Time Series as stochastic processes

Juan Garbayo

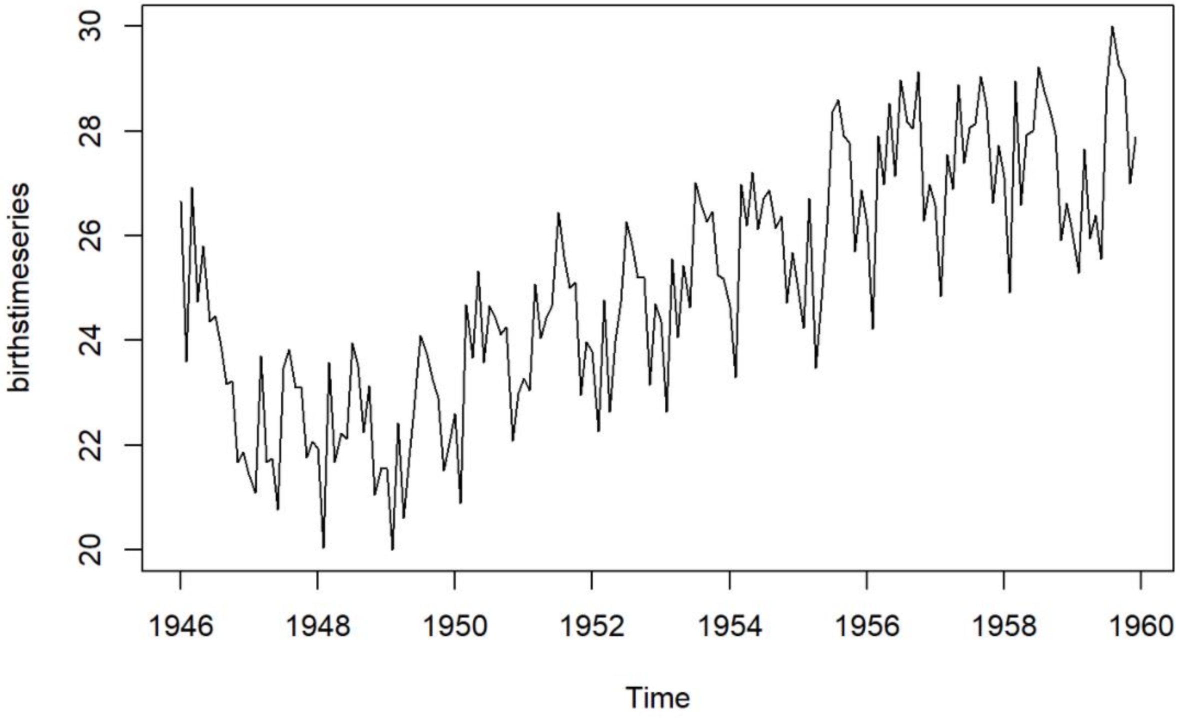
Forecasting & Time Series Analysis

IE UNIVERSITY

Contents

Modelling Time Series as a Stochastic Process

Time Series Data: sequential observations of one or several variables over time



From a mathematical modeling standpoint, we consider the following:

- **Future values of the time series:** a random variable that has not yet been realized.
- **Past values of the time series:** realizations of a random variable

From a mathematical modeling standpoint, we consider the following:

