Statistics of Stochastic Processes

**discrete time:** sequence of r.v. \( \ldots, X_{-1}, X_0, X_1, X_2, \ldots \)

\( X_t \in \mathbb{R}^d \) in general. Here: \( d = 1 \).

**continuous time:** random function \( X(t), a \leq t \leq b \)
\((-\infty \leq a < b \leq \infty)\)

**spatial processes** (e.g. image analysis, \( d = 2, 3 \)):

\( X_z, z \in \mathbb{Z}^d \) (discrete)

\( X(z), z \in \mathbb{R}^d \) (continuous)
**Stationarity** (spatial homogeneity)

\[ \mathcal{L}(X_{t_1}, \ldots, X_{t_n}) = \mathcal{L}(X_{t_1+s}, \ldots, X_{t_n+s}) \quad \text{for all } n \geq 1, t_1, \ldots, t_n, s \in \mathbb{Z} \]

Analogously for \( t_j, s \in \mathbb{R}, z_j, s \in \mathbb{Z}^d, z_j, s \in \mathbb{R}^d \)

Note: Distribution of \( X(t), a \leq t \leq b \), as a random function, e.g. as a random variable with values in \( C[a, b] \), completely determined by the **finite-dimensional distributions**

\[ \mathcal{L}(X(t_1), \ldots, X(t_n)) \quad n \geq 1, t_1, \ldots, t_n \in [a, b] \]

**Gaussian process** if \( \mathcal{L}(X(t_1), \ldots, X(t_n)) \) \( n \)-dimensional normal

**Examples:**

a) autoregressive process of order 1 or AR(1)-process

\[ X_t = aX_{t-1} + Z_t, \quad \text{E}Z_t = 0, \text{var } Z_t = \sigma^2 < \infty \]

\( Z_t \) innovation at time \( t \); uncorrelated or even i.i.d.
\[(X_t - \mu) = a(X_{t-1} - \mu) + Z_t, \quad \mathbb{E}Z_t = 0, \text{var } Z_t = \sigma^2 < \infty\]

if \(\mathbb{E}X_t = \mu \neq 0\).

**Gaussian** if innovations \(Z_t\) i.i.d \(\mathcal{N}(0, \sigma^2)\).

**Stationary** \(\iff |a| < 1\).

For \(a = 1\): **random walk**

\[X_t = X_0 + \sum_{j=1}^{t} Z_j, \quad \text{var } X_t = \text{var } X_0 + t\sigma^2\]

\(\sim\) distribution of \(X_t\) depends on \(t \sim\) not stationary

b) stationary AR(1)-process solves **stochastic difference equation**:

\[\Delta X_t = X_t - X_{t-1} = (a - 1)X_t + Z_t = -bX_t + Z_t\]

with \(b > 0\).
AR(1): \[ X_t = 0.9X_{t-1} + Z_t \]
Gaussian random walks with $\mathbb{E}X_t = 0$ for all $t$
\[ \Delta X_t = X_t - X_{t-1} = -bX_t + Z_t, \quad b > 0 \]

time discretization: \( t - 1 \sim t - \Delta t, \Delta t \to 0 \)

**stochastic differential equation** (Ornstein-Uhlenbeck process)

\[ dX(t) = -bX(t) + \sigma dW(t), \quad b > 0 \]

\( W(t) \) standard **Wiener process** or **Brownian motion**, e.g. on \( [0, \infty) \) with \( W(0) = 0 \):

i) **process with independent increments:**

\( W(t) - W(s) \) independent of \( W(s) \), \( 0 \leq s < t \)

ii) **Gaussian:** increment \( W(t) - W(s) \) is \( \mathcal{N}(0, t - s) \)-distributed

**Wiener process** = continuous-time analogue to and limit of **Gaussian random walk** (used for simulation)
c) nonlinear autoregressive process of order $p$ or NLAR($p$)-process:

$$X_t = m(X_{t-1}, \ldots, X_{t-p}) + Z_t, \quad \mathbb{E}Z_t = 0, \text{var } Z_t = \sigma^2 < \infty$$

innovations $Z_t$ usually i.i.d.

