ASTR 3890 - Selected Topics: Data Science for Large Astronomical Surveys (Spring 2022)

Time Series Analysis: II

Dr. Nina Hernitschek
March 28, 2022

recap: Time Series

A time series is a sequence of random variables $\{\mathbf{X}_t\}_{t=1,2,\cdots}$.

Thus, a time series is a **series of data points ordered in time**. The time of observations provides a source of additional information to be analyzed.

recap: Time Series

A time series is a sequence of random variables $\{\mathbf{X}_t\}_{t=1,2,\cdots}$.

Thus, a time series is a **series of data points ordered in time**. The time of observations provides a source of additional information to be analyzed.

Since there may be an infinite number of random variables, we consider **multivariate distributions of random vectors**, that is, of finite subsets of the sequence $\{\mathbf{X}_t\}_{t=1,2,\cdots}$.

recap: Time Series

A time series is a sequence of random variables $\{\mathbf{X}_t\}_{t=1,2,\cdots}$.

Thus, a time series is a **series of data points ordered in time**. The time of observations provides a source of additional information to be analyzed.

Since there may be an infinite number of random variables, we consider **multivariate distributions of random vectors**, that is, of finite subsets of the sequence $\{\mathbf{X}_t\}_{t=1,2,\cdots}$.

A **time series model** for the observed data $\{x_t\}$ is defined to be a specification of all of the joint distributions of the random vectors $\mathbf{X} = (X_1, \cdots, X_n)^{\mathrm{T}}$, $n = 1, 2, \cdots$ of which $\{x_t\}$ are possible realizations, that is, at all of these probabilities

$$P(X_1 \leq x_1, \cdots, X_n \leq x_n), \ -\infty < x_1, \cdots, x_n < \infty, \\ n = 1, 2, \cdots .$$

Covariance Matrix

For a N-dimensional random vector $\mathbf{X}$ (such as a time series), one can calculate the **covariance matrix** which gives the covariance between each pair of elements of a given random vector:

If the entries in the column vector $\mathbf{X} = (X_1, X_2, ..., X_n)^{\mathrm{T}}$ are random variables, each with finite variance and expected value, then the covariance matrix $K_{\mathbf{XX}}$ is the matrix whose (i, j) entry is the covariance

$$K_{X_i X_j} = \mathrm{cov}[X_i, X_j] = \mathrm{E}[(X_i - \mathrm{E}[X_i])(X_j - \mathrm{E}[X_j])]$$

where the operator E denotes the **expectation value** of its argument.

Cross-Covariance Matrix

By comparison, the notation for the cross-covariance matrix between two vectors $\mathbf{X}, \mathbf{Y}$ is

$$cov(\mathbf{X}, \mathbf{Y}) = \mathrm{K}_{\mathbf{XY}} = E\left[(\mathbf{X} - E[\mathbf{X}])(\mathbf{Y} - E[\mathbf{Y}])^{\mathrm{T}}\right].$$

In detail, the cross-covariance matrix is calculated as:

variance:

$$var(\mathbf{X}) = \frac{\sum_i^N (X_i - \bar{X})^2}{N-1}$$

covariance:

$$cov(\mathbf{X}, \mathbf{Y}) = \frac{\sum_i^N (X_i - \bar{X}) \cdot (Y_i - \bar{Y})^2}{N-1}$$

cross-covariance matrix:

$$C(\mathbf{X}, \mathbf{Y}) = \begin{bmatrix} var(\mathbf{X}) & cov(\mathbf{X}, \mathbf{Y}) \\ cov(\mathbf{X}, \mathbf{Y}) & var(\mathbf{Y})) \end{bmatrix}$$

$$C(\mathbf{X}, \mathbf{Y}, \mathbf{Z}) = \begin{bmatrix} var(\mathbf{X}) & cov(\mathbf{X}, \mathbf{Y}) & cov(\mathbf{X}, \mathbf{Z}) \\ cov(\mathbf{X}, \mathbf{Y}) & var(\mathbf{Y}) & cov(\mathbf{Y}, \mathbf{Z}) \\ cov(\mathbf{X}, \mathbf{Z}) & cov(\mathbf{Y}, \mathbf{Z}z) & var(\mathbf{Z}) \end{bmatrix}$$