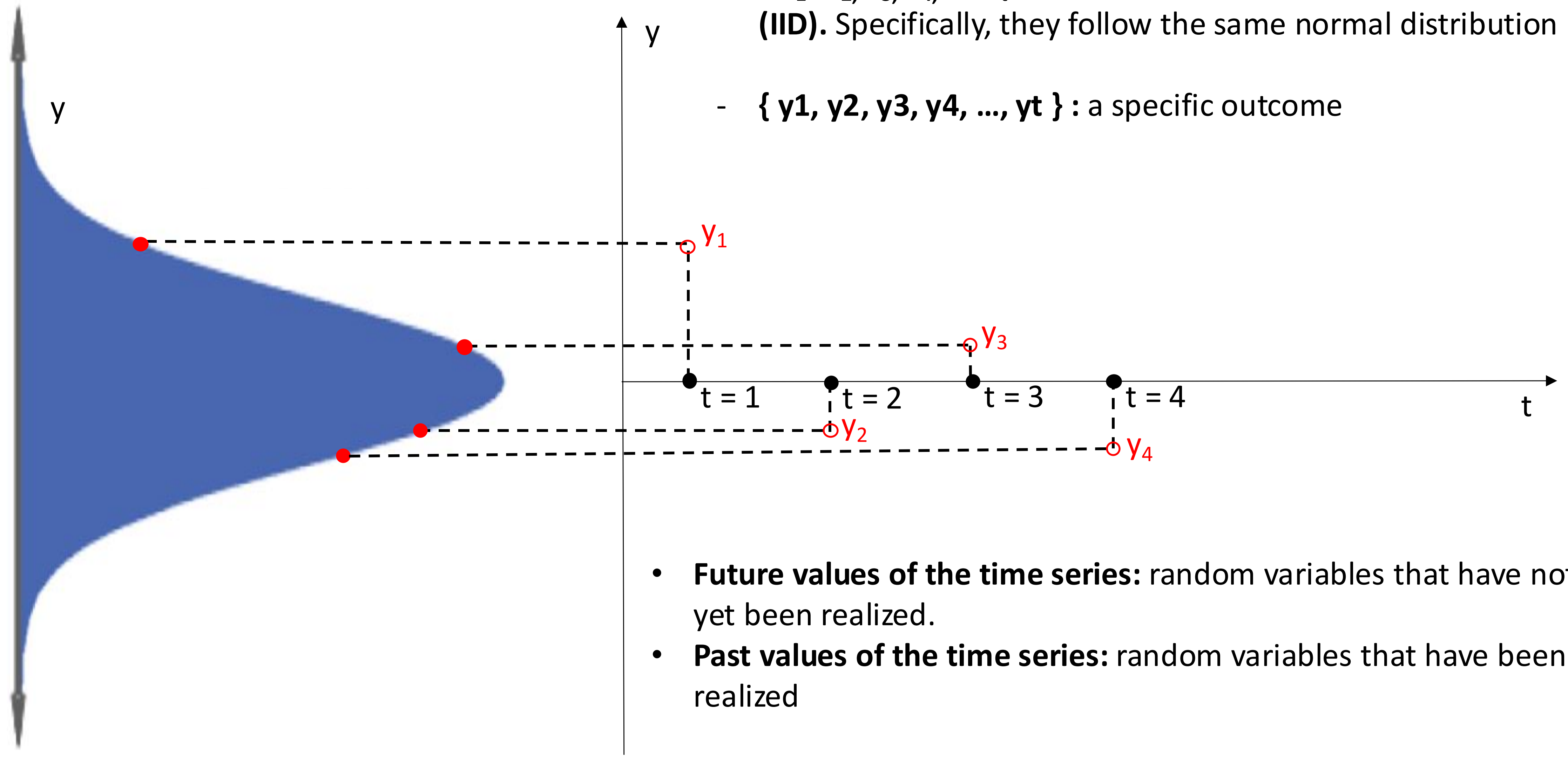
- **Future values of the time series:** a random variable that has not yet been realized.
- **Past values of the time series:** realizations of a random variable

Time Series as a mathematical entity: **a collection of random variables $\{Y_t\}$** **indexed over time**

- **$\{ Y_1, Y_2, Y_3, Y_4, \ldots, Y_t \}$:** uppercase letters refer to random variables. Description of the **PROCESS**

- **$\{ y_1, y_2, y_3, y_4, \ldots, y_t \}$:** lowercase letters refer to a **SPECIFIC OUTCOME** of the process (of each of the random variables)

Example:

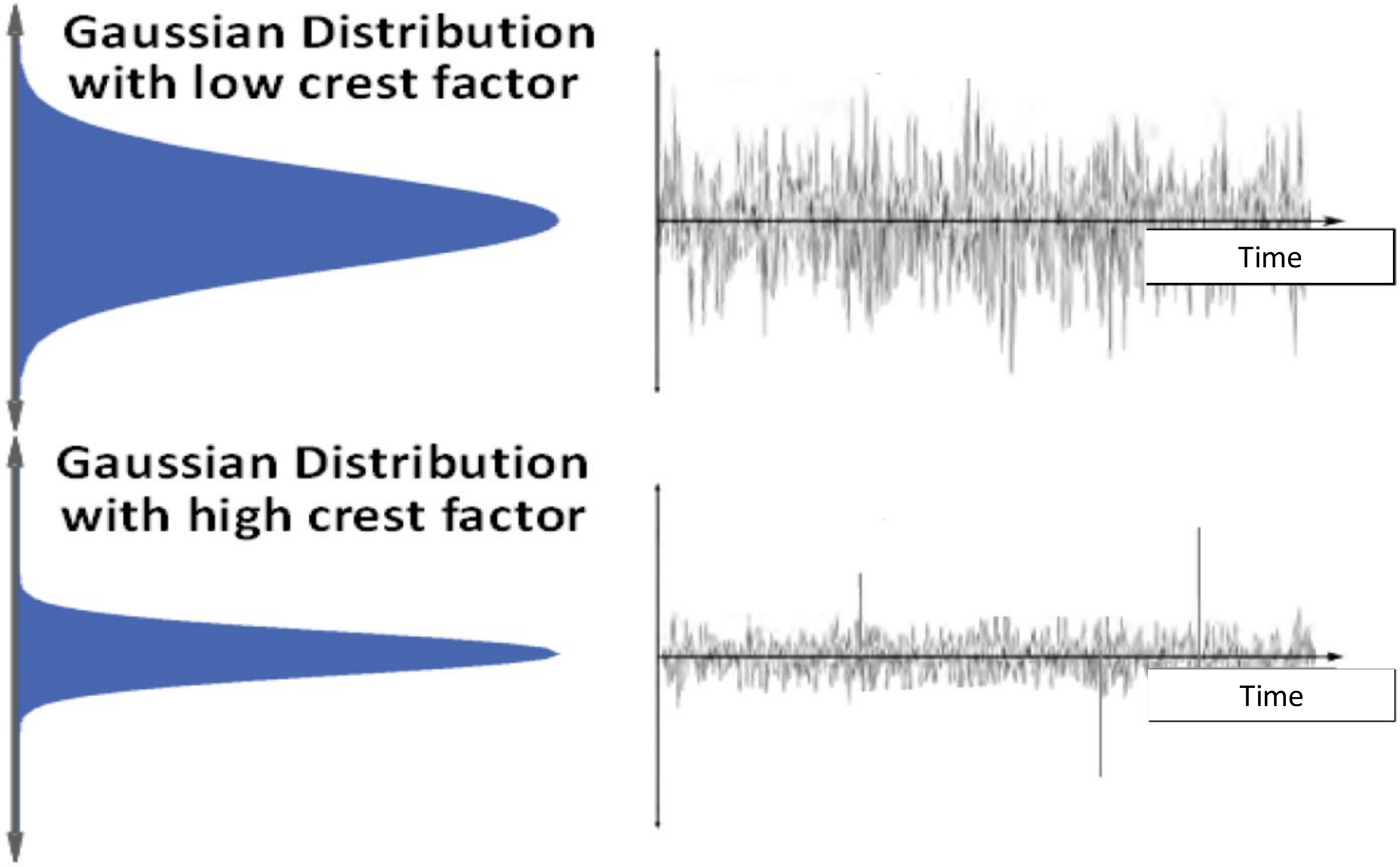


- **{ Y_1, Y_2, Y_3, Y_4, …, Y_t } are Independent and Identically Distributed (IID).** Specifically, they follow the same normal distribution

- **{ y1, y2, y3, y4, …, yt } :** a specific outcome

- **Future values of the time series:** random variables that have not yet been realized.
- **Past values of the time series:** random variables that have been realized

In this particular time series all the random variables are identically distributed (follow the same distribution).


Gaussian Distribution
with low crest factor
Time
Gaussian Distribution
with high crest factor
Time

Modelling Time Series as a Stochastic Process

Stochastic vs Random: these two words are many times used interchangeably. However, it is normally the case that:

- **Random** is reserved to refer to random variables.
- **Stochastic** is reserved to refer to processes involving a family of random variables indexed by a set. In the case of time series, indexed over time.

