Neuroinformatics - Linear Time Series Aanlysis (1) -

2022.10.18

Prof: Boreom Lee

- Major difference btw time series and data sets discussed before (from a purely statistical point of view): temporally consecutive measurements usually highly dependent
 - Violating the assumption of identically and independently distributed observations
 - iid assumption: most of conventional statistical inference relies on
- Independency assumption
 - Not only violated in time series but also in a number of other common test situations
 - The class of mixed models
 - Combine fixed and random effects \rightarrow suited for both nested and longitudinal (i.e., time series) data
 - The assumption of independent observations given up
 - In the context of neuroscience
 - Dependent and nested data frequently occur (other than time series)
 - Recordings from multiple neurons, nested within animals, nested within treatment groups \rightarrow introduce dependency

- Mixed models
 - Much more flexible (parameterized) forms for the involved covariance matrices \rightarrow account for dependency
 - Regression model
 - A full covariance matrix for the error terms (instead of the scalar forms) \rightarrow captures some of the correlations among observations
 - ML estimator for parameter $\boldsymbol{\beta}: \widehat{\boldsymbol{\beta}} = (\mathbf{X}^T \boldsymbol{\Sigma}^{-1} \mathbf{X})^{-1} \mathbf{X}^T \boldsymbol{\Sigma}^{-1} \mathbf{y}$
 - Σ a full covariance structure; under the multivariate normal model
 - Dependency → the likelihood does not factor into the individual observations anymore
 - Still easily obtained with the observations jointly multivariate normal
 - Estimation of the covariance matrices: generally less straightforward
 - In general, no analytical solution for mixed models \rightarrow numerical techniques

- Time series supposedly generated by some underlying <u>dynamical</u> <u>system (a more general, scientific point of view)</u>
 - Recovered from the data
 - Encapsulates the essence of formal understanding of the underlying process
 - Assumption: this dynamical (time series) model captures all the dependencies among consecutive data points → the residuals from this model independent again & hence conventional asymptotic test statistics more ore less directly revoked
 - The simplest class of such time series models: linear
 - (sets of) linear difference or differential equations
 - Pretty much the same mathematical layout as conventional multiple or multivariate regression models
 - Output variables regressed on time-lagged versions of their own (instead of on a different (independent) set of observations) → catch the correlations among temporally consecutive measurements

- In many domain of neuroscience
 - Time series models: the most important class of statistical models
 - fMRI recordings, optical imaging, multiple-/single-unit recordings, EEG or MEG signals: inherently as time series generated by a dynamical system (the brain) with stronger or weaker temporal dependencies among consecutive measurements (depending on the type of signals recorded)
- In behavioral data
 - Time series frequently occur
 - Ex:
 - A learning process that develops across trials
 - The impact of cyclic (e.g., hormonal) variations on behavioral performance

• Autocorrelation

- m/c tools for descriptive characterization of (the linear properties of) time series: autocorrelation function & the power spectrum
- A univariate time series $\{x_t\}$
 - Variable x sampled at discrete times t (cf. x(t) in case of a continuous time function)
 - Auto-covariance (acov) function: $acov(x_t, x_{t+\Delta t}) \equiv \gamma(x_t, x_{t+\Delta t}) \coloneqq E[(x_t \mu_t)(x_{t+\Delta t} \mu_{t+\Delta t})]$
 - The conventional covariance applied to time-lagged versions of x_t
 - μ_t and $\mu_{t+\Delta t}$: the means at times t and $t + \Delta t$, respectively
 - Autocorrelation (acorr): $\operatorname{acorr}(x_t, x_{t+\Delta t}) \equiv \rho(x_t, x_{t+\Delta t}) \coloneqq \frac{\operatorname{acov}(x_t, x_{t+\Delta t})}{\sqrt{\operatorname{var}(x_t)\operatorname{var}(x_{t+\Delta t})}} = \frac{\gamma(x_t, x_{t+\Delta t})}{\sigma_t \sigma_{t+\Delta t}}$
 - Dividing the auto-covariance by the product of s.d.
 - An <u>ensemble of time series</u> drawn from the same underlying process \rightarrow the expectancies and (co-)variances at specific times *t*