Autocovariance Function

As a time series usually involves a large (infinite in theory) number of random variables, there is a very large number of pairs of variables. So the **autocovariance function** is defined as an extension of the covariance matrix. It is usually denoted by the Greek letter γ and we write

$$\begin{aligned}\gamma(x_{t+\tau}, x_t) &= \mathrm{Cov}(X_{t+\tau}, X_t) \\ &= \frac{1}{n}\sum_{t=1}^{n-|\tau|}(x_{t+|\tau|} - \bar{x}),\ -n < \tau < n\end{aligned}$$

for all indices t and lags τ and where

$$\bar{x} = \frac{1}{n}\sum_{t=1}^{n} x_t.$$

Stationary Time Series

Definition

A time series $\{X_t\}$ is called **weakly stationary** or just stationary if

- $\mathrm{E}(X_t) = \mu_{X_t} = \mu < \infty$, that is, the expectation of X_t is finite and is not depending on t and
- $\gamma(x_{t+\tau}, x_t) = \gamma_\tau$, that is, for each τ, the autocovariance of $X_{t+\tau, X_t}$ is not depending on t.

Stationary Time Series

Definition

A time series $\{X_t\}$ is called **weakly stationary** or just stationary if

- $\mathrm{E}(X_t) = \mu_{X_t} = \mu < \infty$, that is, the expectation of X_t is finite and is not depending on t and
- $\gamma(x_{t+\tau}, x_t) = \gamma_\tau$, that is, for each τ, the autocovariance of $X_{t+\tau, X_t}$ is not depending on t.

A more restrictive definition of stationarity involves all the multivariate distributions of the subsets of time series random variables.

Stationary Time Series

A more restrictive definition of stationarity involves all the multivariate distributions of the subsets of time series random variables.

Definition

A time series $\{X_t\}$ is called **strictly stationary** if the random vectors $(X_{t_1}, \cdots, X_{t_n})^{\mathrm{T}}$ and $(X_{t_1+\tau}, \cdots, X_{t_n+\tau})^{\mathrm{T}}$ have the same joint distribution for all sets of indices $\{t_1, \cdots t_n\}$ and for all integers τ and $n > 0$. It is written as

$$(X_{t_1}, \cdots, X_{t_n})^{\mathrm{T}} \stackrel{\mathrm{d}}{=} (X_{t_1+\tau}, \cdots, X_{t_n+\tau})^{\mathrm{T}},$$

where $\stackrel{\mathrm{d}}{=}$ means *equal in distribution*.

Stationary Time Series

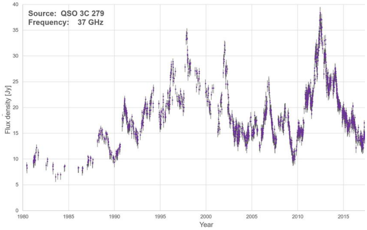
Properties of a Strictly Stationary Time Series:

- The random variables X_t are identically distributed for all t.
- Pairs of random variables $(X_t, X_{t+\tau})^{\mathrm{T}}$ are identically distributed for all t and τ, that is $(X_t, X_{t+\tau})^{\mathrm{T}} \stackrel{d}{=} (X_1, X_{1+\tau})^{\mathrm{T}}$
- The series X_t is a weakly stationary time series if $\mathrm{E}(X_t^2) < \infty$ for all t.
- Weak stationarity does not imply strict stationarity.

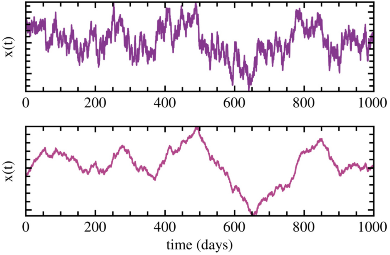
Describing Light Curves as Stochastic Processes

If a system is always variable, but the variability is not (infinitely) predictable, then we have a **stochastic process** (leading to a time series). These processes can also be characterized, but only statistically, not deterministically.

observed quasar light curve:



simulated light curves time series generated by a stochastic process (Moreno et al. 2019):



Stochastic Processes

Although the definition of a stochastic process varies, it is typically characterized as a collection of random variables indexed by some set.

Stochastic Processes

Although the definition of a stochastic process varies, it is typically characterized as a collection of random variables indexed by some set.

Stochastic variability includes behavior that is not predictable forever as in the periodic case, but unlike temporally localized events, variability is always present.