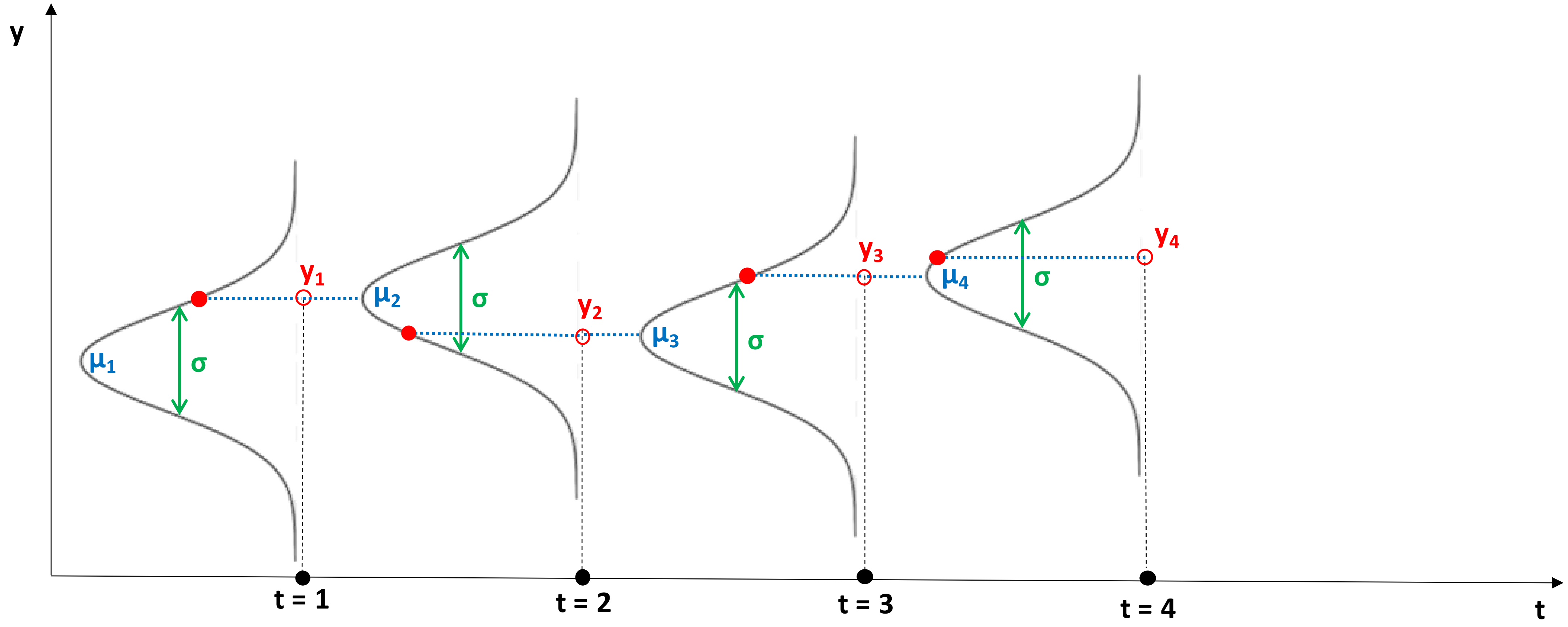
In our previous example:

- **For each point in time** t_1, t_2, t_3… we had a **random variable** Y_1, Y_2, Y_3…
- This **collection of random variables $\{Y_t\}$** **indexed by time** is a **stochastic process**
- The specific outcome of the process is denoted by the collection of outcomes of **$\{Y_t\}$** → $\{y_t\}$ (lowercase notation).

NOTE

- There are **different kinds of stochastic processes**. Time series are just one of the infinitely many stochastic processes we can think of.

