Stationary Stochastic Processes

Lecture 1 - Stationary processes and spectral representations

Mattias Villani 👨

Department of Statistics
Stockholm University

Department of Computer and Information Science
Linköping University

mattiasvillani.com   @matvil   mattiasvillani

Lecture overview

- **Motivation**: flexible statistical models as stochastic processes.
- Decomposing a stochastic process into **random harmonics**
- **Ergodicity**
- **Spectral representation**
- Literature:
  - Lindgren, Rootzén and Sandsten (2014). *Stationary Stochastic Processes for Scientist and Engineers*, CRC Press. (**LRS**)
  - Lindgren (2014). *Stationary Stochastic Processes*, CRC Press.

Stochastic processes

- **Stochastic process**: an indexed family of random variables

$$\{X(t), t \in T\}$$

  - T can be an interval of $\mathbb{R}$ (**continuous time**/parameter)
  - T can be a discrete set $T = \{1, 2, 3, \ldots\}$ (**discrete time**)
  - $t \in T$ can be multi-dimensional.
  - **Spatial process** (**random field**) $X(\boldsymbol{u})$ with $\boldsymbol{u} = (u_1, u_2)$ containing longitude and latitudes. **Images**.
  - **Spatiotemporal process** $X(t, \boldsymbol{u})$.
- **Sample space** $\omega \in \Omega$. Stochastic process: $X(t, \omega)$.
- **Realization**, **sample path** for a given $\omega \in \Omega$: $t \mapsto X(t, \omega)$.
- **Ensemble**: collection of all possible ($\omega \in \Omega$) sample paths.

Stochastic processes for flexible logistic regression

- Stochastic processes = Random functions, $f(t)$.
- Modern flexible/semiparametric statistical models.
- (Linear) logistic regression

$$\Pr(y = 1|x) = \frac{\exp(\beta_0 + \beta_1 x)}{1 + \exp(\beta_0 + \beta_1 x)}$$

- Log-odds is linear in x. Linear decision boundaries.
- Gaussian process logistic regression

$$\Pr(y = 1|x) = \frac{\exp(f(x))}{1 + \exp(f(x))}$$

where $f(x)$ is a random function, i.e. a stochastic process.

Smoothness of a stochastic process

- **Properties of $f(x)$** are crucial for the statistical model.
- We want **flexibility**, but not crazy wiggly stuff.
- How **smooth** are realizations of $f(x)$?
- Three (related) **smoothness characterizations**:
  - How fast does $\text{Corr}\left[X(t), X(t+\tau)\right]$ **decay** with distance τ?
  - Continuity and **differentiability of sample paths** $X(t)$? Quadratic mean convergence.
  - **Decompose**$\{X(t)\}$ as **sum of cosines** with random amplitudes and phases.

Moments

- The LRS book is mainly about second order properties:
  - Variances and (auto)correlations.
  - Spectral decomposition is a variance decomposition.
  - Gaussian processes.
- Mean function

$$m(t) = \mathbb{E}\left(X(t)\right)$$

- Variance function

$$v(t) = \mathbb{V}\left(X(t)\right)$$

- Covariance function

$$r(s,t) = \mathbb{C}\left(X(s), X(t)\right)$$

- Covariance of sums

$$\mathbb{C}\left(\sum_{i=1}^{k} a_i X_i, \sum_{j=1}^{l} b_j Y_j\right) = \sum_{i=1}^{k}\sum_{j=1}^{l} a_i b_j \mathbb{C}(X_i, Y_j)$$

Dependence

- Stochastic process: **dependence** between variables in family.
- Three useful **principles of dependence**:
  - **Markov Principle**:

$$p\left(X(t) \middle| \{X(\tau), \tau \leq s\}\right) = p\left(X(t) \middle| X(s)\right)$$

  - **Martingale Principle**:

$$\mathbb{E}\left(X(t) \middle| \{X(\tau), \tau \leq s\}\right) = X(s) \text{ for } s \leq t$$

  - **Stationarity Principle**: For any choice of time periods $t_1, \ldots, t_n$ and lag τ such that $t_i + \tau \in T$

$$p\left(X(t_1), \ldots, X(t_n)\right) = p\left(X(t_1 + \tau), \ldots, X(t_n + \tau)\right)$$

