Gaussian, Markov and stationary processes

Gonzalo Mateos
Dept. of ECE and Goergen Institute for Data Science
University of Rochester
gmateosb@ece.rochester.edu
http://www.ece.rochester.edu/~gmateosb/

November 15, 2019

Random processes

- Random processes assign a function $X(t)$ to a random event
  - $\Rightarrow$ Without restrictions, there is little to say about them
  - $\Rightarrow$ Markov property simplifies matters and is not too restrictive
- Also constrained ourselves to discrete state spaces
  - $\Rightarrow$ Further simplification but might be too restrictive
- Time t and range of $X(t)$ values continuous in general
  - Time and/or state may be discrete as particular cases
- Restrict attention to (any type or a combination of types)
  - $\Rightarrow$ Markov processes (memoryless)
  - $\Rightarrow$ Gaussian processes (Gaussian probability distributions)
  - $\Rightarrow$ Stationary processes ("limit distribution")

Markov processes

- $X(t)$ is a Markov process when the future is independent of the past
- For all $t > s$ and arbitrary values $x(t)$, $x(s)$ and $x(u)$ for all $u < s$

$$\mathrm{P}\left(X(t) \leq x(t) \,\middle|\, X(s) \leq x(s), X(u) \leq x(u), u < s\right) \\ = \mathrm{P}\left(X(t) \leq x(t) \,\middle|\, X(s) \leq x(s)\right)$$

  $\Rightarrow$ Markov property defined in terms of cdfs, not pmfs

- Markov property useful for same reasons as in discrete time/state
  $\Rightarrow$ But not that useful as in discrete time /state
- More details later

Gaussian processes

- $X(t)$ is a Gaussian process when all prob. distributions are Gaussian
- For arbitrary $n > 0$, times $t_1, t_2, \ldots, t_n$ it holds
  $\Rightarrow$ Values $X(t_1), X(t_2), \ldots, X(t_n)$ are jointly Gaussian RVs
- Simplifies study because Gaussian distribution is simplest possible
  $\Rightarrow$ Suffices to know mean, variances and (cross-)covariances
  $\Rightarrow$ Linear transformation of independent Gaussians is Gaussian
  $\Rightarrow$ Linear transformation of jointly Gaussians is Gaussian
- More details later

Markov processes + Gaussian processes

- Markov (memoryless) and Gaussian properties are different
  - ⇒ Will study cases when both hold
- Brownian motion, also known as Wiener process
  - ⇒ Brownian motion with drift
  - ⇒ White noise ⇒ Linear evolution models
- Geometric brownian motion
  - ⇒ Arbitrages
  - ⇒ Risk neutral measures
  - ⇒ Pricing of stock options (Black-Scholes)

Stationary processes

- Process $X(t)$ is stationary if probabilities are invariant to time shifts
- For arbitrary $n > 0$, times $t_1, t_2, \ldots, t_n$ and arbitrary time shift s

$$P\left(X(t_1+s) \leq x_1, X(t_2+s) \leq x_2, \ldots, X(t_n+s) \leq x_n\right) = P\left(X(t_1) \leq x_1, X(t_2) \leq x_2, \ldots, X(t_n) \leq x_n\right)$$

$\Rightarrow$ System's behavior is independent of time origin

- Follows from our success studying limit probabilities

$\Rightarrow$ Study of stationary process $\approx$ Study of limit distribution

- Will study $\Rightarrow$ Spectral analysis of stationary random processes
  $\Rightarrow$ Linear filtering of stationary random processes
- More details later

Jointly Gaussian random variables

- **Def:** Random variables $X_1, \ldots, X_n$ are jointly Gaussian (normal) if any linear combination of them is Gaussian
  - $\Rightarrow$ Given $n > 0$, for any scalars $a_1, \ldots, a_n$ the RV ($\mathbf{a} = [a_1, \ldots, a_n]^T$)

$$Y = a_1X_1 + a_2X_2 + \ldots + a_nX_n = \mathbf{a}^T\mathbf{X} \text{ is Gaussian distributed}$$

  - $\Rightarrow$ May also say vector RV $\mathbf{X} = [X_1, \ldots, X_n]^T$ is Gaussian
- Consider 2 dimensions $\Rightarrow$ 2 RVs X_1 and X_2 are jointly normal
- To describe joint distribution have to specify
  - $\Rightarrow$ Means: $\mu_1 = \mathbb{E}[X_1]$ and $\mu_2 = \mathbb{E}[X_2]$
  - $\Rightarrow$ Variances: $\sigma_{11}^2 = \text{var}[X_1] = \mathbb{E}\left[(X_1 - \mu_1)^2\right]$ and $\sigma_{22}^2 = \text{var}[X_2]$
  - $\Rightarrow$ Covariance: $\sigma_{12}^2 = \text{cov}(X_1, X_2) = \mathbb{E}[(X_1 - \mu_1)(X_2 - \mu_2)] = \sigma_{21}^2$

Pdf of jointly Gaussian RVs in 2 dimensions

- Define mean vector $\boldsymbol{\mu} = [\mu_1, \mu_2]^T$ and covariance matrix $\mathbf{C} \in \mathbb{R}^{2\times 2}$

$$\mathbf{C} = \begin{pmatrix} \sigma_{11}^2 & \sigma_{12}^2 \\ \sigma_{21}^2 & \sigma_{22}^2 \end{pmatrix}$$