- Autocorrelation
 - A single observed time series $\{x_t\}$ (t = 1, ..., T)
 - Assumptions of *stationarity* and *ergodicity* $\rightarrow \hat{\gamma}(x_t, x_{t+\Delta t})$ and $\hat{\rho}(x_t, x_{t+\Delta t})$
 - Estimates across samples replaced by estimates across time
 - Mean & variance the same across all $t \rightarrow \mu_t = \mu_{t+\Delta t} = \mu$ and $\sigma_t^2 = \sigma_{t+\Delta t}^2 = \sigma^2$
 - μ and σ^2 replaced by sample estimates \bar{x} and s_x^2 respectively
 - acorr and acov functions depend on time lag Δt only
 - $\gamma(x_t, x_{t+\Delta t}) = \gamma(\Delta t)$ and $\rho(x_t, x_{t+\Delta t}) = \rho(\Delta t) = \gamma(\Delta t)/\gamma(0)$
 - Any time lag $\Delta t \neq 0$ cut off Δt values at one end or the other of the empirical time series sample
 - The product of s.d. in the denominator computed across the first $1, \ldots, T \Delta t$ and the last $\Delta t + 1, \ldots, T$ values \rightarrow usually ignored & irrelevant for sufficiently long time series

• Likewise for the <u>means</u>

- Autocorrelation
 - The acorr function $(= \rho(x_t, x_{t+\Delta t}))$
 - The dependencies among temporally neighboring values along a time series & how quickly with time these dependencies die out (i.e., the acorr drops to zero as Δt increases)
 - An important tool to characterize some of the temporal structure in a time series
 - Bounded within in [-1, +1] (just as the standard Pearson correlation, by definition)
 - Symmetrical: $\rho(x_t, x_{t+\Delta t}) = \rho(x_{t+\Delta t}, x_t)$ or $\rho(\Delta t) = \rho(-\Delta t)$ in the stationary case
 - iid random numbers $\{x_t\}$ and some basic conditions \rightarrow asymptotically $\hat{\rho}(\Delta t) \sim N(-1/T, 1/T)$ (or N(0, 1/T) for large T?)
 - o Used to establish confidence bounds or check for significance of the autocorrelations
 - Application on different types of neural time series (Fig. 7.1)
 - Series of interspike intervals (obtained from single-unit recordings) & fMRI BOLD signal traces → quite different autocorrelative properties in different types of data
 - Some important properties of the underlying system
 - Oscillations: periodic increases and decreases in the autocorrelation
 - "long-memory" properties: a very slow decay of the autocorrelation



Fig. 7.1 Illustration of sample autocorrelation functions (left), power spectra (center), and return plots (right) on interspike interval (ISI) series (top row; from rat prefrontal cortex) and BOLD signals (bottom row) from human fMRI recordings. For the spike data, the power spectrum was computed on the original binned (at 10 ms) spike trains, not the ISI series. Spike train data recorded by Christopher Lapish, Indiana University Purdue University Indianapolis (see also Lapish et al. 2008; Balaguer-Ballester et al. 2011). Human fMRI recordings obtained by Florian Ba[°]hner, Central Institute for Mental Health Mannheim (Ba[°]hner et al. 2015). MATL7_1

- Power Spectrum
 - Wiener-Khinchin theorem
 - Weak-sense stationary & certain conditions \rightarrow a 1:1 relationship btw the acorr function and the power spectrum (or spectral density) of a time series
 - Decomposition into a weighted sum of harmonic oscillations (i.e., pure sine and cosine functions)
 - The frequency domain representation of a periodic function x(t): $x(t) = x(t + \Delta t)$ for some fixed Δt and all t
 - Approximation by a series of frequencies (Fourier series):

$$x(t) \approx \frac{a_0}{2} + \sum_{k=1}^{\infty} \left[a_k \cos\left(\omega kt\right) + b_k \sin\left(\omega kt\right) \right] = \sum_{k=-\infty}^{\infty} c_k e^{i\omega kt}$$

- $\omega(=2\pi f)$ = angular frequency; $f(=1/\Delta t)$ = the oscillation frequency in Hz (Δt =oscillation period)
- Dirichlet's condition: Fourier series known to converge to x(t)

