2022 Fall
IE 313 Time Series Analysis

2. Fundamental Concepts

Yongjae Lee
Department of Industrial Engineering
ULSAN NATIONAL INSTITUTE OF SCIENCE AND TECHNOLOGY

Chapter 2. Fundamental Concepts

- 2.1 Time Series and Stochastic Processes
- 2.2 Means, Variances, and Covariances
- 2.3 Stationarity

Chapter 2.1

Time Series and Stochastic Processes

Time series

- **Time series data**

  – A series of data points observed at different points in time

  – Usually, a time series is a sequence taken at successive equally spaced points in time

  – Time series are mostly represented in line charts

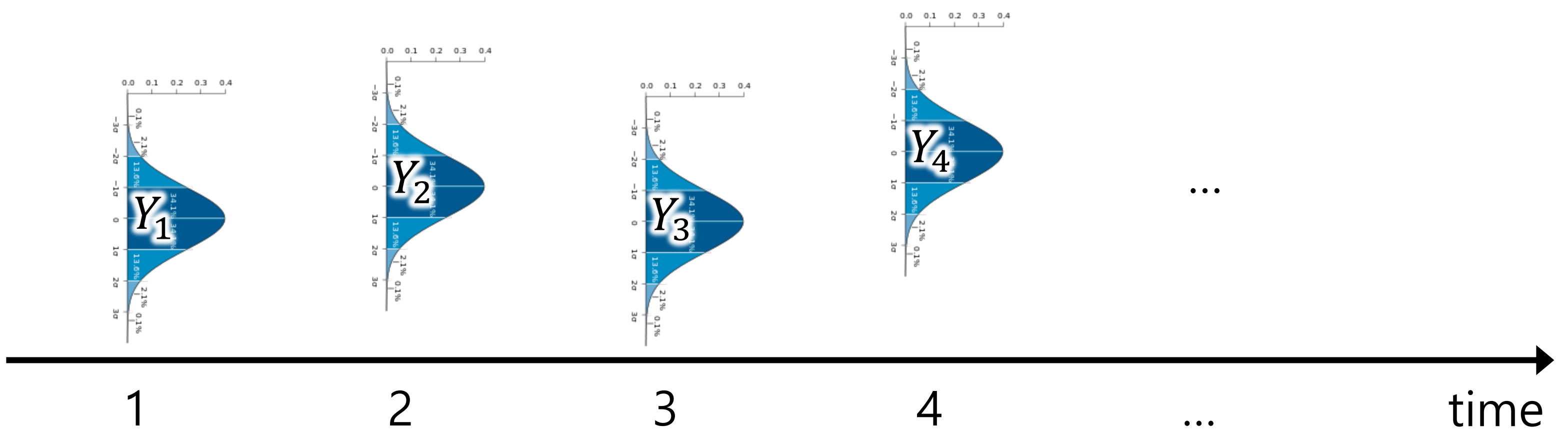
Time series modeling

- **Objective**
  - Find **mathematical models** that provide plausible descriptions for sample data
- **How?**
  - We will model a time series as a **stochastic process**

Stochastic process

Stochastic process

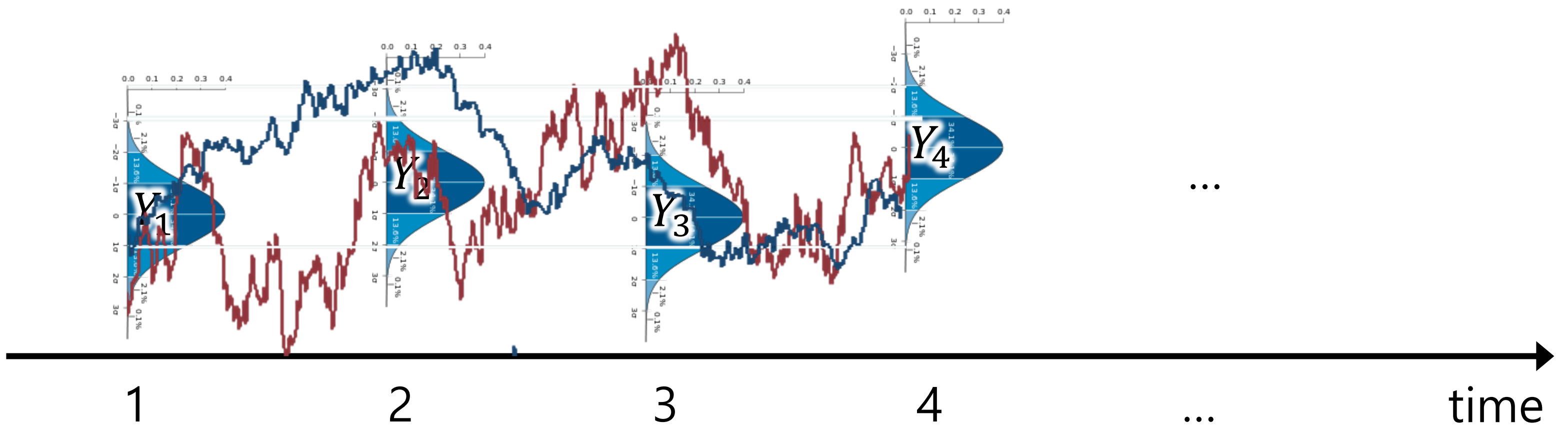
- **Sequence of random variables** $\{Y_t: t = 0, \pm 1, \pm 2, \pm 3, \ldots\}$ indexed by time point t
  - Y_t denotes the value taken by the series at the i th time period