$\Rightarrow$ $\mathbf{C}$ is symmetric, i.e., $\mathbf{C}^T = \mathbf{C}$ because $\sigma_{21}^2 = \sigma_{12}^2$

- Joint pdf of $\mathbf{X} = [X_1, X_2]^T$ is given by

$$f_{\mathbf{X}}(\mathbf{x}) = \frac{1}{2\pi \det^{1/2}(\mathbf{C})} \exp\left(-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^T \mathbf{C}^{-1}(\mathbf{x}-\boldsymbol{\mu})\right)$$

$\Rightarrow$ Assumed that $\mathbf{C}$ is invertible, thus $\det(\mathbf{C}) \neq 0$

- If the pdf of $\mathbf{X}$ is $f_{\mathbf{X}}(\mathbf{x})$ above, can verify $Y = \mathbf{a}^T\mathbf{X}$ is Gaussian

Pdf of jointly Gaussian RVs in n dimensions

- For $\mathbf{X} \in \mathbb{R}^n$ (n dimensions) define $\boldsymbol{\mu} = \mathbb{E}[\mathbf{X}]$ and covariance matrix

$$\mathbf{C} := \mathbb{E}\left[(\mathbf{X}-\boldsymbol{\mu})(\mathbf{X}-\boldsymbol{\mu})^T\right] = \begin{pmatrix} \sigma_{11}^2 & \sigma_{12}^2 & \dots & \sigma_{1n}^2 \\ \sigma_{21}^2 & \sigma_{22}^2 & \dots & \sigma_{2n}^2 \\ \vdots & & \ddots & \vdots \\ \sigma_{n1}^2 & \sigma_{n2}^2 & \dots & \sigma_{nn}^2 \end{pmatrix}$$

  $\Rightarrow$ $\mathbf{C}$ symmetric, (i,j)-th element is $\sigma_{ij}^2 = \text{cov}(X_i, X_j)$

- Joint pdf of $\mathbf{X}$ defined as before (almost, spot the difference)

$$f_{\mathbf{X}}(\mathbf{x}) = \frac{1}{(2\pi)^{n/2}\det^{1/2}(\mathbf{C})}\exp\left(-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^T\mathbf{C}^{-1}(\mathbf{x}-\boldsymbol{\mu})\right)$$

  $\Rightarrow$ $\mathbf{C}$ invertible and $\det(\mathbf{C}) \neq 0$. All linear combinations normal

- To fully specify the probability distribution of a Gaussian vector $\mathbf{X}$

  $\Rightarrow$ The mean vector $\boldsymbol{\mu}$ and covariance matrix $\mathbf{C}$ suffice

Notational aside and independence

- With $\mathbf{x} \in \mathbb{R}^n$, $\boldsymbol{\mu} \in \mathbb{R}^n$ and $\mathbf{C} \in \mathbb{R}^{n \times n}$, define function $\mathcal{N}(\mathbf{x}; \boldsymbol{\mu}, \mathbf{C})$ as

$$\mathcal{N}(\mathbf{x}; \boldsymbol{\mu}, \mathbf{C}) := \frac{1}{(2\pi)^{n/2} \det^{1/2}(\mathbf{C})} \exp\left(-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu})^T \mathbf{C}^{-1} (\mathbf{x} - \boldsymbol{\mu})\right)$$

$\Rightarrow$ $\boldsymbol{\mu}$ and $\mathbf{C}$ are parameters, $\mathbf{x}$ is the argument of the function

- Let $\mathbf{X} \in \mathbb{R}^n$ be a Gaussian vector with mean $\boldsymbol{\mu}$, and covariance $\mathbf{C}$
  $\Rightarrow$ Can write the pdf of $\mathbf{X}$ as $f_{\mathbf{X}}(\mathbf{x}) = \mathcal{N}(\mathbf{x}; \boldsymbol{\mu}, \mathbf{C})$

- If $X_1, \ldots, X_n$ are mutually independent, then $\mathbf{C} = \text{diag}(\sigma_{11}^2, \ldots, \sigma_{nn}^2)$ and

$$f_{\mathbf{X}}(\mathbf{x}) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma_{ii}^2}} \exp\left(-\frac{(x_i - \mu_i)^2}{2\sigma_{ii}^2}\right)$$

Gaussian processes

- Gaussian processes (GP) generalize Gaussian vectors to infinite dimensions
- **Def:** $X(t)$ is a GP if any linear combination of values $X(t)$ is Gaussian

  $\Rightarrow$ For arbitrary $n > 0$, times $t_1, \ldots, t_n$ and constants $a_1, \ldots, a_n$

  $$Y = a_1X(t_1) + a_2X(t_2) + \ldots + a_nX(t_n) \text{ is Gaussian distributed}$$

  $\Rightarrow$ Time index t can be continuous or discrete
- More general, any linear functional of $X(t)$ is normally distributed

  $\Rightarrow$ A functional is a function of a function

Ex: The (random) integral $Y = \int_{t_1}^{t_2} X(t)\, dt$ is Gaussian distributed

$\Rightarrow$ Integral functional is akin to a sum of $X(t_i)$, for all $t_i \in [t_1, t_2]$

