function $\theta(\boldsymbol{x} = X)$ becomes also linear.

Regression learning

We will employ a learning rule that estimates a function $\theta{:}\mathbb{R}^{?}\mapsto\mathbb{R}^{?}$ from representative observations of an independent variable (also termed indicator) $\boldsymbol{\xi}\in\mathbb{R}^{?}$, for which a multivariate approximation problem can be stated by optimizing across $m\in M$ the following framework:

$$\min_{\boldsymbol{\pi}}\mathbb{E}\left\{\|\boldsymbol{\eta}(\boldsymbol{y}_m)-\theta\{\boldsymbol{\xi}(\boldsymbol{x}_m|\boldsymbol{\pi})\}+\boldsymbol{\epsilon})\|_2:\forall m\in M\right\},$$

where $\boldsymbol{\eta}\in\mathbb{R}^{M}$ is the response vector (dependent variable), $\boldsymbol{\epsilon}\in\mathbb{R}^{M}$ is the additive error term that is independent of $\boldsymbol{\xi}$, and $\boldsymbol{\pi}$ is the unknown parameter vector that allows optimization of the approximating function $\theta(;)$ (termed regression), fitting the data as much as close in terms of ℓ_2-norm distance.

$$P(\mathbf{W} \mid \widehat{\mathbf{y}}, \mathbf{x}) = P(\hat{\mathbf{y}} \mid \mathbf{x}, \mathbf{W})\mathbf{P}(\mathbf{W} \mid \mu, \sigma_{\mathbf{0}}^{\mathbf{2}})$$

$$\propto \exp\left(-\frac{(\hat{\mathbf{y}} - \mathbf{W}^{\top}\mathbf{x})^{\mathbf{2}}}{2\sigma_0}\right) \exp\left(-\frac{\mathbf{W}^{2}}{2\sigma_{0}}\right)$$