Stochastic process

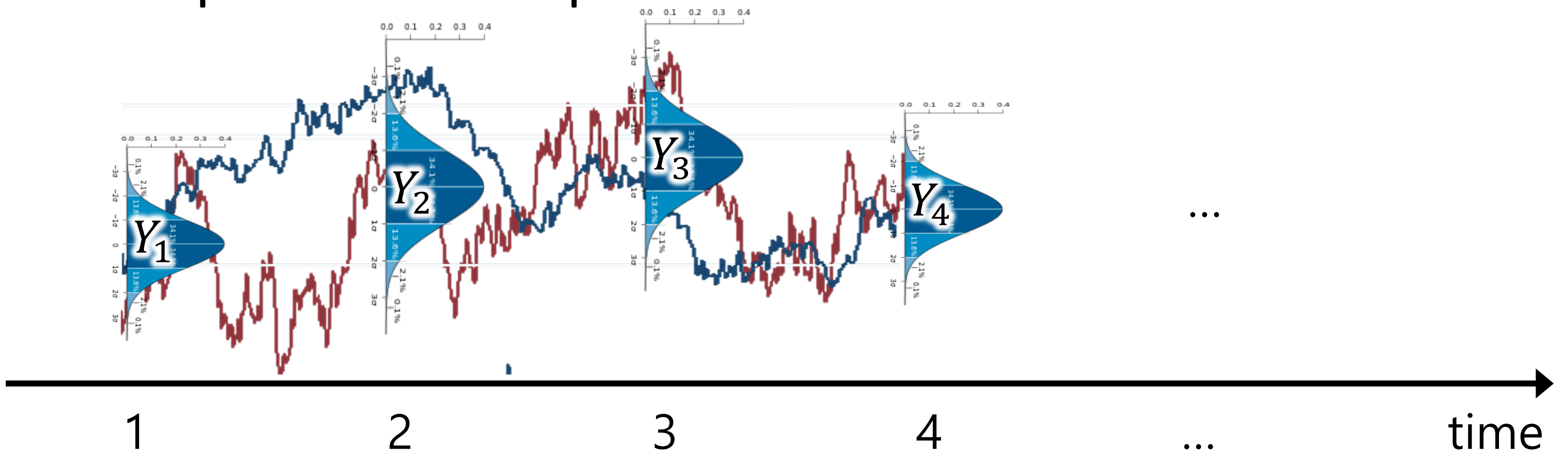
▪ Stochastic process

– **Sequence of random variables** $\{Y_t : t = 0, \pm 1, \pm 2, \pm 3, \dots\}$ indexed by time point t
  - Y_t denotes the value taken by the series at the i th time period
  - The **observed values** of a stochastic process are referred to as a **realization** of the stochastic process
    - › There can be many different realizations based on a single stochastic process

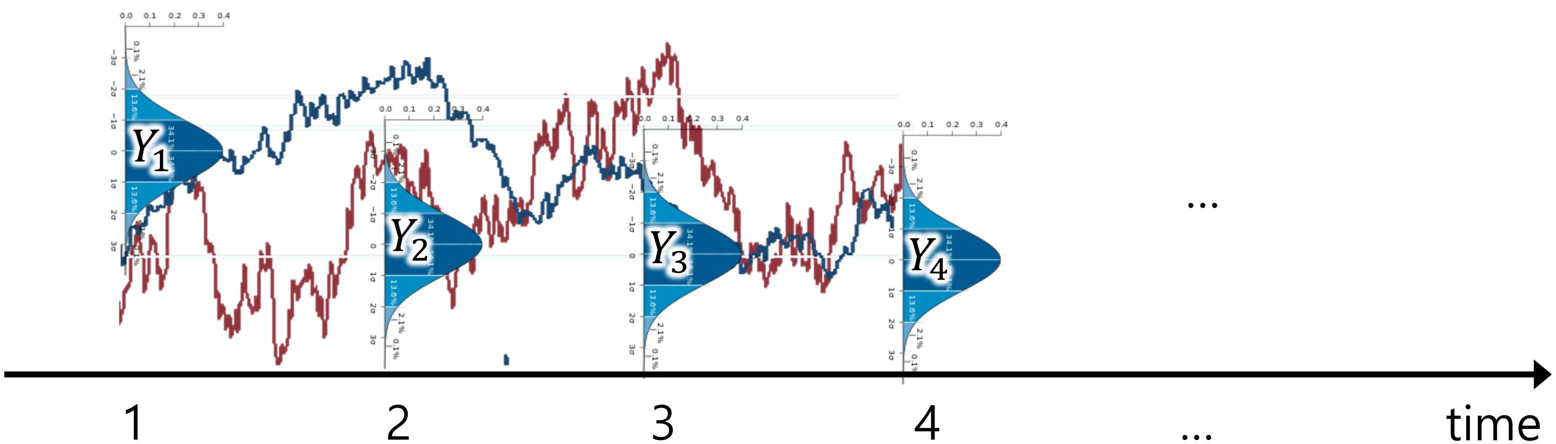


Time series modeling

- **Objective:** Find **stochastic processes** that provide plausible descriptions for sample data