Joint pdfs in a Gaussian process

- Consider times $t_1, \ldots, t_n$. The mean value $\mu(t_i)$ at such times is

$$\mu(t_i) = \mathbb{E}\left[X(t_i)\right]$$

- The covariance between values at times t_i and t_j is

$$C(t_i, t_j) = \mathbb{E}\left[\big(X(t_i) - \mu(t_i)\big)\big(X(t_j) - \mu(t_j)\big)\right]$$

- Covariance matrix for values $X(t_1), \ldots, X(t_n)$ is then

$$\mathbf{C}(t_1, \ldots, t_n) = \begin{pmatrix} C(t_1, t_1) & C(t_1, t_2) & \ldots & C(t_1, t_n) \\ C(t_2, t_1) & C(t_2, t_2) & \ldots & C(t_2, t_n) \\ \vdots & \vdots & \ddots & \vdots \\ C(t_n, t_1) & C(t_n, t_2) & \ldots & C(t_n, t_n) \end{pmatrix}$$

- Joint pdf of $X(t_1), \ldots, X(t_n)$ then given as

$$f_{X(t_1),\ldots,X(t_n)}(x_1, \ldots, x_n) = \mathcal{N}\left(\left[x_1, \ldots, x_n\right]^T; \left[\mu(t_1), \ldots, \mu(t_n)\right]^T, \mathbf{C}(t_1, \ldots, t_n)\right)$$

Mean value and autocorrelation functions

- To specify a Gaussian process, suffices to specify:
  - $\Rightarrow$ Mean value function $\Rightarrow \mu(t) = \mathbb{E}[X(t)]$; and
  - $\Rightarrow$ Autocorrelation function $\Rightarrow R(t_1, t_2) = \mathbb{E}\big[X(t_1)X(t_2)\big]$
- Autocovariance obtained as $C(t_1, t_2) = R(t_1, t_2) - \mu(t_1)\mu(t_2)$
- For simplicity, will mostly consider processes with $\mu(t) = 0$
  - $\Rightarrow$ Otherwise, can define process $Y(t) = X(t) - \mu_X(t)$
  - $\Rightarrow$ In such case $C(t_1, t_2) = R(t_1, t_2)$ because $\mu_Y(t) = 0$
- Autocorrelation is a symmetric function of two variables t_1 and t_2

$$R(t_1, t_2) = R(t_2, t_1)$$

Probabilities in a Gaussian process

- All probs. in a GP can be expressed in terms of $\mu(t)$ and $R(t_1, t_2)$
- For example, pdf of $X(t)$ is

$$f_{X(t)}(x_t) = \frac{1}{\sqrt{2\pi\big(R(t,t) - \mu^2(t)\big)}} \exp\left(-\frac{\big(x_t - \mu(t)\big)^2}{2\big(R(t,t) - \mu^2(t)\big)}\right)$$

- Notice that $\frac{X(t)-\mu(t)}{\sqrt{R(t,t)-\mu^2(t)}}$ is a standard Gaussian random variable

  $\Rightarrow \mathrm{P}\left(X(t) > a\right) = \Phi\left(\frac{a-\mu(t)}{\sqrt{R(t,t)-\mu^2(t)}}\right)$, where

$$\Phi(x) = \int_x^\infty \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right) dx$$

Joint and conditional probabilities in a GP

- For a zero-mean GP $X(t)$ consider two times t_1 and t_2
- The covariance matrix for $X(t_1)$ and $X(t_2)$ is

$$\mathbf{C} = \begin{pmatrix} R(t_1, t_1) & R(t_1, t_2) \\ R(t_1, t_2) & R(t_2, t_2) \end{pmatrix}$$

- Joint pdf of $X(t_1)$ and $X(t_2)$ then given as (recall $\mu(t) = 0$)

$$f_{X(t_1),X(t_2)}(x_{t_1}, x_{t_2}) = \frac{1}{2\pi \det^{1/2}(\mathbf{C})} \exp\left(-\frac{1}{2}[x_{t_1}, x_{t_2}]^T \mathbf{C}^{-1} [x_{t_1}, x_{t_2}]\right)$$

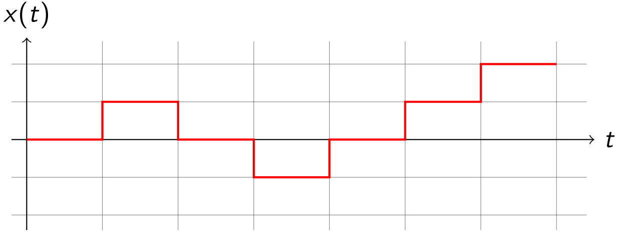
- Conditional pdf of $X(t_1)$ given $X(t_2)$ computed as

$$f_{X(t_1)|X(t_2)}(x_{t_1} \,\big|\, x_{t_2}) = \frac{f_{X(t_1),X(t_2)}(x_{t_1}, x_{t_2})}{f_{X(t_2)}(x_{t_2})}$$

Brownian motion as limit of random walk

- Gaussian processes are natural models due to Central Limit Theorem
- Let us reconsider a symmetric random walk in one dimension

Time interval $= h$

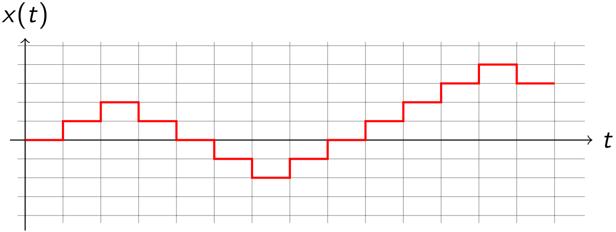


- Walker takes increasingly frequent and increasingly smaller steps

Brownian motion as limit of random walk

- Gaussian processes are natural models due to Central Limit Theorem
- Let us reconsider a symmetric random walk in one dimension

