8. Random processes

Contents

(All variables are real and one-dimensional unless otherwise specified.)

This tutorial is a summary of various special words and concepts that you often encounter in the field of time series analysis.

1. The meaning of "random"

"Random" is a vague word—it can mean whatever you want it to mean. Therefore, other expressions may be used in various contexts to describe how "random" a process is and to emphasize certain properties of the process.

- **Truly random process.** Truly random processes only exist in quantum mechanics (as far as I know). People used to generate random numbers by observing the outcome of physical experiments. They also **wrote down the numbers on books (https://en.wikipedia.org/wiki/A_Million_Random_Digits_with_100,000_Normal_Deviates)** for other people to use.
- **Pseudorandom process.** A pseudorandom process tries to mimic the properties of a truly random process by following some **deterministic algorithms (https://en.wikipedia.org/wiki/Linear_congruential_generator)**. Essentially all random numbers that we can generate with programming languages today are pseudorandom numbers.
- **Stochastic process.** A stochastic process also looks random due to some hidden deterministic rules; but unlike a pseudorandom process, it does not intend to mimic a truly random process. For example, a stock's price is stochastic because its apparent randomness is in fact controlled by traders' collective actions.

- **Stationary process.** Although its value is fluctuating, a process is considered stationary if its properties remain unchanged over time. A stationary process in science (like chemical reactions) is often said to be "in **equilibrium**".

    - **Strongly stationary process.** A process whose underlying probability distribution is independent of time. This is, however, no way a practical definition.
    - **Weakly stationary process.** A process that has (i) a constant **mean**, (ii) a constant **variance**, and (iii) a constant **autoccorelation**. These three properties makes a process stable enough for real-life applications.
        - A process has a high autocorrelation if it changes the direction of its value slowly. In other words, it has a low **frequency**.
        - A process has a low autocorrelation if it changes the direction rapidly. In other words, it has a high frequency.

- **Markov process.** (From *Andrey Markov*.) The crucial feature of a Markov process is **path independence** or **history independence**. The value of a Markov process at the next moment depends only on the value of the process right now but not on its previous values.
- **Wiener process.** (From *Nobert Wiener.*) A Wiener process $W(t)$ satisfies

$$\Delta W = W(t+\Delta t) - W(t) \sim \mathcal{N}(0, \Delta t)\,.$$

  It means that ΔW is a **normally distributed** with mean zero and variance Δt.
    - **Generalization.** A Wiener process can be generally defined as $X(t) = \mu t + \sigma W(t)$ for some **time-independent** parameters μ and σ. The constant μ is usually called the **drift** of the process, whereas σ^2 amounts to the variance of X. The process satisfies

$$\Delta X = X(t+\Delta t) - X(t) = \mu \Delta t + \sigma \Delta W\,.$$

      We can write this as $\mathrm{d}X = \mu \mathrm{d}t + \sigma \mathrm{d}W$ by letting $\Delta t \to 0$.
- **Itô process.** (From *Kiyosi Itô*.) An Itô process is a Wiener process with **time-dependent** drift and variance.

- **Random walk.** A process whose increment is some kind of zero-mean random number. If the random number follows a normal distribution, it is a **Gaussian random walk**, of which Wiener processes are particular examples.
- **Brownian motion.** (From *Robert Brown*.) It is often used interchangeably with "Wiener process" and "random walk" despite their different connotations: "Brownian motion" sounds more related to physics, whereas the other two sounds more related to mathematics.
    - Brown, as a botanist, observed some kind of irregular movement of pollen particles, which is now called "Brownian motion", under a microscope in 1827. While its origin was explained by Einstein in 1905, Wiener (1894–1964) studied its mathematical properties extensively.
- **Exponential Brownian motion.** A process G whose logarithm $\ln G = \mu t + \sigma W$ is a Wiener process. It satisfies

$$\frac{\mathrm{d}G}{G} = \mu \mathrm{d}t + \sigma \mathrm{d}W \quad \Leftrightarrow \quad \mathrm{d}G = \mu G \mathrm{d}t + \sigma G \mathrm{d}W$$

Also called "**geometric Brownian motion**".

2. Modelling a random process

2.1 Time: discrete versus continuous

A process X can have **discrete time** or **continuous time**. When we want to highlight their differences, we may use the superscript notation X_t to represent the former but the bracket notation $X(t)$ to represent the latter.

While discrete time requires us to calculate properties of X_t with summation like $\bar{X} \sim \sum_t X_t$, continuous time enables us to calculate properties of $X(t)$ with integration like $\bar{X} \sim \int X(t) \mathrm{d}t$, which is usually easier. On the contrary, results obtained from the integration may become inapplicable since we can only obtain discrete-time data from experiments. (Can we brutally let $\Delta t \to 0$ when Δt is discrete?)

2.2 Operators: more than a notation?

Consider a discrete process X_t. As we encounter terms like $X_t - X_{t-k}$ (for some integer $k > 0$) a lot, we would like to introduce some symbols called "**operators**" to express them in a more "beautiful" way.