Key property for statistics:

**1 realization** $X_1, \ldots, X_N$ of part of a stationary process suffices for statistical inference (estimation, testing, ...), e.g. (under weak assumptions) **ergodic theorem** implies:

$$\bar{X}_N = \frac{1}{N} \sum_{t=1}^{N} X_t \to \mu = \mathbb{E}X_s \quad \text{(independent of } s)$$

Extension of inference to local stationarity or piecewise stationarity possible
NLAR(2): $X_t = \psi(X_{t-1} + X_{t-2}) + Z_t$, $\psi = \text{logistic function}$
Markov property: distributions of future $X_s, s > t$ given past and present only depend on current value $X_t$

Examples: AR(1), Ornstein-Uhlenbeck, NLAR(1)

Markov random fields (spatial): distributions of $X_z$ beyond a boundary only depend on data on the boundary

Statistical problems:

Estimation, e.g. for parametric Gaussian NLAR(p) by least-squares as in regression:

$$\sum_{t=p+1}^{N} \left( X_t - m(X_{t-1}, \ldots, X_{t-p}, b) \right)^2 = \min_b$$

Forecasting: Fit appropriate autoregressive scheme to the data (e.g. NLAR(p)) and predict $X_{N+1}$ by

$$\hat{X}_{N+1} = m(X_{t-1}, \ldots, X_{t-p})$$
AR(1) with changepoint at 151: mean shift of +2
AR(1) with changepoint at 151: \textit{start of trend} with slope 0.05
AR(1) with changepoint at 151: change of dependence from $a = 0.9$ to $a = 0.5$
Spectral Analysis of Stationary Processes

Here only **stationary time series**: \( \ldots, X_{-1}, X_0, X_1, X_2, \ldots \)

Analogously: **spatial processes**
(stationary \( \iff \) spatially homogeneous)

Extension to nonstationary data: **wavelets**

Goal:

Detect and interpret **periodic features** in the data structure

Use those periodic features to differ, e.g. by **hypothesis tests**, between or to **classify** different kinds of objects
Building blocks of spectral decomposition: sines and cosines
period/wavelength $\ell$, frequency $\nu = \frac{1}{\ell}$, angular frequency $\omega = 2\pi\nu$, phase (shift) $\phi$, amplitude $A$
Deterministic signal: How to reconstruct the cosine-components?

\[ \text{Fourier analysis} \]
Random signals and spectral representation

$X_1, \ldots, X_N$ sample from a random time series $\ldots, X_{-1}, X_0, X_1, X_2, \ldots$

Assumption (weak stationarity of signal):

$$\text{var } X_t < \infty \quad \text{and} \quad \text{cov}(X_t, X_s) = r_{t-s}$$

i.e. linear dependence between $X_t, X_s$ depends on time span $t - s$ between the observations only.

Strict stationarity requires that the whole data generating mechanism does not change with time:

$$(X_{t_1+s}, \ldots, X_{t_n+s}) \text{ has same distribution as } (X_{t_1}, \ldots, X_{t_n})$$

for all $n \geq 1, t_1, \ldots, t_n, s$
Terminology:

$r_t$ **autocovariance** at lag $t$

$f(\omega)$ **power spectral density** at (angular) frequency $\omega$

\[
r_t = \text{cov}(X_{t+s}, X_s), \quad f(\omega) = \sum_{t=-\infty}^{\infty} r_t e^{-it\omega}, \quad -\pi \leq \omega \leq \pi
\]

if $\sum |r_t| < \infty$. Remark: For real-valued signals: $f(\omega) = f(-\omega) \geq 0$.

Both are equivalent characterizations of the stationary signal $X_t$

Example: **White noise**

$X_t, X_s$ uncorrelated for $t \neq s \Rightarrow r_t = 0$ for all $t \neq 0$

$\Rightarrow f(\omega) \equiv r_0 = \text{constant for all } \omega$
Spectral representation „theorem“: Every stationary signal can be arbitrarily well approximated by a random trigonometric polynomial if only $m$ is large enough:

$$X_t \approx \sum_{j=-m}^{m} Z_j e^{i\omega_j t} = A_0 + \sum_{j=1}^{m} A_j \cos(\omega_j t + \Phi_j)$$

with random and independent $A_0 \in \mathbb{R}, A_j \geq 0, 0 \leq \Phi_j < 2\pi, j = 1, \ldots, m$ and Fourier frequencies $\omega_j = \frac{2\pi j}{M}, M = 2m + 1$.

Cramér’s Theorem:

$$X_t = \frac{1}{\sqrt{2\pi}} \int_{-\pi}^{\pi} e^{i\omega t} dZ(\omega)$$

$Z(\omega)$ complex-valued stochastic processes with orthogonal increments, $Z(-\pi) = 0$, e.g. Wiener process,

$$\mathbb{E}|Z(\omega + d\omega) - Z(\omega)|^2 \approx f(\omega) d\omega$$
Two special cases:

a) **discrete spectrum**:

\[
X_t = A_0^* + \sum_{\ell=1}^{n} A_{\ell}^* \cos(\omega_{\ell}^* t + \Phi_{\ell}^*)
\]

with fixed, but unknown \( n \) and \( \omega_{\ell}^* \) not necessarily Fourier frequencies \( \Rightarrow \) **spectral lines** at \( \omega_{\ell}^*, \ell = 1, \ldots, n \)

b) **continuous spectrum**: \( X_t \) has a **spectral density** \( f(\omega) \) which is large in frequency bands dominating the cyclic behaviour of the signal