Stochastic Processes

Although the definition of a stochastic process varies, it is typically characterized as a collection of random variables indexed by some set.

Stochastic variability includes behavior that is not predictable forever as in the periodic case, but unlike temporally localized events, variability is always present.

Typically, the **underlying physics** is so complex that we cannot deterministically predict future values. Despite their seemingly irregular behavior, there are a number of ways on how to **quantify and characterize** the data.

Autoregressive Models

Processes that are not periodic, but that nevertheless *retain memory* of previous states, can be described in terms of **autoregressive models**.

Autoregressive models (AR) with dependencies on k past values are called **autoregressive process of order** k and denoted as AR(k). A generalization is called the continuous autoregressive process, CAR(k).

Autoregressive Models

For **linear regression**, we are predicting the dependent variable from the independent variable

$$y = mx + b.$$

For **auto-regression**, the dependent and independent variable is the same and we are predicting a future value of y based on k past values of y:

$$y_i = a_i y_{i-1} + \ldots = \sum_{j=1}^{k} a_j y_{i-j} + \epsilon_i$$

where a_i is the **lag coefficient**.

Autoregressive Models

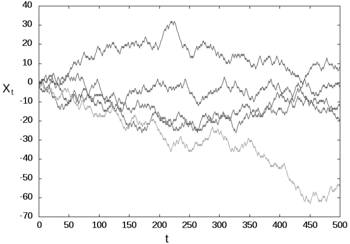
A **random walk** is an example of such a process (with $a_i = 1$, $k = 1$); every value is given by the preceeding value plus noise.

The model can then be written as

$$X_t = X_{t-1} + Z_t,$$

where Z_t is a white noise variable with zero mean and variance σ^2. This model is *not stationary*.

different realizations of a 1D Random Walk time series with 500 time steps



Random Walk

Repeatedly substituting for past variables results in

$$\begin{aligned} X_t &= X_{t-1} + Z_t \\ &= X_{t-2} + Z_{t-1} + Z_t \\ &= X_{t-3} + Z_{t-2} + Z_{t-1} + Z_t \\ &\vdots \\ &= X_0 + \sum_{j=0}^{t-1} Z_{t-j}. \end{aligned}$$

If the initial value X_0 is fixed, then the mean value of X_t is equal to X_0, that is,

$$\mathrm{E}(X_t) = \mathrm{E}\left(X_0 + \sum_{j=0}^{t-1} Z_{t-j}\right) = X_0.$$

Random Walk

The mean is constant, but the variance and covariance depend both on time, not just on the lag.
Since the white noise variables Z_t are uncorrelated, we obtain

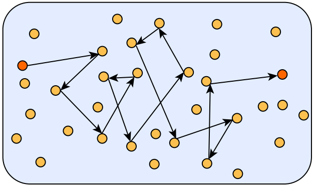
$$\begin{aligned}\mathrm{Var}(X_t) &= \mathrm{Var}\left(X_0 + \sum_{j=0}^{t-1} Z_{t-j}\right) = \mathrm{Var}\left(\sum_{j=0}^{t-1} Z_{t-j}\right) \\ &= \sum_{j=0}^{t-1} \mathrm{Var}\left(Z_{t-j}\right) = t\sigma^2\end{aligned}$$

and

$$\begin{aligned}\mathrm{Cov}(X_t, X_{t-\tau}) &= \mathrm{Cov}\left(\sum_{j=0}^{t-1} Z_{t-j}, \sum_{k=0}^{t-\tau-1} Z_{t-\tau-k}\right) \\ &= \mathrm{E}\left\{\left(\sum_{j=0}^{t-1} Z_{t-j}\right)\left(\sum_{k=0}^{t-\tau-1} Z_{t-\tau-k}\right)\right\} \\ &= \min(t, t-\tau)\sigma^2.\end{aligned}$$

Damped Random Walk

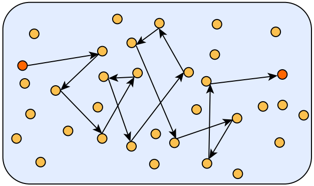
The **Ornstein-Uhlenbeck process** is a stochastic process that originally describes the velocity of a massive Brownian particle under the influence of friction.



Brownian motion is the random motion of particles suspended in a medium.

Damped Random Walk

The **Ornstein-Uhlenbeck process** is a stochastic process that originally describes the velocity of a massive Brownian particle under the influence of friction.