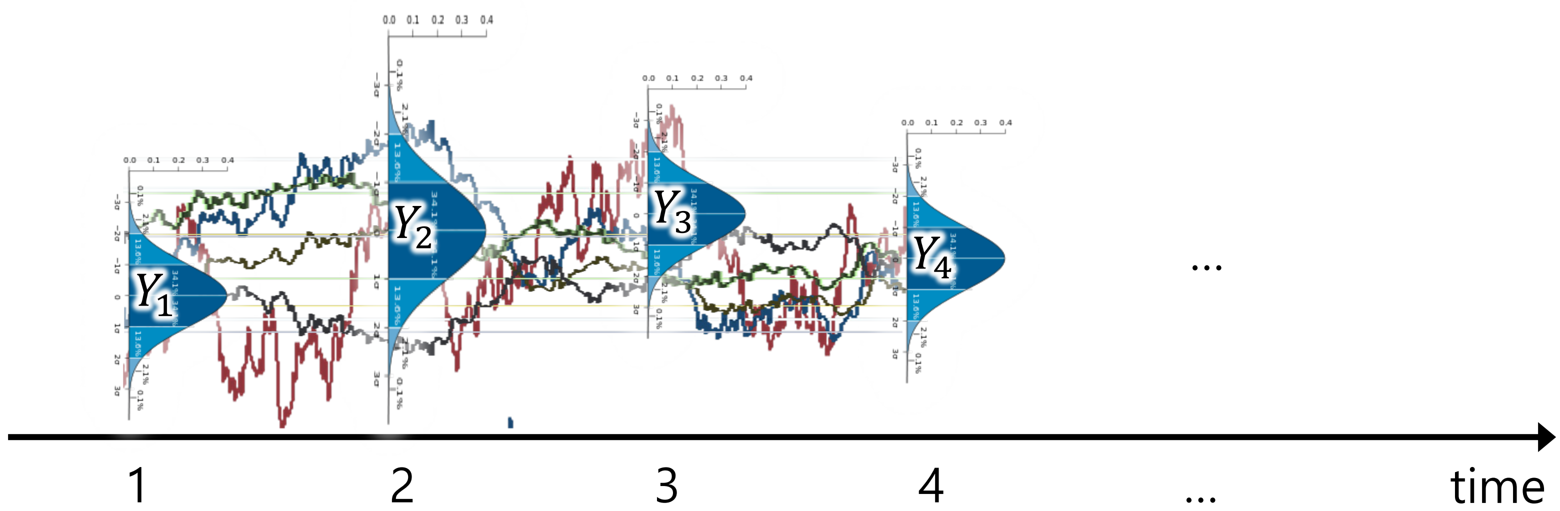


vs



Time series modeling

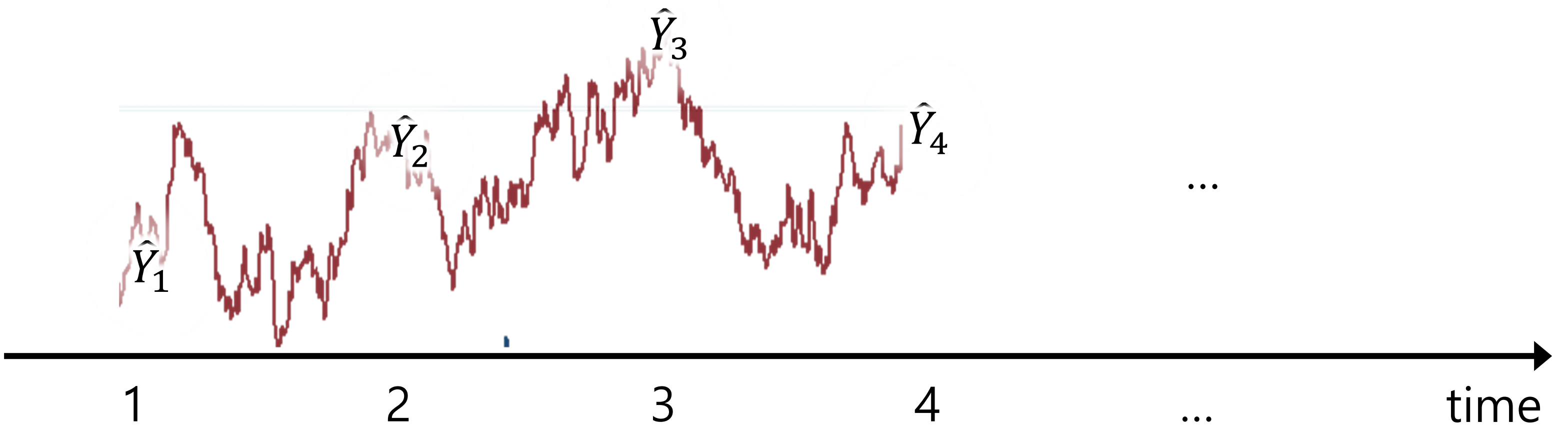
- **Major difficulties** arise in time series modeling



  – If we have a large number of samples, we would have no problem of finding random variables $Y_1, Y_2, Y_3, Y_4, \ldots$

Time series modeling

- **Major difficulties** arise in time series modeling



  – In time series analysis, it is common that we only observe a single realization (or just a few realizations)
  – Then, how should we model random variables $Y_1, Y_2, Y_3, Y_4, \ldots$? *Let's keep this in our mind*

Time series modeling

- Therefore,
  - 1. We are going to study the relationship between

    Y_t and $Y_{t-1}, Y_{t-2}, \ldots$

    instead of

    each of $Y_1, Y_2, Y_3, \ldots$
  - 2. We will assume that the above relationship is consistent throughout the time series
- We will get into much details as we go on

Chapter 2.2

Means, Variances, and Covariances

Stochastic process

Stochastic process

- **Sequence of random variables** $\{Y_t: t = 0, \pm 1, \pm 2, \pm 3, \dots\}$ indexed by time point t
  - Y_t denotes the value taken by the series at the i th time period

- A stochastic process is determined by the set of distributions of all finite collections of the Y_t's

- Fortunately, we will not have to deal explicitly with these multivariate distributions

- Much of the information in these joint distributions can be described in terms of **means, variances, and covariances**

Mean function

- For a stochastic process $\{Y_t: t = 0, \pm 1, \pm 2, \pm 3, \dots\}$,
  - The **mean function** is defined by

$$\mu_t = E(Y_t) \qquad \text{for } t = 0, \pm 1, \pm 2, \dots$$

    - That is, μ_t is just the expected value of the process at time t
    - In general, μ_t can be different at each time point t

Autocovariance function

- For a stochastic process $\{Y_t: t = 0, \pm 1, \pm 2, \pm 3, \dots\}$,

  – The **autocovariance function** is defined by

  $$\gamma_{t,s} = Cov(Y_t, Y_s) \qquad \text{for } t, s = 0, \pm 1, \pm 2, \dots$$

  • Where $Cov(Y_t, Y_s) = E[(Y_t - \mu_t)(Y_s - \mu_s)] = E(Y_t Y_s) - \mu_t \mu_s$

Autocorrelation function (ACF)

- For a stochastic process $\{Y_t : t = 0, \pm 1, \pm 2, \pm 3, \dots\}$,
  - The **autocorrelation function (ACF)** is defined by

$$\rho_{t,s} = Corr(Y_t, Y_s) \qquad \text{for } t, s = 0, \pm 1, \pm 2, \dots$$

- Where $Corr(Y_t, Y_s) = \frac{Cov(Y_t, Y_s)}{\sqrt{Var(Y_t) Var(Y_s)}} = \frac{\gamma_{t,s}}{\sqrt{\gamma_{t,t}\gamma_{s,s}}}$