- Power Spectrum
 - Plots the coefficients (a²_k + b²_k)/2 against frequency ω or f
 → the energy contribution of each frequency f to the "total energy" in the signal
 - The 1st coefficient $a_0/2$: the mean of x(t) across one oscillation period Δt
 - The power $(a_k^2 + b_k^2)/2$ of the *k*th frequency component = the amount of <u>variance</u> in the signal explained by that frequency
 - An estimate of these functions obtained by fast Fourier transform (FFT) algorithm
 - Fourier transformation of x(t) only captures its linear time series properties (as fully specified through the acorr function)

- Power Spectrum
 - Neuroscience
 - The frequency domain representation of neurophysiological signals like LFP or EEG of uttermost importance for characterizing oscillatory neural processes in different frequency bands
 - Ex: theta ($\sim 3 7$ Hz) or gamma ($\sim 30 80$ Hz) band
 - Oscillations: a pivotal role in neural information processing
 - Means for synchronizing the activity and information transfer btw distant brain areas
 - A carrier signal for phase codes of external events or internal representations
 - Ex1: stimulus-specific increases in the power within the gamma or theta frequency band
 - In response to external stimuli (e.g., in the bee olfactory system in response to biologically relevant odors)
 - In conjunction with the internal active maintenance of memory items (e.g., during the delay phase of a working memory task)

- Power Spectrum
 - Neuroscience
 - Ex2: neurons in the hippocampus coding for specific places in an environment align their spiking activity with a specific phase of the hippocampal theta rhythm
 - While the animal moves through the neuron's preferred place field thus encoding environmental information in the relative phase (a phase code) w.r.t an underlying oscillation
 - Ex3: neurons in visual cortex
 - Encode and maintain information about visual patterns in working memory by aligning their spike phase with an underlying theta oscillation during the delay period
 - In a stimulus-specific manner with the phase relationship breaking down for items not preferred by the recorded cell
 - Ex4: the hippocampus and prefrontal cortex phase-lock during working memory tasks
 - During the choice epochs where the animal chooses the response in a twoarm maze based on previous choices or stimuli
 - Oscillations help to organize the information transfer among areas

- White Noise
 - The simplest form of a time series process $\{x_t\}$: a pure random process with <u>zero mean</u> and <u>fixed variance</u> but <u>no temporal</u> <u>correlations</u> at all

•
$$\operatorname{E}[x_t] = 0$$
 for all t

•
$$E[x_t x_{t'}] = \begin{cases} \sigma^2 & \text{for } t = t' \\ 0 & \text{otherwise} \end{cases}$$

- → Called *white noise* processes W(0, σ^2) (not necessarily Gaussian)
 - No distinguished frequency in the frequency domain representation \rightarrow completely flat power spectrum
 - No specific "color" \rightarrow a uniform mixture of all possible colors, giving white
 - The unique setup of <u>autocorrelation coefficients at different time lags</u> $\Delta t \neq 0$ for the oscillatory properties of the time series (in accordance with the Wiener-Khinchin theorem)
 - If they are all zero \rightarrow no (linear) oscillations

- White Noise
 - Most of the statistical inference on time series
 - Assumption: the residuals from a model form a white noise sequence
 - Wold decomposition theorem:
 - Each stationary discrete-time process $x_t = z_t + \eta_t$ split into a systemic (purely deterministic) part z_t and an uncorrelated purely stochastic process $\eta_t = \sum_{k=0}^{\infty} b_k \varepsilon_{t-k}$ with $\varepsilon_t \sim W(0, \sigma^2)$
 - *Gaussian* white noise
 - $\varepsilon_t \sim N(0, \sigma^2)$, $E[\varepsilon_t \varepsilon_{t'}] = 0$ for all $t \neq t'$
 - Explicitly check this assumption by
 - Comparing the empirical \mathcal{E}_t distribution to a Gaussian using common Kolmogorov-Smirnov or χ^2 -based test statistics
 - Evaluating whether any of the autocorrelations significantly deviates from 0 [or -1/T] for $\Delta t \neq 0$
 - Moments up to 2nd order completely specify a white noise process in general and the Gaussian in particular
 - Alternatively evaluate whether the power spectrum conforms to a uniform distribution