- **Lag operator.** Denoted by L,

  $$\mathrm{L} X_t \equiv X_{t-1} \, .$$

  Then we can iterate the formula and write $X_{t-2} = \mathrm{L} X_{t-1} = \mathrm{L} \left(\mathrm{L} X_t \right) \equiv \mathrm{L}^2 X_t$. It means that

  $$\mathrm{L}^k X_t \equiv X_{t-k} \, .$$

- **Backshift operator.** Essentially the same as the lag operator although it is denoted by B, i.e. $\mathrm{B} X_t \equiv X_{t-1}$.

- **Forward difference operator.** Or just "**difference operator**". Denoted by Δ,

$$\Delta X_t \equiv X_{t+1} - X_t \equiv (1 - \mathrm{L})\, X_t \,.$$

We may also write $\Delta_k X_t \equiv X_{t+k} - X_t$, but this is quite rare. A more common notation is $\left(1 - \mathrm{L}^k\right) X_t \equiv X_t - X_{t-k}$.

- **Backward difference operator.** Some people also use ∇, which is defined through

$$\nabla X_t \equiv X_t - X_{t-1} \,.$$

This is even rarer, though. (I never use it.)

You may wonder why we need to redefine the simple concepts with such operators redundantly. Indeed, they are useless unless you do advanced mathematics: operators can often be treated as **polynomials**, on which some cool mathematical tools can be applied to simplify algebra.

2.3 Colours of noise, errors, and residuals

Noise itself is also some kind of random process. We need to know their properties because the signals that we want to observe are exactly distorted by them. Interestingly, properties of noise are commonly associated with **colours** (https://en.wikipedia.org/wiki/Colors_of_noise).

- **White noise.** If ε_t is white noise, it is a bunch of random numbers that has zero mean, constant variance, and zero autocorrelation.

$$\langle \varepsilon_t \rangle = 0 \quad \text{and} \quad \left\langle \varepsilon_t^2 \right\rangle = \sigma^2 \quad \text{and} \quad \langle \varepsilon_t \varepsilon_{t+\tau} \rangle = \begin{cases} \sigma^2 & (t = \tau) \\ 0 & (t \neq \tau) \end{cases}$$

  It is said to be "white" because it resembles white light by containing all kinds of **frequency** components after Fourier transform, which are equally strong.
  - **Gaussian white noise.** Note that the white colour does not specify the distribution of the random numbers. If they follow normal distribution, the white noise is called Gaussian white noise. (Pseudorandom numbers have a uniform distribution instead.)
- Pink noise. The strength of pink noise's Fourier components follows $1/f$, where f is their frequencies. It is said to be "pink" because it contains more lower-frequency components than high-frequency ones. (In physics, red light has a lower frequency than blue light.)
- **Brown noise.** Identical to "Brownian motion"—"Brown" is the scientist's name instead of the colour. However, this pun makes sense because the strength of brown noise's Fourier components scales with $1/f^2$, meaning that it contains even more lower-frequency components and thus looks "redder" than pink noise does.

On one hand, noise distorts our signals undesirably. On the other hand, when we model a random process, the **errors** of our model (aka the **residuals** in the jargon of statistics) should fluctuate like noise (preferably the white noise), otherwise there is probably some **systematic bias**.

2.4 Evolving processes

It is usually more convenient to work with stationary processes. If we have an **evolving** (i.e. non-stationary) process, we would often like to turn it into a stationary process first and deal with its evolving features separately.

Let Y_t be an evolving process and X_t be any stationary process.

- **Trend and detrending.** If

  $$Y_t = f(t) + X_t$$

  for some function f, then Y_t is said to be **trend-stationary**, and the term $f(t)$ is called "trend". Ideally, we can obtain X_t by calculating $Y_t - f(t)$; however, we do not know f in advance, so we must estimate it as $\hat{f}$ after observing Y_t, and this is not easy when f is not linear. We must also check whether $Y_t - \hat{f}(t)$ is indeed stationary: if no, the assumption is invalid, and Y_t may not be trend-stationary at all.
- **Drift and differencing.** If Y_t has a linear trend at for some constant a, the constant a is called "drift" or "drift rate". We may first check whether is $Y_t - at$ is stationary. If yes, then everything is done. If no, we may proceed to check whether $\Delta Y_t - a$ is stationary because

  $$\Delta Y_t = a + X_t \quad \Rightarrow \quad Y_t = at + \sum_t X_t$$

  and the cumulative sum $\sum_t X_t$ of a stationary process is almost surely not stationary.
- **Seasonality and seasonal detrending.** A process may be periodic. (For example, the number of consumers at bars peak every Friday night.) If Y_t has a period k, we should check whether $\left(1 - \mathrm{L}^k\right) Y_t = Y_t - Y_{t-k}$ is stationary.

Sometimes, a process is so complicated that we may need to combine the techniques and apply them several times iteratively before obtaining a stationary process.