- **Second order stationarity**:
  - mean $m(t)$ and variance $v(t)$ constant over time, t.
  - covariance $r(s, t)$ depend only on time lag $\tau = |t - s|$.

$$r(\tau) = r(t, t + \tau)$$

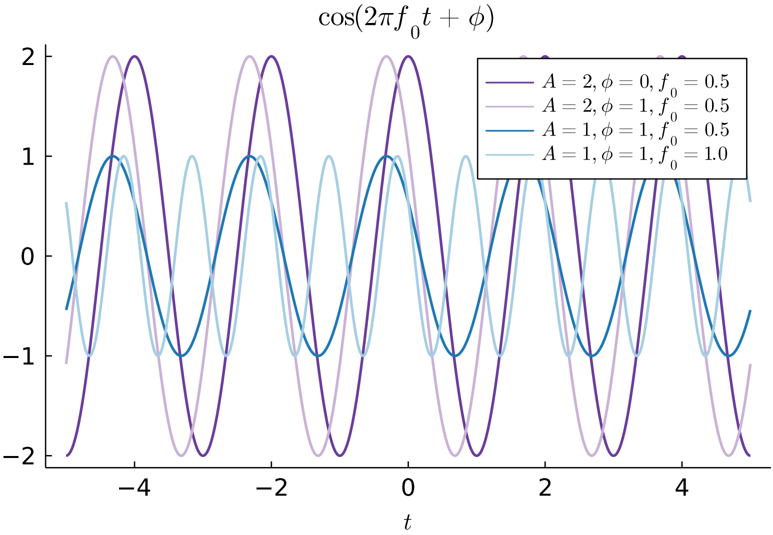
Single harmonic with random amplitude and phase

- Random **phase** and **amplitude** process at fixed **frequency** f_0

$$X(t) = A\cos(2\pi f_0 t + \phi)$$

- Periodic function with
  - **period** $1/f_0$ and **frequency** f_0. Hertz.
  - **amplitude** $A > 0$ independent of the
  - **phase** $\phi \sim \mathrm{Uniform}(0, 2\pi)$.
- Strictly stationary because of the uniformly distributed phase.
- **Angular frequency**: $\omega_0 = 2\pi f_0$. $X(t) = A\cos(\omega_0 t + \phi)$.
- Moments:
  - $\mathbb{E}[X(t)] = 0$
  - $\mathbb{V}[X(t)] = \frac{1}{2}E[A^2] = \sigma^2$
  - $r(\tau) = \sigma^2\cos(2\pi f_0 \tau)$

Single harmonic with random amplitude and phase



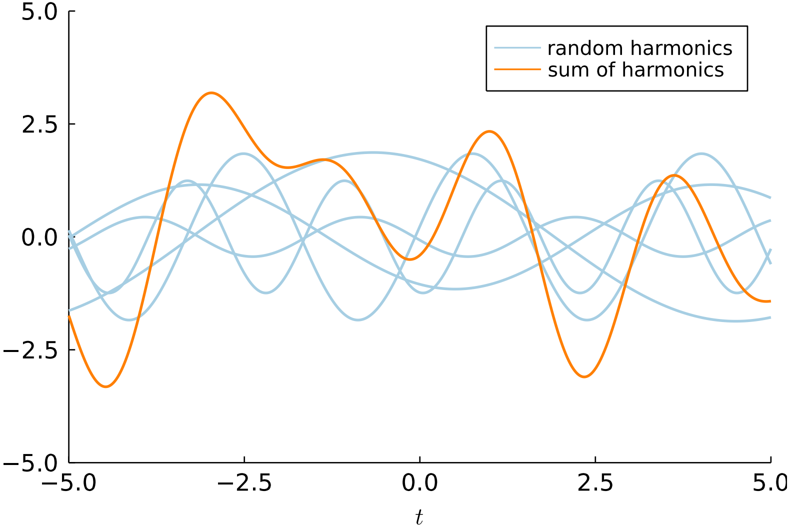
Multiple random harmonics

- **Sum of harmonics** at fixed frequencies $f_1, \ldots, f_n$

$$X(t) = A_0 + \sum_{k=1}^{n} A_k \cos(2\pi f_k t + \phi_k)$$

- **Amplitudes** $A_k > 0$ indep of **phases** $\phi_k \sim \mathrm{Uniform}(0, 2\pi)$.
- Strictly stationary. Moments:
  - $\mathbb{E}[X(t)] = 0$
  - $\mathbb{V}[X(t)] = \sum_{k=0}^{n} \sigma_k^2$, where $\sigma_k^2 = \frac{1}{2} E[A_k^2]$. **ANOVA**.
  - $r(\tau) = \sigma_0^2 + \sum_{k=1}^{n} \sigma_k^2 \cos(2\pi f_k \tau)$
- Infinite number of harmonics - **stochastic convergence**.
- Harmonics are deep: **Cramér's representation** of any stationary process.