Time interval $= h/2$

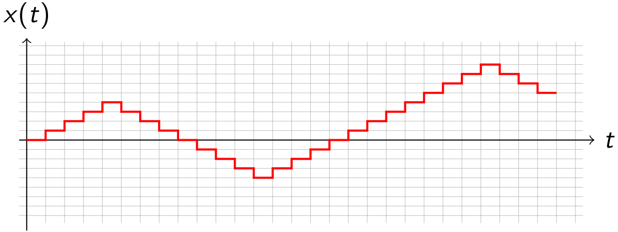


- Walker takes increasingly frequent and increasingly smaller steps

Brownian motion as limit of random walk

- Gaussian processes are natural models due to Central Limit Theorem
- Let us reconsider a symmetric random walk in one dimension

Time interval $= h/4$



- Walker takes increasingly frequent and increasingly smaller steps

Random walk, time step h and step size $\sigma\sqrt{h}$

- Let $X(t)$ be the position at time t with $X(0) = 0$
  - $\Rightarrow$ Time interval is h and $\sigma\sqrt{h}$ is the size of each step
  - $\Rightarrow$ Walker steps right or left w.p. $1/2$ for each direction
- Given $X(t) = x$, prob. distribution of the position at time $t + h$ is

$$\mathrm{P}\left(X(t+h) = x + \sigma\sqrt{h} \,\middle|\, X(t) = x\right) = 1/2$$
$$\mathrm{P}\left(X(t+h) = x - \sigma\sqrt{h} \,\middle|\, X(t) = x\right) = 1/2$$

- Consider time $T = Nh$ and index $n = 1, 2, \ldots, N$
  - $\Rightarrow$ Introduce step RVs $Y_n = \pm 1$, with $\mathrm{P}(Y_n = \pm 1) = 1/2$
  - $\Rightarrow$ Can write $X(nh)$ in terms of $X((n-1)h)$ and Y_n as

$$X(nh) = X((n-1)h) + \left(\sigma\sqrt{h}\right) Y_n$$

Central Limit Theorem as $h \to 0$

- Use recursion to write $X(T) = X(Nh)$ as (recall $X(0) = 0$)

$$X(T) = X(Nh) = X(0) + \left(\sigma\sqrt{h}\right)\sum_{n=1}^{N} Y_n = \left(\sigma\sqrt{h}\right)\sum_{n=1}^{N} Y_n$$

- $Y_1, \ldots, Y_N$ are i.i.d. with zero-mean and variance

$$\text{var}[Y_n] = \mathbb{E}\left[Y_n^2\right] = (1/2) \times 1^2 + (1/2) \times (-1)^2 = 1$$

- As $h \to 0$ we have $N = T/h \to \infty$, and from Central Limit Theorem

$$\sum_{n=1}^{N} Y_n \sim \mathcal{N}(0, N) = \mathcal{N}(0, T/h)$$

$$\Rightarrow X(T) \sim \mathcal{N}\left(0, (\sigma^2 h) \times (T/h)\right) = \mathcal{N}\left(0, \sigma^2 T\right)$$

Conditional distribution of future values

- More generally, consider times $T = Nh$ and $T + S = (N + M)h$
- Let $X(T) = x(T)$ be given. Can write $X(T + S)$ as

$$X(T+S) = x(T) + \left(\sigma\sqrt{h}\right) \sum_{n=N+1}^{N+M} Y_n$$

- From Central Limit Theorem it then follows

$$\sum_{n=N+1}^{N+M} Y_n \sim \mathcal{N}\big(0, (N+M-N)\big) = \mathcal{N}(0, S/h)$$

$$\Rightarrow \Big[X(T+S) \,\Big|\, X(T) = x(T)\Big] \sim \mathcal{N}(x(T), \sigma^2 S)$$

Definition of Brownian motion

- The former analysis was for motivational purposes
- **Def:** A Brownian motion process (a.k.a Wiener process) satisfies

  (i) $X(t)$ is normally distributed with zero mean and variance $\sigma^2 t$

  $$X(t) \sim \mathcal{N}\left(0, \sigma^2 t\right)$$

  (ii) Independent increments $\Rightarrow$ For disjoint intervals (t_1, t_2) and (s_1, s_2) increments $X(t_2) - X(t_1)$ and $X(s_2) - X(s_1)$ are independent RVs

  (iii) Stationary increments $\Rightarrow$ Probability distribution of increment $X(t+s) - X(s)$ is the same as probability distribution of $X(t)$

- Property (ii) $\Rightarrow$ Brownian motion is a Markov process
- Properties (i)-(iii) $\Rightarrow$ Brownian motion is a Gaussian process

Mean and autocorrelation of Brownian motion

- Mean function $\mu(t) = \mathbb{E}[X(t)]$ is null for all times (by definition)

$$\mu(t) = \mathbb{E}[X(t)] = 0$$

- For autocorrelation $R_X(t_1, t_2)$ start with times $t_1 < t_2$
- Use conditional expectations to write

$$R_X(t_1, t_2) = \mathbb{E}[X(t_1)X(t_2)] = \mathbb{E}_{X(t_1)}\Big[\mathbb{E}_{X(t_2)}\big[X(t_1)X(t_2) \,\big|\, X(t_1)\big]\Big]$$