You may refer to the basic properties of expectation, variance, covariance, and correlation in Appendix A on page 24 of our textbook

Measures of dependence

- Both **covariance** and **correlations** are **measures of the (linear) dependence** between random variables, but the unitless correlation is somewhat easier to interpret

- **Some important properties**

$$\gamma_{t,t} = Var(Y_t) \qquad \rho_{t,t} = 1$$

$$\gamma_{t,s} = \gamma_{s,t} \qquad \rho_{t,s} = \rho_{s,t}$$

$$|\gamma_{t,s}| \leq \sqrt{\gamma_{t,t}\gamma_{s,s}} \qquad |\rho_{t,s}| \leq 1$$

  – Values of $\rho_{t,s}$ near ± 1 indicate strong (linear) dependence, whereas values near zero indicate weak (linear) dependence
    - If $\rho_{t,s} = 0$, we say that Y_t and Y_s are uncorrelated

Measures of dependence

- Both **covariance** and **correlations** are **measures of the (linear) dependence** between random variables, but the unitless correlation is somewhat easier to interpret

- **Some important properties**
  - If $c_1, c_2, \dots, c_m$ and $d_1, d_2, \dots, d_n$ are constants and $t_1, t_2, \dots, t_m$ and $s_1, s_2, \dots, s_n$ are time points, then

$$Cov\left[\boxed{\sum_{i=1}^{m} c_i Y_{t_i}}, \boxed{\sum_{j=1}^{n} d_j Y_{s_j}}\right] = \sum_{i=1}^{m}\sum_{j=1}^{n} c_i d_j Cov(Y_{t_i}, Y_{s_j})$$

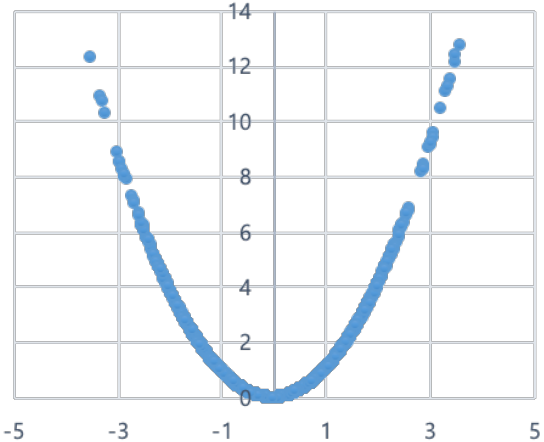
Linear combinations of values in time series are called **filtered** series

Measures of dependence

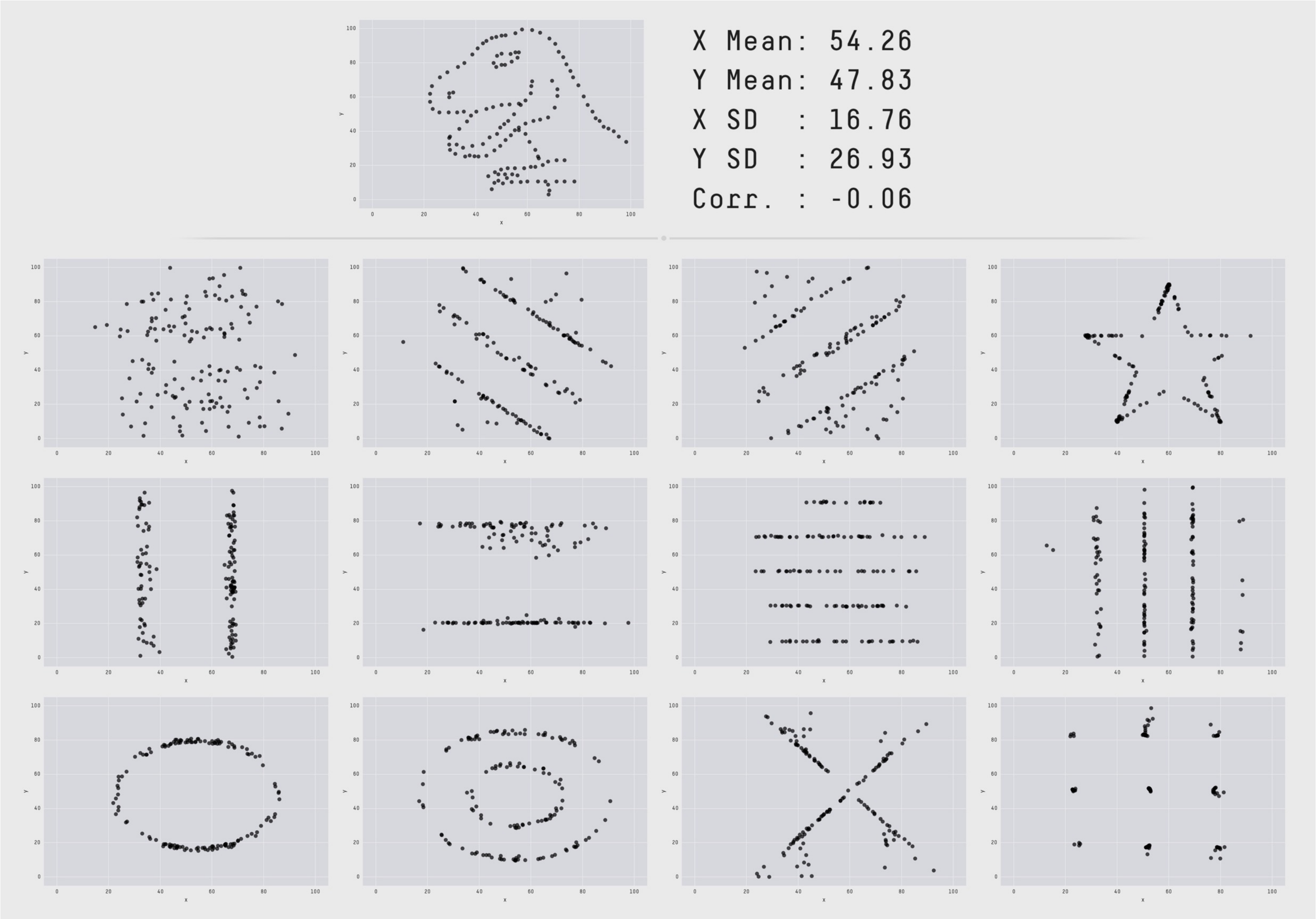
- Both **covariance** and **correlations** are **measures of the (linear) dependence** between random variables, but the unitless correlation is somewhat easier to interpret
  - Very **smooth** series exhibit autocovariance (or autocorrelation) functions that stay large even when the two time points are far apart
  - Whereas **choppy** series tend to have autocovariance (or autocorrelation) functions that are nearly zero for large separations