- White Noise
 - Or employ <u>more general tests for randomness</u> in the time series by checking for any sort of sequential dependencies
 - Discretize (bin) \mathcal{E}_t , chart the transition frequencies among different bins, and compare them to the expected base rates under independence using χ^2 tables
 - Also examine the binned series for unusually long runs of specific bin-values, based on the binomial or multinomial distribution
 - Another possibility:
 - To chart the intervals btw successive maxima (or minima) of a real-valued series
 - The length of an interval I_i btw any two successive maxima independent of the length I_{i-1} of the previous interval for a pure random process: $p(I_i|I_{i-1}) = p(I_i)$
 - Plotting all pairs (I_i, I_{i-1}) (called "first-return plot") & inspecting the graph for systematic trends in the distribution \rightarrow a visual idea of whether this holds
 - Durstewitz and Gabriel (2007) used this to examine
 - Whether single neuron ISI series recorded under different pharmacological conditions exhibit any evidence of deterministic structure or
 - Whether they are indeed largely random as suggested by the common Poisson assumption of neural spiking statistics
 - (more formally) A significant regression coefficient relating I_i to $I_{i-1} \rightarrow$ doubt on the assumption of independence
 - A number of different informal checks or formal tests in this context

- Stationarity and Ergodicity
 - Stationarity
 - A fundamental concept (for model estimation and inference) in time series analysis
 - Roughly means that properties of the time series do not change across time
 - Weak sense versus strong stationarity
 - Weak stationarity:

$$E[x_t] = \mu = \text{const.}, \operatorname{acov}(x_t, x_{t+\Delta t}) = \operatorname{acov}(\Delta t) \forall t, \Delta t$$

(a)

(b)

- Mean: constant & independent of time
- acov (acorr) function: a function of time lag only, but does not change with *t* either
- Stronger form of stationarity: joint distribution F of $\{x_t\}$ time-invariant \rightarrow
- $F(\{x_t | t_0 \le t < t_1\}) = F(\{x_t | t_0 + \Delta t \le t < t_1 + \Delta t\}) \text{ for all } t_0, t_1, \text{ and } \Delta t$
 - All higher-order moments of the $\{x_t\}$ distribution independent of t as well (equivalent to weak stationarity (a) for a purely Gaussian process)

- Stationarity and Ergodicity
 - Stationarity
 - Assuming that we have access to a large sample of time series $\{x_t\}^{(i)}$ generated by the same underlying process \rightarrow expectancies taken across all series i at time t to evaluate the 1st moments $\mathbf{E}_i \left[x_t^{(i)} \right] = \lim_{N \to \infty} \sum_{i=1}^N x_t^{(i)} / N$
 - Thus, the definition does not exclude *conditional dependence* in the series: $E[x_t|x_{t-1}] \neq E[x_t|x'_{t-1}]$ for $x_{t-1} \neq x'_{t-1}$
 - → Central for identifying periodic (like harmonic oscillatory) processes as stationary
 - x_t may indeed systematically change across time
 - A time series generated by the harmonic oscillatory process with noise:

$$x_t^{(i)} = \sin(2\pi ft + \varphi_i) + \varepsilon_t, \varepsilon_t \sim N(0, \sigma^2)$$

- φ_i : a r.v. across different realization $x_t^{(i)}$ of the process
- $E[x_t] = \text{const}$ for all t; consecutive values x_t in time conditionally dependent as defined through the sine function (the systematic part)

- Stationarity and Ergodicity
 - Ergodicity
 - Often employed in access only to one realization of time series process
 - Meaning that estimates across different independent realizations of the same process at fixed t replaced by estimates across time

• Mean:
$$\mathbf{E}_{i}\left[x_{t}^{(i)}\right] = \mathbf{E}_{t}\left[x_{i}^{(t)}\right]$$

• Variance:
$$\mathbf{E}_{i}\left[\left(x_{t}^{(i)}-\bar{x}_{t}^{(i)}\right)^{2}\right] = \mathbf{E}_{t}\left[\left(x_{i}^{(t)}-\bar{x}_{i}^{(t)}\right)^{2}\right]$$

> The 1st expectation – taken across sample series i(fixed t); the 2nd across time points t(fixed i)

- Stationarity and Ergodicity
 - Time series data commonly not iid but governed by autocorrelations → not at all evident that such properties hold
 - A <u>sufficient condition for a stationary process to be ergodic in the mean</u>: the autocorrelations die out to zero as the lag increases
 - Autocorrelation still affect the sampling distribution of a time series mean \bar{x} estimated from <u>a finite series of length *T*</u> with its <u>squared standard error</u>:

$$\mathbf{E}\left[\left(\bar{x}_T - \mu\right)^2\right] = \frac{\sigma^2}{T} \left[1 + 2\sum_{\Delta t=1}^{T-1} \left(1 - \frac{\Delta t}{T}\right)\rho(\Delta t)\right]$$