Brownian motion is the random motion of particles suspended in a medium.

The process can be considered as a modification of the random walk but with a **tendency to move back towards a central location**, with a greater **attraction** when the process is further away from the center. The process is stationary.

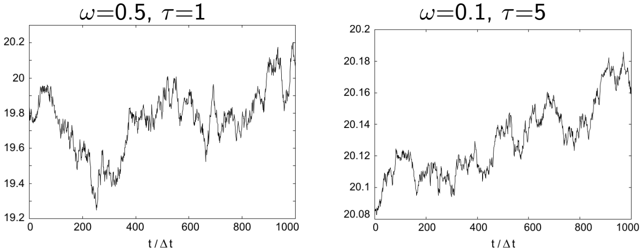
Damped Random Walk

An Ornstein-Uhlenbeck process x_t satisfies the following stochastic differential equation:

$$\mathrm{d}x_t = \tau(\mu - x_t)\mathrm{d}t + \omega\,\mathrm{d}W_t$$

where $\tau > 0$, μ and $\omega > 0$ are parameters and W_t denotes the Wiener process (random walk in continuous time).
The stationary (long-term) variance is given by $\mathrm{Var}(x_t) = \frac{\omega^2}{2\tau}$.

two realizations of a 1D Damped Random Walk, each with a mean of 20



Damped Random Walk

The damped random walk can also be described by its covariance matrix:

$$S_{ij} = \sigma^2 \exp(-|t_{ij}/\tau|)$$

where σ and τ are the model parameters.
σ^2 controls the short timescale covariance ($t_{ij} \ll \tau$), which decays exponentially on a timescale given by τ which is called the characteristic timescale (relaxation time, or damping timescale).

With this, the **autocorrelation function for a damped random walk** is

$$\mathrm{ACF}_{\mathrm{DRW}}(t) = \exp(-t/\tau).$$

Moving Average Process

A **moving average (MA) process** is similar to an AR process, but the value at each time step depends not on the value of previous time step, but rather the **perturbations from previous time steps**.

MA processes are defined by

$$y_i = \epsilon_i + \sum_{j=1}^{q} b_j \epsilon_{i-j}.$$

So, for example, an MA(q=1) process would look like

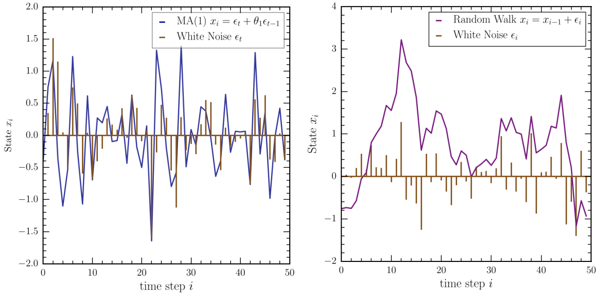
$$y_i = \epsilon_i + b_1 \epsilon_{i-1},$$

whereas an AR(p=2) process would look like

$$y_i = a_1 y_{i-1} + a_2 y_{i-2} + \epsilon_i$$

Moving Average Process

Thus in an MA process a 'shock/impulse' affects only the current value and q values into the future. In an *AR* process a 'shock/impulse' affects all future values. These two plots show the difference between an MA(1) process and an AR(1) (random walk) process:



credit: Moreno et al. (2019)

(Auto-)Correlation Function

problem:

We observe a (stochastically varying) quasar which has both **line and continuum emission** and where the line emission is stimulated by the continuum.

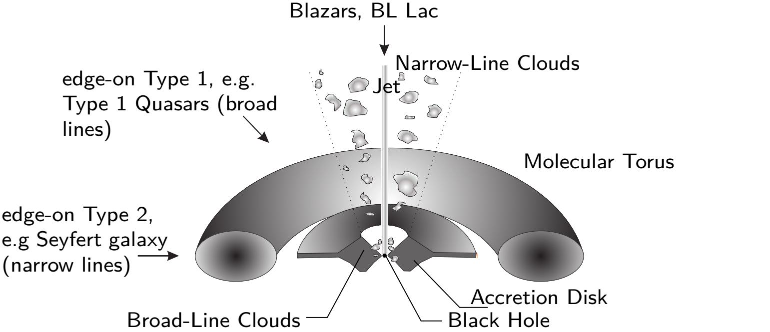
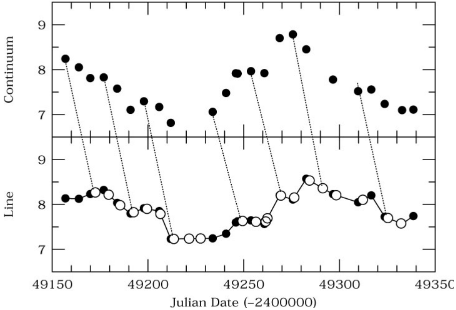