Measures of dependence

- Both **covariance** and **correlations** are **measures of the (linear) dependence** between random variables, but the unitless correlation is somewhat easier to interpret
  - Recall from classical statistics that even when $\rho_{t,s} = 0$, there still may be some dependence structure between them
  - Example of correlation (not autocorrelation)
    - X is a standard normal random variable
    - Let $Y = X^2$
      - Then, ρ_{XY} would be 0
      - But they have obvious dependence



Measures of dependence



Source: http://www.thefunctionalart.com/2016/08/download-datasaurus-never-trust-summary.html

Random walk

Random walk

- Let $e_1, e_2, \ldots$ be a sequence of independent and identically distributed (i.i.d.) random variables with zero mean and variance σ_e^2.
- The observed time series $\{Y_t: t = 1,2, \ldots\}$ is constructed as follows:

$$\begin{aligned} Y_1 &= e_1 \\ Y_2 &= e_1 + e_2 \\ &\vdots \\ Y_t &= e_1 + e_2 + \cdots + e_t \end{aligned}$$

- Alternatively, we can write

$$Y_t = Y_{t-1} + e_t$$

  - With initial condition $Y_1 = e_1$

Random walk

- **Random walk**
  - **Mean**

$$
\begin{aligned}
\mu_t &= E(Y_t) \\
&= E(e_1 + e_2 + \cdots + e_t) \\
&= E(e_1) + E(e_2) + \cdots + E(e_t) \\
&= 0 + 0 + \cdots + 0 \\
&= 0
\end{aligned}
$$

  - Therefore, $\mu_t = 0$ for all t

Random walk

- **Random walk**
  - **Variance**

$$
\begin{aligned}
Var(Y_t) &= Var(e_1 + e_2 + \cdots + e_t) \\
&= Var(e_1) + Var(e_2) + \cdots + Var(e_t) \\
&= \sigma_e^2 + \sigma_e^2 + \cdots + \sigma_e^2 \\
&= t\sigma_e^2
\end{aligned}
$$

  - Therefore, $Var(Y_t) = t\sigma_e^2$
    - Variance of random walk increases linearly with time

Random walk

- **Random walk**
  - **Autocovariance** (for $1 \leq t \leq s$)

$$\begin{aligned}\gamma_{t,s} &= Cov(Y_t, Y_s) \\ &= Cov(e_1 + e_2 + \cdots + e_t, e_1 + e_2 + \cdots + e_s) \\ &= \sum_{i=1}^{s} \sum_{j=1}^{t} Cov(e_i, e_j) \\ &= t\sigma_e^2\end{aligned}$$

From page 16

- Recall that
  - $Cov(e_i, e_j) = \sigma_e^2$ when $i = j$
  - $Cov(e_i, e_j) = 0$ otherwise

– Therefore, $\gamma_{t,s} = t\sigma_e^2$ for $1 \leq t \leq s$

Random walk

- **Random walk**
  - Autocorrelation (for $1 \le t \le s$)

$$\rho_{t,s} = \frac{\gamma_{t,s}}{\sqrt{\gamma_{t,t}\gamma_{s,s}}} = \frac{t\sigma_e^2}{\sqrt{t\sigma_e^2 s\sigma_e^2}} = \sqrt{\frac{t}{s}}$$

  - Note that

$$\rho_{1,2} = \sqrt{\frac{1}{2}} = 0.707 \qquad \rho_{8,9} = \sqrt{\frac{8}{9}} = 0.943$$

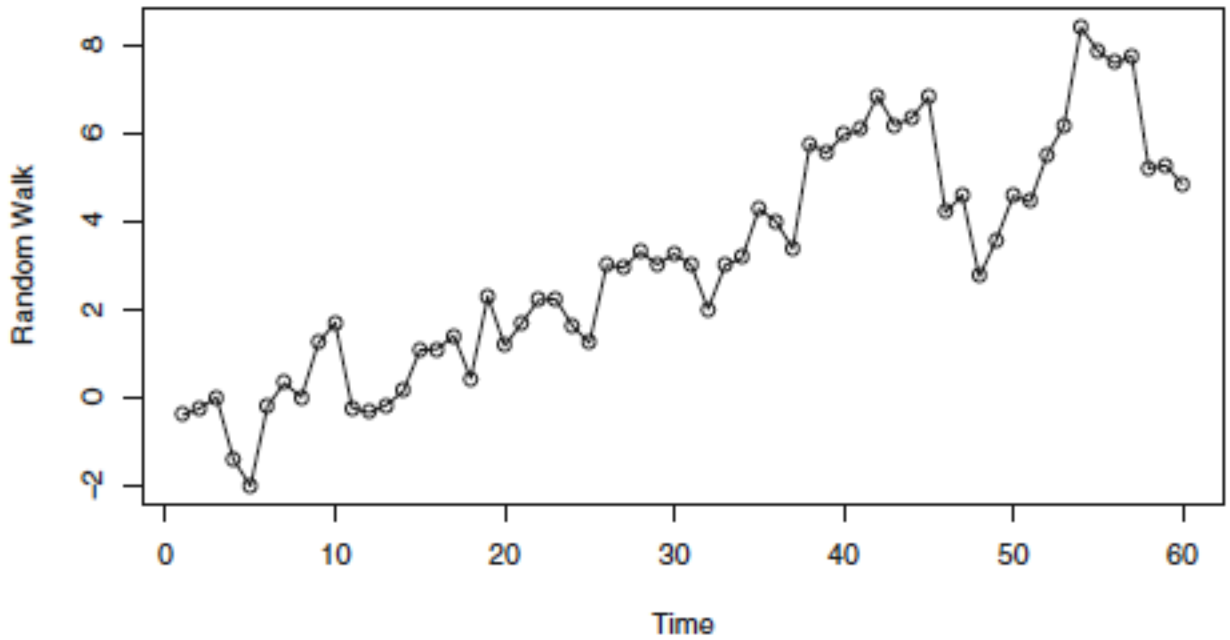
$$\rho_{24,25} = \sqrt{\frac{24}{25}} = 0.980 \qquad \rho_{1,25} = \sqrt{\frac{1}{25}} = 0.200$$