- These autocorrelations \rightarrow an unbiased estimate of the standard error of \bar{x} from a single time series $\{x_t\}$ (unlike the conventional iid case $(=\sigma^2/T)$)
- A reflection of the more general issue: dealing with dependent data in time series → violating a crucial assumption of most conventional statistics

- Stationarity and Ergodicity
 - Another problem
 - What we consider as stationary depends on observation period T
 - Something appearing nonstationary on short-time scales may be stationary on longer scales (*T*: brief compared to the period of an underlying oscillation)
 - Other ways of defining stationarity
 - Stationary if the generating process has time-invariant parameters
 - A process $x_t = f_{\theta}(x_{t-1}) + \varepsilon_t$: the parameter set θ constant
 - Not clear whether such a definition is generally consistent with (a) or (b)
 - Dynamical systems with constant parameters
 - May generate time series which potentially violate the above statistical definition of stationarity
 - o If the dynamical system possesses multiple coexisting attractor states characterized by different distributions among which it may hop d/t perturbations
 - (vice versa) a process with time-varying parameters $\pmb{\theta}$ still stationary according to defs (a) and (b)
 - If the parameters at each point in time are themselves drawn from a stationary distribution

- Stationarity and Ergodicity
 - Experiments
 - Different tests proposed to directly check whether statistical moments of the time series stay within certain confidence limits across time
 - Quiroga-Lombard et al. (2013)
 - A formal test developed based on def. (a)
 - First standardizes and transforms the observed quantities (interspike intervals [ISI]) through the Box-Cox transform to bring their distribution into close agreement with a standard Gaussian
 - Then checks within sliding windows of *k* consecutive variables whether the local average and standardized sum of squares fall outside predefined confidence bounds of the normal and χ^2 -distribution estimated from the full series, respectively (Fig. 7.2)
 - Ignores autocorrelations in the series (which often decay rapidly for ISI series in vivo)
 - Durstewitz and Gabriel (2007)
 - Kolmogorov-Smirnov tests used to check whether distribution across a set of consecutive samples of ISI series significantly deviate from each other



Time (s) Fig. 7.2 Dissecting spike trains into stationary segments. (a) Running estimate of test statistic $T_{m,k}$ comparing the local average to the grand average of the series on sliding windows of ten BoxCox-transformed interspike intervals (ISIs), with [2%, 98%] confidence bands. (b) Running estimate of χ 2-distributed statistic Q_{m,k} evaluating the variation of the local ISIs around the grand average, with [2%, 98%] confidence bands. (c) Original ISI series with resulting set of jointly stationary segments in gray shading. Reprinted from Quiroga-Lombard et al. (2013), Copyright (2013) by The American Physiological Society, with permission

23

• *Stationarity and Ergodicity*

- Non-stationarity also recognized from the estimated coefficients of the model
- One obvious type of non-stationarity = a systematic trend across time
 - Caution: a slow oscillation may look like a trend on shorter-time scales
 - Indicated by having a lot of power in the lowest frequency bands, or equivalently, having very long-term autocorrelations
 - At least three different ways of removing a systematic trend oscillations, or other forms of nonstationarity and undesired confounds

1. A parametric, or nonparametric model fitted to the data (e.g., a linear regression model, a locally linear regression, or a spline model) \rightarrow then work from the residuals after removing the trend, oscillation, or any other systematic component in the data that may spoil the process of interest

2. Designing a filter to take out the slowest frequency bands or any other prominent frequency band \rightarrow remove trends or oscillations in the frequency domain

3. Differencing the time series as often as required

- A nonstationary time series $\{x_t\}$ transformed into a stationary one by considering the series of 1st-order differences $\{x_{t+1} x_t\}$
- Higher-order differencing in some cases, required to make the series stationary
- Transformations of the data to stabilize the variance (e.g., a log-transform) or to move them toward a normal distribution (e.g., Box-Cox transforms)
 - Sometimes also help
 - Used carefully (`` potentially also lead to spurious phenomena (e.g., induce oscillations) or inflate the noise)