Multiple random harmonics



Estimating the mean of a stationary process

- **Ensemble average** $m = \mathbb{E}[x(t)]$ for all t.
- Estimated by **time average**:

$$\hat{m}_n = \frac{1}{n}\sum_{t=1}^{n} X_t$$

- **Unbiased** $\mathbb{E}[\hat{m}_n] = m$.
- **Asymptotic variance** of $\hat{m}_n$ (if $\sum_{\tau=0}^{\infty} r(\tau)$ is convergent)

$$\mathbb{V}[\hat{m}_n] \approx \frac{1}{n}\sum_{\tau=-\infty}^{\infty} r(\tau) = \frac{r(0)}{n} + \frac{2}{n}\sum_{\tau=1}^{\infty} r(\tau)$$

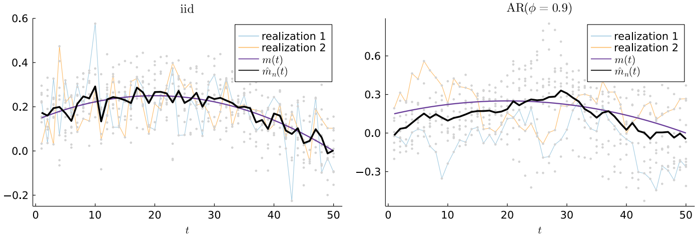
- Compare with iid observations $[r(\tau) = 0 \text{ for } \tau \neq 0]$:

$$\mathbb{V}[\hat{m}_n] \approx \frac{r(0)}{n}$$

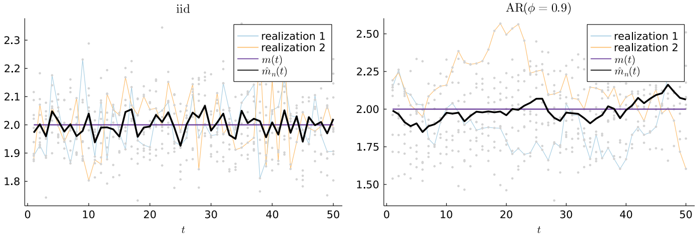
- The estimator $\hat{m}_n$ is therefore **consistent**
  - **in mean square** $\mathbb{E}[(\hat{m}_n - m)^2] \to 0$ as $n \to \infty$ [$\mathrm{MSE} = \mathrm{Bias}^2 + \mathrm{Var}$].
  - **in probability** $\mathrm{P}(|\hat{m}_n - m| > \varepsilon) \to 0$ as $n \to \infty$ [Markov's inequality]

Ergodicity

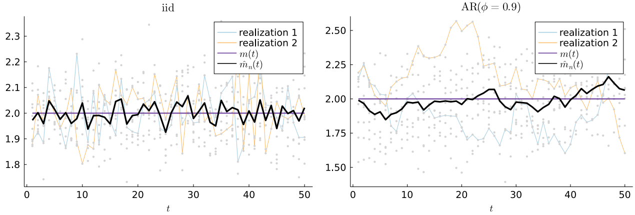
- **Ensemble average** $m(t) = \mathbb{E}[x(t)]$.



- Stationary process, $m(t) = m$



Ergodicity



- A process is **linearly ergodic** when the **time average consistently estimates** the **ensemble average** m

$$\hat{m}_n = \frac{1}{n}\sum_{t=1}^{n} X_t \xrightarrow{p} m \text{ as } n \to \infty$$

- **Sufficient** for ergodicity: $\sum_{\tau=0}^{\infty} r(\tau) < \infty \implies \hat{m}_n \xrightarrow{p} m$.
- **Ergodic process**: time average can consistently estimate any ensemble average $\mathbb{E}[g(X_{t_1}, X_{t_2}, \ldots, X_{t_p})]$. Histograms.
- Stationary Gaussian: ergodic if $\frac{1}{n}\sum_{\tau=1}^{n} r^2(\tau) \to 0$ as $n \to \infty$.