Illustration of the unified AGN model for the radio-quiet case. The classes of AGN are indicated by the viewing angle shown by arrows. Own diagram based on Elvis+2000.

(Auto-)Correlation Function

The physical separation between the regions that produce each type of emission causes a delay between the light curves:



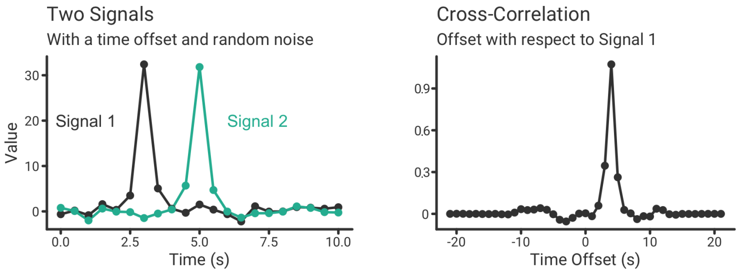
credit: Peterson+2001

We measure line and continuum emission and want to compute the time lag t_{lag}.

Correlation Function

To find a time lag between two time series, we can compute the **correlation function**. If one time series is derived from another simply by shifting the time axis by t_{lag}, then their (cross-)correlation function will have a peak at $\Delta t = t_{\text{lag}}$.

Computing the correlation function is basically the mathematical processes of convolution, i.e., sliding the two curves over each other and computing the degree of similarity for each step in time:



Correlation Function

The **(cross-)correlation function** between time series $f(t)$ and $g(t)$ is defined as

$$\mathrm{CCF}(\Delta t) = \frac{\lim_{T\to\infty} \frac{1}{T} \int_T f(t)\, g(t+\Delta t) dt}{\sigma_f \sigma_g}$$

σ_f and σ_g are the standard deviations of $f(t)$ and $g(t)$, respectively. With this normalization, the correlation function is unity for $\Delta t = 0$.

Correlation Function

For the **autocorrelation function (ACF)**, we take our correlation function from above and set $f(t) = g(t)$.

Correlation Function

For the **autocorrelation function (ACF)**, we take our correlation function from above and set $f(t) = g(t)$.

Whereas the correlation function yields information about a possible time lag between $f(t)$ and $g(t)$, the autocorrelation function yields information about the variable timescales present in a process. When $f(t)$ values are uncorrelated (e.g., due to white noise without any signal), $\text{ACF}(\Delta t) = 0$ except for ACF(0)=1. For processes that *retain memory* of previous states only for some characteristic time τ, the autocorrelation function vanishes for $\Delta t \ll \tau$.

Correlation Function

Let $x(t_i)$ and $y(t_i)$ be two **discrete, evenly sampled time series**.
In this case, the cross-correlation function as a function of time lag τ is

$$\mathrm{CCF}(\tau) = \frac{1}{N} \sum_{i=1}^{N} \frac{\left[x(t_i) - \bar{x}\right]\left[y(t_i - \tau) - \bar{y}\right]}{\sigma_x \sigma_y},$$

Correlation Function

Let $x(t_i)$ and $y(t_i)$ be two **discrete, evenly sampled time series**.

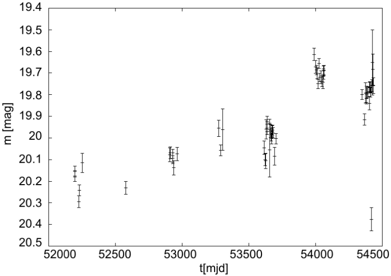
In this case, the cross-correlation function as a function of time lag τ is

$$\mathrm{CCF}(\tau) = \frac{1}{N} \sum_{i=1}^{N} \frac{\left[x(t_i) - \bar{x}\right]\left[y(t_i - \tau) - \bar{y}\right]}{\sigma_x \sigma_y},$$

The **accuracy** with which the peak of a CCF can be determined when the data points are evenly spaced has been dealt with by Tonry+1979. The accuracy which is achieved is proportional to the ratio of the height of the peak to the noise level in the CCF divided by the half-width at half-maximum of the peak in the CF.