    - The values of Y at neighboring time points are more and more strongly and positively correlated as time goes by
    - On the other hand, the values of Y at distant time points are less and less correlated

Random walk

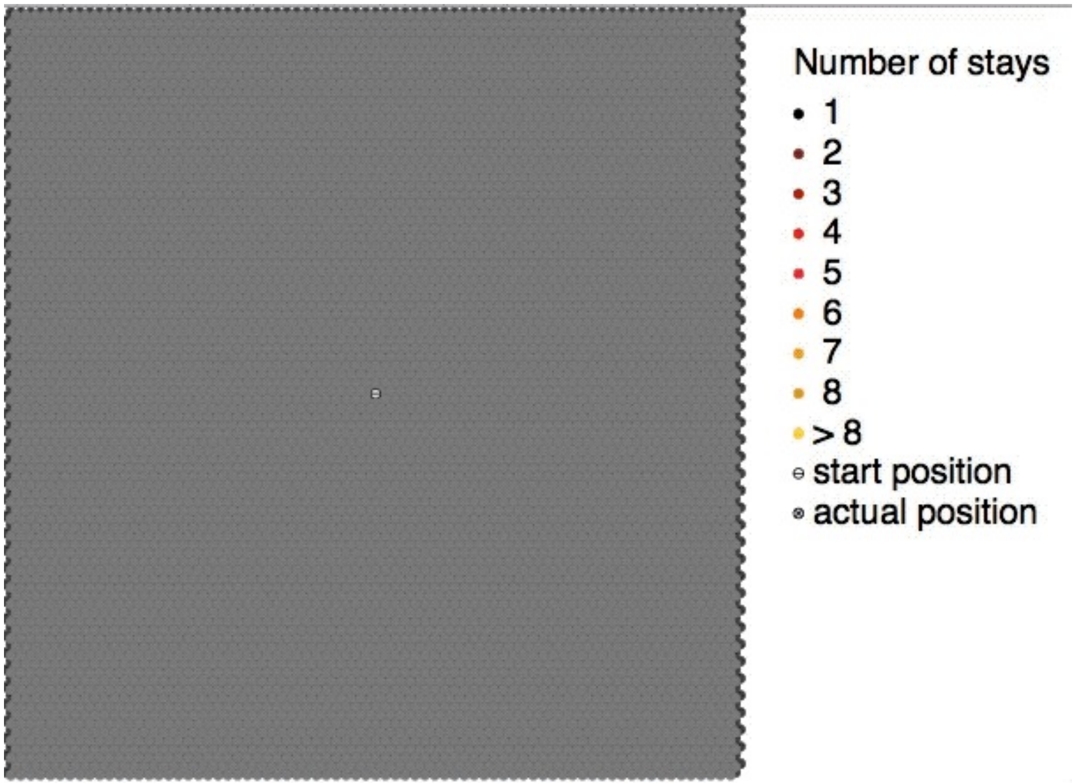
- **Random walk**

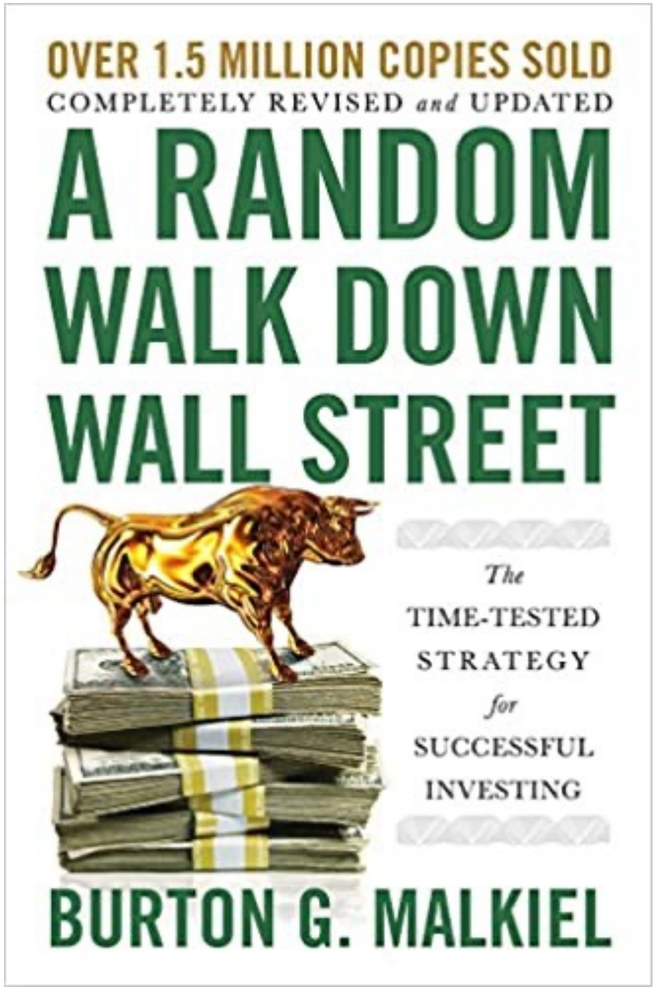
Exhibit 2.1 Time Series Plot of a Random Walk