Spectral density

Theorem (Spectral representation of $r(\tau)$)

*If the covariance function $r(\tau)$ of a stationary process $\{X(t), t \in \mathbb{R}\}$ is **continuous**, there exists a positive, **symmetric** and integrable function $R(f)$ such that*

$$r(\tau) = \int_{-\infty}^{\infty} e^{i2\pi f\tau} R(f) df$$

- Converse holds. Determines a valid covariance function.
- **Complex exponentials**

$$e^{ix} = \cos(x) + i \cdot \sin(x)$$

- The autocovariance function $r(\tau)$ is indeed real:

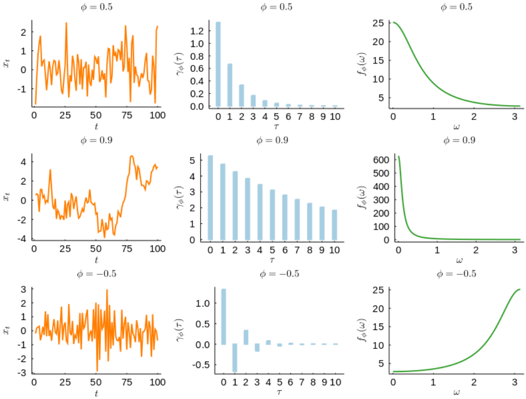
$$\int_{-\infty}^{\infty} e^{i2\pi f\tau} R(f) df = \int_{-\infty}^{\infty} \cos(2\pi f\tau) R(f) df + i \underbrace{\int_{-\infty}^{\infty} \sin(2\pi f\tau) R(f) df}_{=0} = 2\int_{0}^{\infty} \cos(2\pi f\tau) R(f) df$$

- $r(\tau)$ continuous at $\tau = 0$, then $r(\tau)$ continuous for all τ.

AR(1) process example

- AR(1) process

$$x_t = \mu + \phi(x_{t-1} - \mu) + \varepsilon_t, \quad \varepsilon_t \overset{\text{iid}}{\sim} N(0, \sigma_\varepsilon^2)$$



Spectral density as a variance decomposition

- The **variance** of the process

$$\mathbb{V}[X(t)] = r(0) = \int_{-\infty}^{\infty} R(f)df$$

- **Variance contribution** from the **frequency band** $a \leq f \leq b$

$$\int_{-b}^{-a} R(f)df + \int_{a}^{b} R(f)df = 2\int_{a}^{b} R(f)df$$

- $R(f)$ represents a frequency with unit $[\mathrm{timeunit}]^{-1}$.
- Time in seconds: Hz. $f = 2$ Hz, two full cycles per second.
- High/Low frequency.
- **Change of time scale** $X_c(t) = X(ct)$ (e.g. hours to sec).
  - $r_c(\tau) = r(c\tau)$
  - $R_c(f) = c^{-1}R(f/c)$ [*density*, so change-of-variable formula.]

Spectral density from autocovariance function

- The **spectral density** $R(f)$ can contain delta functions

$$\int g(f)\delta_{f_0}(f)df = g(f_0)$$

- **Continuous spectrum**: $R(f)$ continuous except jumps.
- **Discrete spectrum**: $R(f) = \sum_k b_k \delta_{f_k}(f)$.
- If $\int_{-\infty}^{\infty} |r(\tau)| \, d\tau < \infty$ the spectrum is continuous with density

$$R(f) = \int_{-\infty}^{\infty} e^{-i2\pi f\tau} r(\tau) d\tau$$

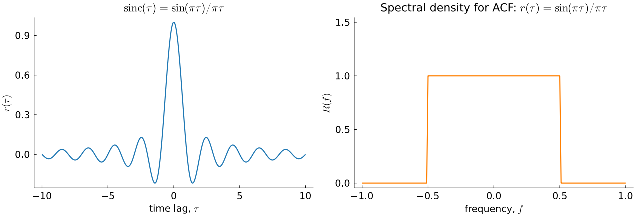
i.e. $r(\tau)$ **and** $R(f)$ **form a Fourier pair**

$$R = \mathcal{F}(r) \text{ and } r = \mathcal{F}^{-1}(R)$$

Example - sinc function

- Autocovariance function **sinc**

$$r(\tau) = \frac{\sin(\pi\tau)}{\pi\tau}$$



- Proof:

$$r(\tau) = \int_{-\infty}^{\infty} e^{i2\pi f\tau} R(f) df = \int_{-1/2}^{1/2} e^{i2\pi f\tau} df = \frac{e^{i\pi\tau} - e^{-i\pi\tau}}{i2\pi\tau} = \frac{\sin(\pi\tau)}{\pi\tau}$$

since for any complex number $z = a + i \cdot b$ with conjugate $\bar{z} = a - i \cdot b$ we have

$$\mathrm{Re}(z) = a = \frac{z + \bar{z}}{2} \text{ and } \mathrm{Im}(z) = b = \frac{z - \bar{z}}{2i}$$

and

$$z = e^{i\pi\tau} = \cos(\pi\tau) + i \cdot \sin(\pi\tau) \text{ and } \bar{z} = e^{-i\pi\tau}.$$

Example - squared exponential

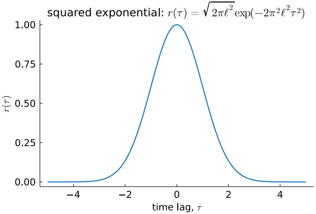
- Squared exponential covariance kernel

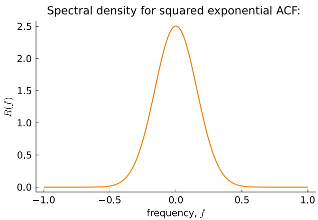
$$r(\tau) = \exp\left(-\frac{\tau^2}{2\ell^2}\right)$$

- Check: $\int_{-\infty}^{\infty} |r(\tau)|\, d\tau = \int_{-\infty}^{\infty} \exp\left(-\frac{\tau^2}{2\ell^2}\right) d\tau = \sqrt{2\pi\ell^2} < \infty$
- Spectral density

$$R(f) = \int_{-\infty}^{\infty} e^{-i2\pi f\tau} r(\tau) d\tau = \int_{-\infty}^{\infty} e^{-[i2\pi f\tau + \tau^2/(2\ell^2)]} r(\tau) d\tau = \sqrt{2\pi\ell^2}\exp(-2\pi^2\ell^2 f^2)$$

by completing the square in the exponent $\Longrightarrow$ Gaussian in τ.





Example - Matérn

- **Matérn** with length scale $\ell > 0$ and degrees of freedom $\nu > 0$

$$r(\tau) = \frac{2^{1-\nu}}{\Gamma(\nu)} \left(\frac{\sqrt{2\nu}\, |\tau|}{\ell} \right)^{\nu} K_{\nu} \left(\frac{\sqrt{2\nu}\, |\tau|}{\ell} \right)$$

- $K_{\nu}(x)$ is modified Bessel function of the first kind of order ν.
- **Spectral density is student-t** with 2ν degrees of freedom.
- As $\nu \to \infty$, Matérn approaches SE.
- When $\nu = p + 1/2$ for integer p: $r(\tau)$ is the product of polynomial and exponential.
- Example: $\nu = 3/2$

$$r(\tau) = \left(1 + \frac{\sqrt{3}\, |\tau|}{\ell} \right) \exp \left(-\frac{\sqrt{3}\, |\tau|}{\ell} \right)$$

The Ornstein-Uhlenback process

- $\nu = 1/2 \Rightarrow$ Exponential kernel = **Ornstein-Uhlenbeck**.
- $R(f)$ is a **Cauchy**. Heavy tails, much mass on **high freq**.
- Ornstein-Uhlenbeck process

$$r(\tau) = \sigma^2 e^{-\alpha|\tau|}$$
$$R(f) = \sigma^2 \frac{2\alpha}{\alpha^2 + (2\pi f)^2}$$

In angular frequencies $\omega = 2\pi f$ (change-of-variables)

$$\tilde{R}(\omega) = \frac{1}{2\pi} R\left(\frac{\omega}{2\pi}\right) = \sigma^2 \frac{\alpha}{\pi\left(\alpha^2 + \omega^2\right)} = \sigma^2 \frac{1}{\pi\alpha\left(1 + \left(\frac{\omega}{\alpha}\right)^2\right)}$$

σ^2 times a Cauchy density with scale parameter $\alpha = 1/\ell$.

Example - Matérn