- In the innermost expectation $X(t_1)$ is a given constant, then

$$R_X(t_1, t_2) = \mathbb{E}_{X(t_1)}\Big[X(t_1)\mathbb{E}_{X(t_2)}\big[X(t_2) \,\big|\, X(t_1)\big]\Big]$$

$\Rightarrow$ Proceed by computing innermost expectation

Autocorrelation of Brownian motion (continued)

- The conditional distribution of $X(t_2)$ given $X(t_1)$ for $t_1 < t_2$ is

$$\Big[X(t_2) \,\Big|\, X(t_1)\Big] \sim \mathcal{N}\Big(X(t_1), \sigma^2(t_2 - t_1)\Big)$$

$\Rightarrow$ Innermost expectation is $\mathbb{E}_{X(t_2)}\big[X(t_2) \,\big|\, X(t_1)\big] = X(t_1)$

- From where autocorrelation follows as

$$R_X(t_1, t_2) = \mathbb{E}_{X(t_1)}\big[X(t_1)X(t_1)\big] = \mathbb{E}_{X(t_1)}\big[X^2(t_1)\big] = \sigma^2 t_1$$

- Repeating steps, if $t_2 < t_1 \;\Rightarrow R_X(t_1, t_2) = \sigma^2 t_2$

- Autocorrelation of Brownian motion $\;\Rightarrow R_X(t_1, t_2) = \sigma^2 \min(t_1, t_2)$

Brownian motion with drift

- Similar to Brownian motion, but start from biased random walk
- Time interval h, step size $\sigma\sqrt{h}$, right or left with different probs.

$$\mathrm{P}\left(X(t+h) = x + \sigma\sqrt{h} \,\middle|\, X(t) = x\right) = \frac{1}{2}\left(1 + \frac{\mu}{\sigma}\sqrt{h}\right)$$
$$\mathrm{P}\left(X(t+h) = x - \sigma\sqrt{h} \,\middle|\, X(t) = x\right) = \frac{1}{2}\left(1 - \frac{\mu}{\sigma}\sqrt{h}\right)$$

$\Rightarrow$ If $\mu > 0$ biased to the right, if $\mu < 0$ biased to the left

- Definition requires h small enough to make $(\mu/\sigma)\sqrt{h} \leq 1$
- Notice that bias vanishes as $\sqrt{h}$, same as step size

Mean and variance of biased steps

- Define step RV $Y_n = \pm 1$, with probabilities

$$P(Y_n = 1) = \frac{1}{2}\left(1 + \frac{\mu}{\sigma}\sqrt{h}\right), \quad P(Y_n = -1) = \frac{1}{2}\left(1 - \frac{\mu}{\sigma}\sqrt{h}\right)$$

- Expected value of Y_n is

$$\begin{aligned}\mathbb{E}[Y_n] &= 1 \times P(Y_n = 1) + (-1) \times P(Y_n = -1) \\ &= \frac{1}{2}\left(1 + \frac{\mu}{\sigma}\sqrt{h}\right) - \frac{1}{2}\left(1 - \frac{\mu}{\sigma}\sqrt{h}\right) = \frac{\mu}{\sigma}\sqrt{h}\end{aligned}$$

- Second moment of Y_n is

$$\mathbb{E}\left[Y_n^2\right] = (1)^2 \times P(Y_n = 1) + (-1)^2 \times P(Y_n = -1) = 1$$

- Variance of Y_n is $\Rightarrow \text{var}[Y_n] = \mathbb{E}\left[Y_n^2\right] - \mathbb{E}^2[Y_n] = 1 - \frac{\mu^2}{\sigma^2}h$

Central Limit Theorem as $h \to 0$

- Consider time $T = Nh$, index $n = 1, 2, \ldots, N$. Write $X(nh)$ as

$$X(nh) = X((n-1)h) + \left(\sigma\sqrt{h}\right) Y_n$$

- Use recursively to write $X(T) = X(Nh)$ as

$$X(T) = X(Nh) = X(0) + \left(\sigma\sqrt{h}\right) \sum_{n=1}^{N} Y_n = \left(\sigma\sqrt{h}\right) \sum_{n=1}^{N} Y_n$$

- As $h \to 0$ we have $N \to \infty$ and $\sum_{n=1}^{N} Y_n$ normally distributed
- As $h \to 0$, $X(T)$ tends to be normally distributed by CLT
  - Need to determine mean and variance (and only mean and variance)

Mean and variance of $X(T)$

- Expected value of $X(T)$ = scaled sum of $\mathbb{E}[Y_n]$ (recall $T = Nh$)

$$\mathbb{E}[X(T)] = \left(\sigma\sqrt{h}\right) \times N \times \mathbb{E}[Y_n] = \left(\sigma\sqrt{h}\right) \times N \times \left(\frac{\mu}{\sigma}\sqrt{h}\right) = \mu T$$

- Variance of $X(T)$ = scaled sum of variances of independent Y_n

$$\begin{aligned} \text{var}[X(T)] &= \left(\sigma\sqrt{h}\right)^2 \times N \times \text{var}[Y_n] \\ &= \left(\sigma^2 h\right) \times N \times \left(1 - \frac{\mu^2}{\sigma^2}h\right) \to \sigma^2 T \end{aligned}$$