Random walk

- **Random walk**





Moving average

Moving average

- Suppose that $\{Y_t\}$ is constructed as

$$Y_t = \frac{e_t + e_{t-1}}{2}$$

An example of **filtered** series

- **Mean**

$$\mu_t = E(Y_t) = E\left\{\frac{e_t + e_{t-1}}{2}\right\} = \frac{E(e_t) + E(e_{t-1})}{2} = 0$$

- **Variance**

$$Var(Y_t) = Var\left\{\frac{e_t + e_{t-1}}{2}\right\} = \frac{Var(e_t) + Var(e_{t-1})}{4} = \frac{1}{2}\sigma_e^2$$

Moving average

- **Moving average**
  - Autocovariance
    - First, let's see

$$Cov(Y_t, Y_{t-1}) = Cov\left\{\frac{e_t+e_{t-1}}{2}, \frac{e_{t-1}+e_{t-2}}{2}\right\}$$
$$= \frac{Cov(e_t,e_{t-1})+Cov(e_t,e_{t-2})+Cov(e_{t-1},e_{t-1})+Cov(e_{t-1},e_{t-2})}{4}$$
$$= \frac{Cov(e_{t-1},e_{t-1})}{4}$$
$$= \frac{1}{4}\sigma_e^2$$

    - Furthermore,

$$Cov(Y_t, Y_{t-2}) = Cov\left\{\frac{e_t+e_{t-1}}{2}, \frac{e_{t-2}+e_{t-3}}{2}\right\} = 0$$

      - Since e_t's are independent for different t's

    - Similarly, $Cov(Y_t, Y_{t-k}) = 0$ for $k > 1$

Moving average

- **Moving average**
  - Autocovariance
    - Therefore,

$$\gamma_{t,s} = \begin{cases} 0.5\sigma_e^2 & \text{for } |t-s| = 0 \\ 0.25\sigma_e^2 & \text{for } |t-s| = 1 \\ 0 & \text{for } |t-s| > 1 \end{cases}$$

  - Autocorrelation
    - Similarly,

$$\rho_{t,s} = \begin{cases} 1 & \text{for } |t-s| = 0 \\ 0.5 & \text{for } |t-s| = 1 \\ 0 & \text{for } |t-s| > 1 \end{cases}$$

Moving average

- **Moving average**
  - **Autocorrelation**
    - Similarly,

$$\rho_{t,s} = \begin{cases} 1 & \text{for } |t-s| = 0 \\ 0.5 & \text{for } |t-s| = 1 \\ 0 & \text{for } |t-s| > 1 \end{cases}$$

    - Unlike random walks,
      - $\rho_{2,1} = \rho_{3,2} = \rho_{4,3} = \rho_{9,8} = 0.5$
      - $\rho_{3,1} = \rho_{4,2} = \rho_{t,t-2}$
      - More generally, $\rho_{t,t-k}$ is the same for all values of t
      - As long as the distance between the two time points are the same, it doesn’t matter where they occur in time

Moving average

Chapter 2.3

Stationarity

Stationarity

- To make statistical inferences about the structure of a stochastic process, it is helpful to make some **simplifying assumptions** about that structure

- The most important such assumption is that of **stationarity**

- The basic idea of **stationarity** is that the **probability laws** that govern the behavior of the process **do not change over time**
  - In a sense, the process is in **statistical equilibrium**

Strict stationarity

Strict stationarity

– A process $\{Y_t\}$ is said to be **strictly stationary**

if the joint distribution of $Y_{t_1}, Y_{t_2}, \dots, Y_{t_n}$ is the same as

the joint distribution of $Y_{t_1-k}, Y_{t_2-k}, \dots, Y_{t_n-k}$

for all choices of time points $t_1, t_2, \dots, t_n$

and all choices of time lag k

Strict stationarity

- **Strict stationarity**
  - For $n = 1$,
    - The (univariate) distribution of Y_t is the same as that of Y_{t-k} for all t and k
      - That is, Y's are (marginally) identically distributed
    - Also, $E(Y_t) = E(Y_{t-k})$ for all t and k
      - Mean function is constant over time
    - And $Var(Y_t) = Var(Y_{t-k})$ for all t and k
      - Variance is also constant over time

Strict stationarity

- **Strict stationarity**
  - For $n = 2$,
    - The bivariate distribution of Y_t and Y_s must be the same as that of Y_{t-k} and Y_{s-k} for all t, s and k
    - It follows that $Cov(Y_t, Y_s) = Cov(Y_{t-k}, Y_{s-k})$ for all t, s and k
      - If we put $k = s$ and then $k = t$, we have

$$
\begin{aligned}
\gamma_{t,s} &= Cov(Y_{t-s}, Y_0) && \leftarrow \boxed{k = s} \\
&= Cov(Y_0, Y_{s-t}) && \leftarrow \boxed{k = t} \\
&= Cov\left(Y_0, Y_{|t-s|}\right) \\
&= \gamma_{0,|t-s|}
\end{aligned}
$$

      - That is, autocovariance depends on time only through the time difference $|t - s|$ and not otherwise on the actual times t and s

Strict stationarity

- **Strict stationarity**
  - For $n = 2$,
    - Thus, for a stationary process, we can simplify our notation and write

$$\gamma_k = Cov(Y_t, Y_{t-k}) \quad \text{and} \quad \rho_k = Corr(Y_t, Y_{t-k})$$

      - Also, $\rho_k = \frac{\gamma_k}{\gamma_0}$
    - General properties for a stationary process

$$\gamma_0 = Var(Y_t) \qquad \rho_0 = 1$$

$$\gamma_k = \gamma_{-k} \qquad \rho_k = \rho_{-k}$$

$$|\gamma_k| \leq \gamma_0 \qquad |\rho_k| \leq 1$$

Strict stationarity

- **One problem**
  - Strict stationarity is **too strong** for most applications
    - Moreover, it is difficult to assess strict stationarity from a single data set
  - Therefore, we would need a **milder version** that imposes conditions **only on the first two moments** of the series
    - Rather than imposing conditions on all possible distributions of a time series