2.5 The ARMA models

A stationary process is usually modelled with the ARMA models. We will go through their properties later.

3. Statistical tests and tools

When we model a process, we need to systematically judge whether our models make sense or not using various tests and tools like the followings.

- ***p*-value.** The conditional probability p of observing the observed data or something more extreme given that a null hypothesis H_0 is true. If it is smaller than a significance level α, then H_0 is rejected because it should have been almost impossible to observe the observed data if H_0 were true, but now we have indeed observed it, implying high probability for H_0 to be false (cf. "modus tollens" in logic (https://en.wikipedia.org/wiki/Modus_tollens)).

- **Dickey-Fuller test.** (From *David Dickey* and *Wayne Fuller*.) The null hypothesis H_0 is "a process X_t is stationary".
  - If it is true, ΔX_t does not depend on X_{t-1}, meaning that when we fit any of the following models to ΔX_t,

    $$\Delta X_t = \gamma X_{t-1} + \varepsilon_t + \begin{cases} 0 \\ \alpha \\ \alpha + \beta t \end{cases}$$

    the coefficient γ of X_{t-1} is zero. While ε_t is white noise, the inclusion of the constant term α and the trend βt are up to us.
  - Suppose we have estimated γ as $\hat{\gamma}$ with **standard error** $\mathrm{SE}(\hat{\gamma})$. If H_0 is true, the quantity $\hat{\gamma}/\mathrm{SE}(\hat{\gamma})$ follows the Dickey-Fuller distribution, which has three different forms according to whether we include α and βt or not.
  - **Augmented Dickey-Fuller test.** In the pth-order augmented test, the model of ΔX_t is replaced by

    $$\Delta X_t = \gamma X_{t-1} + \textstyle\sum_{i=1}^{p} \zeta_i \Delta X_{t-i} + \varepsilon_t + \begin{cases} 0 \\ \alpha \\ \alpha + \beta t \end{cases},$$

    which includes an extra autoregressive part. Then the testing procedure remains the same.
- **Ljung-Box test.** (From *Greta Ljung* and *George Box*.) The null hypothesis H_0 is "a process X_t is not autocorrelated at lag $\tau > 0$". The test is useful when we want to check whether X_t is white noise.
  - Let n be the length of X_t and ρ_k be the autocorrelation of X_t measured at lag k.
  - If H_0 is true,

$$Q = n\left(n+2\right)\sum_{k=1}^{h}\frac{\rho_k^2}{n-k}$$

follows the χ^2 distribution with h degrees of freedom.

- ***F*-test.** (From *Ronald Fisher*.) In modelling, an *F*-test can be used to compare a model with an "expanded" version of the same model, which reduces to the original model by setting some parameters equal to zero. The null hypothesis H_0 is "the expanded model **does not** fit the data more efficiently than the original model".
  - Let n be the length of X_t. Suppose the original model uses p_{original} paramters and results in a **sum of squared errors** (SSE) equal to S_{original}; similarly, the expanded model uses p_{expanded} parameters and results in an SSE euqal to S_{expanded}.
  - If H_0 is true,

$$F = \frac{S_{\text{original}} - S_{\text{expanded}}}{S_{\text{expanded}}}\frac{n - p_{\text{expanded}}}{p_{\text{expanded}} - p_{\text{original}}}$$

    follows the *F*-distribution with $\left(p_{\text{expanded}} - p_{\text{original}}, n - p_{\text{expanded}}\right)$ degrees of freedom. The quantity $p_{\text{expanded}} - p_{\text{original}}$ is the number of extra parameters, and the quantity $n - p_{\text{expanded}} > 0$ is the degree of freedoms of the expanded model.
  - If H_0 is not rejected, the original (and simpler) model should be used. If H_0 is rejected, the expanded model should be used.
- ***t*-test.** (By *William Gosset*.) In modelling, a *t*-test can be used to check whether a term x should exist in a model of y. The null hypothesis H_0 is "the coefficient β of the term x is zero".
  - Let n be the sample size, i.e. the number of observations which are used to build the model. Then suppose we have estimated β as $\hat{\beta}$ with standard error $\mathrm{SE}\left(\hat{\beta}\right)$.
  - If H_0 is true, $\hat{\beta}/\mathrm{SE}\left(\hat{\beta}\right)$ follows the *t*-distribution with $n-2$ degrees of freedom. Because H_0 can be rejected by either $\hat{\beta} \gg 0$ or $\hat{\beta} \ll 0$, we need a **two-tailed test** here.

- **Box-Jenkins method.** (From *George Box* and *Gwilym Jenkins*.) General procedure to fit an ARMA model to a stationary process. It primitively determines the order of the model by checking the shape of **autocorrelation** and **partial autocorrelation** of the process.
- **Akaike information criterion (AIC).** (From *Hirotsugu Akaike*.) When given multiple models are given, the model that yields the smallest (or the most negative) AIC is often regarded as the best model.