Time series, correlation and independence



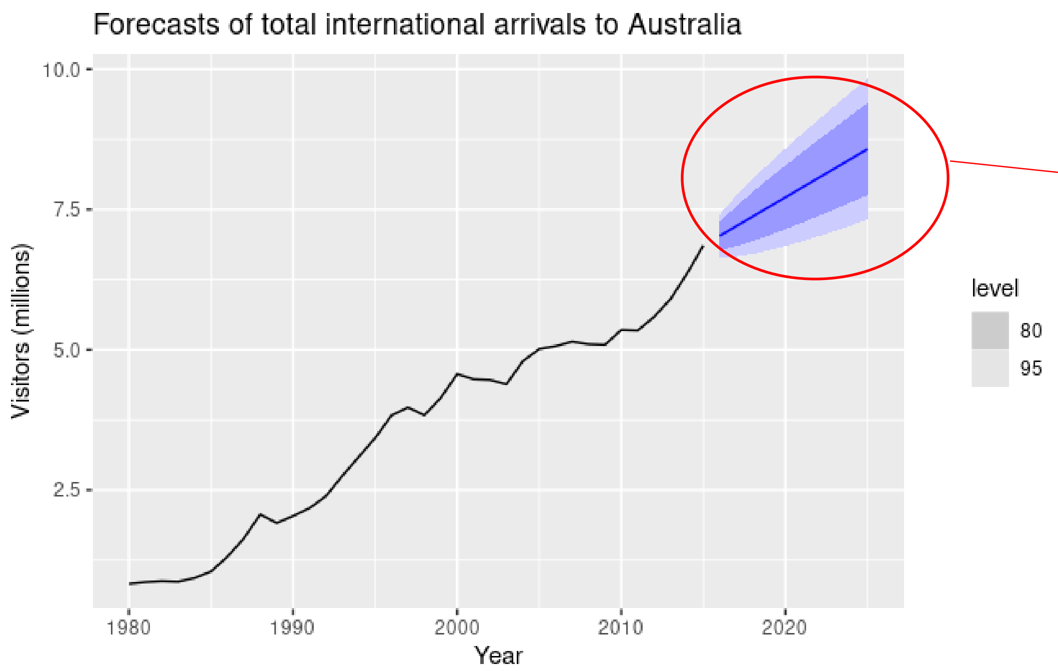
Since **past events** tend to **influence future events**, there is a <u>correlation between the values of the time series at different points in time</u>.

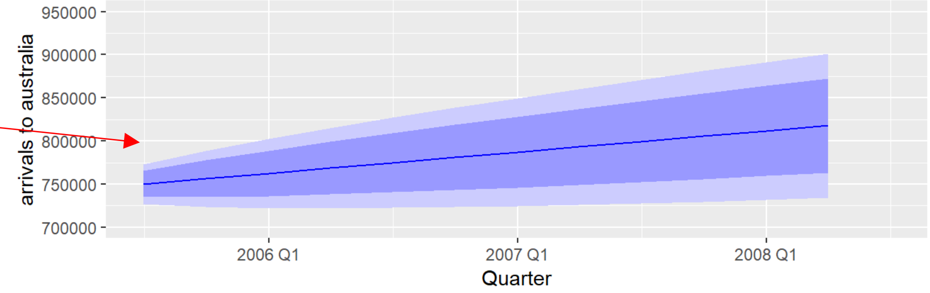
The random variables of the **time series process** are **in general correlated** (and therefore <u>NOT independent</u>).

Key takeway 1

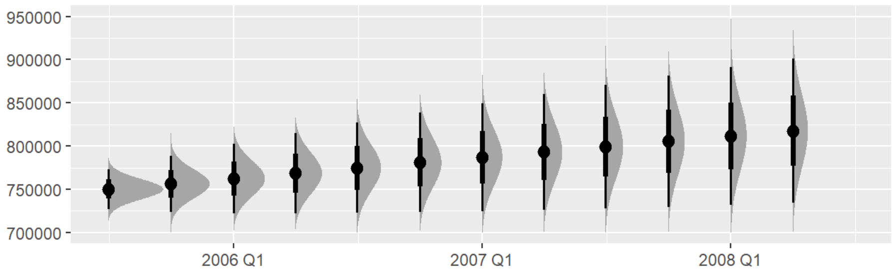
Time series as **stochastic processes** (a collection of random variables indexed over time):