The examples on the slides 4.26-4.27, 4.33-4.36 have a **mixed spectrum**, i.e. the signal is an additive superposition of a periodic signal and another signal (here: white noise) with continuous spectrum.
Fourier frequencies $\omega_k = \frac{2\pi k}{N}$. Assume:

$$X_t = f(t) = \sum_{j=-m}^{m} Z_j e^{i\omega_j t}, \quad t = 1, \ldots, N, \quad N = 2m + 1$$

Discrete Fourier transform (DFT) of data:

$$D_N(\omega_k) = \frac{1}{\sqrt{N}} \sum_{t=1}^{N} X_t e^{-it\omega_k}, \quad k = -m, \ldots, m$$

one-to-one data transformation $X_1, \ldots, X_N \mapsto D_N(\omega_{-m}), \ldots, D_N(\omega_m)$

fast calculation: **FFT algorithm** (fast Fourier transform) and its inverse (MATLAB: `fft` and `ffti`)

$D_N(\omega_k)$ complex, difficult to plot $\leadsto$ periodogram

$$I_N(\omega_k) = |D_N(\omega_k)|^2, \quad k = -m, \ldots, m, \quad I_N(\omega_k) = |Z_k|^2 \quad \text{for} \quad X_t = f(t)$$
\[ I_N(\omega_k) = \frac{1}{N} \left| \sum_{t=1}^{N} X_t e^{-it\omega_k} \right|^2, \quad k = -m, \ldots, m \]

\(X_t\) real-valued \(\sim I_N(\omega_k) = I_N(-\omega_k)\) symmetric.

**Periodogram** for discrete and continuous spectrum, if \(N \to \infty\):

a) \(I_N(\omega_j) \approx \frac{1}{2} A_{\ell}^2 N\) if \(|\omega_j - \omega^*_\ell| \approx 0\), \(I_N(\omega_j) \approx 0\) else.

b) \(E I_N(\omega_j) \approx f(\omega_j)\), but \(\text{var } I_N(\omega_j) \to \text{constant} > 0\)

In case b), the periodogram will not converge for \(N \to \infty\), but it even becomes more and more irregular visually. For estimating the power spectral density, the periodogram has to be smoothed:

\[ \hat{f}(\omega_k) = \sum_{\ell=-L}^{L} w_\ell I_N(\omega_k + \ell), \quad w_{-\ell} = w_\ell, \quad \sum_{\ell=-L}^{L} w_\ell = 1. \]
Überlagerung von 4 Schwingungen mit Frequenzen 1/8, 1/16, 1/32 und 1/64
$\cos(\Omega k/30) + 0.5 \cos(\Omega k/64) - 0.5 \cos(\Omega k/69) + 0.3 \sin(\Omega k/13)$

Periodogramm, Überlagerung von 4 Schwingungen
\[ x(t) = 2 \cos(\omega t), \, t = 1, \ldots, N, \, N = 129, \, \omega = \frac{2\pi}{20} \] - together with continuous signal
Periodogram of previous signal - having maximum $(\hat{A}^2 N/2)$ at $\omega$ with amplitude $A=2$. 
signal $X_t = 2 \cos(t)$ with $N=129$, $\omega = (2\pi \times 0.5)/N$ = not a Fourier frequency
$X_i \sim \cos(\theta_i) + Z_i, \, \theta_i = (2\pi \cdot i) / N, \, Z_i \text{ i.i.d. normal with mean 0, variance 4, N=129}$
Data example: **Wolfer’s sunspot numbers** (since 1700)

The annual data represent an index for the number of large sunspots and clusters of smaller sunspots related to the activity of the sun.

The data here are taken from MATLAB.

a.k.a. **Wolf number**

(A. Wolfer has been the student of R. Wolf, both active in Zürich in the 19th century)
periodogram of raw sunspot numbers $X_t$
periodogram $I^0_N(\omega_k)$ of centered sunspot numbers $X^0_t = X_t - \overline{X}_N$
averaged periodogram \( \hat{f}(\omega_k) = \frac{1}{3}(I_N^0(\omega_{k-1}) + I_N^0(\omega_k) + I_N^0(\omega_{k+1})) \)
averaged periodogram with weights $\frac{1}{15}(1, 2, 3, 3, 3, 2, 1)$
periodogram from last slide smoothed with weights (1,2,3,3,2,1)/15
Data example: \textbf{EEG}-like data

The following sample has been generated from a time series model which has been fitted to 10 seconds of real EEG data of a young child which has been in an idle state of mind during EEG recording.

The simulated data correspond to approximately 1.5 sec of brain activity.

For the original data and the fitted model (an autoregression of order 8), compare:

$\alpha$ peak, $\approx 10$ Hz

$\beta$ peak, $\approx 20$ Hz