$\Rightarrow$ Used $T = Nh$ and $1 - (\mu^2/\sigma^2)h \to 1$

- Brownian motion with drift (BMD) $\Rightarrow X(t) \sim \mathcal{N}\left(\mu t, \sigma^2 t\right)$

$\Rightarrow$ Normal with mean μt and variance $\sigma^2 t$

$\Rightarrow$ Independent and stationary increments

Mean-square derivative of a random process

- Consider a realization $x(t)$ of the random process $X(t)$
- **Def:** The derivative of (lowercase) $x(t)$ is

$$\frac{\partial x(t)}{\partial t} = \lim_{h \to 0} \frac{x(t+h) - x(t)}{h}$$

- When this limit exists $\Rightarrow$ Limit may not exist for all realizations
- Can define sure limit, a.s. limit, in probability, ...

  $\Rightarrow$ Notion of convergence used here is in mean-squared sense

- **Def:** Process $\partial X(t)/\partial t$ is the mean-square sense derivative of $X(t)$ if

$$\lim_{h \to 0} \mathbb{E}\left[\left(\frac{X(t+h) - X(t)}{h} - \frac{\partial X(t)}{\partial t}\right)^2\right] = 0$$

Mean-square integral of a random process

- Likewise consider the integral of a realization $x(t)$ of $X(t)$

$$\int_a^b x(t)dt = \lim_{h \to 0} \sum_{n=1}^{(b-a)/h} hx(a+nh)$$

$\Rightarrow$ Limit need not exist for all realizations

- Can define in sure sense, almost sure sense, in probability sense, ...

  $\Rightarrow$ Again, adopt definition in mean-square sense

- **Def:** Process $\int_a^b X(t)dt$ is the mean square sense integral of $X(t)$ if

$$\lim_{h \to 0} \mathbb{E}\left[\left(\sum_{n=1}^{(b-a)/h} hX(a+nh) - \int_a^b X(t)dt\right)^2\right] = 0$$

- Mean-square sense convergence is convenient to work with GPs

Linear state model example

- **Def:** A random process $X(t)$ follows a linear state model if

$$\frac{\partial X(t)}{\partial t} = aX(t) + W(t)$$

with $W(t)$ WGN, autocorrelation $R_W(t_1, t_2) = \sigma^2 \delta(t_1 - t_2)$

- Discrete-time representation of $X(t)$ $\Rightarrow X(nh)$ with step size h
- Solving differential equation between nh and $(n+1)h$ (h small)

$$X((n+1)h) \approx X(nh)e^{ah} + \int_{nh}^{(n+1)h} W(t)\, dt$$

- Defining $X(n) := X(nh)$ and $W(n) := \int_{nh}^{(n+1)h} W(t)\, dt$ may write

$$X(n+1) \approx (1 + ah)X(n) + W(n)$$

$\Rightarrow$ Where $\mathbb{E}\left[W^2(n)\right] = \sigma^2 h$ and $W(n_1)$ independent of $W(n_2)$

Stationary random processes

- All joint probabilities invariant to time shifts, i.e., for any s

$$\mathrm{P}\left(X(t_1+s) \leq x_1, X(t_2+s) \leq x_2, \ldots, X(t_n+s) \leq x_n\right) = \mathrm{P}\left(X(t_1) \leq x_1, X(t_2) \leq x_2, \ldots, X(t_n) \leq x_n\right)$$

$\Rightarrow$ If above relation holds $X(t)$ is called strictly stationary (SS)

- First-order stationary $\Rightarrow$ probs. of single variables are shift invariant

$$\mathrm{P}\left(X(t+s) \leq x\right) = \mathrm{P}\left(X(t) \leq x\right)$$

- Second-order stationary $\Rightarrow$ joint probs. of pairs are shift invariant

$$\mathrm{P}\left(X(t_1+s) \leq x_1, X(t_2+s) \leq x_2\right) = \mathrm{P}\left(X(t_1) \leq x_1, X(t_2) \leq x_2\right)$$

Pdfs and moments of stationary processes

- For SS process joint cdfs are shift invariant. Hence, pdfs also are

$$f_{X(t+s)}(x) = f_{X(t)}(x) = f_{X(0)}(x) := f_X(x)$$

- As a consequence, the mean of a SS process is constant

$$\mu(t) := \mathbb{E}\left[X(t)\right] = \int_{-\infty}^{\infty} x f_{X(t)}(x)dx = \int_{-\infty}^{\infty} x f_X(x)dx = \mu$$

- The variance of a SS process is also constant

$$\text{var}\left[X(t)\right] := \int_{-\infty}^{\infty} (x-\mu)^2 f_{X(t)}(x)dx = \int_{-\infty}^{\infty} (x-\mu)^2 f_X(x)dx = \sigma^2$$

- The power (second moment) of a SS process is also constant

$$\mathbb{E}\left[X^2(t)\right] := \int_{-\infty}^{\infty} x^2 f_{X(t)}(x)dx = \int_{-\infty}^{\infty} x^2 f_X(x)dx = \sigma^2 + \mu^2$$