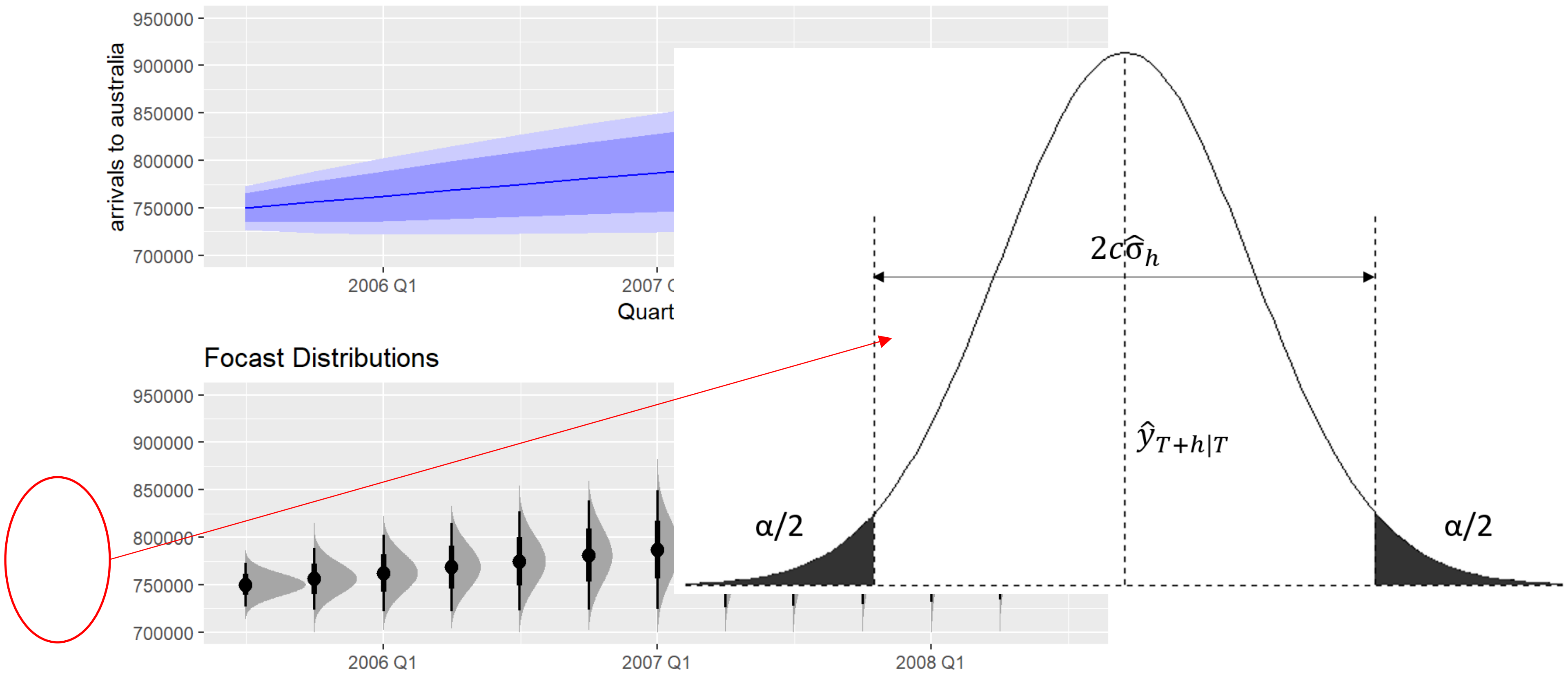
- **Process (R.Variables)** { $Y_1, Y_2, Y_3, Y_4, \ldots Y_t$ } vs **Specific outcome**: { $y_1, y_2, y_3, y_4, \ldots y_t$ }
- **Past values:** realizations of random variables.
  - ➢ They are no longer random, they are a specific number, an outcome of the random variable.
- **Forecast or future values:** forecasting a random variable that has not yet realized.
  - ➢ The concept of **confidence Intervals** therefore applies to time series forecasting.
  - ➢ You should think of the values to forecast as a random variable.





Focast Distributions




arrivals to australia
950000
900000
850000
800000
750000
700000
2006 Q1
2007 Q
Quart
Focast Distributions
950000
900000
850000
800000
750000
700000
2006 Q1
2007 Q1
2008 Q1
$2c\hat{\sigma}_h$
$\hat{y}_{T+h|T}$
α/2
α/2

Key takeway 2

The random variables of a **time series process** are in general:

- **Correlated** (and therefore **not independent**).
  - In most phenomena studied in time series, past values tend to affect future values.
- **Not identically distributed.**

The ***independent and identically distributed hypothesis*** used in most convential statistical methods (e.g. Central Limit Theorem) does **not apply.**

Time Series can be referred to as the *systematic approach to answer the mathematical and statistical questions posed by these time correlations*.