Correlation Function for Unevenly Sampled Time Series

In practice, for time series being **long-time astronomical data** or other measurements, such as those produced by quasar monitoring programs, $x(t_i)$ and $y(t_j)$ are usually **not known at regular intervals of time**. A considerable body of theory has been developed for this case, how one can carries out time series analysis based on CCF even in this cases.



A quasar light curve from SDSS Stripe 82.

Correlation Function for Unevenly Sampled Time Series

There are in general two approaches for dealing with the uneven sampling: the **Discrete Cross Correlation Function (DCCF)** and interpolation methods like the **Interpolated Cross-Correlation Function (ICCF)**.

These two methods are briefly described in the following.

The Discrete Cross Correlation Function (DCCF)

In the DCCF method (Edelson+1988), a contribution to the DCCF is **calculated only using the actual data points**. Each pair of points, one from each of the light curves, gives one correlation value at a lag corresponding to the time separation. For two light curves with N and M data points respectively this gives an Unbinned Cross Correlation Function (UCCF),

$$\mathrm{UCCF}_{ij} = \frac{(x_i - \bar{x})(y_j - \bar{y})}{\sigma_x \sigma_y}.$$

The DCCF is then obtained by averaging the UCCF in time lag bins. This results in

$$\mathrm{UCCF}_{ij} = \frac{1}{N} \frac{(x_i - \bar{x})(y_j - \bar{y})}{\sigma_x \sigma_y}.$$

where N is the number of pairs.

A point to notice is that a large variation in observational coverage over the light curve can have a strong effect on the DCCF amplitudes.

The Interpolated Cross Correlation Function (ICCF)

In the ICCF method (Gaskell+1986), the light curve is **linearly interpolated** and resampled onto a regular grid. It is common to calculate the ICCF twice, where interpolation is done in each of the light curves, one at a time, and average these.

For two time series x_i and y_i, in the case of interpolation in time series y_i, the cross-correlation function is approximated by

$$\mathrm{ICCF}_{xy}(\Delta t) = \frac{1}{N}\sum_{i=1}^{N} \frac{(x_{t_i} - \bar{x})\{L\left[y_{t_i+\Delta t}\right] - \bar{y}\}}{\sqrt{\sum_{i=1}^{N}(x_{t_i} - \bar{x})^2}\sqrt{\sum_{i=1}^{N}(y_{t_i} - \bar{y})^2}}$$

where L indicates a piecewise linear interpolation of series y at time $t_i + \Delta t$ and N is the number of data points in x.

The piecewise linear interpolation of y is done by

$$L\left[y_{t_i+\Delta t}\right] = \begin{cases} y_{t_i+\Delta t} & \text{if observed for } t_i + \Delta t \\ y_{t_{i-1}} \frac{t_{i+1}-(t_i+\Delta t)}{t_{i+1}-t_{i-1}} + y_{t_{i+1}} \frac{(t_i+\Delta t)-t_{i-1}}{t_{i+1}-t_{i-1}} & \text{else} \end{cases}$$

where $y_{t_{i-1}}$ and $y_{t_{i+1}}$ are indicating the nearest data points available.

Structure Function

The **structure function** is another quantity that is frequently used in astronomy and is related to the ACF:

$$\mathrm{SF}(\Delta t) = \mathrm{SF}_\infty[1 - \mathrm{ACF}(\Delta t)]^{1/2},$$

where SF_∞ is the standard deviation of the time series as evaluated on timescales much larger than any charateristic timescale, τ.

The ACF for a Damped Random Walk (DRW) is given by

$$\mathrm{ACF}(t) = \exp(-t/\tau),$$

where τ is the characteristic timescale (i.e., the damping timescale). Remember that a DRW modeled as an AR(1) has $a_1 = \exp(-1/\tau)$. The **structure function for a DRW** can then be written as

$$\mathrm{SF}(t) = \mathrm{SF}_\infty[1 - \exp(-t/\tau)]^{1/2}.$$

Break & Questions

afterwards we continue with `lecture_9.ipynb` from the `github` repository