Joint pdfs of stationary processes

- Joint pdf of two values of a SS random process

$$f_{X(t_1)X(t_2)}(x_1, x_2) = f_{X(0)X(t_2-t_1)}(x_1, x_2)$$

  $\Rightarrow$ Used shift invariance for shift of t_1

  $\Rightarrow$ Note that $t_1 = 0 + t_1$ and $t_2 = (t_2 - t_1) + t_1$

- Result above true for any pair t_1, t_2

  $\Rightarrow$ Joint pdf depends only on time difference $s := t_2 - t_1$

- Writing $t_1 = t$ and $t_2 = t + s$ we equivalently have

$$f_{X(t)X(t+s)}(x_1, x_2) = f_{X(0)X(s)}(x_1, x_2) = f_X(x_1, x_2; s)$$

Stationary processes and limit distributions

- Stationary processes follow the footsteps of limit distributions

- For Markov processes limit distributions exist under mild conditions
  - Limit distributions also exist for some non-Markov processes

- Process somewhat easier to analyze in the limit as $t \to \infty$

  $\Rightarrow$ Properties can be derived from the limit distribution

- Stationary process $\approx$ study of limit distribution

  $\Rightarrow$ Formally initialize at limit distribution

  $\Rightarrow$ In practice results true for time sufficiently large

- Deterministic linear systems $\Rightarrow$ transient + steady-state behavior

  $\Rightarrow$ Stationary systems akin to the study of steady-state

- But steady-state is in a probabilistic sense (probs., not realizations)

Autocorrelation function

- From the definition of autocorrelation function we can write

$$R_X(t_1, t_2) = \mathbb{E}\left[X(t_1)X(t_2)\right] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x_1 x_2 f_{X(t_1)X(t_2)}(x_1, x_2)\, dx_1 dx_2$$

- For SS process $f_{X(t_1)X(t_2)}(\cdot)$ depends on time difference only

$$R_X(t_1, t_2) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x_1 x_2 f_{X(0)X(t_2-t_1)}(x_1, x_2)\, dx_1 dx_2 = \mathbb{E}\left[X(0)X(t_2-t_1)\right]$$

$\Rightarrow$ $R_X(t_1, t_2)$ is a function of $s = t_2 - t_1$ only

$$R_X(t_1, t_2) = R_X(0, t_2 - t_1) := R_X(s)$$

- The autocorrelation function of a SS random process $X(t)$ is $R_X(s)$
  - $\Rightarrow$ Variable s denotes a time difference / shift / lag
  - $\Rightarrow$ $R_X(s)$ specifies correlation between values $X(t)$ spaced s in time

Autocovariance function

- Similarly to autocorrelation, define the autocovariance function as

$$C_X(t_1, t_2) = \mathbb{E}\left[\big(X(t_1) - \mu(t_1)\big)\big(X(t_2) - \mu(t_2)\big)\right]$$

- Expand product to write $C_X(t_1, t_2)$ as

$$C_X(t_1, t_2) = \mathbb{E}\left[X(t_1)X(t_2)\right] + \mu(t_1)\mu(t_2) - \mathbb{E}\left[X(t_1)\right]\mu(t_2) - \mathbb{E}\left[X(t_2)\right]\mu(t_1)$$

- For SS process $\mu(t_1) = \mu(t_2) = \mu$ and $\mathbb{E}\left[X(t_1)X(t_2)\right] = R_X(t_2 - t_1)$

$$C_X(t_1, t_2) = R_X(t_2 - t_1) - \mu^2 = C_X(t_2 - t_1)$$

$\Rightarrow$ Autocovariance function depends only on the shift $s = t_2 - t_1$

- We will typically assume that $\mu = 0$ in which case

$$R_X(s) = C_X(s)$$

$\Rightarrow$ If $\mu \neq 0$ can study process $X(t) - \mu$ whose mean is null

Wide-sense stationary processes

- **Def:** A process is wide-sense stationary (WSS) when its
  - $\Rightarrow$ Mean is constant $\Rightarrow \mu(t) = \mu$ for all t
  - $\Rightarrow$ Autocorrelation is shift invariant $\Rightarrow R_X(t_1, t_2) = R_X(t_2 - t_1)$
- Consequently, autocovariance of WSS process is also shift invariant

$$\begin{aligned} C_X(t_1, t_2) &= \mathbb{E}\left[X(t_1)X(t_2)\right] + \mu(t_1)\mu(t_2) - \mathbb{E}\left[X(t_1)\right]\mu(t_2) - \mathbb{E}\left[X(t_2)\right]\mu(t_1) \\ &= R_X(t_2 - t_1) - \mu^2 \end{aligned}$$

- Most of the analysis of stationary processes is based on $R_X(t_2 - t_1)$
  - $\Rightarrow$ Thus, such analysis does not require SS, WSS suffices

Wide-sense stationarity versus strict stationarity

- SS processes have shift-invariant pdfs
  - $\Rightarrow$ Mean function is constant
  - $\Rightarrow$ Autocorrelation is shift-invariant
- Then, a SS process is also WSS
  - $\Rightarrow$ For that reason WSS is also called weak-sense stationary
- The opposite is obviously not true in general
- But if Gaussian, process determined by mean and autocorrelation
  - $\Rightarrow$ WSS implies SS for Gaussian process
- WSS and SS are equivalent for Gaussian processes (More coming)