(Weak) Stationarity

(Weak) Stationarity

- A process $\{Y_t\}$ is said to be **weakly** (or **second-order**) **stationary** if
  - The mean function is constant over time
  - $\gamma_{t,t-k} = \gamma_{0,k}$ for all time t and lag k

- Henceforth, the term **'stationary'** when used alone will always refer to this weaker form of stationarity
  - But if the joint distributions for the process are all multi-variate normal distributions, the two definitions coincide

White noise

White noise

- A very important example of a stationary process is the so-called **white noise** process
  - Defined as a sequence of i.i.d. random variables $\{e_t\}$
    - Usually assume that it has mean zero and variance $Var(e_t) = \sigma_e^2$
  - Its strict stationarity is easy to see

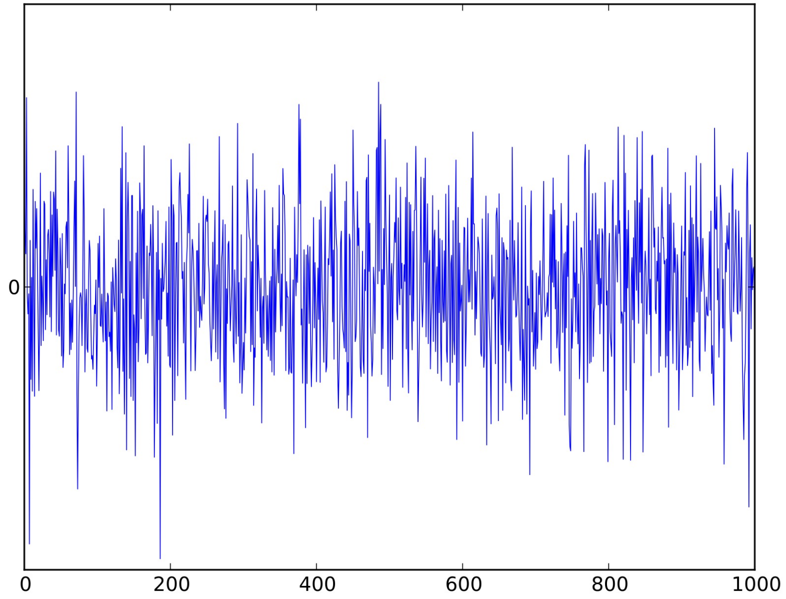
$$
\begin{aligned}
&\Pr\left(e_{t_1} \le x_1, e_{t_2} \le x_2, \dots, e_{t_n} \le x_n\right) && \text{by independence} \\
&= \Pr\left(e_{t_1} \le x_1\right) \Pr\left(e_{t_2} \le x_2\right) \cdots \Pr\left(e_{t_n} \le x_n\right) && \text{identical distributions} \\
&= \Pr\left(e_{t_1-k} \le x_1\right) \Pr\left(e_{t_2-k} \le x_2\right) \cdots \Pr\left(e_{t_n-k} \le x_n\right) && \text{by independence} \\
&= \Pr\left(e_{t_1-k} \le x_1, e_{t_2-k} \le x_2, \dots, e_{t_n-k} \le x_n\right)
\end{aligned}
$$

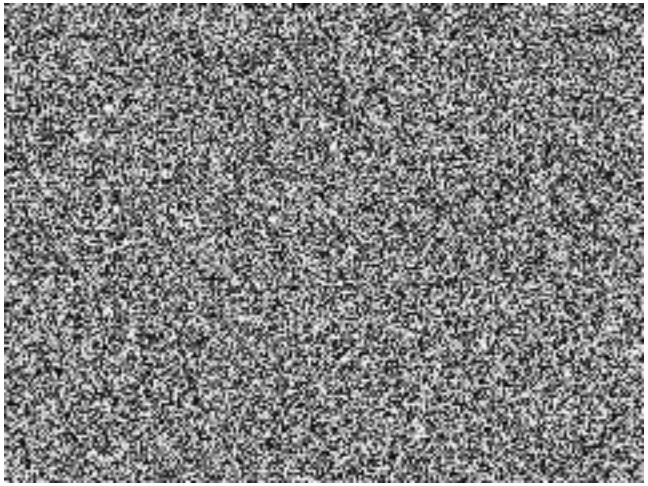
    - Also, $\mu_t = E(e_t)$ is constant and $\gamma_k = \begin{cases} Var(e_t) & \text{for } k = 0 \\ 0 & \text{for } k \neq 0 \end{cases}$
    - Alternatively, $\rho_k = \begin{cases} 1 & \text{for } k = 0 \\ 0 & \text{for } k \neq 0 \end{cases}$

White noise

White noise

- A very important example of a stationary process is the so-called **white noise** process
  - Defined as a sequence of i.i.d. random variables $\{e_t\}$
    - Usually assume that it has mean zero and variance $Var(e_t) = \sigma_e^2$





White noise

▪ White noise

– A very important example of a stationary process is the so-called **white noise** process

  - Many useful processes can be constructed from white noise

    › **Example: moving average**

$$Y_t = \frac{e_t + e_{t-1}}{2}$$

      » In our new notation, the process has

$$\rho_k = \begin{cases} 1 & \text{for } k = 0 \\ 0.5 & \text{for } |k| = 1 \\ 0 & \text{for } |k| \geq 2 \end{cases}$$

White noise

▪ White noise

– A very important example of a stationary process is the so-called **white noise** process

- Many useful processes can be constructed from white noise

› **Example: random walk**

$$Y_t = Y_{t-1} + e_t, \qquad Y_1 = e_1$$

» Note that random walk is constructed from white noise but is **not stationary**

- $Var(Y_t) = t\sigma_e^2$ is not constant
- $\gamma_{t,s} = Cov(Y_t, Y_s) = t\sigma_e^2$ for $0 \leq t \leq 2$ does not depend only on time lag $|t - s|$