Gaussian wide-sense stationary process

- WSS Gaussian process $X(t)$ with mean 0 and autocorrelation $R(s)$
- The covariance matrix for $X(t_1+s), X(t_2+s), \ldots, X(t_n+s)$ is

$$\mathbf{C}(t_1{+}s, \ldots, t_n{+}s) = \begin{pmatrix} R(t_1+s, t_1+s) & R(t_1+s, t_2+s) & \ldots & R(t_1+s, t_n+s) \\ R(t_2+s, t_1+s) & R(t_2+s, t_2+s) & \ldots & R(t_2+s, t_n+s) \\ \vdots & \vdots & \ddots & \vdots \\ R(t_n+s, t_1+s) & R(t_n+s, t_2+s) & \ldots & R(t_n+s, t_n+s) \end{pmatrix}$$

- For WSS process, autocorrelations depend only on time differences

$$\mathbf{C}(t_1+s, \ldots, t_n+s) = \begin{pmatrix} R(t_1-t_1) & R(t_2-t_1) & \ldots & R(t_n-t_1) \\ R(t_1-t_2) & R(t_2-t_2) & \ldots & R(t_n-t_2) \\ \vdots & \vdots & \ddots & \vdots \\ R(t_1-t_n) & R(t_2-t_n) & \ldots & R(t_n-t_n) \end{pmatrix} = \mathbf{C}(t_1, \ldots, t_n)$$

$\Rightarrow$ Covariance matrices $\mathbf{C}(t_1, \ldots, t_n)$ are shift invariant

Gaussian wide-sense stationary process (continued)

- The joint pdf of $X(t_1+s), X(t_2+s), \ldots, X(t_n+s)$ is

$$f_{X(t_1+s),\ldots,X(t_n+s)}(x_1,\ldots,x_n) = \mathcal{N}(\mathbf{0}, \mathbf{C}(t_1+s,\ldots,t_n+s); [x_1,\ldots,x_n]^T)$$

$\Rightarrow$ Completely determined by $\mathbf{C}(t_1+s,\ldots,t_n+s)$

- Since covariance matrix is shift invariant can write

$$f_{X(t_1+s),\ldots,X(t_n+s)}(x_1,\ldots,x_n) = \mathcal{N}(\mathbf{0}, \mathbf{C}(t_1,\ldots,t_n); [x_1,\ldots,x_n]^T)$$

- Expression on the right is the pdf of $X(t_1), X(t_2), \ldots, X(t_n)$. Then

$$f_{X(t_1+s),\ldots,X(t_n+s)}(x_1,\ldots,x_n) = f_{X(t_1),\ldots,X(t_n)}(x_1,\ldots,x_n)$$

- Joint pdf of $X(t_1), X(t_2), \ldots, X(t_n)$ is shift invariant

$\Rightarrow$ Proving that WSS is equivalent to SS for Gaussian processes

Brownian motion and white Gaussian noise

Ex: Brownian motion $X(t)$ with variance parameter σ^2

$\Rightarrow$ Mean function is $\mu(t) = 0$ for all $t \geq 0$

$\Rightarrow$ Autocorrelation is $R_X(t_1, t_2) = \sigma^2 \min(t_1, t_2)$

- While the mean is constant, autocorrelation is not shift invariant

$\Rightarrow$ Brownian motion is not WSS (hence not SS)

Ex: White Gaussian noise $W(t)$ with variance parameter σ^2

$\Rightarrow$ Mean function is $\mu(t) = 0$ for all t

$\Rightarrow$ Autocorrelation is $R_W(t_1, t_2) = \sigma^2 \delta(t_2 - t_1)$

- The mean is constant and the autocorrelation is shift invariant

$\Rightarrow$ White Gaussian noise is WSS

$\Rightarrow$ Also SS because white Gaussian noise is a GP

Properties of autocorrelation function

For WSS processes:

(i) The autocorrelation for $s = 0$ is the power of the process

$$R_X(0) = \mathbb{E}\left[X^2(t)\right] = \mathbb{E}\left[X(t)X(t+0)\right]$$

(ii) The autocorrelation function is symmetric $\Rightarrow R_X(s) = R_X(-s)$

Proof.
Commutative property of product and shift invariance of $R_X(t_1, t_2)$

$$\begin{aligned} R_X(s) &= R_X(t, t+s) \\ &= \mathbb{E}\left[X(t)X(t+s)\right] \\ &= \mathbb{E}\left[X(t+s)X(t)\right] \\ &= R_X(t+s, t) = R_X(-s) \end{aligned}$$

□

Properties of autocorrelation function (continued)

For WSS processes:

(iii) Maximum absolute value of the autocorrelation function is for $s = 0$

$$\left|R_X(s)\right| \leq R_X(0)$$

Proof.
Expand the square $\mathbb{E}\left[\left(X(t+s) \pm X(t)\right)^2\right]$

$$\begin{aligned}\mathbb{E}\left[\left(X(t+s) \pm X(t)\right)^2\right] &= \mathbb{E}\left[X^2(t+s)\right] + \mathbb{E}\left[X^2(t)\right] \pm 2\mathbb{E}\left[X(t+s)X(t)\right] \\ &= R_X(0) + R_X(0) \pm 2R_X(s)\end{aligned}$$

Square $\mathbb{E}\left[\left(X(t+s) \pm X(t)\right)^2\right]$ is always nonnegative, then

$$0 \leq \mathbb{E}\left[\left(X(t+s) \pm X(t)\right)^2\right] = 2R_X(0) \pm 2R_X(s)$$

Rearranging terms $\Rightarrow R_X(0) \geq \mp R_X(s)$

□